Bernoulli **20**(1), 2014, 304–333
DOI: 10.3150/12-BEJ488

Testing over a continuum of null hypotheses with False Discovery Rate control

GILLES BLANCHARD[1], SYLVAIN DELATTRE[2] and ETIENNE ROQUAIN[3]

[1]*Universität Potsdam, Institut für Mathematik, Am Neuen Palais 10, 14469 Potsdam, Germany.*
E-mail: gilles.blanchard@math.uni-potsdam.de
[2]*Université Paris Diderot, LPMA, 4, Place Jussieu, 75252 Paris cedex 05, France.*
E-mail: sylvain.delattre@univ-paris-diderot.fr
[3]*UPMC Université Paris Diderot, LPMA, 4, Place Jussieu, 75252 Paris cedex 05, France.*
E-mail: etienne.roquain@upmc.fr

We consider statistical hypothesis testing simultaneously over a fairly general, possibly uncountably infinite, set of null hypotheses, under the assumption that a suitable single test (and corresponding p-value) is known for each individual hypothesis. We extend to this setting the notion of false discovery rate (FDR) as a measure of type I error. Our main result studies specific procedures based on the observation of the p-value process. Control of the FDR at a nominal level is ensured either under arbitrary dependence of p-values, or under the assumption that the finite dimensional distributions of the p-value process have positive correlations of a specific type (weak PRDS). Both cases generalize existing results established in the finite setting. Its interest is demonstrated in several non-parametric examples: testing the mean/signal in a Gaussian white noise model, testing the intensity of a Poisson process and testing the c.d.f. of i.i.d. random variables.

Keywords: continuous testing; false discovery rate; multiple testing; positive correlation; step-up; stochastic process

1. Introduction

1.1. Motivations

Multiple testing is a long-established topic in statistics which has seen a surge of interest in the past two decades. This renewed popularity is due to a growing range of applications (such as bioinformatics and medical imaging) enabled by modern computational possibilities, through which collecting, manipulating and processing massive amounts of data in very high dimension has become commonplace. Multiple testing is in essence a multiple decision problem: each individual test output is a yes/no (or accept/reject) decision about a particular question (or null hypothesis) concerning the generating distribution of some random observed data.

The standard framework for multiple testing is to consider a finite family of hypotheses and associated tests. However, in many cases of interest, it is natural to interpret the observed data as the discretization of an underlying continuously-indexed random process; each decision (test) is then associated to one of the discretization points. A first example is that of detecting unusually frequent words in DNA sequences: a classical model is to consider a Poisson model for the (non-overlapping) word occurrence process (Robin [13]), the observed data being interpreted as a discretized version of this process. A second example is given in the context of medical imaging, where the observed pixelized image can be interpreted as a sampled random process, and the decision to take is, for each pixel, whether the observed value is due to pure noise or reveals some relevant activity (pertaining to this setting, see in particular the work of Perone Pacifico *et al.* [11,12]; see also Schwartzman, Gavrilov and Adler [15]).

Therefore, the present paper explores multiple testing for a (possibly) uncountably infinite set of hypothesis. With some abuse of language, we will refer to this as the *continuous setting* and use loosely the word "continuous" in reference to sets in order to mean: possibly uncountably infinite. For the above examples, this corresponds to perform testing for the underlying continuously-indexed process in a direct manner, without explicit discretization.

1.2. Contribution and presentation of this work

The principal contributions of the present work are the following. We first define a precise, but general in scope, mathematical setting for multiple testing over a continuous set of hypotheses, taking particular attention to measurability issues. Specifically, we focus on the extension to continuously-indexed observation (and decision) processes of so-called *step-up* multiple testing procedures, and the control of the (continuous analogue of) their *false discovery rate (FDR)*, a type I error measure which has gained massive acceptance in the last 15 years for testing in high-throughput applications. To this end, we use the tools and analysis developed by (Blanchard and Roquain [6]) (a programmatic sketch of the present work can be found in Section 4.4 of the latter paper). In particular, we extend suitably to the continuous setting the notion of positive regressively dependent on a subset (PRDS) condition, which plays a crucial role in the analysis. The latter is a general type of dependence condition on the individual tests' p-values allowing to ensure FDR control. An important difference between the continuous and finite setting is that the continuous case precludes the possibility of independent p-values, which is the simplest reference setting considered in the finite case, so that a more general assumption on dependence structure is necessary (on this point, see the discussion at the end of Section 2.2).

We have tried as much as possible to make this work self-contained, and accessible to readers having little background knowledge in multiple testing. We begin in the next section with an extended informal discussion of the framework considered in this paper in relation to existing literature on non-parametric testing. Sections 2 and 3 of the paper introduce the necessary notions for multiple testing with an angle towards stochastic processes, and some specific examples for the introduced setting. The main result is exposed in Section 4, followed by its applications to the examples introduced in Section 2. The proof for the main theorem is found in Section 5. Extensions and discussions come in Section 6, while some technical results are deferred to Appendix and to the supplementary material (Blanchard, Delattre and Roquain [5]). Throughout the paper, the numbering of the sections and results of this supplement are preceded by "S-" for clarity (by writing, e.g., Section S-1).

1.3. Relationship to non-parametric statistics

Multiple testing over a continuous set of hypotheses has natural ties with non-parametric statistics. In the present section, we discuss this link and introduce informally our goals. We do not attempt a comprehensive survey of the very broad field of non-parametric testing, but rather emphasize some key specificities of the point of view adopted in the current work.

Non-parametric testing. In order to be more concrete, consider the classical white noise model $dZ_t = f(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, where B is a Wiener process, $t \in [0,1]$ (this model will be more formally considered as Example 2.3), with unknown drift function $f \in \mathcal{F}$, where $\mathcal{F}$ is some a priori smoothness class. The problem of testing various hypotheses about f against non-parametric alternatives has received considerable attention since the seminal work of Ingster [9,10]. The most common goal is to test one single qualitative null hypothesis, for instance: f is identically zero; f is non-negative; f is monotone; f is convex. For concreteness, consider the first possibility, denoted $H^0_* := \{f \equiv 0\}$. A common strategy to approach this goal is to consider a collection of test statistics of the form $T_\psi := \int \psi(t)\,\mathrm{d}Z_t$ for some well chosen family Ψ of test functions ψ (normalized so that $\|\psi\|_2 = 1$). For each individual test function ψ, we have $T_\psi \sim \mathcal{N}(\int f\psi, \sigma^2)$, so that this test statistic can be used for a Gauss test of the "local" null hypothesis $H^0_\psi := \{f \in \mathcal{F}; \int f\psi = 0\} \supset H^0_*$.

Intuitively, each statistic T_ψ will have power against a certain type of alternative. Taking into account simultaneously all test statistics T_ψ over $\psi \in \Psi$ now constitutes a multiple test problem. The simplest way to combine these is to consider $T_\Psi := \sup_{\psi\in\Psi} |T_\psi|$ and reject H^0_* whenever T_Ψ exceeds a certain quantile τ of its distribution under the null H^0_*. In multiple testing parlance, this is interpreted as testing over the hypothesis family $(H^0_\psi)_{\psi\in\Psi}$ with *weak* control of the *family-wise error rate* (FWER), which is defined as the probability to reject falsely one or more of the considered null hypotheses. Here a local hypothesis H^0_ψ is interpreted as rejected if $|T_\psi|$ exceeds τ. The qualifier "weak" refers to the fact that this probability is controlled at a nominal level only under the global null H^0_* rather than for all $f \in \mathcal{F}$.

This connection has been noted and discussed in the literature; for instance, Dümbgen and Spokoiny [8], using the above type construction and multiple testing interpretation, observe that "whenever the null hypothesis H^0_* is rejected, we have some information about where this violation occurs". To formalize this idea more precisely, define the rejection set $R_\tau := \{\psi \in \Psi\colon |T_\psi| > \tau\}$ as the set of (indices of) hypotheses from the family $(H^0_\psi)_{\psi\in\Psi}$ which are deemed false. In order for R_τ to be interpretable as intended, the threshold τ should now be chosen so that for *any* $f \in \mathcal{F}$, the probability that the rejection set has non-empty intersection with (indices of) the set of hypotheses satisfied by f, $\{\psi \in \Psi\colon f \in H^0_\psi\}$, is less than a nominal level. This is called *strong* control of the FWER. This point of view appears to have been only seldom considered explicitly in non-parametric testing literature; a recent example is the work of Schmidt-Hieber, Munk and Duembgen [14] (in the framework of density deconvolution), wherein each individual null hypothesis H^0_ψ has a qualitative interpretation in terms of f being monotone on some subinterval.

Type I *error criteria*. For adequate (weak or strong) control of the FWER in the example above, it is clear that a stochastic control of the deviations of the supremum statistic T_Ψ is necessary; this, in turn, depends on the complexity of the set Ψ (typically measured through L_2 metric entropy). As a consequence, FWER-controlling procedures will be more conservative as the complexity of the family of the underlying test increases; for instance, in a d-dimensional version of the above example, or if we simply consider a longer observation interval, the threshold τ would have to be more conservative (larger) to maintain FWER at the same level.

Motivated by the multiple testing point of view, we consider alternative, less stringent type I error criteria. Let us still consider the white noise example, in a slightly modified setup where we are specifically interested in the family of null hypotheses $(H_t^0)_{t\in[0,1]}$ with $H_t^0 := \{f \in \mathcal{F};\ f(t) = 0\}$. Each individual null hypothesis H_t^0 can be tested using a statistic T_{ψ_t} as defined above; more precisely, this statistic can test for the null $H_{\psi_t}^{\delta} := \{f \in \mathcal{F}\colon \int f\psi_t \in [-\delta, \delta]\}$ (ψ_t and δ being chosen adequately so that $H_t^0 \subset H_{\psi_t}^{\delta}$ holds). Define similarly to earlier $R_\tau := \{t \in [0,1]\colon |T_{\psi_t}| > \tau\}$, and introduce $F_\tau(f) := R_\tau \cap \{t \in [0,1]\colon f \in H_t^0\}$ the set of incorrectly rejected hypotheses when the true drift is f. To reiterate, FWER control is the requirement that $\mathbb{P}_f[F_\tau(f) \neq \varnothing]$ is bounded at a nominal level for all $f \in \mathcal{F}$. Consider now the weaker requirement that the average size of $F_\tau(f)$ (measured through its Lebesgue measure, denoted $|\cdot|$) is bounded at some nominal level. We observe via an application of Fubini's theorem:

$$\mathbb{E}_f\big[\big|F_\tau(f)\big|\big] = \int_0^1 \mathbb{P}_f\big[|T_{\psi_t}| > \tau\big]\mathbf{1}\big\{f \in H_t^0\big\}\,\mathrm{d}t;$$

the averaged false reject size is simply the individual test error level, integrated over the null hypothesis family. It is the continuous analogue of the so-called *per-comparison error rate* (PCER) in multiple testing. To control this quantity, clearly no multiple testing correction is necessary, and it is sufficient to choose τ so that each individual test has Type I error controlled at the desired level. This criterion, however, is not very useful in practice: only if the volume the rejection set is much larger than the nominal expected volume of errors can we have some trust that the rejection set contains interesting information. To address this issue, introduce the average *volume proportion* of falsely rejected hypotheses:

$$\mathrm{FDR}(R_\tau, f) := \mathbb{E}_f\big[\mathrm{FDP}(R_\tau, f)\big] \qquad \text{where } \mathrm{FDP}(R_\tau, f) = \frac{|F_\tau(f)|}{|R_\tau|}, \tag{1}$$

with the convention $\frac{0}{0} = 1$. The acronyms FDP and FDR stand for *false discovery proportion* and *rate*, respectively, and the above are the continuous generalization of corresponding criteria introduced for finite hypothesis spaces by Benjamini and Hochberg [1] for a finite number of hypotheses, and which have gained widespread acceptance since. Controlling the FDR is a more difficult task than for the PCER, because of the random denominator $|R_\tau|$ inside the expectation, and is the main aim of this paper.

As we shall see, a crucial difference of the FDR criterion with respect to the FWER is that, if a family of tests of the individual hypotheses is known, there exist relatively generic procedures called *step-up* to combine individual tests into a FDR-controlled multiple testing, by finding an adequate, data-dependent rejection threshold. In particular, these procedures do not depend on an intrinsic complexity measure of the initial family, nor of the control of deviations of suprema of statistics.

To conclude these considerations, it is worth noting that a FDR-controlled procedure can also be used for the goal of testing the single "global null" hypothesis $H_*^0 = \bigcap_{t\in[0,1]} H_t^0$, as in the opening discussion. Namely, if R_τ is a procedure whose FDR is controlled at level α, we can reject the global null hypothesis H_*^0 whenever $|R_\tau| > 0$. Under the global null, the FDP takes only the values 0 or 1 and precisely coincides with $\mathbf{1}\{|R_\tau| > 0\}$; thus, its expectation is the probability of type I error for testing the global null this way, and is bounded by α. That this can be achieved by a generic procedure without explicitly considering the deviations of a supremum process can seem surprising at first. Since the focus of this paper is centered on the multiple hypothesis testing point of view, we will not elaborate on this issue further, although a power comparison with the "standard" approaches would certainly be of interest.

Related work. The continuous FDR criterion using volume ratios was introduced before by Perone Pacifico *et al.* [11,12] to test non-negativity of the mean of a Gaussian field. In that work, the authors follow a two-step approach where the first step consists in a FWER-controlled multiple testing based on suprema statistics, as delineated above. This is then used to define an upper bound on the FDP holding with large probability, following the principle of so-called *augmentation procedures* (van der Laan, Dudoit and Pollard [17]) for multiple tests. An advantage of this approach is that a control of the FDP holding with large probability is obtained (which is stronger than a bound on the FDR, its expectation); on the other hand, the authors observe that since the first step is inherently based on FWER control, it is more conservative than a step-up procedure. In the present work, we focus on step-up procedures, for which a probabilistic control of the deviations of suprema statistics is not needed; this allows us also to address directly a broader range of applications.

2. Setting

2.1. Multiple testing: Mathematical framework

Let X be a random variable defined from a measurable space $(\Omega, \mathfrak{F})$ to some observation space $(\mathcal{X}, \mathfrak{X})$. We assume that there is a family of probability distributions on $(\Omega, \mathfrak{F})$ that induces a subset $\mathcal{P}$ of probability distributions on $(\mathcal{X}, \mathfrak{X})$, which is called the model. The distribution of X on $(\mathcal{X}, \mathfrak{X})$ is denoted by P; for each $P \in \mathcal{P}$ there exists a distribution on $(\Omega, \mathfrak{F})$ for which $X \sim P$; it is referred to as $\mathbb{P}_{X\sim P}$ or simply by $\mathbb{P}$ when unambiguous. The corresponding expectation operator is denoted $\mathbb{E}_{X\sim P}$ or $\mathbb{E}$ for short.

We consider a general multiple testing problem for P, defined as follows. Let $\mathcal{H}$ denote an index space for (null) hypotheses. To each $h \in \mathcal{H}$ is associated a known subset $H_h \subset \mathcal{P}$ of probability measures on $(\mathcal{X}, \mathfrak{X})$. Multiple hypothesis testing consists in taking a decision, based on a single realization of the variable X, of whether for each $h \in \mathcal{H}$ it holds or not that $P \in H_h$ (which is read "P satisfies H_h", or "H_h is true"). We denote by $\mathcal{H}_0(P) := \{h \in \mathcal{H}\colon\ P \text{ satisfies } H_h\}$ the set of true null hypotheses, and by its complementary $\mathcal{H}_1(P) := \mathcal{H} \setminus \mathcal{H}_0(P)$ the set of false nulls. These sets are of course unknown because they depend on the unknown distribution P. For short, we will write sometimes $\mathcal{H}_0$ and $\mathcal{H}_1$ instead of $\mathcal{H}_0(P)$ and $\mathcal{H}_1(P)$, respectively.

As an illustration, if we observe a continuous Gaussian process $X = (X_h)_{h\in[0,1]^d}$ with a continuous mean function $\mu : h \in [0, 1]^d \mapsto \mu(t) := \mathbb{E}X_t$, then P is the distribution of this process, $(\mathcal{X}, \mathfrak{X})$ is the Wiener space and $\mathcal{P}$ is the set of distributions generated by continuous Gaussian processes having a continuous mean function. Typically, $\mathcal{H} = [0, 1]^d$ and, for any h, we choose H_h equal to the set of distributions in $\mathcal{P}$ for which the mean function μ satisfies $\mu(h) \leq 0$. This is usually denoted H_h: "$\mu(h) \leq 0$". Finally, the set $\mathcal{H}_0(P) = \{h \in [0, 1]^d\colon\ \mu(h) \leq 0\}$ corresponds to the true null hypotheses. Several other examples are provided below in Section 2.4.

Next, for a more formal definition of a multiple testing procedure, we first assume the following:

$$\text{The index space } \mathcal{H} \text{ is endowed with a } \sigma\text{-algebra } \mathfrak{H} \text{ and for all } P \in \mathcal{P}, \text{ the set } \mathcal{H}_0(P) \text{ of true nulls is assumed to be measurable, that is, } \mathcal{H}_0(P) \in \mathfrak{H}. \tag{A1}$$

Definition 2.1 (Multiple testing procedure). *Let $X : (\Omega, \mathfrak{F}) \to (\mathcal{X}, \mathfrak{X})$ be a random variable, $\mathcal{P}$ a model of distributions of $\mathcal{X}$, and $\mathcal{H}$ an index set of null hypotheses. Assume* (A1) *holds. A multiple testing procedure on $\mathcal{H}$ is a function $R : X(\Omega) \subset \mathcal{X} \to \mathfrak{H}$ such that the set*

$$\{(\omega, h) \in \Omega \times \mathcal{H}:\ h \in R(X(\omega))\}$$

is a $\mathfrak{F} \otimes \mathfrak{H}$-measurable set; or in other terms, that the process $(\mathbf{1}\{h \in R(X)\})_{h\in\mathcal{H}}$ is a measurable process.

The fact that R need only be defined on the image $X(\Omega)$, rather than on the full space $\mathcal{X}$, is a technical detail necessary for later coherence; this introduces no restriction since R will only be ever applied to possible observed values of X.

A multiple testing procedure R is interpreted as follows: based on the observation $x = X(\omega)$, $R(x)$ is the set of null hypotheses that are deemed to be false, also called set of *rejected* hypotheses. The set $\mathcal{H}_0(P) \cap R(x)$ formed of true null hypotheses that are rejected in error is called the set of *type* I *errors*. Similarly, the set $\mathcal{H}_1(P) \cap R^c(x)$ is that of *type* II *errors*.

2.2. The p-value functional and process

We will consider a very common framework for multiple testing, where the decision for each null hypothesis $H_h, h \in \mathcal{H}$, is taken based on a scalar statistic $p_h(x) \in [0, 1]$ called a p-value. The characteristic property of a p-value statistic is that if the generating distribution P is such that the corresponding null hypothesis is true (i.e., $h \in \mathcal{H}_0(P)$), then the random variable $p_h(X)$ should be stochastically lower bounded by a uniform random variable. Conversely, this statistic is generally constructed in such a way that if the null hypothesis H_h is false, its distribution will be more concentrated towards the value 0. Therefore, a p-value close to 0 is interpreted as evidence from the data against the validity of the null hypothesis, and one will want to reject hypotheses having lower p-values. Informally speaking, based on observation x, the construction of a multiple testing procedure generally proceeds as follows:

(i) compute the p-value $p_h(x)$ for each individual null index $h \in \mathcal{H}$;
(ii) determine a threshold $t_h(x)$ for each $h \in \mathcal{H}$, depending on the whole family $(p_h(x))_{h\in\mathcal{H}}$;
(iii) put $R(x) = \{h \in \mathcal{H}:\ p_h(x) \le t_h(x)\}$.

To summarize, the rejection set consists of hypotheses whose p-values are lower than a certain threshold, this threshold being itself random, depending on the observation x and possibly depending on h. This will be elaborated in more detail in Section 3.2, in particular how the threshold function $t_h(x)$ is chosen. For now, we focus on properly defining the p-value functional itself, the associated process, and the assumptions we make on them.

Formally, we define the *p-value functional* as a mapping $\mathbf{p} : \mathcal{X} \to [0,1]^{\mathcal{H}}$, or equivalently as a collection of functions $\mathbf{p} = (p_h(x))_{h\in\mathcal{H}}$, each of the functions $p_h : \mathcal{X} \to [0,1]$, $h \in \mathcal{H}$, being considered as a scalar statistic that can be computed from the observed data $x \in \mathcal{X}$.

We will consider correspondingly the random p-values $\omega \in \Omega \mapsto p_h(X(\omega))$, and p-value process $\omega \in \Omega \mapsto \mathbf{p}(X(\omega))$. With some notation overload, we will sometimes drop the dependence on X and use the notation p_h and $\mathbf{p}$ to also denote the *random variables* $p_h(X)$ and $\mathbf{p}(X)$ (the meaning – function of x, or random variable on Ω – should be clear from the context).

We shall make the following assumptions on the p-value process:

- Joint measurability over $\Omega \times \mathcal{H}$: we assume that the random process $(p_h(X))_{h\in\mathcal{H}}$ is a measurable process, that is:

$$(\omega, h) \in (\Omega \times \mathcal{H}, \mathfrak{F} \otimes \mathfrak{H}) \mapsto p_h\big(X(\omega)\big) \in [0,1] \qquad \text{is (jointly) measurable.} \tag{A2}$$

- For any $P \in \mathcal{P}$, the marginal distributions of the p-values corresponding to true nulls are stochastically lower bounded by a uniform random variable on $[0,1]$:

$$\forall h \in \mathcal{H}_0(P) \qquad \forall u \in [0,1], \qquad \mathbb{P}_{X\sim P}\big(p_h(X) \le u\big) \le u. \tag{A3}$$

(The distribution of p_h if h lies in $\mathcal{H}_1(P)$ can be arbitrary).

Condition (A2) is specific to the continuous setting considered here and will be discussed in more detail in the next section. Condition (A3) is the standard characterization of a single p-value statistic in classical (single or multiple) hypothesis testing. In general, an arbitrary scalar statistic used to take the rejection decision on hypothesis H_h can be monotonically normalized into a p-value as follows: assume $S_h(x)$ is a scalar test statistic, then

$$p_h(x) = \sup_{P\in H_h} F_{h,P}\big(S_h(x)\big)$$

is a p-value in the sense of (A3), where $F_{h,P}(t) = \mathbb{P}_{X\sim P}(S_h(X) \ge t)$ (and where the supremum is assumed to maintain the measurability in x, for any fixed h). If the scalar statistic $S_h(x)$ is constructed so that it tends to be stochastically larger when hypothesis H_h is false, the corresponding p-value indeed has the desirable property that it is more concentrated towards 0 in this case. Such test statistics abound in the (single) testing literature, and a few examples will be given below.

2.3. Discussion on measurability assumptions

Since the focus of the present work is to be able to deal with uncountable spaces of hypotheses $\mathcal{H}$, we have to be somewhat careful with corresponding measurability assumptions over $\mathcal{H}$ (a problem that does not arise when $\mathcal{H}$ is finite or countable). The main assumption needed in this regard in order to state properly the results to come is the joint measurability assumption appearing in either Definition 2.1 (for the multiple testing procedure) or in (A2) (for the p-value process), both of which are specific to the uncountable setting. Essentially, joint measurability will be necessary in order to use Fubini's theorem on the space $(\Omega \times \mathcal{H}, \mathfrak{F} \otimes \mathfrak{H})$, and have the expectation operator w.r.t. ω and the integral operator over $\mathcal{H}$ commute.

If $\mathcal{H}$ has at most countable cardinality, and is endowed with the trivial σ-field comprising all subsets of $\mathcal{H}$, then (A2) is automatically satisfied whenever all individual p-value functions $p_h : \mathcal{X} \to [0, 1]$, $h \in \mathcal{H}$, are separately measurable, which is the standard setting in multiple testing.

If $\mathcal{H}$ is uncountable, a sufficient condition ensuring (A2) is the joint measurability of the p-value *functional*,

$$(x, h) \in (\mathcal{X} \times \mathcal{H}, \mathfrak{X} \otimes \mathfrak{H}) \mapsto p_h(x) \in [0, 1] \qquad \text{is (jointly) measurable,} \tag{A2$'$}$$

which implies (A2) by composition. Unfortunately, (A2′) might not always hold. To see this, consider the following canonical example. Assume the observation takes the form of a stochastic process indexed by the hypothesis space itself, $X = \{X_h, h \in \mathcal{H}\}$. In this case, the observation space $\mathcal{X}$ is included in $\mathbb{R}^{\mathcal{H}}$. Furthermore, assume the p-value function $p_h(x)$ is given by a fixed measurable mapping ψ of the value of x at point h, that is, $p_h(x) = \psi(x_h), \forall h \in \mathcal{H}$. In this case assumption (A2′) boils down to the joint measurability of the evaluation mapping $(x, h) \in \mathcal{X} \times \mathcal{H} \mapsto x_h$. Whether this holds depends on the nature of the space $\mathcal{X}$. We give some classical examples in the next section where the assumption holds; for example, it is true if $\mathcal{X}$ is the Wiener space.

However, the joint measurability of the evaluation mapping does not hold if $\mathcal{X}$ is taken to be the product space $\mathbb{R}^{\mathcal{H}}$ endowed with the canonical product σ-field (indeed, this would imply that any $x \in \mathbb{R}^{\mathcal{H}}$, i.e., any function from $\mathcal{H}$ into $\mathbb{R}$, is measurable). The more general assumption (A2) may still hold, though, but it generally requires some additional regularity or structural assumptions on the paths of the process X. In particular, in the above example if $X = \{X_h, h \in \mathcal{H}\}$ is a stochastic process having a (jointly) measurable modification (and more generally for other examples, if there exists a modification of X such that (A2) is satisfied), we will always assume that we observe such a modification, so that assumption (A2) holds.

We have gathered in Section S-1 of the supplementary material Blanchard, Delattre and Roquain [5] some auxiliary (mostly classical) results related to the existence and properties of such modifications. Lemma S-1.2 shows that such a (jointly) measurable modification exists as soon as the process is continuous in probability. The latter is not an iff condition, but is certainly much weaker than having continuous paths.

On the other hand, it is important to observe here that a jointly measurable modification of X, or, for that matter, of the p-value process, might not exist. Lemma S-1.1 reproduces a classical argument showing that for $\mathcal{H} = [0, 1]$, assumption (A2) is violated for any modification of a mutually independent p-value process. Therefore, for an uncountable space of hypotheses $\mathcal{H}$, assumption (A2) precludes the possibility that the p-values $\{p_h, h \in \mathcal{H}\}$ are mutually independent. This contrasts strongly with the situation of a finite hypothesis set $\mathcal{H}$, where mutual independence of the p-values is generally considered the reference case.

A final issue is to which extent the results exposed in the remainder of this work depend on the (jointly) measurable modification chosen for the underlying stochastic process. Lemma S-1.4 elucidates this issue by showing that this is not the case, because the FDR (the main measure of type I error, which will be formally defined in Section 3.1) is identical for two such modifications.

2.4. Examples

To illustrate the above generic setting, let us consider the following examples.

Example 2.2 (Testing the mean of a process). Assume that we observe the realization of a real-valued process $X = (X_t)_{t\in[0,1]^d}$ with an unknown (measurable) mean function $\mu : t \in [0,1]^d \mapsto \mu(t) := \mathbb{E}X_t$. We take $\mathcal{H} = [0,1]^d$ and want to test simultaneously for each $t \in [0,1]^d$ the null hypothesis H_t: "$\mu(t) \leq 0$". Assume that for each t the marginal distribution of $(X_t - \mu(t))$ is known, does not depend on t and has upper-tail function G (e.g., X is a Gaussian process with marginals $X_t \sim \mathcal{N}(\mu(t), 1)$). We correspondingly define the p-value process $\forall t \in [0,1]^d$, $p_t(X) = G(X_t)$, which satisfies (A3). Next, the measurability assumption (A2) follows from a regularity assumption on X:

- if we assume that the process X has continuous paths, $X : \omega \mapsto (X_t(\omega))_t$ can be seen as taking values in the Wiener space $\mathcal{X} = \mathcal{C}_{[0,1]^d} = C([0,1]^d, \mathbb{R})$ of continuous functions from $[0,1]^d$ to $\mathbb{R}$. (In this case, the Borel σ-field corresponding to the supremum norm topology on $\mathcal{C}_{[0,1]^d}$ is the trace of the product σ-field on $\mathcal{C}_{[0,1]^d}$, and X is measurable iff all its coordinate projections are.) Furthermore, the p-value function can be written as

$$(x,t) \in \mathcal{C}_{[0,1]^d} \times [0,1]^d \mapsto p_t(x) = G\big(x(t)\big) \in [0,1].$$

  The evaluation functional $(x,t) \in \mathcal{C}_{[0,1]^d} \times [0,1]^d \mapsto x(t)$ is jointly measurable because it is continuous, thus $p_t(x)$ is jointly measurable by composition and (A2′) holds, hence also (A2);
- if $d = 1$ and the process X is càdlàg , the random variable X can be seen as taking values in the Skorohod space $\mathcal{X} = \mathcal{D} := D([0,1], \mathbb{R})$ of càdlàg functions from $[0,1]$ to $\mathbb{R}$. In this case, the Borel σ-field generated by the Skorohod topology is also the trace of the product σ-field on $\mathcal{D}$ (see, e.g., Theorem 14.5 page 121 of Billingsley [4]). Moreover, the evaluation functional $(x,t) \mapsto x(t)$ is jointly measurable, as for any càdlàg function x, it is the pointwise limit of the jointly measurable functions ζ_n: $(x,t) \mapsto \zeta_n(x,t) := \sum_{k=1}^{2^n} x(k2^{-n})\mathbf{1}\{(k-1)2^{-n} \leq t < k2^{-n}\} + x(1)\mathbf{1}\{t = 1\}$, therefore (A2′) is fulfilled by composition, hence also (A2);
- assume that X is a Gaussian process defined on the space $\mathcal{X} = \mathbb{R}^{[0,1]^d}$ endowed with the canonical product σ-field, and with a covariance function $\Sigma(t,t')$ such that Σ is continuous on all points (t,t) of the diagonal and takes a constant (known) value σ^2 on those points.

  This assumption is not sufficient to ensure that X has a continuous version, but it ensures that (X_t) is continuous in L^2 and hence in probability; Lemma S-1.2 then states that X has a modification such that the evaluation functional is jointly measurable. Assuming that such a jointly measurable modification is observed, we deduce that (A2) holds for the associated p-value process.

Example 2.3 (Testing the signal in a Gaussian white noise model). Let us consider the Gaussian white noise model $dZ_t = f(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, $t \in [0,1]$, where B is a Wiener process on $[0,1]$ and $f \in C([0,1])$ is a continuous signal function. For simplicity, the standard deviation σ is assumed to be equal to 1. Equivalently, we assume that we can observe the stochastic integral of Z_t against any test function in $L^2([0,1])$, that is, that we observe the Gaussian process $(X_g)_{g\in L^2([0,1])}$ defined by

$$X_g := \int_0^1 g(t) f(t)\,\mathrm{d}t + \int_0^1 g(t)\,\mathrm{d}B_t, \qquad g \in L^2\big([0,1]\big).$$

Formally, the observation space is the whole space $\mathcal{X} = \mathbb{R}^{L^2([0,1])}$, endowed with the product σ-field. However, in the sequel, we will use the observation of the process X only against a "small" subspace of functions of $L^2([0,1])$.

Let us consider $\mathcal{H} = [0,1]$ and the problem of testing for each $t \in \mathcal{H}$, the null H_t: "$f(t) \le 0$" (signal non-positive). We can build p-values based upon a kernel estimator in the following way. Consider a kernel function $K \in L^2(\mathbb{R})$, assumed positive on $[-1,1]$ and zero elsewhere, and denote by $K_t \in L^2([0,1])$ the function $K_t(s) := K((t-s)/\eta)$, where $0 < \eta \le 1$ is a bandwidth to be chosen. Let us consider the process $\widetilde{X}_t := X_{K_t}, t \in [0,1]$. From Lemma S-1.3, $\widetilde{X}$ has a modification which is jointly measurable in (ω, t). Clearly, this implies that there exists a modification of the original process X such that $\widetilde{X}$ is jointly measurable in (ω, t), and we assume that we observe such a modification. For any $t \in [0,1]$, letting $c_{K,t} := \int_0^1 K((t-s)/\eta)\,\mathrm{d}s > 0$ and $v_{K,t} := \int_0^1 K^2((t-s)/\eta)\,\mathrm{d}s \ge c_{K,t}^2 > 0$, we can consider the following standard estimate of $f(t)$:

$$\begin{aligned}\widehat{f_\eta}(t) &:= c_{K,t}^{-1} X_{K_t} \\ &= c_{K,t}^{-1}\int_0^1 K\left(\frac{t-s}{\eta}\right) f(s)\,\mathrm{d}s + c_{K,t}^{-1}\int_0^1 K\left(\frac{t-s}{\eta}\right)\mathrm{d}B_s.\end{aligned} \tag{2}$$

Assume that there is a known $\delta_{t,\eta} > 0$ such that for any t with $f(t) \le 0$, we have the upper-bound

$$\mathbb{E}\widehat{f_\eta}(t) = c_{K,t}^{-1}\int_0^1 K\left(\frac{t-s}{\eta}\right) f(s)\,\mathrm{d}s \le \delta_{t,\eta}. \tag{3}$$

For instance, this holds if we can assume a priori knowledge on the regularity of f, of the form $\sup_{s:|s-t|\le\eta} |f(s) - f(t)| \le \delta_{t,\eta}$. Then, the statistics $(\widehat{f_\eta}(t))_t$ can be transformed into a p-value process in the following way:

$$p_t(X) = \overline{\Phi}\left(\frac{\widehat{f_\eta}(t) - \delta_{t,\eta}}{v_{K,t}^{1/2}/c_{K,t}}\right), \tag{4}$$

where $\overline{\Phi}(w) := \mathbb{P}(W \ge w)$, $W \sim \mathcal{N}(0,1)$, is the upper tail distribution of a standard Gaussian distribution. The p-value process (4) satisfies (A3), because for any t with $f(t) \le 0$ and any $u \in [0,1]$,

$$\begin{aligned}\mathbb{P}\big(p_t(X) \le u\big) &= \mathbb{P}\big(\widehat{f_\eta}(t) - \delta_{t,\eta} \ge v_{K,t}^{1/2}/c_{K,t}\,\overline{\Phi}^{-1}(u)\big) \\ &\le \mathbb{P}\big(c_{K,t}\big(\widehat{f_\eta}(t) - \mathbb{E}\widehat{f_\eta}(t)\big)/v_{K,t}^{1/2} \ge \overline{\Phi}^{-1}(u)\big) \\ &= u,\end{aligned}$$

because $\int_0^1 K_t(s)\,\mathrm{d}B_s \sim \mathcal{N}(0, v_{K,t})$. Moreover, the p-value process (4) satisfies (A2), since we assumed $(X_{K_t})_t \in [0,1]$ to be jointly measurable in (ω, t).

Example 2.4 (Testing the c.d.f.). Let $X = (X_1, \dots, X_m) \in \mathcal{X} = \mathbb{R}^m$ be a m-uple of i.i.d. real random variables of common continuous c.d.f. F. For $\mathcal{H} = I$ an interval of $\mathbb{R}$ and a given benchmark c.d.f. F_0, we aim to test simultaneously for all $t \in I$ the null H_t: "$F(t) \le F_0(t)$". The individual hypothesis H_t may be tested using the p-value

$$p_t(X) = G_t\big(m\mathbb{F}_m(X,t)\big), \tag{5}$$

where $\mathbb{F}_m(X,t) = m^{-1}\sum_{i=1}^m \mathbf{1}\{X_i \le t\}$ is the empirical c.d.f. of $X_1, \dots, X_m$ and where $G_t(k) = \mathbb{P}[Z_t \ge k]$, $Z_t \sim \mathcal{B}(m, F_0(t))$, is the upper-tail function of a binomial distribution of parameter $(m, F_0(t))$. The conditions (A2) and (A3) are both clearly satisfied.

Figure 1 provides a realization of the p-value process (5) when testing for all $t \in [0,1]$ the null H_t: "$F(t) \le t$" when F comes from a mixture of beta distributions. The correct/erroneous rejections are also pictured for the simple procedure $R(X) = \{t \in [0,1]\colon\ p_t(X) \le 0.4\}$.

Example 2.5 (Testing the intensity of a Poisson process). Assume we observe $(N_t)_{t\in[0,1]} \in \mathcal{X} = D([0,1],\mathbb{R})$ a Poisson process with intensity $\lambda : [0,1] \to \mathbb{R}^+ \in L^1(d\Lambda)$, where Λ denotes the Lebesgue measure on $[0,1]$. For each $t \in [0,1]$, we aim to test H_t: "$\lambda(t) \le \lambda_0(t)$" where $\lambda_0(\cdot) > 0$ is a given benchmark intensity. Assume that for a given bandwidth $\eta \in (0,1]$, there is a known upper bound $\delta_{t,\eta}$ for $\int_{(t-\eta)\vee 0}^{(t+\eta)\wedge 1} \lambda(s)\,\mathrm{d}s$ that holds true for any t such that $\lambda(t) \le \lambda_0(t)$.

For instance, we can choose $\delta_{t,\eta} = ((t+\eta)\wedge 1 - (t-\eta)\vee 0)(\lambda_0(t) + \sup_{s:|t-s|\le\eta} |\lambda(t) - \lambda(s)|)$ (assuming knowledge on the regularity of λ is available a priori). For any $t \in [0,1]$, the variable $N_{(t+\eta)\wedge 1} - N_{(t-\eta)\vee 0}$ follows a Poisson variable of parameter $\int_{(t-\eta)\vee 0}^{(t+\eta)\wedge 1} \lambda(s)\,\mathrm{d}s$. Since the latter parameter is smaller than $\delta_{t,\eta}$ as soon as $\lambda(t) \le \lambda_0(t)$, the following p-value process satisfies (A3):

$$p_t(X) = G_t(N_{(t+\eta)\wedge 1} - N_{(t-\eta)\vee 0}), \tag{6}$$

where for any $k \in \mathbb{N}$, $G_t(k)$ denotes $\mathbb{P}[Z \ge k]$ for Z a Poisson distribution of parameter $\delta_{t,\eta}$. Moreover, the p-value process fulfills condition (A2′), because (N_t) is a càdlàg process, so that arguments similar to those of Example 2.2 apply. Thus, (A2) also holds.

3. Main concepts and tools

3.1. False discovery rate

Following the usual philosophy of hypothesis testing, one wants to ensure some control over type I errors committed by the procedure. As discussed in Section 1.3, in the present work we focus on a generalization to a continuum of hypotheses of the *False Discovery Rate* (FDR). For a finite number of null hypotheses, the FDR, as introduced by Benjamini and Hochberg [1] (see also Seeger [16]), is defined as the average proportion of type I errors in the set of all rejected hypotheses. To extend this definition to a possibly uncountable space, following Perone Pacifico

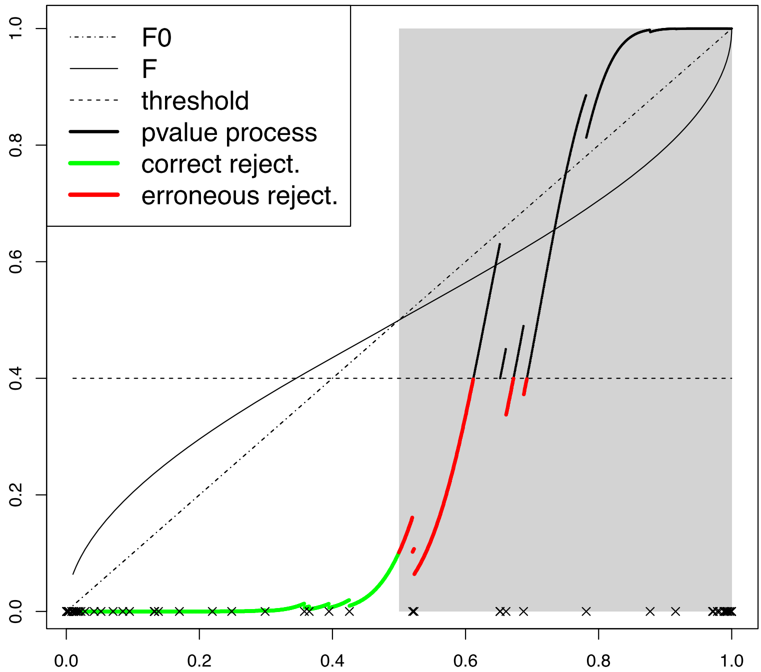

Figure 1. Plot of a realization of the p-value process as defined in (5) for the c.d.f. testing, together with F_0 and F, for $F_0(t) = t$ and $F(t) = 0.5F_1(t) + 0.5F_2(t)$, where F_1 (resp., F_2) is the c.d.f. of a beta distribution of parameter (0.5, 1.5) (resp., (1.5, 0.5)). The region where the null hypothesis "$F(t) \leq F_0(t)$" is true is depicted in grey color. The crosses correspond to the elements of $\{X_i, 1 \leq i \leq m\}$; $m = 50$. The correct/erroneous rejections refer to the procedure $R(X) = \{t \in [0,1]\colon\ p_t(X) \leq 0.4\}$ using the threshold 0.4.

et al. [11,12], we quantify this proportion by a volume ratio, defined with respect to a finite measure Λ on $(\mathcal{H}, \mathfrak{H})$ (the usual definition over a finite space is recovered by taking Λ equal to the counting measure).

Definition 3.1 (False discovery proportion, false discovery rate). *Let Λ be a finite positive measure on $(\mathcal{H}, \mathfrak{H})$. Let R be a multiple testing procedure on $\mathcal{H}$. The false discovery rate* (*FDR*) *of R is defined as the average of the false discovery proportion* (*FDP*):

$$\forall P \in \mathcal{P}, \forall x \in X(\Omega), \qquad \mathrm{FDP}\big(R(x), P\big) := \frac{\Lambda(R(x) \cap \mathcal{H}_0(P))}{\Lambda(R(x))} \mathbf{1}\big\{\Lambda\big(R(x)\big) > 0\big\} \tag{7}$$

and

$$\forall P \in \mathcal{P}, \qquad \mathrm{FDR}(R, P) := \mathbb{E}_{X \sim P}\big[\mathrm{FDP}\big(R(X), P\big)\big]. \tag{8}$$

The indicator function in (7) means that the ratio is taken equal to zero whenever the denominator is zero. Observe that, due to the joint measurability assumption in Definition 2.1 of a multiple testing procedure, both of the above quantities are well-defined (the FDP is only formally defined over the image of Ω through X since only on this set is the measurability of $R(x)$ guaranteed by the definition. In particular, it is defined for P-almost all $x \in \mathcal{X}$).

As illustration, in the particular realization of the p-value process pictured in Figure 1, if we denote by "Red" (resp., "Green") the length of the interval corresponding to the projection of the red (resp., green) part of the p-value process on the X-axis, the FDP of the procedure $R(X) = \{t \in [0, 1]\colon\ p_t(X) \le 0.4\}$ is Red/(Red + Green). A similar interpretation for the FDP holds in Figure 2.

Finding a procedure R with a FDR smaller than or equal to α has the following interpretation: on average, the volume proportion of type I errors among the rejected hypotheses is smaller than α. This means that the procedure is allowed to reject in error some true nulls but in a small (average) proportion among the rejections. For a pre-specified level α, the goal is then to determine multiple testing procedures R such that for any $P \in \mathcal{P}$, it holds that $\mathrm{FDR}(R, P) \le \alpha$. (In fact, the statement need only hold for $P \in \mathcal{P} \cap \bigcup_{h\in\mathcal{H}} H_h$, since outside of this set $\mathcal{H}_0(P) = \varnothing$ and the FDR is 0.) The rest of the paper will concentrate on establishing sufficient conditions under which the FDR is controlled at a fixed level α. Under this constraint, in order to get a procedure with good power properties (i.e., low type II error), it is, generally speaking, desirable that R rejects as many nulls as possible, that is, has volume $\Lambda(R)$ as large as possible.

3.2. Step-up procedures

In what follows, we will focus on a particular form of multiple testing procedures which can be written as function of the p-value family $\mathbf{p}(x) = (p_h(x))_{h\in\mathcal{H}}$.

First, we define a parametrized family of possible rejection sets having the following form: for a given *threshold function* $\Delta\colon (h, r) \in \mathcal{H} \times \mathbb{R}^+ \mapsto \mathbb{R}^+$, we define for any $r \ge 0$ the sub-level set

$$\forall x \in \mathcal{X}, \qquad L_\Delta(x, r) := \big\{h \in \mathcal{H}\colon\ p_h(x) \le \Delta(h, r)\big\} \subset \mathcal{H}. \tag{9}$$

For short, we sometimes write $L_\Delta(r)$ instead of $L_\Delta(x, r)$ when unambiguous. We will more particularly focus on threshold functions Δ of the product form $\Delta(h, r) = \alpha\pi(h)\beta(r)$, where $\alpha \in (0, 1)$ is a positive scalar (*level*), $\pi : \mathcal{H} \to \mathbb{R}^+$ is measurable (*weight function*), and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ is non-decreasing and right-continuous (*shape function*). Clearly, this decomposition is not unique, but will be practical for the formulation of the main result.

Given a threshold function Δ of the above form, we will be interested in a particular, data-dependent choice of the parameter r determining the rejection set, called *step-up procedure*.

Definition 3.2 (Step-up procedure). *Let $\Delta(h, r) = \alpha\pi(h)\beta(r)$ a threshold function with $\alpha \in (0, 1)$; $\pi : \mathcal{H} \to \mathbb{R}^+$ measurable and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ non-decreasing and right-continuous. Then the step-up multiple testing procedure R on $(\mathcal{H}, \Lambda)$ associated to Δ, is defined by*

$$\forall x \in X(\Omega), \qquad R(x) = L_\Delta\big(x, \widehat{r}(x)\big) \qquad \text{where } \widehat{r}(x) := \max\big\{r \ge 0\colon\ \Lambda\big(L_\Delta(x, r)\big) \ge r\big\}. \tag{10}$$

Note that $\widehat{r}$ above is well-defined: first, since $x \in X(\Omega)$ and from assumption (A2), the function $h \mapsto p_h(x) - \alpha\pi(h)\beta(r)$ is measurable; thus $L_\Delta(x,r)$ is a measurable set of $\mathcal{H}$, which in turn implies that $\Lambda(L_\Delta(x,r))$ is well-defined. Secondly, the supremum of $\{r \geq 0\colon\ \Lambda(L_\Delta(x,r)) \geq r\}$ exists because $r = 0$ belongs to this set and $M = \Lambda(\mathcal{H})$ is an upper bound. Third, this supremum is a maximum because the function $r \mapsto \Lambda(L_\Delta(x,r))$ is non-decreasing (right-continuity is not needed for this).

We should ensure in Definition 3.2 that a step-up procedure satisfies the measurability requirements of Definition 2.1. This is proved separately in Section 5.2. In that section, we also check that the equality $\Lambda(L_\Delta(x,\widehat{r}(x))) = \widehat{r}(x)$ always holds. Hence, $\widehat{r}(x)$ is the largest intersection point between the function $r \mapsto \Lambda(L_\Delta(x,r))$ giving the volume of the candidate rejection sets as a function of r, and the identity line $r \mapsto r$.

To give some basic intuition behind the principle of a step-up procedure, consider for simplicity that π is a constant function, so that the family defined by (9) are ordinary sub-level sets of the p-value family. The goal is to find a suitable common rejection threshold t giving rise to rejection set R_t. Assume also without loss of generality that $\Lambda(\mathcal{H}) = 1$. Now consider the following heuristic. If the threshold t is deterministic, any p-value associated to a true null hypothesis, being stochastically lower bounded by a uniform variable, has probability less than t of being rejected in error. Thus, we expect on average a volume $t\Lambda(\mathcal{H}_0) \leq t$ of erroneously rejected null hypotheses. If we therefore use t as a rough upper bound of the numerator in the definition (7) of the FDP or FDR, and we want the latter to be less than α, we obtain the constraint $t/\Lambda(R_t) \leq \alpha$, or equivalently $\Lambda(R_t) \geq \alpha^{-1}t$. Choosing the largest t satisfying this heuristic constraint is equivalent to the step-up procedure wherein $\beta(u) = u$. The choice of a different shape function with $\beta(u) \leq u$ can be interpreted roughly as a pessimistic discount to compensate for various inaccuracies in the above heuristic argument (in particular the fact that the obtained threshold is really a random quantity).

In the case where $\mathcal{H}$ is finite and Λ is the counting measure, it can be seen that the above definition recovers the usual notion of step-up procedures (see, e.g., Blanchard and Roquain [6]); in particular, the linear shape function $\beta(u) = u$ gives rise to the celebrated linear step-up procedure of Benjamini and Hochberg [1].

3.3. PRDS conditions

To ensure control of the FDR criterion, an important role is played by structural assumptions on the dependence of the p-values. While the case of independent p-values is considered as the reference setting in the case where $\mathcal{H}$ is finite, we recall that for an uncountable set $\mathcal{H}$, we cannot assume mutual independence of the p-values since this would contradict our measurability assumptions (see concluding discussion of Section 2.3).

We will consider two different situations in our main result: first, if the dependence of the p-values can be totally arbitrary, and secondly, if a form of positive dependence is assumed. This is the latter condition which we define more precisely now. We consider a generalization to the case of infinite, possibly uncountable space $\mathcal{H}$, of the notion of positive regression dependence on each one from a subset (PRDS) introduced by Benjamini and Yekutieli [3] in the case of a finite set of hypotheses.

For any finite set $\mathcal{I}$, a subset $D \subset [0,1]^{\mathcal{I}}$ is called *non-decreasing* if for all $\mathbf{z}, \mathbf{z}' \in [0,1]^{\mathcal{I}}$ such that $\mathbf{z} \leq \mathbf{z}'$ (i.e., $\forall h \in \mathcal{I}, z_h \leq z'_h$), we have $\mathbf{z} \in D \Rightarrow \mathbf{z}' \in D$.

Definition 3.3 (PRDS conditions for a finite p-value family). *Assume $\mathcal{H}$ to be finite. For $\mathcal{H}'$ a subset of $\mathcal{H}$, the p-value family $\mathbf{p}(X) = (p_h(X))_{h\in\mathcal{H}}$ is said to be weak PRDS on $\mathcal{H}'$ for the distribution P, if for any $h \in \mathcal{H}'$, for any measurable non-decreasing set D in $[0,1]^{\mathcal{H}}$, the function $u \in [0,1] \mapsto \mathbb{P}(\mathbf{p}(X) \in D | p_h(X) \leq u)$ is non-decreasing on $\{u \in [0,1]\colon \mathbb{P}(p_h(X) \leq u) > 0\}$; it is said to be strong PRDS if the function $u \mapsto \mathbb{P}(\mathbf{p}(X) \in D | p_h(X) = u)$ is non-decreasing.*

To be completely rigorous, observe that the conditional probability with respect to the event $\{p_h(X) \leq u\}$ is defined pointwise unequivocally whenever this event has positive probability, using a ratio of probabilities; while the conditional probability with respect to $p_h(X) = u$ can only be defined via conditional expectation, and is therefore only defined up to a $p_h(X)$-negligible set. Hence, in the definition of strong PRDS, strictly speaking, we only require that the conditional probability coincides $p_h(X)$-a.s. with a non-decreasing function.

Definition 3.4 (Finite dimensional PRDS conditions for a p-value process). *For $\mathcal{H}'$ a subset of $\mathcal{H}$, the p-value process $\mathbf{p}(X) = (p_h(X))_{h\in\mathcal{H}}$ is said to be finite dimensional weak PRDS on $\mathcal{H}'$ (resp., finite dimensional strong PRDS on $\mathcal{H}'$) for the distribution P, if for any finite subset $\mathcal{S}$ of $\mathcal{H}$, the finite p-value family $\mathbf{p}_{\mathcal{S}}(X) = (p_h(X))_{h\in\mathcal{H}\cap\mathcal{S}}$ is weak PRDS on $\mathcal{H}' \cap \mathcal{S}$ (resp., strong PRDS on $\mathcal{H}' \cap \mathcal{S}$) for the distribution P.*

While the finite dimensional weak PRDS property will be sufficient to state our main result, the strong PRDS property is sometimes easier to handle. Hence, it is important to note that the finite dimensional strong PRDS property implies the finite dimensional weak PRDS property, as we establish in Lemma S-2.2, by using a standard argument pertaining to classical multiple testing theory for a finite set of hypotheses.

Finally, Benjamini and Yekutieli [3] (Section 3.1 therein) proved that the p-value family corresponding to a finite Gaussian random vector are (finite) strong PRDS as soon as all the coefficient of the covariance matrix are non-negative. This equivalently proves the following result.

Lemma 3.5. *Let $\mathbf{p}(X) = (p_h(X))_{h\in\mathcal{H}}$ be a p-value process of the form $p_h(X) = G(X_h)$, $h \in \mathcal{H}$, where $X = (X_h)_{h\in\mathcal{H}}$ is a Gaussian process and where G is continuous decreasing from $\mathbb{R}$ to $[0,1]$. Assume that the covariance function Σ of X satisfies*

$$\forall h, h' \in \mathcal{H}, \qquad \Sigma(h, h') \geq 0. \tag{11}$$

Then the p-value process is finite dimensional strong PRDS (on any subset).

4. Control of the FDR

In this section, our main result is stated and then illustrated with several examples.

4.1. Main result

The following theorem establishes our main result on sufficient conditions to ensure FDR control at a specified level for step-up procedures. It is proved in Section 5.

Theorem 4.1. *Assume that the hypothesis space $\mathcal{H}$ satisfies* (A1) *and is endowed with a finite measure Λ. Let $\mathbf{p}(X) = (p_h(X))_{h\in\mathcal{H}}$ be a p-value process satisfying the conditions* (A2) *and* (A3). *Denote R the step-up procedure on $(\mathcal{H}, \Lambda)$ associated to a threshold function of the product form $\Delta(h, r) = \alpha\pi(h)\beta(r)$, with $\alpha \in (0, 1)$, β a non-decreasing right-continuous shape function and π a probability density function on $\mathcal{H}$ with respect to Λ. Then for any $P \in \mathcal{P}$, letting $\Pi(\mathcal{H}_0(P)) := \int_{h\in\mathcal{H}_0(P)} \pi(h)\,\mathrm{d}\Lambda(h)$, the inequality*

$$\mathrm{FDR}(R, P) \le \alpha\Pi\big(\mathcal{H}_0(P)\big) \qquad (\le \alpha) \tag{12}$$

holds in either of the two following cases:

1. *$\beta(x) = x$ and the p-value process $\mathbf{p}$ is finite dimensional weak PRDS on $\mathcal{H}_0(P)$ for the distribution P*;
2. *the function β is of the form*

$$\beta_\nu(x) = \int_0^x u\,\mathrm{d}\nu(u), \tag{13}$$

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

Since π is taken as a probability density function on $\mathcal{H}$ with respect to Λ, the FDR in (12) is upper bounded by $\alpha\Pi(\mathcal{H}_0) \le \alpha\Pi(\mathcal{H}) = \alpha$, so that the corresponding step-up procedure provides FDR control at level α. As an illustration, a typical choice for π is the constant probability density function $\forall h \in \mathcal{H}, \pi(h) = 1/\Lambda(\mathcal{H}) = M^{-1}$.

According to the standard philosophy of (multiple) testing, while the FDR is controlled at level α as in (12), we aim to have a procedure that rejects a volume of hypotheses as large as possible. In that sense, choosing a step-up procedure with $\beta(x) = x$ always leads to a better step-up procedure than choosing $\beta(x)$ of the form (13), because $\int_0^x u\,\mathrm{d}\nu(u) \le x$. Hence, in Theorem 4.1, the PRDS assumption allows us to get a result which is less conservative (i.e., rejecting more) than under arbitrary dependencies. Therefore, when we want to apply Theorem 4.1, an important issue is to obtain, if possible, the finite dimensional PRDS condition, see the examples of Section 4.2. When the PRDS assumption does not hold, we refer to Blanchard and Roquain [6] for an extended discussion on choices of the shape function β of the form (13) (which can be suitably adapted to the uncountable case).

4.2. Applications

4.2.1. *FDR control for testing the mean of a Gaussian process*

Consider the multiple testing setting of Example 2.2. More specifically, we consider here the particular case where we observe $\{X_t, t \in [0, 1]^d\}$ a Gaussian process with measurable mean μ, with unit variance and covariance function Σ. Recall that the problem is to test for all $t \in [0,1]^d$ the hypothesis H_t: "$\mu(t) \le 0$". Taking for Λ the d-dimensional Lebesgue measure, the FDR control at level α of a step-up procedure of shape function β and weight function $\pi(h) = 1$ can be rewritten as

$$\mathbb{E}\left[\frac{\Lambda(\{t \in [0,1]^d \colon \mu(t) \le 0, \overline{\Phi}(X_t) \le \alpha\beta(\hat{r}(X))\})}{\Lambda(\{t \in [0,1]^d \colon \overline{\Phi}(X_t) \le \alpha\beta(\hat{r}(X))\})}\right] \le \alpha, \tag{14}$$

where $\overline{\Phi}$ is the upper-tail distribution function of a standard Gaussian variable and

$$\hat{r}(X) = \max\bigl\{r \in [0,1] \colon \Lambda\bigl(\bigl\{t \in [0,1]^d \colon \overline{\Phi}(X_t) \le \alpha\beta(r)\bigr\}\bigr) \ge r\bigr\}.$$

Thus, Theorem 4.1 and Lemma 3.5 entail the following result.

Corollary 4.2. *For any jointly measurable Gaussian process* $\{X_t\}_{t\in[0,1]^d}$ *over* $[0,1]^d$ *with a measurable mean* μ *and unit variances, the FDR control* (14) *holds in either of the two following cases*:

- $\beta(x) = x$ *and the covariance function of the process is coordinates-wise non-negative, that is, satisfies* (11);
- β *is of the form* (13), *under no assumption on the covariance function.*

For instance, any Gaussian process with continuous paths is measurable and thus can be used in Corollary 4.2. More generally, Lemma S-1.2 states that any Gaussian process with a covariance function $\Sigma(t,t')$ such that

$$\forall t \in [0,1]^d, \qquad \lim_{t'\to t} \Sigma\bigl(t',t\bigr) = \Sigma(t,t) \quad \text{and} \quad \lim_{t'\to t} \Sigma\bigl(t',t'\bigr) = \Sigma(t,t) = 1$$

has a measurable modification and hence can be used in Corollary 4.2.

4.2.2. *FDR control for testing the signal in a Gaussian white noise model*

We continue Example 2.3, in which we observe the Gaussian process X defined by $X_g = \int_0^1 g(t)f(t)\,\mathrm{d}t + \int_0^1 g(t)\,\mathrm{d}B_t$, $g \in L^2([0,1])$, where B is a Wiener process on $[0,1]$ and $f \in C([0,1])$ is a continuous signal function. Remember that we aim at testing H_t: "$f(t) \le 0$" for any $t \in [0,1]$, using the integration of the process against a smoothing kernel K_t. Assuming condition (3) holds, the p-value process is obtained via (4) as $p_t(X) = \overline{\Phi}^{-1}(Y_t)$, where $Y_t = v_{K,t}^{-1/2}(X_{K_t} - \delta_{t,\eta}c_{K,t})$ is a Gaussian process. Applying Lemma 3.5, we can prove that the p-value process defined by (4) is finite dimensional strong PRDS (on any subset) by checking that the covariance function of $(Y_t)_t$ has non-negative values: the latter holds because the kernel K has been taken non-negative and $\forall t, s$, $\mathrm{Cov}(Y_t, Y_s) = c\int_0^1 K((t-u)/\eta)K((s-u)/\eta)\,\mathrm{d}u$, for a non-negative constant c. As a consequence, Theorem 4.1 shows that a step-up procedure using $\beta(x) = x$ controls the FDR.

To illustrate this result, let us consider a simple particular case where the kernel K is rectangular, that is, $K(s) = \mathbf{1}\{|s| \le 1\}/2$ and f is L-Lipschitz. Also, to avoid the boundary effects due to the kernel smoothing, we assume that the observation X is made against functions of $L^2([-1, 2])$ while the test of H_t: "$f(t) \le 0$" has only to be performed for $t \in [0, 1]$ only. In that case, for $t \in [0, 1]$, $\delta_{t,\eta} = L\eta$, $c_{K,t} = \eta$, $v_{K,t} = \eta/2$, so that $Y_t = (2\eta)^{-1/2}(Z_{t+\eta} - Z_{t-\eta} - L\eta^2)$. Therefore, the following statement holds.

Corollary 4.3. *Let us consider the Gaussian process* $Z_t = \int_{-1}^{t} f(s)\,\mathrm{d}s + B_t$, $t \in [-1, 2]$, *where* B *is a Wiener process on* $[-1, 2]$ *and* f *is a* L*-Lipschitz function on* $[-1, 2]$ $(L > 0)$. *Let* $\eta \in (0, 1]$ *and* $Y_t = (2\eta)^{-1/2}(Z_{t+\eta} - Z_{t-\eta} - L\eta^2)$. *Denote the Lebesgue measure on* $[0, 1]$ *by* Λ. *Consider the volume rejection of the step-up procedure using* $\pi(t) = 1$ *and* $\beta(x) = x$, *that is*,

$$\hat{r}(X) = \max\Big\{r \in [0, 1]:\ \Lambda\Big(\Big\{t \in [0, 1]:\ \overline{\Phi}(Y_t) \le \alpha r\Big\}\Big) \ge r\Big\},$$

where $\overline{\Phi}$ *denotes the upper-tail distribution function of a standard Gaussian variable. Then the following FDR control holds*:

$$\mathbb{E}\left[\frac{\Lambda(\{t \in [0, 1]:\ f(t) \le 0, \overline{\Phi}(Y_t) \le \alpha\hat{r}(X)\})}{\Lambda(\{t \in [0, 1]:\ \overline{\Phi}(Y_t) \le \alpha\hat{r}(X)\})}\right] \le \alpha. \tag{15}$$

4.2.3. *FDR control for testing the c.d.f.*

Consider the testing setting of Example 2.4 where we aim at testing whether "$F(t) \le F_0(t)$" for any t in an interval $I \subset \mathbb{R}$. Lemma A.1 states that the p-value process defined by (5) is finite dimensional weak PRDS (on any subset). As a consequence, Theorem 4.1 applies and leads to a control of the FDR.

For instance, let us consider the simple case where $I = [0, 1]$, $F_0(t) = t$ and Λ is the Lebesgue measure on $[0, 1]$. In this case, for any $k \in \{1, \dots, m\}$, the function $G_t(k) = \mathbb{P}(Z_t \ge k)$, with $Z_t \sim \mathcal{B}(m, t)$, is continuous increasing in the variable $t \in [0, 1]$. Moreover, for any $t \in (0, 1)$, the function $G_t(k)$ is decreasing in $k = 0, \dots, m$. Therefore, denoting $0 = X_{(0)} \le X_{(1)} \le \cdots \le X_{(m)} \le X_{(m+1)} = 1$ the order statistics of $X_1, \dots, X_m$, the p-value process $t \mapsto p_t(X) = G_t(|\{1 \le i \le m\colon X_i \le t\}|)$ is equal to 1 on $[0, X_{(1)})$, is increasing on each interval $(X_{(j)}, X_{(j+1)}]$, $j = 1, \dots, m$, and is left-discontinuous and right-continuous in each $X_{(j)}$, $1 \le j \le m$, with a left limit larger than $p_{X_{(j)}}(X) = G_{X_{(j)}}(j)$ (see Figure 1).

As a consequence, for any threshold $u \in (0, 1)$, we obtain the following relation for the Lebesgue measure $\gamma(u)$ of the level set $\{t \in [0, 1]:\ p_t(X) \le u\}$:

$$\begin{aligned}\gamma(u) &= \sum_{j=0}^{m} \mathbf{1}\big\{G_{X_{(j)}}(j) \le u\big\}\Lambda\Big(\big\{t \ge X_{(j)}:\ G_t(j) \le u \text{ and } t < X_{(j+1)}\big\}\Big),\\ &= \sum_{j=0}^{m}\big(X_{(j+1)} \wedge t_j(u) - X_{(j)}\big)_+,\end{aligned} \tag{16}$$

where $t_j(u)$, $j = 0, \dots, m$ is the unique solution of the equation $G_t(j) = u$, which can be easily computed numerically. Choosing for simplicity a uniform weighting $\pi(x) \equiv 1$, the choice of the rejection threshold given by the linear step-up procedure is then $\widehat{u} = \alpha\widehat{r}$, where $\widehat{r}$ is the largest solution of the equation $\gamma(\alpha r) = r$. To sum up, we have shown the following result.

Corollary 4.4. *Let $X = (X_1, \dots, X_m)$ be a vector of m i.i.d. real random variables of common continuous c.d.f. F. Consider $(p_t(X))_{t\in[0,1]}$ the p-value process $p_t(X) = G_t(|\{1 \le i \le m\colon X_i \le t\}|)$ for $G_t(k) = \mathbb{P}(Z_t \ge k)$, where Z_t is a binomial variable of parameters (m, t). Assume that the hypothesis space* [0, 1] *is endowed with the Lebesgue measure Λ. Consider the volume rejection of the step-up procedure given by*

$$\widehat{r}(X) = \max\bigl\{r \in [0, 1]\colon \gamma(\alpha r) \ge r\bigr\}, \tag{17}$$

where $\gamma(\cdot)$ is defined by (16). *Then the following FDR control holds*:

$$\mathbb{E}\left[\frac{\Lambda(\{t \in [0, 1]\colon F(t) \le t, p_t(X) \le \alpha\hat{r}(X)\})}{\Lambda(\{t \in [0, 1]\colon p_t(X) \le \alpha\hat{r}(X)\})}\right] \le \alpha. \tag{18}$$

4.2.4. *FDR control for testing the intensity of a Poisson process*

Let us consider the testing setting of Example 2.5. Lemma A.2 states that the p-values process is finite dimensional strong PRDS (on any subset). Thus, it is also finite dimensional weak PRDS (on any subset) by Lemma S-2.2, and Theorem 4.1 leads to a control of the FDR.

Now, we aim at finding a closed formula for the linear step-up procedure ($\beta(x) = x$) using the p-value process $(p_t(X))_t$. Let us consider the particular case where the benchmark intensity $\lambda_0(\cdot)$ is constantly equal to some $\lambda_0 > 0$ while $\lambda(\cdot)$ is L-Lipschitz. Also, to avoid the boundary effects, assume that the process $(N_t)_t$ is observed for $t \in [-1, 2]$ while H_t: "$\lambda(t) \le \lambda_0$" is tested only for $t \in [0, 1]$. In this case, the p-value process is simply given by

$$p_t(X) = G(N_{t+\eta} - N_{t-\eta}), \tag{19}$$

where for any $k \in \mathbb{N}$, $G(k)$ denotes $\mathbb{P}[Z \ge k]$ for Z a Poisson distribution of parameter $2\eta\lambda_0 + L\eta^2$ (note that $G(\cdot)$ is independent of t). Consider the jumps $\{T_j\}_j$ of the process $(N_t)_{t\in[-1,2]}$ and the set $\mathcal{S} = \{s_i\}_{2\le i\le m}$ of the distinct and ordered values of the set $\bigcup_j \{T_j - \eta, T_j + \eta\} \cap (0, 1)$. Moreover, we let $s_1 = 0$ and $s_{m+1} = 1$. Next, since the p-value process is constant on each interval $[s_i, s_{i+1})$, $1 \le i \le m$, we have for any $u \ge 0$,

$$\begin{aligned}\Lambda\bigl(\bigl\{t \in [0, 1]\colon p_t(X) \le u\bigr\}\bigr) &= \sum_{i=1}^{m} (s_{i+1} - s_i)\mathbf{1}\bigl\{p_{s_i}(X) \le u\bigr\} \\ &= \sum_{k=1}^{m} w_k \mathbf{1}\bigl\{q_{\sigma(k)}(X) \le u\bigr\},\end{aligned}$$

where we let $q_i(X) = p_{s_i}(X)$, where σ is a permutation of $\{1, \dots, m\}$ such that $q_{\sigma(1)} \le \cdots \le q_{\sigma(m)}$ and where $w_k = s_{\sigma(k)+1} - s_{\sigma(k)} > 0$ can be interpreted as a "weighting" associated to

$q_{\sigma(k)}$. As a consequence, we get

$$\begin{aligned}\widehat{r}(X) &= \max\left\{r \in [0,1]\colon \sum_{\ell=0}^{m} w_\ell \mathbf{1}\left\{q_{\sigma(\ell)}(X) \le \alpha r\right\} \ge r\right\}\\ &= \max\left\{\sum_{\ell=0}^{k} w_\ell,\ \text{for } k \in \{0,\dots,m\} \text{ s.t. } q_{\sigma(k)}(X) \le \alpha \sum_{\ell=0}^{k} w_\ell\right\},\end{aligned} \tag{20}$$

because since $\widehat{r}(X)$ is a maximum, it is of the form $\sum_{\ell=0}^{k} w_\ell$, $k \in \{0,\dots,m\}$. Note that we should let $q_{\sigma(0)} = 0$ and $w_0 = 0$ to cover the case $\widehat{r}(X) = 0$. Relation (20) only involves a finite number of variables. Thus, $\widehat{r}(X)$ can be easily computed in practice. This is illustrated in Figure 2.

We have proved the following result.

Corollary 4.5. *Let $X = (N_t)_{t\in[-1,2]}$ be a Poisson process with an intensity $\lambda\colon[-1,2] \to \mathbb{R}^+$ L-Lipschitz ($L > 0$) and let $\lambda_0 > 0$. For $\eta \in (0,1]$, consider the p-value process $\{p_t(X)\}_{t\in[0,1]}$ given by* (19). *Assume that the hypothesis space* $[0,1]$ *is endowed with the Lebesgue measure Λ. Then $\widehat{r}(X)$ defined by* (20) *satisfies the following*:

$$\mathbb{E}\left[\frac{\Lambda(\{t \in [0,1]\colon \lambda(t) \le \lambda_0, p_t(X) \le \alpha\hat{r}(X)\})}{\Lambda(\{t \in [0,1]\colon p_t(X) \le \alpha\hat{r}(X)\})}\right] \le \alpha. \tag{21}$$

To illustrate Corollary 4.5, Figure 2 displays the case where $\lambda(t)$ is a truncated triangular signal. The choice of the bandwidth η has been made manually, see Section 6.2 for a discussion on this point.

Remark 4.6. Up to increase the set $\mathcal{S} = \{s_i\}_i$ so that $t \mapsto \mathbf{1}\{\lambda(t) \le \lambda_0\}$ is constant over each $[s_i, s_{i+1})$, the FDR control (21) can be rewritten as

$$\mathbb{E}\left[\frac{\sum_{i=1}^{m}(s_{i+1}-s_i)\mathbf{1}\{\lambda(s_i) \le \lambda_0\}\mathbf{1}\{p_{s_i}(X) \le \alpha\hat{r}(X)\}}{\sum_{i=1}^{m}(s_{i+1}-s_i)\mathbf{1}\{p_{s_i}(X) \le \alpha\hat{r}(X)\}}\right] \le \alpha. \tag{22}$$

Hence, the procedure (20) appears as controlling the discrete FDR-weighting on $\{1,\dots,m\}$ where the weight for rejecting "$\lambda(s_i) \le \lambda_0$" is $(s_{i+1} - s_i)$ and where the initial p-values are $q_i(X) = p_{s_i}(X)$. The rationale behind this is that if $q_i(X) = p_{s_i}(X)$ is below $\hat{r}(X)$, then so are all $p_t(X)$, $t \in [s_i, s_{i+1})$. Hence, a rejection for a p-value $q_i(X) = p_{s_i}(X)$ accounts for the length of the entire interval in the FDR. From an intuitive point of view, this means that the type I error importance in the FDR is larger for "isolated" points of the process. This bears some similarity with discrete multiple testing with weights, Benjamini and Hochberg [2] and Blanchard and Roquain [6], but those results would not apply here since the weights themselves are data-dependent.

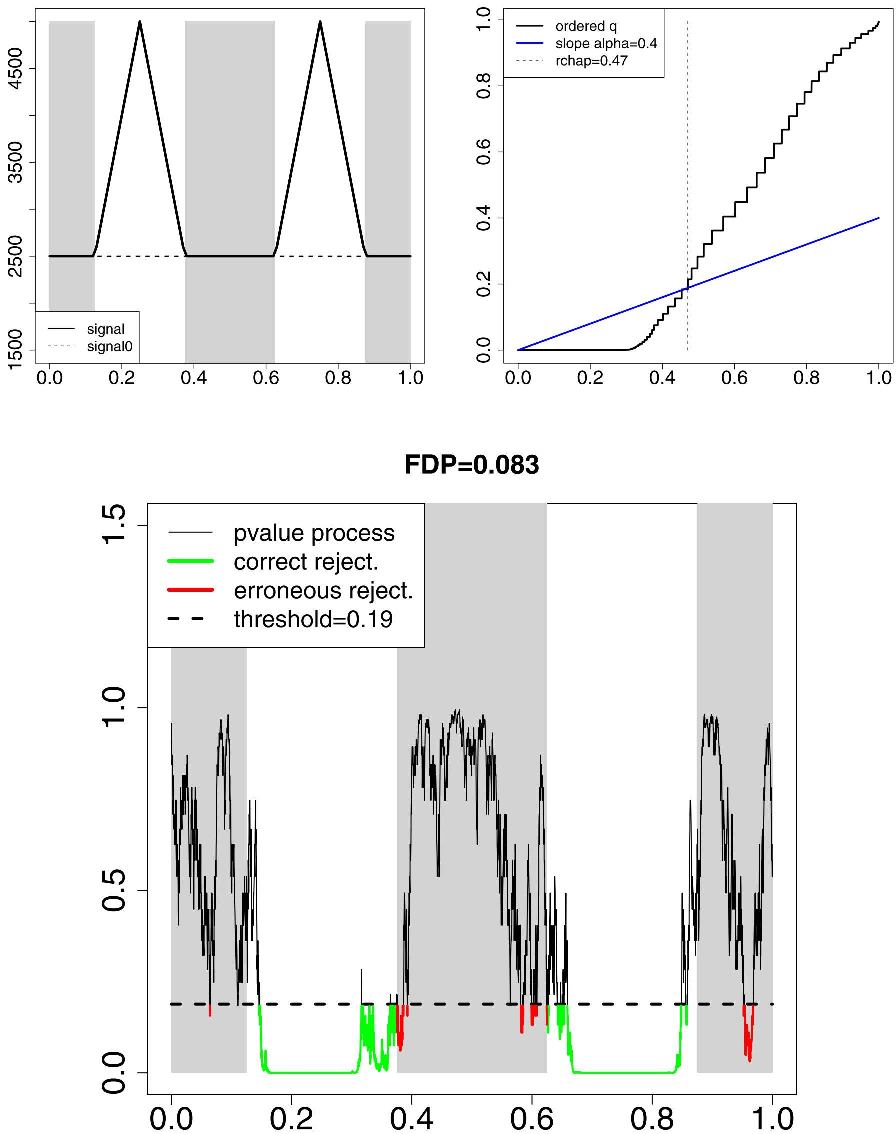

Figure 2. Several plots versus $t \in [0, 1]$. Top left: $\lambda(t)$ (solid) and λ_0 (dashed). Top right: $q_{\sigma(k)}(X)$ and $\alpha \sum_{\ell=0}^{k} w_\ell$ in function of $\sum_{\ell=0}^{k} w_\ell$, for $k = 1, \dots, m$. Bottom: p-value process $p_t(X)$ defined by (19). $\eta = 0.015$, $\alpha = 0.4$. The grey areas indicate regions where the null hypotheses are true.

5. Proof of Theorem 4.1

5.1. Two conditions for controlling the FDR

Similarly to Proposition 2.7 of Blanchard and Roquain [6] (which we refer to as BR08 for short from now on), we can prove that the FDR control $\mathrm{FDR}(R, P) \le \alpha \Pi(\mathcal{H}_0(P))$ holds true for any $P \in \mathcal{P}$ as soon as the two following sufficient conditions hold for any $P \in \mathcal{P}$:

- the multiple testing procedure R satisfies the "self-consistency condition"

$$R(x) \subset \big\{h \in \mathcal{H}\colon\ p_h(x) \le \alpha\pi(h)\beta\big(\Lambda\big(R(x)\big)\big)\big\} \qquad \text{for } P\text{-almost all } x \in \mathcal{X} \quad (\mathrm{SC}(\alpha, \pi, \beta))$$

- for any $h \in \mathcal{H}_0(P)$ the couple of real random variables $(U_h, V) := (p_h(X), \Lambda(R(X)))$ satisfies the "dependence control condition"

$$\forall c > 0, \qquad \mathbb{E}\left[\frac{\mathbf{1}\{U_h \le c\beta(V)\}}{V}\mathbf{1}\{V > 0\}\right] \le c. \qquad (\mathrm{DC}(\beta))$$

The proof is as follows: by definition and by using Fubini's theorem, we have

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

where we have used the shortened notation R for $R(X)$ and p_h for $p_h(X)$, and used successively conditions ($\mathrm{SC}(\alpha, \pi, \beta)$) and ($\mathrm{DC}(\beta)$) for the two above inequalities. Observe that the use of Fubini's theorem is granted by the measurability assumption of Definition 2.1.

Therefore, to obtain the FDR bound of Theorem 4.1 in each case, we simply have to check conditions ($\mathrm{SC}(\alpha, \pi, \beta)$) and ($\mathrm{DC}(\beta)$) in the different settings.

5.2. Any step-up procedure satisfies ($\mathrm{SC}(\alpha, \pi, \beta)$)

From the definition of a step-up procedure, for all $\varepsilon > 0$, we have $\Lambda(L_\Delta(\widehat{r})) \le \Lambda(L_\Delta(\widehat{r} + \varepsilon)) < \widehat{r} + \varepsilon$. This entails that $\widehat{r}$ satisfies $\Lambda(L_\Delta(\widehat{r})) = \widehat{r}$. Hence, the step-up procedure R satisfies $\mathrm{SC}(\alpha, \pi, \beta)$ with equality.

We now check that any step-up procedure is a multiple testing procedure, that is, that $(\omega, h) \mapsto \mathbf{1}\{h \in R(X(\omega))\} = \mathbf{1}\{p_h(X(\omega)) \le \alpha\pi(h)\beta(\widehat{r}(X(\omega)))\}$ is (jointly) measurable. From (A2) and since β and π are measurable, it is enough to check that $\omega \mapsto \widehat{r}(X(\omega))$ is measurable. For any $x \in X(\Omega)$, let us consider the function

$$f: r \in \mathbb{R}^+ \mapsto \Lambda\big(L_\Delta(x,r)\big) = \int_{\mathcal{H}} \mathbf{1}\big\{p_h(x) \le \alpha\pi(h)\beta(r)\big\}\, \mathrm{d}\Lambda(h).$$

We observe that f is right-continuous and non-decreasing (because β is) and bounded, and that $\widehat{r} = \max\{r \ge 0\colon\ f(r) \ge r\}$. Applying Lemma S-2.3, we deduce that for any $x \in X(\Omega)$,

$$\begin{aligned} \widehat{r}(x) &= \inf_{\varepsilon>0, \varepsilon\in\mathbb{Q}} \sup\big\{r \in \mathbb{Q}^+\colon\ \Lambda\big(L_\Delta(x,r)\big) \ge r - \varepsilon\big\} \\ &= \inf_{\varepsilon>0, \varepsilon\in\mathbb{Q}} \sup_{r\in\mathbb{Q}^+} \big(r\mathbf{1}\big\{\Lambda\big(L_\Delta(x,r)\big) \ge r-\varepsilon\big\}\big). \end{aligned} \tag{23}$$

Since from (A2), for all $\varepsilon > 0$, $\varepsilon \in \mathbb{Q}$ and $r \in \mathbb{Q}^+$, the function

$$\omega \mapsto r\mathbf{1}\big\{\Lambda\big(L_\Delta\big(X(\omega), r\big)\big) \ge r - \varepsilon\big\} = r\mathbf{1}\big\{\Lambda\big(\big\{h \in \mathcal{H}\colon\ p_h\big(X(\omega)\big) \le \alpha\pi(h)\beta(r)\big\}\big) \ge r - \varepsilon\big\}$$

is measurable, expression (23) implies that $\omega \mapsto \widehat{r}(X(\omega))$ is measurable. Hence, a step-up procedure satisfies the measurability requirements of Definition 2.1.

5.3. Conditions implying (DC(β))

We use the following lemma which was proved in (BR08) (see Lemma 3.2, items (ii, iii) therein):

Lemma 5.1. *Let (U, V) be a couple of non-negative random variables such that U is stochastically lower bounded by a uniform variable on $[0,1]$, that is, $\forall t \in [0,1], \mathbb{P}(U \le t) \le t$. Then the dependence control condition DC(β) is satisfied by (U, V) in either one of the following situations*:

(i) $\beta(x) = x$ *and*

$$\forall r \in \mathbb{R}^+, \qquad u \mapsto \mathbb{P}(V < r | U \le u) \qquad \text{is non-decreasing on } \big\{u\colon\ P(U \le u) > 0\big\}. \tag{24}$$

(ii) *The shape function β is of the form* (13).

The point (ii) above, together with the results of the two previous sections, establishes point 2 of Theorem 4.1. To establish point 1 and finish the proof, we have to prove that (24) holds in the finite dimensional weak PRDS dependence context, which is done in the following proposition:

Proposition 5.2. *Assume that the p-values process $\mathbf{p} = (p_h, h \in \mathcal{H})$ is finite dimensional weak PRDS on $\mathcal{H}_0(P)$ for any $P \in \mathcal{P}$. Consider R the step-up procedure defined by Definition 3.2 with $\beta(x) = x$. Then for any $P \in \mathcal{P}$, for any $h \in \mathcal{H}_0(P)$, the couple of variables $(U_h, V) = (p_h, \Lambda(R))$ satisfies* (24) *and thus* DC(β) *holds for $\beta(x) = x$.*

Proof. In the above statement and the present proof, we use the shortened notation R, p_h, and $L_\Delta(r)$ for the random quantities $R(X)$, $p_h(X)$, and $L_\Delta(X,r)$, respectively. The goal of the proof is to establish (24), that is for any $h_0 \in \mathcal{H}_0$ (h_0 is assumed to be fixed in $\mathcal{H}_0$ in the rest of the proof), for any t, and $0 \le u \le u'$ with $\mathbb{P}(p_{h_0} \le u) > 0$:

$$\mathbb{P}\big[\Lambda(R) < t \,|\, p_{h_0} \le u\big] \le \mathbb{P}\big[\Lambda(R) < t \,|\, p_{h_0} \le u'\big];$$

From Definition 3.2, the real random variable $\Lambda(R)$ can be rewritten as $\Lambda(R) = \widehat{r} = \max\{r\colon\ f(r) \ge r\}$ with $f : r \mapsto \Lambda(L_\Delta(r))$. Furthermore, denoting $G_u = \frac{\mathbf{1}\{p_{h_0} \le u\}}{\mathbb{P}[p_{h_0} \le u]}$, we are equivalently aiming at proving that for any t and $0 \le u \le u'$ with $\mathbb{P}(p_{h_0} \le u) > 0$:

$$\mathbb{E}\big[\mathbf{1}\{\widehat{r} < t\} G_u\big] \le \mathbb{E}\big[\mathbf{1}\{\widehat{r} < t\} G_{u'}\big]. \tag{25}$$

By using Lemma 5.3 (and the notation therein), there exists a fixed sequence of finitely supported measures Λ_n on $\mathcal{H}$ such that, denoting $\widehat{r}_{n,k} = \max\{r \ge 0\colon\ \Lambda_n(L_\Delta(r)) \ge r - k^{-1}\}$, it holds that

$$\widehat{r} = \lim_{k\to\infty} \widehat{r}_k^{+} = \lim_{k\to\infty} \widehat{r}_k^{-} \qquad \text{almost surely}, \tag{26}$$

where we let $\widehat{r}_k^{+} = \limsup_{n\to\infty} \widehat{r}_{n,k}$ and $\widehat{r}_k^{-} = \liminf_{n\to\infty} \widehat{r}_{n,k}$.

Let $\mathcal{S}_n$ be the (finite) support of Λ_n and $\mathcal{S}'_n = \mathcal{S}_n \cup \{h_0\}$. Writing $\widehat{r}_{n,k}$ as a function of the finite p-value set $\{p_h, h \in \mathcal{S}'_n\}$, the function $\widehat{r}_{n,k}\colon\ \mathbf{z} = (z_h)_{h\in\mathcal{S}'_n} \in [0,1]^{\mathcal{S}'_n} \mapsto \widehat{r}_{n,k}(\mathbf{z})$ is measurable (where the space $[0,1]^{\mathcal{S}'_n}$ is endowed with the standard product Borel σ-field), and is additionally non-increasing in each p-value. Hence, the set $\{\mathbf{z} = (z_h)_{h\in\mathcal{S}'_n}\colon\ \widehat{r}_{n,k}(\mathbf{z}) < t + k^{-1}\}$ is a non-decreasing measurable subset of $[0,1]^{\mathcal{S}'_n}$. Using that the p-value process $\mathbf{p} = (p_h, h \in \mathcal{H})$ is finite dimensional weak PRDS on $\mathcal{H}_0$, the p-values $(p_h, h \in \mathcal{S}'_n)$ are PRDS on $\mathcal{H}_0 \cap \mathcal{S}'_n$, which implies that for any $t \ge 0$ and $u \le u'$ with $\mathbb{P}(p_{h_0} \le u) > 0$,

$$\mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_u\big] \le \mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_{u'}\big]. \tag{27}$$

Now, to prove (25), it suffices to carefully make n and k tend to infinity. By Fatou's lemma and by (27), we have for all $k \ge 1$:

$$\begin{aligned}
\mathbb{E}\Big[\liminf_n \mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_u\Big] &\le \liminf_n \mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_u\big] \\
&\le \limsup_n \mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_{u'}\big] \\
&\le \mathbb{E}\Big[\limsup_n \mathbf{1}\big\{\widehat{r}_{n,k} - k^{-1} < t\big\} G_{u'}\Big].
\end{aligned}$$

Notice that the following inclusions of events hold: $\{\widehat{r}_k^{+} < t + k^{-1}\} \subset \liminf_n \{\widehat{r}_{n,k} < t + k^{-1}\}$, $\limsup_n \{\widehat{r}_{n,k} < t + k^{-1}\} \subset \{\widehat{r}_k^{-} \le t + k^{-1}\}$. Hence, we obtain for all k:

$$\mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_k^{+} - k^{-1} < t\big\} G_u\big] \le \mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_k^{-} - k^{-1} \le t\big\} G_{u'}\big].$$

Then, if t is such that $\mathbb{P}[\widehat{r}=t]=0$, the above expression can be rewritten as

$$\mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_k^+ - k^{-1} < t\big\} G_u \mathbf{1}\{\widehat{r}\neq t\}\big] \le \mathbb{E}\big[\mathbf{1}\big\{\widehat{r}_k^- - k^{-1} \le t\big\} G_{u'} \mathbf{1}\{\widehat{r}\neq t\}\big].$$

We now let $k\to\infty$ in the above expression by using (26) and the dominated convergence theorem: for any $u\le u'$ with $\mathbb{P}(p_{h_0}\le u)>0$, and any $t\notin D:=\{s\ge 0\colon \mathbb{P}[\widehat{r}=s]>0\}$, we have

$$\mathbb{E}\big[\mathbf{1}\{\widehat{r}<t\} G_u\big] \le \mathbb{E}\big[\mathbf{1}\{\widehat{r}<t\} G_{u'}\big]. \tag{28}$$

Since the above expectations may be interpreted as (conditional) probabilities, the LHS and RHS in (28) are left-continuous functions of t. Using that $\mathbb{R}^+\cap D^c$ is dense in $\mathbb{R}^+$ (because D is at most countable), we obtain that (28) holds for any t. Finally, the condition (DC(β)) comes from Lemma 5.1. □

5.4. Finite approximation of step-up procedures

As usual, to lighten notation $R, p_h, L_\Delta(r), \widehat{r}$ denote the random quantities $R(X), p_h(X), L_\Delta(X,r), \widehat{r}(X)$. The following result shows how to derive the continuous step-up procedure (see Definition 3.2) from a limit of finite step-up procedures. It is used in the proof of Proposition 5.2.

Lemma 5.3. *Consider the step-up procedure $R=L_\Delta(\widehat{r})$ on $\mathcal{H}$ using Λ and with $\widehat{r}$ defined in Definition 3.2. Then there exists a sequence of finitely supported measures Λ_n on $\mathcal{H}$ such that, denoting*

$$\widehat{r}_{n,k} = \max\big\{r\ge 0\colon \Lambda_n\big(L_\Delta(r)\big) \ge r - k^{-1}\big\},$$

we have

$$\widehat{r} = \lim_{k\to\infty}\Big(\limsup_{n\to\infty} \widehat{r}_{n,k}\Big) = \lim_{k\to\infty}\Big(\liminf_{n\to\infty} \widehat{r}_{n,k}\Big) \qquad \textit{almost surely.}$$

Proof. We start with the following observation. Consider (Λ_n) some sequence of measures on $\mathcal{H}$ such that $\Lambda_n(\mathcal{H})\equiv M$. For a fixed realization $x\in X(\Omega)$ of X, we consider $f: r\in\mathbb{R}^+\mapsto \Lambda(L_\Delta(x,r))$ and $f_{\Lambda_n}: r\in\mathbb{R}^+\mapsto \Lambda_n(L_\Delta(x,r))$. Clearly, f and f_{Λ_n} are non-decreasing right-continuous functions. Using Lemma S-2.4, we conclude that the desired result holds provided that, for P-almost all $x\in\mathcal{X}$, f_{Λ_n} converges uniformly to f over $[0, M+1]$.

It remains thus to prove that there exists a sequence of finitely supported measures Λ_n on $\mathcal{H}$ such that for P-almost all $x\in\mathcal{X}$,

$$\limsup_{n\to\infty}\Big\{\sup_{r\in[0,M+1]}\big|\Lambda_n\big(L_\Delta(x,r)\big) - \Lambda\big(L_\Delta(x,r)\big)\big|\Big\} = 0. \tag{29}$$

Denote $\mathcal{Y}$ the product space $\mathcal{H}^{\mathbb{N}}$, endowed with the product sigma-algebra. For $y:=(h_i)_{i\ge 1}\in\mathcal{Y}$ some sequence of hypotheses, denote $\Lambda_n^{[y]} = Mn^{-1}\sum_{i=1}^n \delta_{h_i}$ the suitably scaled uniform atomic measure on $(h_1,\dots,h_n)$.

Consider now $Y := (H_i)_{i\geq 1} \in \mathcal{Y}$ an i.i.d. sequence of hypotheses drawn independently of X according to the probability distribution Λ/M on $\mathcal{H}$. Observe that for any fixed $x \in X(\Omega)$, $L_\Delta(x,r) = \{h \in \mathcal{H}\colon\ p_h(x) \leq \alpha\pi(h)\beta(r)\} = \{h \in \mathcal{H}\colon\ q(h,x) \leq \alpha\beta(r)\}$, where

$$q(h,x) := \begin{cases} p_h(x)/\pi(h), & \text{if } \pi(h) > 0; \\ 0, & \text{if } \pi(h) = 0 \text{ and } p_h(x) = 0; \\ \alpha\beta(M+1)+1, & \text{if } \pi(h) = 0 \text{ and } p_h(x) > 0. \end{cases}$$

Thus, applying the Glivenko–Cantelli theorem to the i.i.d. variables $(q(H_i,x))_i$, we deduce that for any $x \in \mathcal{X}(\Omega)$, $\zeta(x,y) = \limsup_{n\to\infty} \sup_{r\in[0,M+1]} |\Lambda_n^{[y]}(L_\Delta(x,r)) - \Lambda(L_\Delta(x,r))| = 0$ for P_Y-almost all realizations y of Y. Observe furthermore that for any fixed r, the function

$$(\omega,y) \in \Omega \times \mathcal{H}^{\mathbb{N}} \mapsto \Lambda_n^{[y]}\big(L_\Delta\big(X(\omega),r\big)\big) = Mn^{-1}\sum_{i=1}^{n} \mathbf{1}\big\{p_{h_i}\big(X(\omega)\big) \leq \alpha\pi(h_i)\beta(r)\big\}$$

is a (jointly) measurable function of (ω,y) by assumption (A2). The inside supremum in (29) can be restricted to rational numbers since the functions involved are right-continuous. Therefore, $(\omega,y) \mapsto \zeta(X(\omega),y)$ is a jointly measurable function in its variables. By Fubini's theorem, this implies that $\mathbb{E}_{X,Y}[\zeta(X,Y)] = 0$; and thus also, for P_Y-almost all $y \in \mathcal{Y}$, $\zeta(x,y) = 0$ for P-almost all $x \in \mathcal{X}$. Since an event of probability 1 is non-empty, there exists a fixed $y \in \mathcal{Y}$ such that $\zeta(x,y) = 0$ for P-almost all $x \in \mathcal{X}$, which gives rise to a sequence of finitely supported measures Λ_n satisfying (29). □

6. Discussion

6.1. FDR control for self-consistent, non-step-up procedures

In some cases, for instance, after a discretization in r or under a global constraint over the admissible geometry of sets of rejected hypotheses, the procedure of interest may not be of the step-up form, while still satisfying the more general condition (SC(α,π,β)) (called self-consistency, see Section 5.1). In that situation, Theorem 4.1 does not apply, because the procedure is not step-up. We proved an extension of Theorem 4.1 holding more generally for (non-increasing) self-consistent procedures, but point 1 of the theorem is established only under a stronger PRDS condition called general PRDS. (On the other hand, the fact that point 2 of Theorem 4.1 remains valid under the more general condition (SC(α,π,β)) is quite immediate.) The general PRDS condition is defined in terms of the entire process X and not only its finite dimensional projections. Therefore, it is substantially more technical than finite dimensional PRDS. In particular, it is an open question to characterize when does finite dimensional PRDS imply general PRDS (we provide some sufficient conditions). For simplicity, we deferred the corresponding study in part II of the supplementary material (Blanchard, Delattre and Roquain [5]).

6.2. Power and adaptive procedures

This work has focused on procedures ensuring control of the type I error as measured by the FDR. Under this constraint, one would like to maximize power. We do not address this issue in the present work; a specific multiple testing power criterion would have to be defined to begin with, for instance the average number of correct rejections. We briefly discuss possible future directions in this regard, in particular adaptivity properties with respect to different types of underlying regularity structure.

Adaptivity of single tests. The power of a multiple testing procedure depends primarily on the power of the underlying single tests and p-values it is built upon. It is of course desirable to design individual tests that are as powerful as possible in the first place. While this issue actually pertains to the domain of single hypothesis testing, and is to this extent quite independent of the methodology studied here, we briefly discuss this issue in the light of the specific example of the Gaussian white noise model $dZ_t = f(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$. For designing a test of the hypothesis $f(t_0) = 0$, we have assumed known regularity of f and considered a test based on a simple kernel estimator. Could this be improved?

There is an abundance of literature on adaptive testing of a global qualitative hypothesis on f (the simplest example being testing that f is identically zero), where adaptation is understood with respect to the (Hölder or Besov) regularity of the alternative and separation from the null is generally measured in some L^p norm. This might give some hope that some form of regularity adaptation is possible also for testing the local hypothesis $f(t_0) = 0$ (and the separation distance $|f(t_0)|$), but the situation is in fact quite different and possibilities for this are severely limited. This is in essence the same phenomenon as for the existence of regularity-adaptive confidence intervals for pointwise estimation of a function, as studied by Cai and Low [7]. We sketch the main arguments here. First, following the discussion in Dümbgen and Spokoiny [8], Section 2, observe that for testing of f_0 against g_0, the power of the optimal NP test is $\Phi(\sigma\|f_0 - g_0\|_2)$, where Φ is a non-decreasing function. Thus, the optimal power of a composite null H_0 against an alternative H_1 is upper bounded (with equality if H_0 and H_1 are convex) by $\Phi(\sigma \inf_{f\in\mathcal{H}_0, g\in\mathcal{H}_1} \|f - g\|_2)$. In the case where $H_0 = \{f \in \mathcal{F}, f(t_0) = 0\}$ and $H_1 = \{f \in \mathcal{F}_1, f(t_0) = \varepsilon\}$, where $\mathcal{F}_1 \subset \mathcal{F}$ are Hölder regularity classes, Cai and Low [7] (Example 1 there) establish that the rate behavior as $\varepsilon \to 0$ of this infimum distance is determined by $\mathcal{F}$ and not by $\mathcal{F}_1$. Therefore, no adaptation to the regularity of the alternative is possible in this configuration, and it is necessary to assume some a priori known regularity class $\mathcal{F}$. On the other hand, these authors show that adaptive confidence intervals (and hence tests) exist in this setting provided some additional shape restrictions, such as monotonicity, are assumed to hold.

Adaptivity to $\Pi(\mathcal{H}_0)$ *and to the dependence structure*. For multiple testing over a finite hypothesis space, recent research has focused on improving step-up procedures to take into account, on the one hand, the (unknown) volume $\Pi(\mathcal{H}_0(P))$ of true null hypotheses – which comes as a nuisance parameter reducing the effective level, see (12), and on the other hand, the dependence structure of the p-values. Both directions suggest further possible developments in the continuous setting as well.

Appendix: PRDS statements

Lemma A.1. *The* p*-value process* $\mathbf{p}(X) = \{p_t(X), t \in I\}$ *defined by* (5) *is finite dimensional weak PRDS* (*on any subset*).

Proof. Let us consider a finite subset $(t_j)_{0\le j\le N-1}$ of I and D a non-decreasing measurable subset of $[0,1]^N$. Let us prove that the function $u \mapsto \mathbb{P}[\mathbf{p}(X) \in D | p_{t_0}(X) \le u]$ is non-decreasing on $\{u \in [0,1]\colon\ \mathbb{P}(p_{t_0}(X) \le u) > 0\}$. If $F(t_0) \in \{0,1\}$, the result is trivial. We thus assume that $F(t_0) \in (0,1)$, so that $\mathcal{U}_{t_0} = \{G_{t_0}(k), k = m, m-1, \dots, 0\}$ contains only increasing points of $(0,1]$. Without loss of generality, we only have to prove the non-decreasing property for $u \in \mathcal{U}_{t_0}$. Since G_{t_0} is decreasing from $\{0,\dots,m\}$ to $\mathcal{U}_{t_0}$, we have $p_{t_0}(X) \le G_{t_0}(k) \Longleftrightarrow m\mathbb{F}_m(X,t_0) \ge k \Longleftrightarrow X_{(k)} \le t_0$ (letting $X_{(0)} = -\infty$). We thus have to prove that for any k, $1 \le k \le m$,

$$\mathbb{P}\big[(X_{(1)},\dots,X_{(m)}) \in D' | X_{(k-1)} \le t_0\big] \ge \mathbb{P}\big[(X_{(1)},\dots,X_{(m)}) \in D' | X_{(k)} \le t_0\big], \qquad (30)$$

where $D' = \{x \in \mathbb{R}^m\colon\ (p_{t_j}(x))_{0\le j\le N-1} \in D\}$ is a non-decreasing subset of $\mathbb{R}^m$ (because $\mathbf{p}$ is coordinate wise non-decreasing, that is, $x \le x' \Rightarrow \forall t,\ p_t(x) \le p_t(x')$). Using that the family of order statistics $\{X_{(i)}\}_i$ has positive regression dependency (see Lemma S-2.1), we derive that the function $f(a,b) = \mathbb{E}[(X_{(1)},\dots,X_{(m)}) \in D' | X_{(k-1)} = a, X_{(k)} = b]$ is non-decreasing in a and b. Therefore, denoting $\gamma = \mathbb{P}[X_{(k)} \le t_0 | X_{(k-1)} \le t_0]$, we get

$$\begin{aligned}\mathbb{P}\big[(X_{(1)},\dots,X_{(m)}) \in D' | X_{(k-1)} \le t_0\big] &= \gamma\mathbb{E}\big[f(X_{(k-1)},X_{(k)}) | X_{(k-1)} \le t_0, X_{(k)} \le t_0\big] \\ &\quad + (1-\gamma)\mathbb{E}\big[f(X_{(k-1)},X_{(k)}) | X_{(k-1)} \le t_0 < X_{(k)}\big] \\ &\ge \mathbb{E}\big[f(X_{(k-1)},X_{(k)}) | X_{(k-1)} \le t_0, X_{(k)} \le t_0\big],\end{aligned}$$

which provides (30) and concludes the proof. □

Lemma A.2. *The* p*-value process* $\mathbf{p}(X) = \{p_t(X), t \in [0,1]\}$ *defined by* (6) *is finite dimensional strong PRDS* (*on any subset*).

Proof. Let $M_t = N_{(t+\eta)\wedge 1} - N_{(t-\eta)\vee 0}$ for any $t \in [0,1]$. Fix $(t_j)_{0\le j\le q-1} \in [0,1]^q$ and assume $t_0 \in [\eta, 1-\eta]$ (the other case can be proved similarly). Take a non-decreasing measurable set $D \subset [0,1]^q$ and consider the set $D' = \{(M_{t_j})_{0\le j\le q-1} \in \mathbb{N}^q\colon\ (G_{t_j}(M_{t_j}))_{0\le j\le q-1} \in D\}$, which is non-increasing on $\mathbb{N}^q$ and measurable. We thus aim to prove that for any $n \ge 0$,

$$\mathbb{P}\big[(M_{t_j})_{0\le j\le q-1} \in D' | M_{t_0} = n+1\big] \le \mathbb{P}\big[(M_{t_j})_{0\le j\le q-1} \in D' | M_{t_0} = n\big]. \qquad (31)$$

Denote by $X_1 < \cdots < X_{k_X}$, $Y_1 < \cdots < Y_{k_Y}$ and $Z_1 < \cdots < Z_{k_Z}$ the jump times of the process $(N_t)_{t\in[0,1]}$ within the (disjoint) subsets $[0, t_0-\eta)$, $[t_0-\eta, t_0+\eta]$ and $(t_0+\eta, 1]$, respectively. Remark that $k_Y = M_{t_0}$ with our notation. Since $(N_t)_{t\in[0,1]}$ is a Poisson process, the family $\{(X_i, 1 \le i \le k_X, k_X), (Y_i, 1 \le i \le k_Y, k_Y), (Z_i, 1 \le i \le k_Z, k_Z)\}$, contains mutually independent elements. Furthermore, the distribution of $(Y_1,\dots,Y_{k_Y})$ conditionally on $k_Y = n$ is equal to the distribution of the order statistics of a sample $(Y_1',\dots,Y_n')$ of i.i.d. random variables with common density $t \mapsto \lambda(t)/\int_{[t_0-\eta,t_0+\eta]} \lambda(s)\,\mathrm{d}s$ on $[t_0-\eta, t_0+\eta]$ (w.r.t. the Lebesgue measure). Next, denoting $I_t = [(t-\eta)\vee 0, (t+\eta)\wedge 1]$, for any $t \in [0,1]$, we can write:

$$\begin{aligned}
&\mathbb{P}\bigl[(M_{t_j})_{0\le j\le q-1} \in D' | M_{t_0} = n+1\bigr] \\
&\quad= \mathbb{P}\Biggl[\Biggl(\sum_{i=1}^{k_X} \mathbf{1}\{X_i \in I_{t_j}\} + \sum_{i=1}^{n+1} \mathbf{1}\bigl\{Y_i' \in I_{t_j}\bigr\} + \sum_{i=1}^{k_Z} \mathbf{1}\{Z_i \in I_{t_j}\}\Biggr)_{0\le j\le q-1} \in D'\Biggr] \\
&\quad= \mathbb{P}\Biggl[\Biggl(\sum_{i=1}^{k_X} \mathbf{1}\{X_i \in I_{t_j}\} + \sum_{i=1}^{n} \mathbf{1}\bigl\{Y_i' \in I_{t_j}\bigr\} + \sum_{i=1}^{k_Z} \mathbf{1}\{Z_i \in I_{t_j}\}\Biggr)_{0\le j\le q-1} \\
&\qquad\qquad \in D' - \bigl(\mathbf{1}\bigl\{Y_{n+1}' \in I_{t_j}\bigr\}\bigr)_j\Biggr] \\
&\quad\le \mathbb{P}\Biggl[\Biggl(\sum_{i=1}^{k_X} \mathbf{1}\{X_i \in I_{t_j}\} + \sum_{i=1}^{n} \mathbf{1}\bigl\{Y_i' \in I_{t_j}\bigr\} + \sum_{i=1}^{k_Z} \mathbf{1}\{Z_i \in I_{t_j}\}\Biggr)_{0\le j\le q-1} \in D'\Biggr] \\
&\quad= \mathbb{P}\bigl[(M_{t_j})_{0\le j\le q-1} \in D' | M_{t_0} = n\bigr],
\end{aligned}$$

the inequality coming from $D' - (\mathbf{1}\{Y_{n+1}' \in I_{t_j}\})_j \subset D'$, because D' is non-increasing. This proves (31) and concludes the proof. □

Acknowledgements

This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. The third author was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01). The three authors where supported by the French ministry of foreign and european affairs (EGIDE – PROCOPE project number 21887 NJ).

Supplementary Material

Supplement to: "Testing over a continuum of null hypotheses with False Discovery Rate control" (DOI: 10.3150/12-BEJ488SUPP; .pdf). This supplement provides some technical results and introduces the so-called general PRDS condition, which is a stronger assumption than the finite dimensional PRDS condition. This condition is useful to prove FDR control for procedures which are not necessarily of the step-up type.

References

[1] 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** 289–300. MR1325392

[2] Benjamini, Y. and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.* **24** 407–418. MR1481424

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

[4] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1700749

[5] Blanchard, G., Delattre, S. and Roquain, E. (2014). Supplement to "Testing over a continuum of null hypotheses with False Discovery Rate control." DOI:10.3150/12-BEJ488SUPP.

[6] Blanchard, G. and Roquain, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.* **2** 963–992. MR2448601

[7] Cai, T.T. and Low, M.G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494

[8] Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[9] Ingster, Y.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Probl. Inf. Transm.* **18** 130–140.

[10] Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I–III. *Math. Methods Statist.* **2** 85–114; 171–189; 249–268.

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

[12] Perone Pacifico, M., Genovese, C., Verdinelli, I. and Wasserman, L. (2007). Scan clustering: A false discovery approach. *J. Multivariate Anal.* **98** 1441–1469. MR2364129

[13] Robin, S. (2002). A compound Poisson model for word occurrences in DNA sequences. *J. Roy. Statist. Soc. Ser. C* **51** 437–451. MR1974071

[14] Schmidt-Hieber, J., Munk, A. and Duembgen, L. (2011). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. Available at arXiv:1107.1404.

[15] Schwartzman, A., Gavrilov, Y. and Adler, R.J. (2011). Multiple testing of local maxima for detection of peaks in 1D. *Ann. Statist.* **39** 3290–3319.

[16] Seeger, P. (1968). A note on a method for the analysis of significances en masse. *Technometrics* **10** 586–593.

[17] van der Laan, M.J., Dudoit, S. and Pollard, K.S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.* **3** 27 pp. (electronic). MR2101464

Received November 2011 and revised September 2012