Two simple sufficient conditions for FDR control

Gilles Blanchard, Etienne Roquain

Electronic Journal of Statistics
Vol. 2 (2008) 963–992
ISSN: 1935-7524
DOI: 10.1214/08-EJS180

Two simple sufficient conditions for FDR control

Gilles Blanchard[†,*]

Fraunhofer-Institut FIRST
Kekuléstrasse 7, 12489 Berlin, Germany
e-mail: blanchar@first.fraunhofer.de

and

Etienne Roquain[‡,*]

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

Abstract: We show that the control of the false discovery rate (FDR) for a multiple testing procedure is implied by two coupled simple sufficient conditions. The first one, which we call "self-consistency condition", concerns the algorithm itself, and the second, called "dependency control condition" is related to the dependency assumptions on the p-value family. Many standard multiple testing procedures are self-consistent (e.g. step-up, step-down or step-up-down procedures), and we prove that the dependency control condition can be fulfilled when choosing correspondingly appropriate rejection functions, in three classical types of dependency: independence, positive dependency (PRDS) and unspecified dependency. As a consequence, we recover earlier results through simple and unifying proofs while extending their scope to several regards: weighted FDR, p-value reweighting, new family of step-up procedures under unspecified p-value dependency and adaptive step-up procedures. We give additional examples of other possible applications. This framework also allows for defining and studying FDR control for multiple testing procedures over a continuous, uncountable space of hypotheses.

AMS 2000 subject classifications: Primary 62J15; secondary 62G10.
Keywords and phrases: False Discovery Rate, multiple testing, step-up, step-down, step-up-down, weighted p-values, PRDS condition.

Received February 2008.

[*]This work was supported in part by the IST and ICT Programmes of the European Community, successively under the PASCAL (IST-2002-506778) and PASCAL2 (ICT-216886) networks of excellence.

[†]Part of this work was done while the first author held an invited position at the statistics department of the University of Chicago, which is gratefully acknowledged.

[‡]Research carried out at the French institute INRA-Jouy and at the Free University of Amsterdam.

1. Introduction

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, which is a subset of the set $\mathcal{H}$ of initial candidate null hypotheses. The false discovery rate (FDR) of the procedure is then defined as the expected proportion of null hypotheses in $R(X)$ which are in fact true and thus incorrectly rejected. Following its introduction by Benjamini and Hochberg (1995), the FDR criterion has emerged recently as a widely used standard for a majority of applications involving simultaneous testing of a large number of hypotheses. It is generally required that a multiple testing procedure R has its FDR bounded by a certain fixed in advance level α.

Our main point in this work is to show that FDR control is implied by two simple conditions. The first one, which we call *self-consistency condition*, requires that any rejected hypothesis $h \in R(X)$ should have its p-value $p_h(X)$ smaller than a threshold $\Delta_\beta(|R(X)|)$ which itself depends on the volume of rejected hypothesis $|R(X)|$, and on a fixed functional parameter β. The second one, called *dependency control condition*, requires that for each true null hypothesis h, the couple of real variables $(U, V) = (p_h, |R(X)|)$ satisfies the inequality (for any $c > 0$, and the same function β as in the first condition):

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

The first condition only concerns how the data is processed to produce the decision, and is hence purely algorithmic. It can easily be checked for several classical multiple testing procedures, such as step-down, step-up or more generally step-up-down procedures. In this condition, the function β controls how the threshold increases with respect to the volume of rejected hypotheses. In particular, for step-wise procedures, β corresponds (up to proportionality constant) to the rejection function used to cut the curve of ordered p-values. The second condition, on the other hand, is essentially probabilistic in nature. More precisely, we can show that (1) can be satisfied under relatively broad assumptions on the dependency of (U, V). In turn, as will be shown in more detail in the paper, this implies that the second condition is largely independent of the exact procedure R, but rather is related to the dependency assumptions between the p-values.

The two conditions are not independent of each other: they are coupled through the same functional parameter β, appearing in (1) as well as in the definition of the threshold Δ_β. The function β, called *shape function*, is assumed to be nondecreasing but otherwise arbitrary; if there exists β such that the two corresponding conditions are satisfied, this entails FDR control.

The main advantage of this approach when controlling the FDR is that it allows us to abstract the particulars of a specific multiple testing procedure, in order to concentrate on proving the bound (1). This results in short proofs which in particular do not resort explicitly to p-values reordering.

We then present different types of applications of the result. This approach is first used to show that several well-known results on FDR control (mainly concerning step-up or step-down procedures based on a linear rejection function) are recovered in a synthetic way (e.g., results of Benjamini and Hochberg, 1995, 1997; Benjamini and Yekutieli, 2001; Sarkar, 2002; Genovese et al., 2006). We also derive the following new results:

- some classical results on step-up procedures are extended to weighted procedures (weighted-FDR and/or p-value weighting), under independence or dependence of the p-values;
- a new family of step-up procedures which control the FDR is presented, under unspecified dependencies between the p-values;
- we present a simple, exemplary application of this approach to the problem of *adaptive* procedures, where an estimate of the proportion π_0 of true null hypotheses in $\mathcal{H}$ is included in the procedure with the aim of increasing power;
- the case of a continuous space of hypotheses is briefly investigated (which can be relevant for instance when the underlying obervation is modelled as a stochastic process);
- the results of Benjamini and Liu (1999a) and Romano and Shaikh (2006a) on a specific type of step-down procedure are extended to the cases of positive dependencies (under a PRDS-type condition) and unspecified dependencies.

To put some perspective, let us emphasize here again that the conditions proposed here are only *sufficient* and certainly not necessary: naturally, there are many examples of multiple testing procedures that are known to have controlled FDR but do not satisfy the coupled conditions presented here (including some particular step-up and step-down procedures). The message that we nevertheless want to convey is that these conditions are able to cover at once an interesting range of classical existing results on FDR control as well as provide a useful technical tool. It was pointed out to us that a result similar in spirit to ours will appear in the forthcoming paper by Finner et al. (2008); this is discussed in more detail in Section 5.1.

This paper is organized as follows: in Section 2, we introduce the framework, the two conditions and we prove that taken together, they imply FDR control. The self-consistency and dependency control conditions are then studied separately in Section 3, leading to specific assumptions, repectively, on the procedure itself (e.g. step-down, step-up) and on the dependency between the p-values (independence, PRDS, unspecified dependencies). The applications summarized above are detailed in Section 4. Some technical proofs are postponed in the appendix.

2. Two sufficient conditions for FDR control

2.1. Preliminaries and notations

Let $(\mathcal{X}, \mathfrak{X}, P)$ be a probability space, with P belonging to a set or "model" $\mathfrak{P}$ of distributions, which can be parametric or non-parametric. Formally, a *null hypothesis* is a subset $h \subset \mathfrak{P}$ 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 determining simultaneously whether or not P satisfies distinct null hypotheses belonging to a certain set $\mathcal{H}$ of candidate hypotheses. Below, we will always assume that $\mathcal{H}$ is at most countable (except specifically in Section 4.4, where we mention extensions to continuous sets of hypotheses). We denote by $\mathcal{H}_0(P) = \{h \in \mathcal{H} \mid P \text{ satisfies } h\}$ the set of null hypotheses satisfied by P, called the *set of true null hypotheses*. We denote by $\mathcal{H}_1(P) = \mathcal{H} \setminus \mathcal{H}_0(P)$ the *set of 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 the *rejected null hypotheses* of the procedure R.

We will only 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 $\mathbf{p} = (p_h, h \in \mathcal{H})$. For this, we must assume 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 $[0, 1]$:

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

A *type I error* occurs when a true null hypothesis h is wrongly rejected i.e. when $h \in R(x) \cap \mathcal{H}_0(P)$. There are several different ways to measure quantitatively the collective type I error of a multiple testing procedure. In this paper, we will exclusively focus on the false discovery rate (FDR) criterion, introduced by Benjamini and Hochberg (1995) and which has since become a widely used standard.

The FDR is defined as the averaged proportion of type I errors in the set of all the rejected hypotheses. This "error proportion" will be defined in terms of a volume ratio, and to this end 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

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

Throughout this paper we will use the following notational 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.

For the sake of simplifying the exposition, we will (as is usually the accepted convention) most often drop in the notation a certain number of dependencies, such as writing R or p_h instead of $R(X)$, $p_h(X)$ and $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathrm{FDR}(R)$ instead of $\mathcal{H}_0(P)$, $\mathcal{H}_1(P)$, $\mathrm{FDR}(R,P)$. We will also omit the fact that the probabilities or expectations are performed with respect to $X \sim P$. Generally speaking, we will implicitly assume that P is fixed, but that all relevant assumptions and results should in fact hold for any $P \in \mathfrak{P}$. For example, our main goal will be to derive upper bounds on $\mathrm{FDR}(R,P)$ valid for all $P \in \mathfrak{P}$; this will be formulated simply as a bound on $\mathrm{FDR}(R)$.

Remark 2.3. (Role of Λ and weighted FDR in the finite case) When the space of hypotheses is finite, the "standard" FDR in multiple testing literature is the one defined using $|.|$ equal to the counting measure (cardinality) on a finite space and will be referred to as "standard Λ weigthing". The notation $|.|$ was kept here to allow notation compatibility with this case and to alleviate some notational burden. We stress however that in the case $\mathcal{H}$ is countably infinite, the volume measure Λ cannot be the cardinality, since we assume it to be finite.

The possibility of using different weights $\Lambda(\{h\})$ for particular hypotheses h leads to the so-called "weighted FDR". In general, the measure Λ represents the relative importance, or criticality, of committing an error about different hypotheses, and can be dictated by external constraints. As discussed by Benjamini and Hochberg (1997) and Benjamini and Heller (2007), controlling the "weighted FDR" can be of interest in some specific applications. For instance, in the situation where each hypothesis concerns a whole cluster of voxels in a brain map, it can be relevant to increase the importance of large discovered clusters when counting the discoveries in the FDR. Note finally that Λ can be rescaled arbitrarily since only volume ratios matter in the FDR.

2.2. Self-consistency, dependency control and the false discovery rate

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 (possibly data-dependent) threshold. We will be more particularly interested in thresholds that specifically depend on a real parameter r and possibly on the hypothesis h itself, as introduced in the next definition.

Definition 2.4 (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 nondecreasing 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 nondecreasing function called the *shape function*. Given a threshold collection Δ, the Δ-thresholding-based multiple testing procedure at rejection volume r is defined as

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

Let us discuss the role of the parameter r and proceed to the first of the two announced sufficient conditions. Remember our goal is to upper bound $\mathrm{FDR}(R)$, where the volume of rejected hypotheses $|R|$ appears as the denominator in the expectation. Hence, intuitively, whenever this volume gets larger, we can globally allow more type I errors, and thus take a larger threshold for the p-values. Therefore, the rejection volume parameter r in the definition above should be picked as an (increasing) function of $|R|$. Formally, this leads to the following "self-referring" property:

Definition 2.5 (Self-consistency condition). Given a factorized threshold collection of the form $\Delta(h,r) = \alpha\pi(h)\beta(r)$, a multiple testing procedure R satisfies the self-consistency condition with respect to the threshold collection Δ if the following inclusion holds a.s.:

$$R \subset L_\Delta(|R|). \tag{$\mathbf{SC}(\alpha,\pi,\beta)$}$$

Next, we introduce the following probabilistic condition on two dependent real variables:

Definition 2.6 (Dependency control condition). Let $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a nondecreasing function. A couple (U,V) of (possibly dependent) nonnegative real random variables is said to satisfy the dependency control condition with shape function β if the following inequalities hold:

$$\forall c > 0, \quad \mathbb{E}\left[\frac{\mathbf{1}\{U \leq c\beta(V)\}}{V}\right] \leq c\,. \tag{$\mathbf{DC}(\beta)$}$$

The following elementary but fundamental result is the main cornerstone linking the FDR control to conditions **SC** and **DC**.

Proposition 2.7. *Let $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a (nondecreasing) shape function, $\pi : \mathcal{H} \to [0,1]$ a weight function and α a positive number. Assume that the multiple testing procedure R is such that:*

(i) the self-consistency condition $\mathbf{SC}(\alpha,\pi,\beta)$ is satisfied;
(ii) for any $h \in \mathcal{H}_0$ the couple $(p_h, |R|)$ satisfies $\mathbf{DC}(\beta)$.

Then $\mathrm{FDR}(R) \leq \alpha\Pi(\mathcal{H}_0)$, *where* $d\Pi = \pi d\Lambda$, *i.e.,* $\Pi(\mathcal{H}_0) := \sum_{h\in\mathcal{H}_0} \Lambda(\{h\})\pi(h)$.

Proof. From (2),

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

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

Let us point out the important difference in nature between the two sufficient conditions: for a fixed shape function β, the self-consistency condition (i) concerns only the algorithm itself (and not the random structure of the problem). On the other hand, the dependency control condition (ii) seems to involve both the algorithm and the statistical nature of the problem. However, we will show below in Section 3.2 that this latter condition can be checked under a weak, general and quite natural assumption on the algorithm itself (namely that $|R(\mathbf{p})|$ is nonincreasing function of the p-values), and primarily depends on the dependency structure of the p-values. (Moreover, in the case of arbitrary dependencies, we will consider a special family of βs which satisfy the condition without any assumptions on the algorithm.) Hence, the interest of the above proposition is that it effectively separates the problem of FDR control into a purely *algorithmic* and an (almost) purely *probabilistic* sufficient condition. The link between the two conditions is the common shape function β: the dependency assumptions between the p-values will determine for which shape function the condition **DC**(β) is valid; in turn, this will impose constraints on the algorithm through condition **SC**(α,π,β).

Remark 2.8. (Role of π and p-value weighting in the finite case) To understand intuitively the role of the weight function π, assume $\mathcal{H}$ is of finite cardinality m and take for simplification $\beta(r)=1$ for now. Consider the corresponding testing procedure L_Δ: the rejected hypotheses are those for which $p'_h := p_h/(m\pi(h)) \le \alpha/m$, where p'_h is the *weighted p-value* of h. If $\pi(h)$ is constant equal to $1/m$, we have $p'_h = p_h$ and the above is just Bonferroni's procedure, which has family-wise error rate (FWER) controlled by α. If $\pi(h)$ is, more generally, an arbitrary probability distribution on $\mathcal{H}$, the above is a weighted Bonferroni's procedure and has also FWER less than α (see, e.g., Wasserman and Roeder, 2006). In this example, π represents the relative importance, or weight of evidence, that is given *a priori* to p-values, and thus plays the role of a prior that can be fixed arbitrarily by the user. Its role in the control of FDR is very similar; the use of weighted p-values for FDR control has been proposed earlier, for example by Genovese et al. (2006). When $\mathcal{H}$ is of finite cardinality m, we will refer to the choice $\pi(h)\equiv 1/m$, in conjunction with Λ being the cardinality measure, as the "standard $\Lambda-\pi$ weighting".

More generally, following Proposition 2.7, control of the FDR at level α is ensured as soon as the weight function π is chosen as a probability density with respect to Λ (i.e. $\sum_{h\in\mathcal{H}}\Lambda(\{h\})\pi(h)=1$). When $\mathcal{H}$ is of finite cardinality m and with the "standard $\Lambda-\pi$ weighting" defined above, we obtain $\text{FDR}\leq \alpha m_0/m\leq \alpha$ (where m_0 denotes the number of true null hypotheses).

Remark 2.9. Proposition 2.7 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 not 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.

3. Study of the two sufficient conditions

In this section, we give a closer look to conditions $\mathbf{SC}(\alpha,\pi,\beta)$ and $\mathbf{DC}(\beta)$, and study typical situations where they are statisfied.

3.1. Self-consistency condition and step-up procedures

The main examples of self-consistent procedures are step-up procedures. In fact, for a fixed choice of parameters (α,β,π), step-up procedures output the largest set of rejected hypotheses such that $\mathbf{SC}(\alpha,\pi,\beta)$ is satisfied, and are in this sense optimal with respect to that condition. Here, we define step-up procedures by this characterizing property, thus avoiding the usual definition using the reordering of the p-values.

Definition 3.1 (Step-up procedure). Let Δ be a factorized threshold collection of the form $\Delta(h,r)=\alpha\pi(h)\beta(r)$. The *step-up multiple testing procedure* R associated to Δ, is given by either of the following equivalent definitions:

$$(i)\quad R=L_\Delta(\hat{r}),\ \text{ where } \hat{r}:=\max\{r\geq 0 \mid |L_\Delta(r)|\geq r\}$$
$$(ii)\quad R=\bigcup\{A\subset\mathcal{H}\mid A \text{ satisfies } \mathbf{SC}(\alpha,\pi,\beta)\ \}\,.$$

Additionally, $\widehat{r}$ satisfies $|L_\Delta(\widehat{r})|=\widehat{r}$; equivalently, the step-up procedure R satisfies $\mathbf{SC}(\alpha,\pi,\beta)$ 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 $r\geq 0$. Therefore, $|L_\Delta(r)|$ is a nondecreasing function of r and the supremum appearing in (i) is a maximum i.e. $|L_\Delta(\hat{r})|\geq\hat{r}$. It is easy to see that $|L_\Delta(\hat{r})|=\hat{r}$ because 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 $\mathbf{SC}(\alpha,\pi,\beta)$ (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 $A\subset L_\Delta(\hat{r})$. □

When $\mathcal{H}$ is finite of cardinal m endowed with the standard Λ-weighting $\Lambda(\cdot)=\text{Card}(\cdot)$, Definition 3.1 is equivalent to the classical definition of a

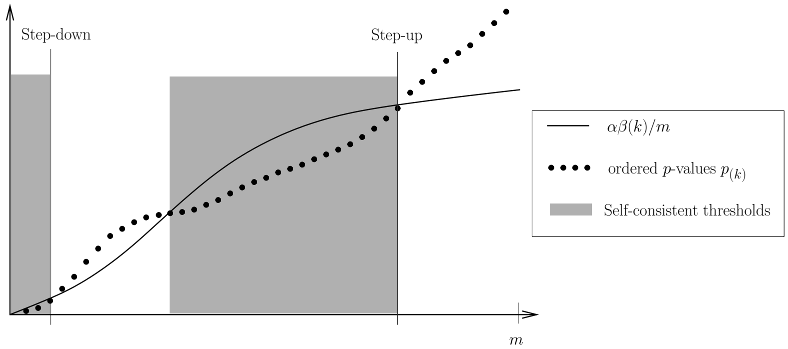

Fig 1. *Pictorial representation of the step-up (and step-down) thresholds, and (in grey) of all thresholds $r \in \{1,\dots,m\}$ for which $L_\Delta(r)$ satisfies the self-consistency condition. The p-values and the rejection function represented here have been picked arbitrarily and in a deliberately exaggerated fashion in order to illustrate the different procedures; they are not meant to represent a realistic data or model. This picture corresponds to the standard Λ-π weighting only.*

step-up procedure, based on reordering the p-values: 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

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

Since $L_\Delta(r) = \{h \in \mathcal{H} \mid p'_h \le \alpha\beta(r)/m\}$, the condition $|L_\Delta(r)| \ge r$ is equivalent to $p'_{(r)} \le \alpha\beta(r)/m$. Hence, the step-up procedure associated to Δ defined in Definition 3.1 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 scaled shape function $\alpha\beta(\cdot)/m$:

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

with $p'_{(0)} := 0$; see Figure 1 for an illustration. For the standard π-weighting $\pi(h) = 1/m$, the weighted p-values are simply the p-values. In particular:

- The step-up procedure associated to the linear shape function $\beta(r) = r$ is the well-known linear step-up procedure of Benjamini and Hochberg (1995).
- The step-up procedure associated to the linear shape function $\beta(r) = r\left(\sum_{i=1}^m \frac{1}{i}\right)^{-1}$ is the distribution-free linear step-up procedure of Benjamini and Yekutieli (2001).

Finally, let us point out that *step-down* and more generally *step-up-down* procedures are also self-consistent. The latter class of step-wise procedures have been introduced by Tamhane et al. (1998), and contains step-up and step-down procedures as particular cases. These procedures select in a certain way among the "crossing points" between the p-value function and some fixed rejection function (for example, on Figure 1, there are only two non-zero crossing points to choose from). More formally, and under arbitrary weighting, given a parameter $\lambda \in [0, |\mathcal{H}|]$, the step-up-down procedure with threshold collection Δ and of order λ is defined as $L_\Delta(\widehat{r}_\lambda)$, where either $\widehat{r}_\lambda := \max\{r \geq \lambda \mid \forall r', \lambda \leq r' \leq r, |L_\Delta(r')| \geq r'\}$ if $|L_\Delta(\lambda)| \geq \lambda$; or $\widehat{r}_\lambda := \max\{r < \lambda \mid |L_\Delta(r)| \geq r\}$ otherwise. In words, assuming the standard weighting case and λ an integer, if $p_{(\lambda)}$ is smaller than the rejection function at λ, the closest crossing point to the right of λ is picked, otherwise the closest crossing point to the left. In particular, the step-up-down procedure of order $\lambda = |\mathcal{H}|$ is simply the step-up procedure (based on the same threshold collection). The case $\lambda = 0$ is the *step-down* procedure. Although generalized step-up-down procedures are not maximal with respect to condition **SC** like the plain step-up, the fact that they still satisfy that condition is worth noticing.

3.2. Dependency control condition

In this section, we show that condition (ii) of Proposition 2.7 holds under different types of assumptions on the dependency of the p-values. We will follow the different types of dependencies considered by Benjamini and Yekutieli (2001), namely independent, positive dependency under the PRDS condition and arbitrarily dependent p-values. In each case, we have to prove **DC**(β) for specific conditions on the variables (U, V), resulting in specific choices for the shape function β.

We start the section with a probabilistic lemma collecting the technical tools used to deal with each situation.

Lemma 3.2. *Let (U, V) be a couple of nonnegative random variables such that U is stochastically lower bounded by a uniform variable on $[0, 1]$, i.e. $\forall t \in [0, 1], \mathbb{P}(U \leq t) \leq t$. Then the dependency control condition* **DC**(β) *is satisfied by (U, V) under any of the following situations:*

(i) $\beta(x) = x$ and $V = g(U)$, where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a nonincreasing function.

(ii) $\beta(x) = x$ and the conditional distribution of V given $U \leq u$ is stochastically decreasing in u, that is,

$$\text{for any } r \geq 0, \text{ the function } u \mapsto \mathbb{P}(V < r \mid U \leq u) \text{ is nondecreasing}. \quad (5)$$

(iii) The shape function is of the form

$$\beta_\nu(r) = \int_0^r x d\nu(x), \quad (6)$$

where ν is an arbitrary probability distribution on $(0, \infty)$, and V is arbitrary.

The proof is found in appendix. Note that there is some redundancy in the lemma since (i) is a particular case of (ii), but this subcase has a particularly simple proof and is of self interest because it corresponds to the case of independent p-values (as will be detailed below).

We now apply this result to prove that for any $h \in \mathcal{H}_0$, the couple of variables $(p_h, |R|)$ satisfies **DC**(β), under the different dependency assumptions on the p-values, and for the correspondingly appropriate functions β given by the lemma. 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).

3.2.1. Independent case

Proposition 3.3. *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 nonincreasing in each p-value p_h such that $h \in \mathcal{H}_0$. For any $h \in \mathcal{H}$, denote $\mathbf{p}_{-h}$ the collection of p-values $(p_g : g \in \mathcal{H}, g \neq h)$.*

Then for any $h \in \mathcal{H}_0$ and for the linear shape function $\beta(x) = x$, the couple of variables $(p_h, |R|)$ satisfies **DC**(β)*, in which the expectation is taken conditionally to the p-values of $\mathbf{p}_{-h}$. As a consequence, it also satisfies* **DC**(β) *unconditionally.*

Proof. By the independence assumption, the distribution of $U = p_h$ 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 g of p_h by the assumption on R. We conclude by part (i) of Lemma 3.2. □

Remark 3.4. Note that Proposition 3.3 is still valid under the slightly weaker assumption that for all $h \in \mathcal{H}_0$, p_h is independent of the family $(p_g, g \neq h)$ (in particular, the p-values of $(p_h, h \in \mathcal{H}_1)$ need not be mutually independent).

3.2.2. Positive dependencies (PRDS)

From point (ii) of Lemma 3.2, each couple $(p_h, |R|)$ satisfies **DC**(β) with $\beta(x) = x$ under the following condition (weaker than independence):

$$\text{for any } r \geq 0\,, \text{ the function } u \mapsto \mathbb{P}(|R| < r \mid p_h \leq u) \text{ is nondecreasing}\,. \quad (7)$$

Following Benjamini and Yekutieli (2001), we state a dependency condition ensuring that $(p_h, |R|)$ satisfies (7). For this, we recall the definition of positive regression dependency on each one from a subset (PRDS) (introduced by Benjamini and Yekutieli, 2001, 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 3.5. 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 $h \in \mathcal{H}'$, for any measurable nondecreasing set $D \subset [0,1]^{\mathcal{H}}$, the function $u \mapsto \mathbb{P}(\mathbf{p} \in D \mid p_h = u)$ is nondecreasing.

We can now state the following proposition:

Proposition 3.6. *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 for any* $h \in \mathcal{H}_0$, *the couple of variables* $(p_h, |R|)$ *satisfies* $\mathbf{DC}(\beta)$ *for the linear shape function* $\beta(x) = x$.

Proof. We merely check that condition (7) 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 by elementary considerations, the PRDS condition (applied using the set D defined above) implies (7). The latter argument was also used by Benjamini and Yekutieli (2001) with a reference to Lehmann (1966). We provide here a succinct proof of this fact in the interest of remaining self-contained.

Under the PRDS condition, for all $u \leq u'$, putting $\gamma = \mathbb{P}\left[p_h \leq u \mid p_h \leq u'\right]$,

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

where we have used the definition of PRDS for the last inequality. □

3.2.3. Unspecified dependencies

We now consider a totally generic setting with no assumption on the dependency structure between the p-values nor on the structure of the multiple testing procedure R. Using point (iii) of Lemma 3.2, we obtain immediately the following result:

Proposition 3.7. *Let* β_ν *be a shape function of the form* (6). *Then for any* $h \in \mathcal{H}_0$, *the couple of variables* $(p_h, |R|)$ *satisfies* $\mathbf{DC}(\beta)$, *for any multiple testing procedure* R.

Note that a shape function of the form (6) must satisfy $\beta_\nu(r) \leq r$, with strict inequality except for at most one point beside zero (some examples will be discussed below in Section 4.2). Therefore, the price to pay here is a more conservative dependency control inequality, in turn resulting in a more restrictive self-consistency condition when using this shape function. This form of shape function was initially introduced by Blanchard and Fleuret (2007), where some ties were exposed between multiple testing and statistical learning theory.

4. Applications

4.1. The linear step-up procedure with $\Lambda - \pi$ weighting

We have seen earlier in Section 3.1 that step-up procedures satisfy the self-consistency condition. Furthermore, is is easy to see that step-up procedures are nonincreasing as a function of the p-values. Using this in conjunction with Proposition 3.3 (resp. Proposition 3.6) and Proposition 2.7, we obtain the following result for the (Λ-weighted) FDR control of the (π-weighted) linear step-up procedure – that is, the step-up procedure associated to the threshold collection $\Delta(h, r) = \alpha\pi(h)r$.

Theorem 4.1. *For any finite and positive volume measure Λ , the (π-weighted) linear step-up procedure R has its (Λ-weighted) FDR upper bounded by $\Pi(\mathcal{H}_0)\alpha$, 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 $\mathcal{H}_0$.*

Again, the statement is redundant since independence is a particular case of PRDS, and we just wanted to recall that the treatment of the independent case is particularly simple. This theorem essentially recovers and unifies some known results concerning particular cases: the two points of the theorem were respectively proved by Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001), with a uniform π, and Λ the cardinality measure. For a general volume measure Λ and a uniform π, the above result in the independent case was proved by Benjamini and Hochberg (1997). A proof with a general π, Λ the cardinality measure and in the independent case was investigated by Genovese et al. (2006).

The interest of the present framework is to allow for a general and unified version of these results with a concise proof (avoiding in particular to consider explicitly p-value reordering). We distinguish clearly between the two different ways to obtain "weighted" versions of step-up procedures, by changing respectively the choice of the volume measure Λ or the weight function π. Both types of weighting are of interest and of different nature; using weighted p-values can have a large impact on power (Genovese et al., 2006; Roquain and van de Wiel, 2008; see also above Remark 2.8), while using a volume Λ different from the cardinality measure can be of relevance for some application cases (see Benjamini and Hochberg, 1997; Benjamini and Heller, 2007; and Remark 2.3 above). Up to our knowledge, the two types of weighting had not been considered simultaneouly before; in particular and as noticed earlier (see Remark 2.8), in order to ensure FDR control at level α under an arbitrary volume measure Λ , the appropriate choice for a weight function π is to take a density function with respect to Λ .

4.2. An extended family of step-up procedures under unspecified dependencies

Similarly, in the case where the p-values have unspecified dependencies, we use Proposition 3.7 instead of Proposition 3.6 to derive the following theorem:

Theorem 4.2. *Consider R the step-up procedure associated to the factorized threshold collection $\Delta(h, r) = \alpha\pi(h)\beta_\nu(r)$, where the shape function β_ν can be written in the form (6). Then R has its (Λ-weighted) FDR controlled at level $\Pi(\mathcal{H}_0)\alpha$.*

Theorem 4.2 can be seen as an extension to the FDR of a celebrated inequality due to Hommel (1983) for the family-wise error rate (FWER), which has been widely used in the multiple testing literature (see, e.g., Lehmann and Romano, 2005; Romano and Shaikh, 2006a,b). Namely, when ν has its support in $\{1, \ldots, m\}$ and $\mathcal{H} = \mathcal{H}_0$, the above result recovers Hommel's inequality. Note that the latter special 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 4.2 generalizes without modification to a possibly continuous hypothesis space, as will be mentioned in Section 4.4. The result of Theorem 4.2 initially appeared in a paper of Blanchard and Fleuret (2007), in a somewhat different setting.

4.2.1. Discussion of the family of new shape functions

Theorem 4.2 establishes that, under arbitrary dependencies between the p-values, there exists a family of step-up procedures with controlled false discovery rate. This family is parametrized by the free choice of a distribution ν on the positive real line, which determines the shape function β_ν.

In the remaining of Section 4.2, we assume $\mathcal{H}$ to be finite of cardinal m, endowed with the standard Λ weighting, i.e., the counting measure. In this situation, let us first remark that it is always preferable to choose ν with support in $\{1, \ldots, m\}$. To see this, notice that only the values of β at integer values $k, 1 \leq k \leq m$ matter for the output of the algorithm. Replacing an arbitrary distribution ν by the discretized distribution $\nu'(\{k\}) = \nu((k-1, k])$ for $k < m$ and $\nu'(\{m\}) = \nu((m-1, +\infty))$ results in a shape function β' which is larger than β on the relevant integer range, hence the associated step-up procedure is more powerful. This discretization operation will however generally result in minute improvements only; sometimes continuous distributions can be easier to handle and avoid cumbersomeness in theoretical considerations.

Here are some simple possible choices for (discrete) ν based on power functions $\nu(\{k\}) \propto k^\gamma$, $\gamma \in \{-1, 0, 1\}$:

- $\nu(\{k\}) = \gamma_m^{-1} k^{-1}$ for $k \in \{1, \ldots, m\}$ with the normalization constant $\gamma_m = \sum_{1 \leq i \leq m} \frac{1}{i}$. This yields $\beta(r) = \gamma_m^{-1} r$, and we recover the distribution-free procedure of Benjamini and Yekutieli (2001).
- ν is the uniform on $\{1, \ldots, m\}$, giving rise to the quadratic shape function $\beta(r) = r(r+1)/2m$. The obtained step-up procedure was proposed by Sarkar (2008).
- $\nu(\{k\}) = 2k/(m(m+1))$ for $k \in \{1, \ldots, m\}$ leads to $\beta(r) = r(r+1)(2r+1)/(3m(m+1))$.

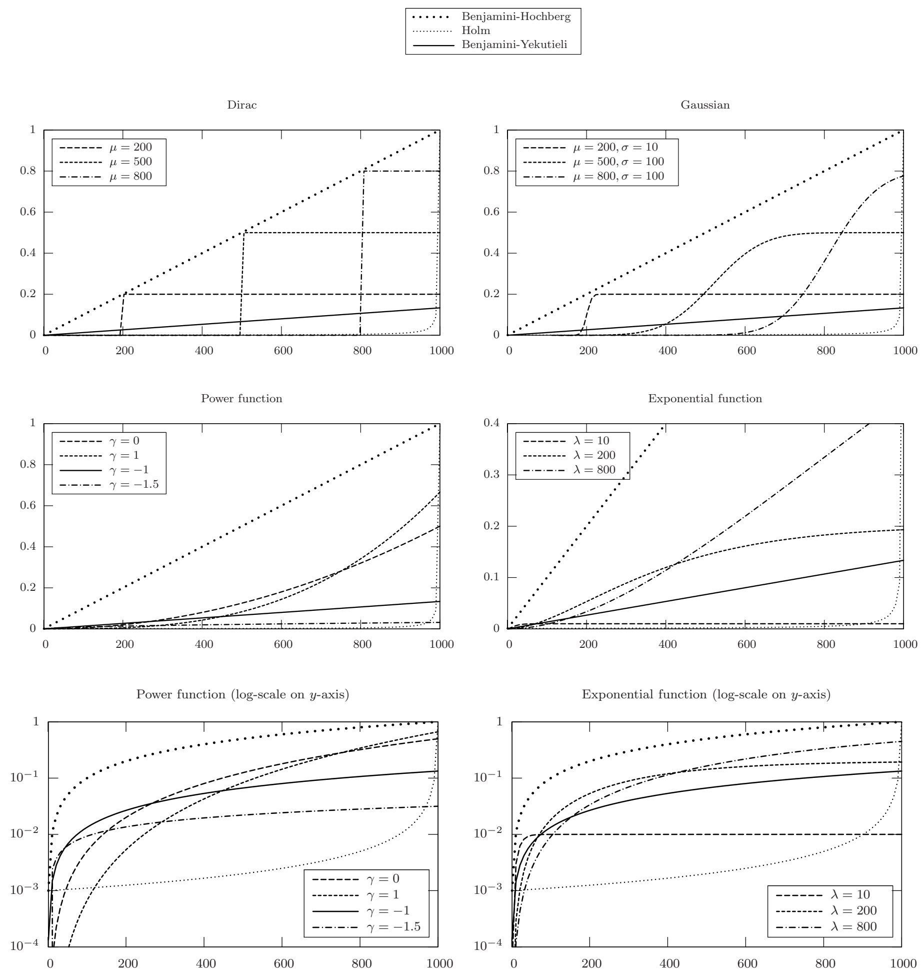

Fig 2. *For the standard* Λ*-weighting and* $m = 1000$ *hypotheses, this figure shows several (normalized) shape functions* $m^{-1}\beta$ *associated to different distributions* ν *on* $\mathbb{R}^+$ *(according to expression* (6)*): Dirac distribution:* $\nu = \delta_\mu$*, with* $\mu > 0$*. (Truncated-) Gaussian distribution:* ν *is the distribution of* $\max(X, 1)$*, where* $X \sim \mathcal{N}(\mu, \sigma^2)$*. Power distribution:* $d\nu(r) = r^\gamma \mathbf{1}\{r \in [1, m]\}dr / \int_1^m u^\gamma du$*,* $\gamma \in \mathbb{R}$*. (Truncated-) Exponential distribution:* $d\nu(r) = (1/\lambda)\exp(-r/\lambda)\mathbf{1}\{r \in [0, m]\}dr$*, with* $\lambda > 0$*. On each graph, for comparison purposes we added the threshold function for Holm's step-down* $m^{-1}\beta(x) = 1/(m - x + 1)$*, (small dots), and the linear thresholds* $\beta(x) = x$ *(large dots) and* $\beta(x) = (\sum_{i \le m} i^{-1})^{-1} x$ *(solid – also corresponding to the power distribution with* $\gamma = -1$*), corresponding to the standard linear step-up and to the distribution-free linear step-up of* Benjamini and Yekutieli (2001)*, respectively.*

On Figure 2, we plotted the shape functions corresponding to different choices of distributions ν (which are actually continuous, *i.e.*, without applying the discretization procedure mentioned above). It is clear that the choice of ν has a large impact on the final number of rejections of the procedure. However, since no shape function uniformly dominates the others, there is no universally optimal choice of ν: the respective performances of these different procedures will depend on the exact distribution P, and in particular on the number of non-true hypotheses.

We like to think of ν as a kind of "prior" on the possible volumes of rejected hypotheses. If we expect to have only a few rejected hypotheses, ν should be concentrated on small values, and more spread out if we expect a significant rejection proportion. This intuition is in accordance with a case of equality in Hommel's inequality established by Lehmann and Romano (2005, Lemma 3.1 (ii)). In the situation studied there (a specifically crafted distribution P), it can be checked that the distribution of the cardinality of the step-up procedure R using the shape function β_ν, conditionally to $R \neq \emptyset$, is precisely ν in our notation, while $\mathrm{FDR}(R)$ is exactly α.

As mentioned previously in Section 3.2.3, for any choice of ν, the shape function β_ν is always upper bounded by the linear shape function $\beta(x) = x$. The only cases of equality are attained if ν is equal to a Dirac measure δ_{x_0} in a point $x_0 \in \{1, \ldots, m\}$: in this case $\beta_{\delta_{x_0}}(x_0) = x_0$ but $\beta_{\delta_{x_0}}(x) < x$ for any $x \neq x_0$. Therefore, these procedures always reject less (or at most as many) hypotheses than the linear step-up. Admittedly, this probably limits the practical implications of this result, as we expect practitioners to prefer using the standard linear step-up even if the theoretical conditions for its validity cannot be formally checked in general. Additional conservativeness is the "price to pay" for validity under arbitrary dependencies, although the above result shows that one has, so to say, the choice in the way this price is to be paid.

Finally, from the examples of shape functions drawn on Figure 2, the shape functions based on *exponential* distributions ν seem particularly interesting; they appear to exhibit a qualitatively diverse range of possible shape functions, offering more flexibility than the Benjamini-Yekutieli procedure while not being as committed as the Dirac distributions to a specific prior belief on the number of rejected hypotheses.

4.2.2. Comparison to Bonferroni's and Holm's procedures

Observe that Bonferroni's procedure also belongs to the family presented here (taking $\nu = \delta_1$) – in the sense that a single-step procedure using a fixed threshold can be technically considered as a step-up procedure. It is well-known, however, that its control on type I error is much stronger than bounded FDR, namely bounded FWER. To this extent, it is worth considering the question of whether other rejections functions in the family – for which only the FDR is controlled – are of interest at all As remarked earlier, no shape function in the family can *uniformly* dominate the others, and consequently there exist particular situations where Bonferroni's procedure can be more powerful (i.e. reject more hypotheses) than other members of the family. However, this case appears only when there is indeed a very small number of rejections (i.e., when the signal is extremely "sparse"). For instance, comparing the three examples mentioned above to Bonferroni asymptotically as $m \to \infty$, we see that the corresponding step-up procedures have a rejection function larger than Bonferroni's threshold — and are therefore *a posteriori* more powerful than Bonferroni — provided their number of rejections $|R|$ is larger than:

- $\Theta(\log m)$ for $\nu(k) \propto k^{-1}$ (Benjamini-Yekutieli procedure);
- $\Theta(\sqrt{m})$ for ν uniform;
- $\Theta\left((m)^{\frac{2}{3}}\right)$ for $\nu(k) \propto k$.

(Recalling here that $\Theta()$ means asymptotic order of magnitude, in other terms "asymptotically lower and upper bounded, up to a constant factor".) In each of the above cases, the largest proportion $u_m = |R|/m$ of rejections for which Bonferroni's procedure would *a posteriori* have been more powerful tends to zero as $m \to \infty$. An identical conclusion will hold if we compare these rejection functions to that of Holm's step-down (Holm, 1979), since the latter is equivalent to Bonferroni when $u_m \to 0$ (in addition, Holm's procedure is step-down while the above procedures are step-up).

More generally, let us exhibit a generic family of shape functions β such that u_m tends to zero as $m \to \infty$. We first define the proportion u_m for a given shape function β more formally, as $u_m = r_m/m$, where r_m is the first point of $\{1, \ldots, m\}$ for which $\beta(\cdot)$ is above 1 (Bonferroni's shape function). Introduce the family of *scale invariant* shape functions β, that is, the βs that can be rewritten under the form $\beta(r) = m\widetilde{\beta}(\frac{r}{m})$ for some *fixed* function $\widetilde{\beta}(u) = \int_0^u v d\widetilde{\nu}(v)$ and *fixed* probability measure $\widetilde{\nu}$ on $(0,1]$. In the latter, $\widetilde{\nu}$ should be taken independently of m as a "prior" on the proportion of rejections. (Equivalently, $\widetilde{\nu}$ takes the role of ν if we consider the following alternate scaling of the standard Λ-π weighting: Λ is the uniform probability measure on $\mathcal{H}$ and $\pi \equiv 1$.) It is then straightforward to check that u_m tends to 0 as $m \to \infty$ if we choose $\widetilde{\nu}$ such that $\widetilde{\beta}(u) > 0$ for all $u > 0$ (i.e. the origin is an accumulation point of the support of $\widetilde{\nu}$). This gives many examples of shape functions which outperform Bonferroni's and Holm's procedures as m grows to infinity in the "non-sparse" case. For example, the "power function" choice $d\widetilde{\nu}(u) = \mathbf{1}\{u \in [0,1]\}(\gamma+1)u^{\gamma}dx$ for $\gamma > -1$ gives rise to the rescaled shape function $\widetilde{\beta}(u) = \frac{\gamma+1}{\gamma+2}u^{\gamma+2}$ and thus $\beta(r) = \frac{\gamma+1}{\gamma+2}\frac{r^{\gamma+2}}{m^{\gamma+1}}$. In the cases $\gamma = 0, 1$, note that the latter corresponds to the functions β considered earlier (up to discretization).

By contrast, one can easily check that there is no scale-invariant *linear* rejection function satisfying (6): the Benjamini-Yekutieli procedure would correspond (up to lower order terms introduced by discretization) to the "truncated" prior $d\widetilde{\nu}(u) = (\log m)^{-1}\mathbf{1}\{m^{-1} \le u \le 1\}x^{-1}du$, which cannot be extended to the origin independently of m since $u \mapsto u^{-1}$ is not integrable in 0. We have seen above that $u_m \to 0$ nevertheless also holds for this procedure: hence scale-invariant shape functions are certainly not the only candidates in the family to asymptotically outperform Bonferroni's and Holm's procedures in the "non-sparse" case.

For comparison w.r.t. several other possible choices of ν, (and for a finite $m = 1000$) we have systematically added Holm's rejection function on the plots of Figure 2. This leads to a qualitatively similar conclusion.

4.3. Adaptive step-up procedures

We now give a very simple application of our results in the framework of *adaptive* step-up procedures. Observe that the FDR control obtained for classical step-up procedures is in fact not at the target level α , but rather at the level $\pi_0\alpha$, where $\pi_0 = \Pi(\mathcal{H}_0)$ is the "weighted volume" of the set of true null hypotheses (equal to the proportion of true null hypotheses m_0/m in the standard case). This motivates the idea of first estimating π_0^{-1} from the data using some estimator $G(\mathbf{p})$, then applying the step-up procedure with the modified shape function $\widetilde{\beta} = G(\mathbf{p})\beta$. Because this function is now data-dependent, establishing FDR control for the resulting procedure is more delicate; it is the subject of numerous recent works (see, e.g., Black, 2004; Benjamini et al., 2006; Finner et al., 2008; see also Gavrilov et al., 2008 for an adaptive step-down procedure).

In this context we prove the following simple result, which is valid under the different types of dependency conditions:

Lemma 4.3. *Assume either of the following conditions is satisfied:*

- *the p-values* $(p_h, h \in \mathcal{H})$ *are PRDS on* $\mathcal{H}_0$, β *is the identity function.*
- *the p-values have unspecified dependencies and* β *is a function of the form* (6).

Define R *as an adaptive step-up procedure using the data-dependent threshold collection* $\Delta(h, r, \mathbf{p}) = \alpha_1\pi(h)G(\mathbf{p})\beta(r)$ *, where* $G(\mathbf{p})$ *is some estimator of* π_0^{-1} *, assumed to be nondecreasing as a function of the p-values. Then the following inequality holds:*

$$\mathrm{FDR}(R) \leq \alpha_1 + \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{G(\mathbf{p}) > \pi_0^{-1}\}\right] . \tag{8}$$

Proof. Consider $\widetilde{R}$ the modified step-up procedure using the data-dependent threshold collection $\alpha_1\pi(h)\max(\pi_0^{-1}, G(\mathbf{p}))\beta(r)$. Then it is easy to check that $\widetilde{R}$ satisfies the self-consistency condition $\mathbf{SC}(\alpha_1\pi_0^{-1}, \pi, \beta)$. Furthermore, $\widetilde{R}$ is a nondecreasing set as a function of the p-values, by the hypothesis on G . Therefore, by combining Proposition 2.7 with Proposition 3.6 (resp. Proposition 3.7), $\widetilde{R}$ has its FDR controlled at level $\pi_0(\alpha_1\pi_0^{-1}) = \alpha_1$ in both dependency situations and we have

$$\begin{aligned}
\mathrm{FDR}(R) &= \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{|R| > 0\}\right] \\
&\leq \mathbb{E}\left[\frac{|\widetilde{R} \cap \mathcal{H}_0|}{|\widetilde{R}|}\mathbf{1}\{|\widetilde{R}| > 0\}\right] + \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{G(\mathbf{p}) > \pi_0^{-1}\}\right] \\
&\leq \alpha_1 + \mathbb{E}\left[\frac{|R \cap \mathcal{H}_0|}{|R|}\mathbf{1}\{G(\mathbf{p}) > \pi_0^{-1}\}\right] .
\end{aligned}$$

□

Incidentally, the above proof illustrates a technical use of the main result where the inclusion in the self-consistency condition is generally not an equality.

We can apply Lemma 4.3 when considering a so-called *two-stage* procedure, where π_0 is estimated using a preliminary multiple testing procedure R_0. We assume here that this first stage has controlled FWER (e.g. Holm's step-down).

Corollary 4.4. *Let R_0 be a multiple testing procedure with $FWER(R_0) := \mathbb{P}(\mathcal{H}_0 \cap R_0 \neq \emptyset)$ controlled at level α_0. Estimate π_0 by $\widehat{\pi}_0 = \Pi((R_0)^c) = \sum_{h \notin R_0} \pi(h)\Lambda(\{h\})$ the π-volume of hypotheses non rejected by the first stage, and put $G(\mathbf{p}) = \widehat{\pi}_0^{-1}$ (defined as $+\infty$ when $\widehat{\pi}_0 = 0$).*

Then the adaptive step-up procedure R using the data-dependent threshold collection $\Delta(h, r, \mathbf{p}) = \alpha_1 \pi(h) G(\mathbf{p}) \beta(r)$ satisfies

$$\mathrm{FDR}(R) \leq \alpha_0 + \alpha_1 .$$

The proof is a direct application of Lemma 4.3: the second term in (8) is upper bounded by $\mathbb{P}(G(\mathbf{p}) > \pi_0^{-1}) = \mathbb{P}(\Pi((R_0)^c) < \Pi(\mathcal{H}_0))$, which is itself smaller than or equal to $\mathbb{P}(\mathcal{H}_0 \cap R_0 \neq \emptyset)$, the FWER of the first stage. Note that in the standard situation where $\Lambda = |.|$ is the counting measure and π is uniform, the above estimator of $\pi_0^{-1} = m/m_0$ is simply $m/\widehat{m}_0$, where $\widehat{m}_0$ is the number of non rejected hypotheses by the first stage.

Because of the loss in the level introduced by the first stage, the latter result is admittedly not extremely sharp: for example, putting $\alpha_0 = \alpha_1 = \alpha/2$, a theoretical improvement over the non-adaptive version at level α is obtained only when more than 50% of hypotheses are rejected in the first stage. However, while sharper results are available under the assumption of independent p-values (see, e.g., Benjamini et al., 2006), up to our knowledge, there are almost no results addressing the case of dependent p-values (as is the case in the above result). The results we know of for this case are found in works of Sarkar (2008) and Farcomeni (2007). The latter reference establishes a result similar to the above one, but seems to make the implicit assumption that the two stages are independent, which we are not assuming here. A more extensive treatment of the question of adaptive procedures when following the general principles exposed in the present work, including other applications of Lemma 4.3, is proposed by Blanchard and Roquain (2008a) (see also the second author's PhD thesis, Roquain, 2007, Chap. 11).

4.4. FDR control over a continuous space of hypotheses

An interesting feature of the approach advocated here for proving FDR control is that it can be readily adapted to the case where $\mathcal{H}$ is a continuous set of hypotheses. A simple example where this situation arises theoretically is when when the underlying observation is modelled as a random process W over a continuous space $\mathcal{T}$, and the goal is to test for each $t \in \mathcal{T}$ whether $\mathbb{E}[W(t)] = 0$. In this case we can identify $\mathcal{H}$ to $\mathcal{T}$. Such a setting was considered for example by Perone Pacifico et al. (2004).

In order to avoid straying too far from our main message in the present work, it was decided to postpone the detailed exposition of this point to a separate note. We refer the interested reader to the Section 5 of the technical report of Blanchard and Roquain (2008b), and restrict ourselves here to a brief overview. First, under appropriate (and tame) measurability assumptions, the framework developed in this paper carries over without change: in the FDR definition, instead of using the cardinality measure (which is of course not adapted in the continuous case), we are able to deal with an arbitrary "volume measure" Λ on $\mathcal{H}$ (such as the Lebesgue measure if $\mathcal{H}$ is a compact subset of $\mathbb{R}^d$). Also, while it seems considerably more difficult to define rigorously step-up procedures in the traditional sense via reordering of the p-values, Definition 3.1 of a step-up procedure carries over in a continuous setting.

Secondly, our main tool, Proposition 2.7, remains true when $\mathcal{H}$ is continuous, by replacing each sum over $\mathcal{H}$ by the corresponding integral (with respect to the measure Λ). Thirdly comes the question of how to adapt the three types of dependency conditions considered in Section 3.2 to a continuous setting. Under unspecified dependencies, there is nothing to change as our arguments are not specific to the discrete setting. The independent case, on the other hand, cannot be adapted to the continuous setting as it conflicts with some measurability assumptions. However, this setting is mainly irrelevant in a continuous setting as continuous families of independent random variables are not usually considered. Finally, in the case of positive dependencies, condition (7) still ensures the dependency control condition since Lemma 3.2 is valid for arbitrary variables, not necessarily discrete. The main difficulty is therefore to suitably adapt the PRDS assumption in the continuous setting. We propose two extensions of the PRDS condition, namely the "strong continuous PRDS", which is a direct adaptation of the finite PRDS definition to a continuous setting, and the "weak continuous PRDS", which states that any finite subfamily of p-values should be (finite) PRDS. The strong continuous PRDS condition is sufficient but arguably possibly not easy to check, while the weak PRDS condition is easier but requires some additional requirements on the procedure R to ensure condition **DC**. An example of a process satisfying either type of condition is a continuous Gaussian process with a positive covariance operator.

4.5. Other types of procedures

We want to point out that the approach advocated here also provides FDR control for procedures more general than step-up. For example, as mentioned at the end of Section 3.1, generalized step-up-down procedures satisfy a self-consistency property. Therefore, combining Proposition 2.7 with Proposition 3.6 (PRDS case) and Proposition 3.7 (unspecified dependencies), we obtain the following result:

Theorem 4.5. *Assume either of the following conditions is satisfied:*

- *the p-values $(p_h, h \in \mathcal{H})$ are PRDS on $\mathcal{H}_0$, β is the identity function.*
- *the p-values have unspecified dependencies and β is of the form (6).*

Then the generalized step-up-down procedure of any order $\lambda \in [0, |\mathcal{H}|]$ *and associated to the threshold collection* $\Delta(h,r) = \alpha\pi(h)\beta(r)$ *has its FDR controlled at level* $\alpha\Pi(\mathcal{H}_0)$ *.*

In the PRDS case and with the standard Γ - π weighting, the first point of the above result has been first proved by Sarkar (2002) (see also Finner et al., 2008, where an approach related to ours is used to prove the same result; this is discussed in more detail below in Section 5.1). The contribution of the above result is to deal with possible Γ - π weighting and with the general dependent case (in particular, note that this theorem contains both Theorem 4.1 and Theorem 4.2). We emphasize that the latter result does not come trivially from the fact that a step-up-down procedure is always a subset of the step-up procedure using the same threshold collection, because in the FDR expression the numerator and the denominator inside the expectation *both* decrease with the rejection set size.

It could however legitimately be objected that only step-up procedures are really of interest in the present context, since they are less conservative than step-up-down, and even the less conservative possible under the **SC** condition, as argued in Section 3.1. But one interest of the self-consistency condition is to allow more flexibility, in particular if there are *additional* constraints to be taken into account. Consider the following plausible scenario: in a medical imaging context, the user wants to enforce additional geometrical constraints on the set R of rejected hypotheses, represented as a 2D set of pixels. For example, one could demand that R be convex or have only a limited number of connected components. If such additional constraints come into play, the step-up may not be admissible, and has to be replaced by a subset satisfying the constraints. In this case, the flexibility introduced by the **SC** condition will be useful in order to give a simple criterion sufficient to establish FDR control without necessarily having to engineer a new proof for each new specific algorithm. Note in particular that in such a scenario, one would probably like to choose a maximal rejection set satisfying both the geometric constraints and self-consistency condition; in this case the resulting procedure cannot be characterized in general as a step-up-down procedure, and the **SC** condition might hold without equality, i.e. $R \subsetneq L_\Delta(|R|)$.

4.6. Another application of condition DC(β)

In this section, we step outside of the framework used in Proposition 2.7; more precisely, we present another application of condition **DC**(β) to study the FDR of a step-down procedure that does not satisfy the self-consistency condition with respect to the adequate shape function. We will prove that the step-down procedure proposed by Benjamini and Liu (1999a) and Romano and Shaikh (2006a) has a controlled FDR under a PRDS-type assumption of $\mathcal{H}_0$ on $\mathcal{H}_1$; we also deduce a straightforward generalization to the unspecified dependencies case. In this section, we only consider Λ equal to the counting measure, so that the aim is to control the standard FDR.

Benjamini and Liu (1999a) and Romano and Shaikh (2006a) introduced the step-down procedure based on the threshold collection $\Delta(i) = \frac{\alpha m}{(m-i+1)^2}$, showed that it has controlled FDR at level α if for each $h_0 \in \mathcal{H}_0$, p_{h_0} is independent of the collection of p-values $(p_h, h \in \mathcal{H}_1)$ (in fact Romano and Shaikh, 2006a used a slightly weaker assumption, but it reduces to independence when the p-values of true null hypotheses are uniform on $[0,1]$). Here, we prove this result under a weaker assumption, namely a positive regression depency assumption of p-values of $\mathcal{H}_1$ from those of $\mathcal{H}_0$. Let us reformulate slightly the notion of "PRDS on $\mathcal{H}_0$" given in Definition 3.5. We say that the p-values of $(p_h, h \in \mathcal{H}_1)$ are positively regression dependent from each one in a separate set $\mathcal{H}_0$ (for short: $\mathcal{H}_1$ PRDSS on $\mathcal{H}_0$) if for any measurable nondecreasing set $D \subset [0,1]^{\mathcal{H}_1}$ and for all $h_0 \in \mathcal{H}_0$, the function

$$u \mapsto \mathbb{P}\big((p_h)_{h\in\mathcal{H}_1} \in D \mid p_{h_0} = u\big)$$

is nondecreasing. Note that the latter condition is obviously satisfied when for all $h_0 \in \mathcal{H}_0$, p_{h_0} is independent of $(p_h, h \in \mathcal{H}_1)$. We chose to introduce a new acronym only to emphasize the fact that, contrarily to the standard PRDS , this assumption does not put constraints on the inner dependency structure of the p-value vector of true hypotheses.

Theorem 4.6. *Suppose that the p-values of $\mathcal{H}_1$ are PRDSS on $\mathcal{H}_0$. 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 (6), 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 a FDR less than or equal to α .

The proof is found in appendix. Essentially, we followed the proof of Benjamini and Liu (1999a) and identified the point where the condition **DC**(β) (along with the results of Lemma 3.2) can be used instead of their argument.

Benjamini and Liu (1999b) proposed a slightly less conservative step-down procedure: the step-down procedure with the threshold collection $\Delta(i) = 1-\big[1-\min\big(1, \frac{\alpha m}{m-i+1}\big)\big]^{1/(m-i+1)}$. It was proved by Benjamini and Liu (1999b) 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 by Sarkar (2002) 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 PRDSS assumption of Theorem 4.6.

The procedure of Theorem 4.6 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 larger on a substantial range. However, in some specific cases (m small and large number of rejections), the threshold collection of Theorem 4.6 can be larger than the one of the LSU procedure. A similar argument can be made when comparing the proposed modified step-down under unspecified dependencies to (for example) the modified LSU procedure of Benjamini and Yekutieli (2001).

In order to use Theorem 4.6 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 yields $\Delta(i) = \alpha\frac{1}{m-i+1}\left(\frac{1}{m-i+1} + \cdots + \frac{1}{m}\right)$,
- taking $\nu\left(\frac{1}{k}\right) \propto k$ results in the threshold function $\Delta(i) = \frac{\alpha}{m+1}\frac{2i}{(m-i+1)}$,
- taking $\nu\left(\frac{1}{k}\right) \propto \frac{1}{k}$ results in $\Delta(i)$ equal to $\gamma_m^{-1}\alpha\frac{m}{m-i+1}\left(\frac{1}{(m-i+1)^2} + \cdots + \frac{1}{m^2}\right) \simeq \gamma_m^{-1}\alpha\frac{i}{(m-i+1)^2}$, with $\gamma_m = \sum_{i\leq m}\frac{1}{i}$.

5. Discussion and conclusion

5.1. The self-consistency condition and connection with other works

The self-consistency condition with a linear shape function can be related to the following heuristic motivation: consider the problem of choosing a threshold for rejected p-values, which we reformulate equivalently as choosing r such that $L_\Delta(r)$ has a FDR smaller than α (for the linear threshold collection $\Delta(h,r) = \alpha r/m$). If the final number of rejections $|L_\Delta(r)|$ was equal to a *deterministic* constant $C(r)$, we would have a FDR bounded by

$$\mathbb{E}\left[\,|\{h \in \mathcal{H}_0 \mid p_h \leq \alpha r/m\}|\,\right]/C(r) \leq \alpha r/C(r)\,,$$

so that the desired FDR control would be attained if $r \leq C(r) = |L_\Delta(r)|$, that is, when $L_\Delta(r)$ satisfies the self-consistency condition. This reasoning is, of course, unrigorous since $L_\Delta(r)$ is in fact a random variable (and we need other arguments to correctly prove the FDR control, e.g. Lemma 3.2). This point of view is in the same spirit as the *post-hoc* interpretation of the classical linear step-up procedure proposed in Section 3.3 of Benjamini and Hochberg (1995), where the authors remarked that the linear step-up procedure maximizes the number of rejected hypotheses under the above constraint, which is the property we used in Definition 3.1.

As mentioned in the introduction and in Section 4.5, the forthcoming paper of Finner et al. (2008) introduces a condition quite similar to the self-consistency condition (although formulated differently). Precisely, condition (T2) of Finner et al. (2008) can be seen to be equivalent to $R = L_\Delta(|R|)$ in our notation (in the specific case of a linear threshold collection Δ and for the standard Λ-π weighting). It is proved in Theorem 4.1 of Finner et al. (2008) that (T2) implies FDR control in the PRDS case (or more precisely, when (7) holds). The authors note that the corresponding proof unifies and simplifies classical results and proofs. The present work, developed independently, led to a very similar conclusion. In particular, Finner et al. (2008) note that their result covers in general the step-up-down procedures satisfying (T2), which is essentially the same as the first point of the present Theorem 4.5 (for the standard Λ and π-weighting).

As an additional contribution, we introduced the "abstract" dependency condition **DC**, which allowed us to increase the range where the self-consistency condition can be used, in particular when the p-values have unspecified dependencies. We also included Λ and π-weighting in our results; the formulation we adopted allows in particular for an easy extension to infinite, possibly continuous hypothesis spaces. Other original applications were exposed in Section 4.

Conversely, Finner et al. (2008) used their approach for different applications of interest, based on an asymptotically optimal rejection curve. Several step-up or step-up-down procedures are proposed by Finner et al. (2008) based on variations on this rejection curve and shown to have a an asymptotic and adaptive (in the sense of Section 4.3) control of the FDR (related to this is also the step-down procedure of Gavrilov et al. (2008), based on the same curve and shown to enjoy non-asymptotic control of the FDR). These results do not fit directly into the framework delineated in the present paper, but some of the technical tools used in their proof are of a similar spirit. A full technical development on this topic is out of the scope of the present work, but we demonstrate in a separate work (Blanchard and Roquain, 2008a) that the two conditions we presented here (along with some additional key ideas coming from Benjamini et al., 2006) can be used to prove (non-asymptotic) FDR control under independence, for an adaptive procedure based on a rejection curve analogous to that considered by Finner et al. (2008) and Gavrilov et al. (2008). To this regard, let us also mention the recent work of Neuvial (2008), which compares a number of these related procedures in terms of their asymptotical power.

Finally, we mention that the self-consistency condition presented here has a slightly weaker form than condition (T2) of Finner et al. (2008), namely it is $R \subset L_\Delta(|R|)$ instead of $R = L_\Delta(|R|)$. From a technical point of view, we note here that the argument of Finner et al. (2008) can actually be adapted straightforwardly to accomodate the weaker condition. Is the weaker form of the condition of interest at all? While the stricter condition is sufficient to cover the case of step-up and step-up-down procedures, in the present work we have also tried to demonstrate that the weaker form is not purely anectodical but useful in some other applications: first for truncated threshold collections (proof of Lemma 4.3), and secondly in Section 4.5 where we mentioned plausible practical scenarios where equality might not hold due to additional constraints.

5.2. Conclusion

The approach advocated in this paper to establish FDR bounds introduced 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, the (essentially probabilistic) dependency control condition. The two conditions are effectively coupled via the common choice of the shape function β appearing in both. The fundamental result of this paper is that these two conditions suffice for FDR control, but part of our message is that this point of view also introduced some relevant technical tools, which, abstracting some key arguments present in previous works, can be of use in various other settings.

While these conditions are only sufficient and hence certainly not universal, we illustrated their interest by recovering in Sections 4.1, 4.2 and 4.5 several existing results of the FDR multiple testing literature in an unified way, as 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 (6).
- in the procedure algorithm, arbitrary choice between "step-down" and "step-up", "step-up-down", and more generally arbitrary choice among the possible orders λ in a "step-up-down" procedure.

In the past literature, many results have been established for specific combinations of the above variations; here we were 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). Several other applications were proposed.

An interesting direction for future work 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.

Appendix

Appendix A: Proof of Lemma 3.2

Part (i). We want to establish 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 $C^* = \inf\{g(u) \mid u \in \mathcal{U}\}$. It is not difficult to check that $u^* \leq cC^*$ (for instance take any nondecreasing 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

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

Part (ii). The proof uses a similar telescopic sum argument as developed by Benjamini and Yekutieli (2001) for proving FDR control of the linear step-up under the PRDS assumption; the goal of the lemma presented here is to isolate this argument in order to specifically concentrate on condition **DC**, and to extend it to arbitrary (non-discrete) variables.

We want to prove the inequality

$$\mathbb{E}\left[\frac{\mathbf{1}\{U \le cV\}}{V}\right] \le c$$

for U, V two nonnegative real variables such that U is stochastically lower bounded by a uniform distribution, and the conditional distribution of V given $U \le u$ 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 $1 \le i \le 2K+1$. The following chain of inequalities holds:

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

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

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

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

□

Appendix B: Proof of Theorem 4.6

To establish the first assertion of the Theorem, remember we assume the p-values of $\mathcal{H}_1$ are PRDSS on $\mathcal{H}_0$, and the threshold collection is $\Delta(i) = \alpha m/(m-i+1)^2$. Assume $m_0 > 0$ (otherwise the result is trivial) and consider $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ the ordered p-values of $(p_h, h \in \mathcal{H})$. 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 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

$$|R \cap \mathcal{H}_0| > 0 \Rightarrow \exists h \in \mathcal{H}_0 \, : p_h \leq \frac{\alpha m}{(m - |R_1|)^2} \leq \frac{\alpha m}{m_0(m - |R_1|)} \,. \tag{9}$$

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 (9) 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 that the couple $(p_h, (m-|R_1|)^{-1})$ satisfies condition **DC**(β) with $c = \alpha m/m_0$ and $\beta(x) = x$; this holds in the present case from part (ii) of Lemma 3.2 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 Proposition 3.6).

For the second part of the theorem, we follow exactly the same proof as above with the modified threshold function and part (iii) of Lemma 3.2 instead of part (ii). □

Acknowledgements

The authors wish to thank the anonymous reviewer and AE for their constructive criticism of previous versions of the manuscript, which allowed to improve the overall organization and focus of the paper.

References

Benjamini, Y. and Heller, R. (2007). False discovery rates for spatial signals. *J. Amer. Statist. Assoc.*, 102(480):1272–1281.

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

Benjamini, Y. and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Statist.*, 24(3):407–418. MR1481424

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507. MR2261438

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

Benjamini, Y. and Liu, W. (1999b). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Inference*, 82(1-2):163–170. MR1736441

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

Black, M. A. (2004). A note on the adaptive control of false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(2):297–304. MR2062377

Blanchard, G. and Fleuret, F. (2007). Occam's hammer. In Bshouty, N. and Gentile, C., editors, *Proceedings of the 20th. conference on learning theory (COLT 2007)*, volume 4539 of *Springer Lecture Notes on Computer Science*, pages 112–126. MR2397582

Blanchard, G. and Roquain, E. (2008a). Adaptive FDR control under independence and dependence. ArXiV preprint math.ST/0707.0536v2.

Blanchard, G. and Roquain, E. (2008b). Self-consistent multiple testing procedures. ArXiV preprint math.ST/0802.1406v1.

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics*, 34(2):275–297. MR2346640

Finner, H., Dickhaus, T., and Roters, M. (2008). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* To appear.

Gavrilov, Y., Benjamini, Y., and Sarkar, S. K. (2008). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.* To appear.

Genovese, C., Roeder, K., and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika*, 93(3):509–524. MR2261439

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

Hommel, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical J.*, 25(5):423–430. MR0735888

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.*, 37:1137–1153. MR0202228

Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *The Annals of Statistics*, 33:1138–1154. MR2195631

Neuvial, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. ArXiv preprint: math.ST/0803.2111v1.

Perone Pacifico, M., Genovese, C., Verdinelli, I., and Wasserman, L. (2004). False discovery control for random fields. *J. Amer. Statist. Assoc.*, 99(468):1002–1014. MR2109490

Romano, J. P. and Shaikh, A. M. (2006a). On stepdown control of the false discovery proportion. In Rojo, J., editor, *Optimality : the 2nd Lehmann Symposium*, volume 49 of *Lecture Notes-Monograph Series*, pages 33–50. Institute of Mathematical statistics. MR2337829

Romano, J. P. and Shaikh, A. M. (2006b). Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.*, 34(4):1850–1873. MR2283720

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

Roquain, E. and van de Wiel, M. (2008). Multi-weighting for fdr control. ArXiV preprint math.ST/0807.4081.

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

Sarkar, S. K. (2008). Two-stage stepup procedures controlling FDR. *Journal of Statistical Planning and Inference*, 138(4):1072–1084. MR2384506

Tamhane, A. C., Liu, W., and Dunnett, C. W. (1998). A generalized step-up-down multiple test procedure. *Canad. J. Statist.*, 26(2):353–363. MR1648451

Wasserman, L. and Roeder, K. (2006). Weighted hypothesis testing. Technical report, Dept. of statistics, Carnegie Mellon University. ArXiv preprint math/0604172v1.