The Annals of Statistics
2015, Vol. 43, No. 3, 1141–1177
DOI: 10.1214/14-AOS1302

NEW PROCEDURES CONTROLLING THE FALSE DISCOVERY PROPORTION VIA ROMANO–WOLF'S HEURISTIC

BY SYLVAIN DELATTRE AND ETIENNE ROQUAIN[1]

Université Paris Diderot and Sorbonne Universités

The false discovery proportion (FDP) is a convenient way to account for false positives when a large number m of tests are performed simultaneously. Romano and Wolf [*Ann. Statist.* **35** (2007) 1378–1408] have proposed a general principle that builds FDP controlling procedures from k-family-wise error rate controlling procedures while incorporating dependencies in an appropriate manner; see Korn et al. [*J. Statist. Plann. Inference* **124** (2004) 379–398]; Romano and Wolf (2007). However, the theoretical validity of the latter is still largely unknown. This paper provides a careful study of this heuristic: first, we extend this approach by using a notion of "bounding device" that allows us to cover a wide range of critical values, including those that adapt to m_0, the number of true null hypotheses. Second, the theoretical validity of the latter is investigated both nonasymptotically and asymptotically. Third, we introduce suitable modifications of this heuristic that provide new methods, overcoming the existing procedures with a proven FDP control.

1. Introduction.

1.1. *Motivation.* Assessing significance in massive data is an important challenge of contemporary statistics, which becomes especially difficult when the underlying errors are correlated. Pertaining to this class of high-dimensional problems, a common issue is to make simultaneously a huge number m of $0/1$ decisions with a valid control of the overall amount of false discoveries (items declared to be wrongly significant). In this context, a convenient way to account for false discoveries is the false discovery proportion (FDP) that corresponds to the proportion of errors among the items declared as significant (i.e., "1") by the procedure.

The Benjamini and Hochberg (BH) procedure has been widely popularized after the celebrated paper Benjamini and Hochberg (1995) and is shown to control the *expectation* of the FDP, called the false discovery rate (FDR), either theoretically under constrained dependency structures [see Benjamini and Yekutieli (2001)] or with simulations; see Kim and van de Wiel (2008). However, many authors have noticed that the distribution of the FDP of BH procedure can be affected by the

Received June 2014; revised December 2014.

[1]Supported by the French Agence Nationale de la Recherche (ANR Grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01).

MSC2010 subject classifications. Primary 62H15; secondary 60F17.

Key words and phrases. Multiple testing, false discovery rate, equi-correlation, Gaussian multivariate distribution, positive dependence, Simes's inequality.

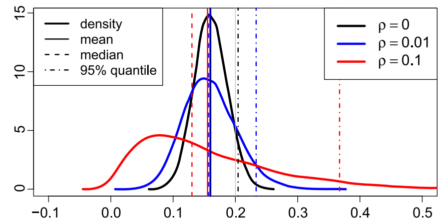

FIG. 1. *Fitted density of the false discovery proportion of the BH procedure when increasing the dependence.* $m = 1000$, $m_0 = 800$ (*number of true null hypotheses*), 10^4 *simulations*, *Gaussian one-sided equicorrelated model.*

dependencies [see, e.g., Delattre and Roquain (2011), Guo, He and Sarkar (2014), Korn et al. (2004)], which makes the use of the BH procedure questionable.

To illustrate further this phenomenon, Figure 1 displays the distribution of the FDP of the BH procedure in the classical one-sided Gaussian multiple testing framework, when the m test statistics are all ρ-equicorrelated. As ρ increases, the distribution of the FDP becomes less concentrated and turns out to be drastically skewed for $\rho = 0.1$ (in particular, it falls outside the Gaussian regime). Clearly, in this case, the mean fails to describe accurately the overall behavior of the FDP distribution. In particular, although the mean of the FDP is below 0.2 [as proved in Benjamini and Yekutieli (2001)], the true value of FDP is not ensured to be small in this case.

An alternative proposed in Genovese and Wasserman (2004), Lehmann and Romano (2005), Perone Pacifico et al. (2004) is to control the $(1 - \zeta)$-quantile of the FDP distribution at level α, that is, to assert

$$\mathbb{P}(\mathrm{FDP} > \alpha) \leq \zeta. \tag{1}$$

While taking $\zeta = 1/2$ into (1) provides a control of the median of the FDP, taking $\zeta = 0.05$ ensures that the FDP does not exceed α with probability at least 95%. Markedly, Figure 1 shows that the $(1 - \zeta)$-quantiles of the FDP distribution are substantially affected by the dependencies, but not equally for all the ζ's: while the 95%-quantile gets substantially larger, the median gets slightly smaller. This suggests that the BH procedure is much too optimistic for a 95%-quantile control, but is actually too conservative for a FDP median control. Overall, this reinforces the fact that in the presence of strong dependence, controlling the $(1 - \zeta)$-quantile of the FDP is an essential task, not covered by the BH procedure.

1.2. *RW's heuristic and main contributions of this paper.* The problem of finding multiple testing procedures ensuring the control (1) has received growing attention in the last decades; see, for instance, Chi and Tan (2008), Dudoit and van der Laan (2008), Guo, He and Sarkar (2014), Guo and Romano (2007), Lehmann and Romano (2005), Romano and Shaikh (2006a, 2006b), Romano and Wolf (2007), Roquain (2011), Roquain and Villers (2011). However, existing procedures with a proven FDP control are in general too conservative. This increases the interest of simple and general heuristics that work "fairly." Romano and Wolf (2007), themselves referring to Korn et al. (2004), have proposed such a heuristic. It is called RW's heuristic in the sequel and can be formulated as follows.

Start from a family $\mathcal{R}_k$, $k \in \{1, \dots, m\}$, of procedures such that for all k, with probability at least $1-\zeta$, the procedure $\mathcal{R}_k$ makes less than $k-1$ false discoveries. Then, choose some $\hat{k}$ such that $(\hat{k}-1)/R_{\hat{k}} \leq \alpha$, where R_k denotes the number of rejections of $\mathcal{R}_k$. Finally use $\mathcal{R}_{\hat{k}}$.

Note that, in the original formulation, $\hat{k}$ was constrained to be chosen such that any k' with $k' < \hat{k}$ should also satisfy $(k'-1)/R_{k'} \leq \alpha$ ("step-down" approach). This constraint is not necessarily applied here (e.g., "step-up" approach is allowed). The rationale behind this principle is that, for each k, the FDP of $\mathcal{R}_k$ is bounded by $(k-1)/R_k$ with probability $1-\zeta$, so that the FDP of $\mathcal{R}_{\hat{k}}$ should be smaller than $(\hat{k}-1)/R_{\hat{k}} \leq \alpha$ with probability $1-\zeta$, which entails (1). However, as it is, this argument is not rigorous because it does not take into account the fluctuations of $\hat{k}$.

This heuristic has been theoretically justified (in the step-down form) in settings where the p-values under the null are independent of the p-values under the alternative [full independence in Guo and Romano (2007); alternative p-values all equal to 0 in Romano and Wolf (2007)]. Since these situations rely on an independence assumption, and since the FDP is particularly interesting under dependence, it seems appropriate to study the precise behavior of this method in "simple" dependent cases. Thus our study is guided by the case where the dependencies are *known*, Gaussian multivariate or carried by latent variables.

In a nutshell, this paper makes the following main contributions:

- It provides a general framework in which RW's heuristic can be investigated, by building the initial k-FWE critical values with "bounding devices": a strong interest is the possibility to build critical values that "adapt" to m_0, the number of true nulls. This allows to encompass many procedures, either new or previously known.
- We show that RW's heuristic may fail to control the FDP nonasymptotically (even under its step-down form). Two corrections that provably control the FDP are introduced. By using simulations, we show that the resulting procedures are more powerful than those previously existing.
- We provide some asymptotic properties of RW's heuristic (in its step-up form): first, we show that it is valid under weak dependence. In addition, we argue that the interest of the latter is only moderate by proving that the simple BH procedure is also valid in this case. Second, we provide particular types of strong dependence for which RW's heuristic can be justified. As a simple illustration, in a ρ-equicorrelated one-sided Gaussian framework, we prove the asymptotic FDP control holds for the step-up procedure using the following new critical

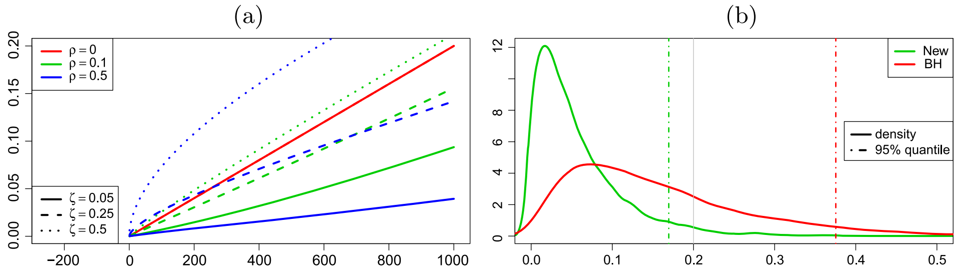

FIG. 2. (a) *plot of the critical values* (2) *in function of* ℓ. (b) *same as Figure* 1 *but only for* $\rho = 0.1$ *and by adding the new step-up procedure using* (2). $m = 1000$, $\alpha = 0.2$, *Gaussian one-sided* ρ*-equi-correlated model.*

values:

$$\tau_\ell = \bar{\Phi}\big(\rho^{1/2}\bar{\Phi}^{-1}(\zeta) + (1-\rho)^{1/2}\bar{\Phi}^{-1}(\alpha\ell/m)\big), \qquad 1 \le \ell \le m, \tag{2}$$

where $\bar{\Phi}$ is the upper-tail of the standard normal distribution.

Finally, let us emphasize that the critical values (2) allow us to describe how the quantities α, ζ and ρ come into play when controlling (asymptotically) the FDP. Taking $\rho = 0$ just gives Simes's critical values, and thus the BH procedure, whatever ζ is. The asymptotic FDP control can be explained in this case by the fast concentration of the FDP of BH around its expectation as m grows to infinity under independence; see, for example, Neuvial (2008). Now, for $\rho > 0$, the new critical values are markedly different from the BH critical values: taking $\zeta = 1/2$ leads to less conservative critical values (if $\alpha \le 1/2$), while taking ζ smaller can lead to more conservativeness (as expected); see Figure 2(a) for an illustration. Finally, we plot in Figure 2(b) the density of the FDP of the step-up procedure using the new critical values (2) for $\zeta = 0.05$. As one might expect, compared to the BH procedure, the density has been shifted to the left so that the 95%-quantile of the FDP of the novel procedure is below α.

1.3. *Multiple testing framework.* We observe a random variable X, whose distribution belongs to some set $\mathcal{P}$. For $m \ge 2$, we define a setting for performing m tests simultaneously by introducing a true/false null parameter $H \in \{0,1\}^m$ and a set of associated distributions $\mathcal{P}_H \subset \mathcal{P}$ which are candidates to be the distribution of X under the configuration H. We denote $\mathcal{H}_0(H) = \{i : H_i = 0\}$, $m_0(H) = \sum_{i=1}^m (1 - H_i)$ and $\mathcal{H}_1(H) = \{i : H_i = 1\}$, $m_1(H) = \sum_{i=1}^m H_i$ the set/number of true and false nulls, respectively. The basic assumption is the following: for all $i \in \{1, \dots, m\}$, there is a p-value $p_i(X)$ satisfying the following assumption:

$$\forall H \in \{0,1\}^m \text{ with } H_i = 0, \forall P \in \mathcal{P}_H, \forall t \in [0,1], \qquad \mathbb{P}_{X\sim P}(p_i(X) \le t) \le t.$$

In this paper, a leading example is the one-sided location model,

$$X_i = H_i \mu_i + Y_i, \qquad 1 \le i \le m, \tag{3}$$

where $H \in \{0, 1\}^m$, $\mu \in (\mathbb{R}_+ \setminus \{0\})^m$ and Y is a m-dimensional centered random vector with identically distributed components. Then the p-values are given by $p_i(X) = \bar{F}(X_i)$, where $\bar{F}(x) = \mathbb{P}(Y_1 \ge x)$, $x \in \mathbb{R}$. Note that this model implicitly assumes that the p-values under the null are uniformly distributed. In this paper, we will often assume that the joint distribution of the noise Y is known, and we consider the two following models for Y:

- Gaussian: Y is a Gaussian vector with covariance matrix Γ (such that $\Gamma_{i,i} = 1$ for simplicity), in which case $Y_1 \sim \mathcal{N}(0, 1)$ and $\bar{F}$ is denoted by $\bar{\Phi}$. A simple particular case is the equi-correlated case,

$$\Gamma_{i,j} = \rho \qquad \text{for all } i \ne j, \text{ where } \rho \in \left[-(m-1)^{-1}, 1\right]. \tag{Gauss-ρ-equi}$$

- Mixture of (1-)factor models: the distribution of Y is given by

$$Y_i = c_i W + \xi_i, \qquad 1 \le i \le m, \tag{facmod}$$

  where c_i, $1 \le i \le m$, are i.i.d., ξ_i, $1 \le i \le m$, are i.i.d., W is a random variable and $(c_i)_{1 \le i \le m}$, $(\xi_i)_{1 \le i \le m}$ and W are independent. Also, the distributions of W, c_1 and ξ_1 are assumed to be known, so that the function $\bar{F}$ is known, and the p-values can be computed. A simple particular case is obtained as follows: for $\rho \in [0, 1]$,

$$Y_i = \varepsilon_i \rho^{1/2} W + (1-\rho)^{1/2} \zeta_i, \qquad 1 \le i \le m, \tag{alt-ρ-equi}$$

  where $W, \zeta_1, \dots, \zeta_m$ are i.i.d. $\mathcal{N}(0, 1)$ and are independent of $\varepsilon_1, \dots, \varepsilon_m$ which are i.i.d. random signs following the distribution $(1-a)\delta_{-1} + a\delta_1$, for a parameter $a \in [0, 1]$.

While the Gaussian model is classical and widely used, (facmod) is useful to model a strong dependence, through the factor W. When the c_i's are deterministic, the latter is often referred to as a *one factor model* in the literature, see, for example, Fan, Han and Gu (2012), Friguet, Kloareg and Causeur (2009), Leek and Storey (2008). Here, the c_i's are unknown and taken randomly with a prescribed distribution. From an intuitive point of view, (facmod) is modeling situations where some of the measurements have been deteriorated by unknown nuisance factors $c_i W$, $1 \le i \le m$. For instance, choosing $c_i \in \{0, 1\}$ corresponds to simultaneously deteriorate the measurements of some unknown subgroup $\{1 \le i \le m : c_i = 1\} \subset \{1, \dots, m\}$. Furthermore, note that while the model (alt-ρ-equi) covers (Gauss-ρ-equi) when $\rho \ge 0$ by taking $a = 0$, (alt-ρ-equi) is able to include negative dependence between some of the Y_i's.

In (facmod), a quantity of interest throughout the paper is the probability that a p-value is below t conditionally on $W = w$ (under the null). According to the particular setting that is at hand, this probability can be written as follows: for $\rho \in [0, 1)$, $w \in \mathbb{R}$,

$$(F_0\text{-facmod}) \qquad F_0(t, w) = \mathbb{P}\big(\bar{F}(c_1 w + \xi_1) \le t\big) = \mathbb{E}\big[\bar{F}_\xi\big(\bar{F}^{-1}(t) - c_1 w\big)\big];$$

$$(F_0\text{-alt-}\rho\text{-equi}) \qquad F_0(t, w) = (1 - a) f(t, -w, \rho) + a f(t, w, \rho);$$

$$(F_0\text{-Gauss-}\rho\text{-equi}) \qquad F_0(t, w) = f(t, w, \rho),$$

where $\bar{F}_\xi(x) = \mathbb{P}(\xi_1 \ge x)$ and $f(t, w, \rho) = \bar{\Phi}((\bar{\Phi}^{-1}(t) - \rho^{1/2} w)/(1 - \rho)^{1/2})$.

REMARK 1.1 (Modifications of the test statistics). Let us consider the model (facmod), where c_i is equal to some known constant; (Gauss-ρ-equi) is one typical instance. Then, as noted by a referee, applying a re-centering operation to the X_i's makes the factor W disappear, and thus can lead to better test statistics (if the bias induced by this operation is not too large); see Section S-1 in the Supplementary Material for more details on this issue. In this respect, our work is particularly relevant in cases where W cannot be estimated (but has a known distribution). On the other hand, we believe that model (Gauss-ρ-equi) keeps the virtue of simplicity and hence remains interesting when studying procedures that are supposed to deal with strong dependencies. Hence while our procedures will in general not be restricted to model (Gauss-ρ-equi), we will also use this model for illustrative purposes throughout the paper.

In the Gaussian case, the joint distribution of the p-values under the null $(p_i, i \in \mathcal{H}_0(H))$ depends, in general, on the subset $\mathcal{H}_0(H)$. Obviously, in this case, we do not want to explore the $\binom{m}{m_0(H)}$ possible subsets of $\{1, \dots, m\}$ in our inference, which inevitably should arise when our procedure fits to such a dependence structure. To circumvent this technical difficulty, we can add random effects to our model. This makes H become random. More formally, we distinguish between the two following models:

- Fixed mixture model: the parameter H is fixed by advance and unknown. Overall, the parameters of the model are given by $\theta = (H, P)$ to be chosen in the set
$$\Theta^F = \big\{(H, P) : H \in \{0, 1\}^m, P \in \mathcal{P}_H\big\}.$$
- Uniform mixture model: the number of true null $m_0 \in \{0, 1, \dots, m\}$ is unknown and fixed by advance, while H is a random vector distributed in such a way that $\mathcal{H}_0(H)$ is randomly generated (independently and previously of the other variables), uniformly in the subsets of $\{1, \dots, m\}$ of cardinal m_0. The parameters of the model are given by $\theta = (m_0, (P_H)_{H : m_0(H) = m_0})$, to be chosen in the set
$$\Theta^U = \big\{(m_0, (P_H)_{H : m_0(H) = m_0}) : m_0 \in \{0, 1, \dots, m\},$$
$$P_H \in \mathcal{P}_H \text{for all } H : m_0(H) = m_0\big\}.$$

  In this model, the distribution of X conditionally on H is P_H.

While the fixed mixture model is the most commonly used model for multiple testing, the uniform mixture model is new to our knowledge and follows the general philosophy of models with random effects; see Efron et al. (2001). It is convenient for the adaptation issue w.r.t. m_0, as we will see later on. With some abuse, we denote $m_0(\theta)$, $\mathcal{H}_0(\theta)$ (or m_0, $\mathcal{H}_0$ when not ambiguous) the number of true nulls in the fixed/uniform mixture models. In the sequel, Θ denotes either Θ^F or Θ^U.

1.4. *Type I error rates.* First, for $t \in [0, 1]$, denote by $V_m(t) = \sum_{i=1}^m (1 - H_i)\mathbf{1}\{p_i(X) \le t\}$ and $R_m(t) = \sum_{i=1}^m \mathbf{1}\{p_i(X) \le t\}$ the number of false discoveries and the number of discoveries (at threshold t), respectively. For some pre-specified $k \in \{1, \dots, m\}$ and some thresholding method $\hat{t}_m \in [0, 1]$ (potentially depending on the data), the k-family-wise error rate (k-FWER) is defined as the probability that more than k true nulls have a p-value smaller than $\hat{t}_m$; see, for example, Hommel and Hoffman (1988), Lehmann and Romano (2005). Formally, for $\theta \in \Theta$ (in one of the models defined in Section 1.3 and Θ being the corresponding parameter space),

$$k\text{-FWER}(\hat{t}_m) = \mathbb{P}_\theta\big(V_m(\hat{t}_m) \ge k\big). \tag{4}$$

Note that $k = 1$ corresponds to the traditional family-wise error rate (FWER). From (4), providing k-FWER$(\hat{t}_m) \le \zeta$ (for all $\theta \in \Theta$), ensures that, with probability at least $1 - \zeta$, less than $k - 1$ false discoveries are made by the thresholding procedure $\hat{t}_m$.

Next, for some threshold $t \in [0, 1]$, define the false discovery proportion at threshold t as follows:

$$\mathrm{FDP}_m(t) = \frac{V_m(t)}{R_m(t) \vee 1}. \tag{5}$$

Note that the quantity $\mathrm{FDP}_m(t)$ is random and not observable because it depends on the unknown process $V_m(t)$. Controlling the FDP via a threshold $t = \hat{t}_m$ (potentially depending on the data) corresponds to the following probabilistic bound:

$$\forall \theta \in \Theta \qquad \mathbb{P}_\theta\big(\mathrm{FDP}_m(\hat{t}_m) \le \alpha\big) \ge 1 - \zeta, \tag{6}$$

for some pre-specified values $\alpha, \zeta \in (0, 1)$. As mentioned before, (6) corresponds to upper-bounding the $(1 - \zeta)$-quantile of the distribution of $\mathrm{FDP}_m(\hat{t}_m)$ by α. Since $\mathrm{FDP}_m(t) > \alpha$ is equivalent to $V_m(t) \ge \lfloor \alpha R_m(t) \rfloor + 1$, the FDP control and the k-FWER control are intrinsically linked.

From a historical point of view, the introduction of the FDP goes back to Eklund in the 1960s [as reported in Seeger (1968)], who has presented the FDP as a solution to the “mass-significance problem.” Much later, the seminal paper of Benjamini and Hochberg (1995) has widely popularized the use of the FDP in practical problems by introducing and studying the false discovery rate (FDR), which corresponds to the expectation of the FDP.

1.5. *Step-up and step-down procedures.* Let us consider the ordered p-values $p_{(1)} \le \cdots \le p_{(m)}$. Consider a nondecreasing sequence $(\tau_\ell)_{1\le\ell\le m}$ of nonnegative values, referred to as the *critical values*. The corresponding step-up (resp., step-down) procedure is defined as rejecting the p-values smaller than $\tau_{\hat{\ell}}$, where $\hat{\ell}$ is defined by either of the two following quantities (with the convention $p_{(0)} = 0$, $\tau_0 = 0$):

$$\text{(SU)} \qquad \max\{\ell \in \{0, 1, \dots, m\} \text{ such that } p_{(\ell)} \le \tau_\ell\};$$

$$\text{(SD)} \qquad \max\{\ell \in \{0, 1, \dots, m\} \text{ such that } \forall \ell' \in \{0, 1, \dots, \ell\}, p_{(\ell')} \le \tau_{\ell'}\}.$$

Let us also recall the so-called *switching relation*: $p_{(\ell)} \le \tau_\ell$ is equivalent to $R_m(\tau_\ell) \ge \ell$. This entails $R_m(\tau_{\hat{\ell}}) = \hat{\ell}$ both in the step-up and step-down cases.

2. Building k-FWE-based critical values.

2.1. *Revisiting RW's heuristic.* Starting from arbitrary critical values $(\tau_\ell)_{1\le\ell\le m}$, and by taking an integer $\hat{\ell}$ such that $R_m(\tau_{\hat{\ell}}) = \hat{\ell}$, we have

$$\begin{aligned} \mathbb{P}_\theta(\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha) &= \mathbb{P}_\theta(V_m(\tau_{\hat{\ell}}) > \alpha R_m(\tau_{\hat{\ell}})) \\ &= \mathbb{P}_\theta(V_m(\tau_{\hat{\ell}}) \ge \lfloor \alpha \hat{\ell} \rfloor + 1). \end{aligned} \tag{7}$$

Hence, by taking τ_ℓ such that $(\lfloor \alpha\ell \rfloor + 1)$-FWER$(\tau_\ell) \le \zeta$ for all ℓ, we should get that (7) is below ζ. However, as already mentioned, the above reasoning does not rigorously establish (6) (with $\hat{t}_m = \tau_{\hat{\ell}}$) because it implicitly assumes that $\hat{\ell}$ is deterministic. Nevertheless, this heuristic is a suitable starting point for building critical values related to the FDP control.

2.2. *Bounding device.* Let us consider either the fixed model $\Theta = \Theta^F$ or the uniform model $\Theta = \Theta^U$. First, let us define a bounding device as any function $B_m^0 : (t, k, u) \mapsto B_m^0(t, k, u) \in [0, 1]$, defined for $t \in [0, 1]$, $k \in \{1, \dots, m\}$ and $u \in \{0, \dots, m\}$, which is nonincreasing in k, with $B_m^0(0, k, u) = 0$ for all u, k, $B_m^0(t, k, u) = 0$ for all $t \in [0, 1]$ whenever $u < k$, and such that for all $t \in [0, 1]$, $k \in \{1, \dots, m\}$ and $u \in \{k, \dots, m\}$, we have

$$\text{(Bound)} \qquad B_m^0(t, k, u) \ge \sup_{\substack{\theta\in\Theta \\ m_0(\theta)=u}} \{\mathbb{P}_\theta(V_m(t) \ge k)\}.$$

Now, define for $t \in [0, 1]$, $k \in \{1, \dots, m\}$ and $\ell \in \{k, \dots, m\}$, the quantities

$$\text{(Bound-nonadapt)} \qquad \bar{B}_m(t, k) = \sup_{0\le u\le m} \{B_m^0(t, k, u)\};$$

$$\text{(Bound-adapt)} \qquad \tilde{B}_m(t, k, \ell) = \sup_{k\le k'\le \ell} \Big\{ \sup_{0\le u\le m-\ell+k'} B_m^0(t, k', u)\Big\},$$

which are additionally assumed to be nondecreasing and left-continuous in t. Note that $\bar{B}_m(t, k)$ and $\tilde{B}_m(t, k, \ell)$ are both nonincreasing in k.

DEFINITION 2.1. Let us consider a bounding device $B_m^0(t,k,u)$ and the above associated quantities $\bar{B}_m(t,k)$ and $\tilde{B}_m(t,k,\ell)$. Then the nonadaptive (resp., adaptive, oracle) k-FWE-based critical values associated to the bounding function B_m^0 are defined as follows (resp.):

$$\bar{\tau}_\ell = \max\{t \in [0,1] : \bar{B}_m(t, \lfloor \alpha\ell \rfloor + 1) \le \zeta\}, \qquad 1 \le \ell \le m; \tag{8}$$

$$\tilde{\tau}_\ell = \max\{t \in [0,1] : \tilde{B}_m(t, \lfloor \alpha\ell \rfloor + 1, \ell) \le \zeta\}, \qquad 1 \le \ell \le m; \tag{9}$$

$$\tau_\ell^0 = \max\{t \in [0,1] : B_m^0(t, \lfloor \alpha\ell \rfloor + 1, m_0) \le \zeta\}, \qquad 1 \le \ell \le m. \tag{10}$$

The above definition implies that $(\bar{\tau}_\ell)_{1\le\ell\le m}$, $(\tilde{\tau}_\ell)_{1\le\ell\le m}$ and $(\tau_\ell^0)_{1\le\ell\le m}$ are nondecreasing sequences, so that they can be used as critical values. The critical values $\tilde{\tau}_\ell$, $\ell = 1,\dots,m$, are said to be *adaptive* because they implicitly (over-)estimate m_0 by

$$m(\ell) = m - \ell + \lfloor \alpha\ell \rfloor + 1. \tag{11}$$

In the literature, this way to adapt to π_0 is often referred to as *one-stage* [in contrast to *two-stage*; see Benjamini, Krieger and Yekutieli (2006), Blanchard and Roquain (2009), Sarkar (2008)]. It has been proved to be asymptotically optimal in a specific sense; see Finner, Dickhaus and Roters (2009). Also, $\bar{\tau}_\ell \le \tilde{\tau}_\ell$ for all ℓ; that is, adaptation always leads to less conservative critical values. Finally, it is worth to check that $\bar{\tau}_m \le \tilde{\tau}_m < 1$ [this comes from $B_m^0(1,k,u) = 1$ for all $u \ge k$] so that the output $\hat{\ell}$ of the step-up algorithm is not identically equal to m.

2.3. *Examples*. We provide below three examples of bounding devices: Markov, K-Markov and Exact. Instances of resulting critical values are displayed in Figure 3 under Gaussian equi-correlation (see Figure S-2 for similar pictures under alternate equi-correlation). As we will see, while the exact bounding device leads to the largest critical values, the Markov-type devices are still useful because they can offer finite sample controls. Also note that in all these examples, we have $\tilde{B}_m(t,k,\ell) = B_m^0(t,k,m-\ell+k)$.

Markov. By Markov's inequality, we have

$$\mathbb{P}_\theta(V_m(t) \ge k) \le \frac{\mathbb{E}_\theta(V_m(t))}{k} = \frac{m_0 t}{k} =: B_m^0(t,k,m_0). \tag{12}$$

Since $\bar{B}_m(t,k) = mt/k$ and $\tilde{B}_m(t,k,\ell) = (m-\ell+k)t/k$, this gives rise to the critical values

$$\bar{\tau}_\ell = \frac{\zeta(\lfloor \alpha\ell \rfloor + 1)}{m}; \qquad \tilde{\tau}_\ell = \frac{\zeta(\lfloor \alpha\ell \rfloor + 1)}{m(\ell)}, \qquad 1 \le \ell \le m, \tag{13}$$

where $m(\ell)$ is defined by (11). The adaptive critical values $(\tilde{\tau}_\ell)_{1\le\ell\le m}$ are those proposed by Lehmann and Romano (2005). Note that these critical values do not adapt to the underlying dependence structure of the p-values.

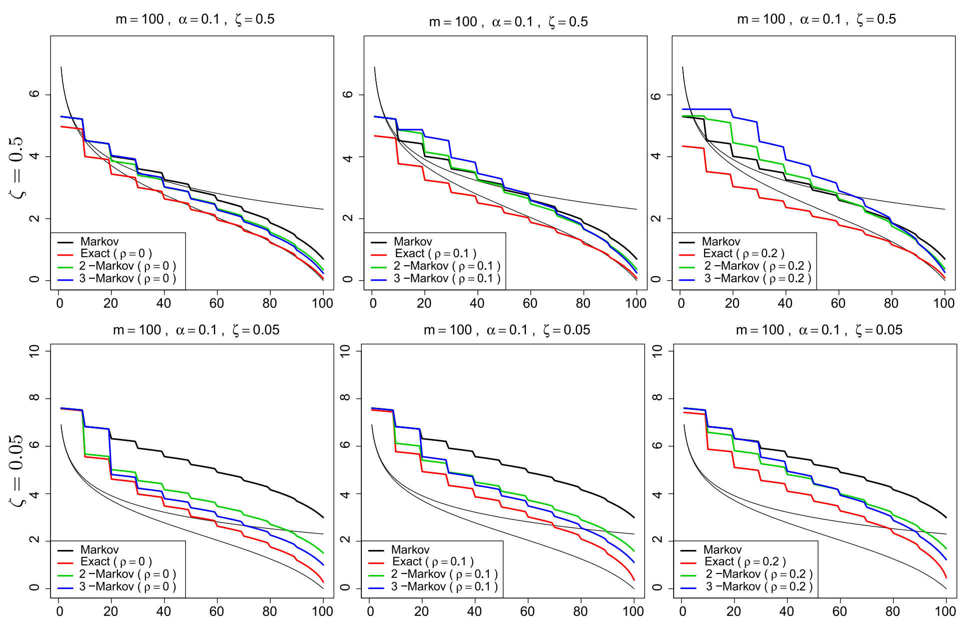

FIG. 3. *Plot of* $-\log(\tilde{\tau}_\ell)$ *in function of* $\ell \in \{1, \dots, m\}$, *for k-FWE-based critical values obtained with several types of bounding devices and assuming* (Gauss-ρ-equi). *For comparison, the solid thin black line corresponds either to the BH critical values* $-\log(\alpha\ell/m)$ (*nonadaptive*) *or the AORC critical values* $-\log(\alpha\ell/(m-(1-\alpha)\ell))$ *defined in* Finner, Dickhaus and Roters (2009) (*adaptive*).

K-Markov. When ζ is small, the Markov device can be too conservative, and we might want to use a sharper tool. Let $K \geq 1$ be an integer. As suggested in Guo, He and Sarkar (2014) (for $K = 2$), we can use the following bound: for $k \geq K$,

$$\mathbf{1}\{V_m(t) \geq k\} \leq \frac{1}{\binom{k}{K}} \sum_{X \subset \mathcal{H}_0 : |X| = K} \mathbf{1}\Big\{\max_{i \in X}\{p_i\} \leq t\Big\}, \tag{14}$$

which leads to upper-bounding $\sup\{\mathbb{P}_\theta(V_m(t) \geq k), \theta \in \Theta, m_0(\theta) = u\}$ by

$$B_m^0(t, k, u) = \frac{u(u-1)\cdots(u-K+1)}{k(k-1)\cdots(k-K+1)} \sup_{\substack{\theta \in \Theta \\ m_0(\theta) = u}} \Big\{P_\theta\Big(\max_{i \in X_0}\{p_i\} \leq t\Big)\Big\}, \tag{15}$$

where $X_0 \subset \mathcal{H}_0$ with $|X_0| = K$ is let arbitrary. In the latter, we implicitly assume that for all θ, the probability $P_\theta(\max_{i \in X}\{p_i\} \leq t)$ depends on X only through $|X| = K$. When $k < K$, bound (15) is useless in essence, so we replace it by the simple Markov device, by letting $B_m^0(t, k, u) = (ut/k) \vee B_m^0(t, K, m)$ if $1 \leq k \leq K - 1$. Note that the operator "$\vee$" in the last display is added to keep the nonincreasing property w.r.t. k. As a first illustration, in the one-sided location model (3), if

(Indep) $\quad p_i, i \in \mathcal{H}_0$, are mutually independent (cond. on H in model Θ^U), we have $P_\theta(\max_{i\in X_0}\{p_i\}\le t)=t^K$, which entails

$$(\tilde\tau_\ell)^K=\zeta\frac{(\lfloor\alpha\ell\rfloor+1)(\lfloor\alpha\ell\rfloor+1-1)\cdots(\lfloor\alpha\ell\rfloor+1-K+1)}{m(\ell)(m(\ell)-1)\cdots(m(\ell)-K+1)}$$

for $\ell\ge\lceil(K-1)/\alpha\rceil$. A second illustration is the one-factor model (facmod), for which $P_\theta(\max_{i\in X_0}\{p_i\}\le t)=\mathbb{E}[F_0(t,W)^K]$ where F_0 is defined by (F_0-facmod). Hence, inverting $B_m^0(t,\lfloor\alpha\ell\rfloor+1,m(\ell))=\zeta$ gives rise to critical values $\bar\tau_\ell$ and $\tilde\tau_\ell$, which both take into account the dependence induced by the common factor W.

Exact. In some cases, closed-formulas can be derived for the RHS of (Bound). First, by assuming (Indep), the distribution of $V_m(t)$ is a binomial with parameters (u,t). Hence, $B_m^0(t,k,u)=\sum_{j=k}^u\binom{u}{j}t^j(1-t)^{u-j}$. The corresponding adaptive critical values can be obtained by a numerical inversion [these critical values were already proposed in Guo and Romano (2007)]. Second, the following exact formula can be used in model (facmod):

$$B_m^0(t,k,u)=\mathbb{E}_W\left[\sum_{j=k}^u\binom{u}{j}\left(F_0(t,W)\right)^j\left(1-F_0(t,W)\right)^{u-j}\right], \tag{16}$$

where F_0 is defined by (F_0-facmod). Third, in a more general manner, nonadaptive threshold can be obtained in the one-sided location model (3), provided that the full joint distribution of Y is known: for this, observe that $V_m(t)$ is upper-bounded by the full null process

$$V_m'(t)=m^{-1}\sum_{i=1}^m\mathbf{1}\{\bar\Phi(Y_i)\le t\}, \tag{17}$$

whose distribution can be approximated by a Monte Carlo method. Finally, to obtain an adaptive threshold, we can make use of the uniform model Θ^U: the added random effect on $\mathcal{H}_0$ entails that $V_m(t)$ can be easily generated for each value of $u=m_0(\theta)$. This leads to nonadaptive and adaptive critical values that incorporate any pre-specified joint distribution for the noise Y (e.g., Γ in the Gaussian case).

3. Finite sample results.

3.1. *Preliminary results.* The following theorem gathers the only existing cases where RW's heuristic has been proved to provide FDP control (to our knowledge).

PROPOSITION 3.1 [Guo, He and Sarkar (2014), Guo and Romano (2007), Romano and Wolf (2007)]. *Consider some bounding device B_m^0 and the associated k-FWE-based critical values $(\tau_\ell)_{1\le\ell\le m}$, being either adaptive or not and computed either in the fixed mixture model ($\Theta=\Theta^F$) or in the uniform mixture model ($\Theta=\Theta^U$). Let us consider the corresponding number of rejections $\hat\ell$ of the associated step-down* (SD) *or step-up* (SU) *procedure. Then the FDP control* (6) *holds* (*with $\hat t_m=\tau_{\hat\ell}$*) *in the following cases*:

(i) *step-down algorithm and the null* p*-values* $(p_i, i : H_i = 0)$ *are independent of the alternative* p*-values* $(p_i, i : H_i = 1)$;

(ii) *step-down or step-up algorithm with the Lehmann–Romano critical values, that is, with* $(\tau_\ell)_{1\le\ell\le m}$ *given by* (13), *and assuming that Simes's inequality is valid,*

$$\forall\theta \in \Theta \qquad \mathbb{P}_\theta\left(\bigcup_{k=1}^{m_0}\{q_{(k)} \le \zeta k/m_0\}\right) \le \zeta, \tag{18}$$

where $q_{(1)} \le q_{(2)} \le \cdots \le q_{(m_0)}$ *denote the ordered* p*-values under the null.*

Case (i) comes from inequalities established in Lehmann and Romano (2005), Romano and Wolf (2007), that we recall in Section 7.1 under an unified form; see also Theorem 5.2 in Roquain (2011). Note that it contains the case where all the p-values under the alternative are equal to zero (Dirac configuration). Case (ii) has been solved more recently in Guo, He and Sarkar (2014). Here, it can be seen as a consequence of the following general inequality; see Section 7.3 for a proof.

PROPOSITION 3.2. *Consider the setting of Proposition* 3.1 *in the step-down or step-up case. Then we have for all* $\theta \in \Theta$

$$\begin{aligned}\mathbb{P}_\theta\big(\mathrm{FDP}_m(\tau_{\hat\ell}) > \alpha\big) &\le \mathbb{P}_\theta\big(V_m(\nu^0_{\hat k}) \ge \hat k \ge 1\big)\\ &= \mathbb{P}_\theta\big(q_{(\hat k)} \le \nu^0_{\hat k}, \hat k \ge 1\big),\end{aligned} \tag{19}$$

for $\hat k = V_m(\tau_{\hat\ell})$, *where*

$$\nu^0_k = \max\{t \in [0,1] : B^0_m(t, k, m_0) \le \zeta\}. \tag{20}$$

Proposition 3.1(ii) thus follows from Proposition 3.2, used with the adaptive Markov bounding device; see (12). Markedly, Proposition 3.2 establishes that the FDP control for adaptive k-FWE based critical values is linked to a specific inequality between the null p-values and the bounding device using the true value of m_0.

Further note that (19) in Proposition 3.2 is sharp whenever $m_0(\theta) = m$: in this case, the LHS and RHS are both equal to the probability that $\hat k (= \hat\ell)$ is not zero, that is, that at least one $\ell \in \{1, \dots, m\}$ is such that $p_{(\ell)} \le \tau_\ell$. For instance, in the independent case and using the exact device (16), when $m = m_0 = 2$ and $\alpha = 0.5$, we have $\bar\tau_1 = \tilde\tau_1 = 1 - (1-\zeta)^{1/2}$ and $\bar\tau_2 = \tilde\tau_2 = \zeta^{1/2}$ and $\mathbb{P}_\theta(\mathrm{FDP}_m(\tau_{\hat\ell}) > \alpha) = 2\zeta - (1 - (1-\zeta)^{1/2})(2\zeta^{1/2} - 1 + (1-\zeta)^{1/2})$. We merely check that the latter is larger than ζ for all $\zeta \in (0, 1)$. Also, simulations of Section S-5 in the Supplementary Material [Delattre and Roquain (2015b)] indicate that this exceeding can hold for a larger value of m. This establishes the following:

FACT 3.3. RW's heuristic does not always provide a valid FDP control for finite m in its step-up form under independence.

Now, an important question is to know whether RW's heuristic always provides a valid FDP control for finite m in its *step-down* form. First, we can merely check that the following cases can be added in Proposition 3.1 in the step-down case:

(iii) for all $\theta \in \Theta$, $\lfloor \alpha b_\alpha(m_0(\theta)) \rfloor = 0$ (e.g., $m_0(\theta) \in \{1, m\}$ or $\lfloor \alpha m \rfloor = 0$);
(iv) under (Gauss-ρ-equi) when $\rho = 1$.

Note that (iii) contains the case $m_0 = m$ which is problematic in the step-up case. A consequence is that any configuration for which the FDP control fails should be searched outside cases (i), (ii), (iii) and (iv). As a matter of fact, we found a numerical example under equi-correlation when using the critical values $(\tau_\ell^0)_\ell$ defined by (10), with the exact device. To this end, we have evaluated the exceedance probability of the FDP by the *exact calculations* proposed in Blanchard et al. (2014), Roquain and Villers (2011). This method is time consuming for large m but avoids the undesirable fluctuations due to the Monte Carlo approximation while performing simulations. Precisely, in model (Gauss-ρ-equi), when $m = 30$, $\alpha = 0.2$, $\zeta = 0.05$, $\rho = 0.3$, $m_0 = 15$, $\mu_i = 1.5$, $1 \leq i \leq m$, we obtain $\mathbb{P}_\theta(\mathrm{FDP}_m(\tau_{\hat{\ell}}^0) > \alpha) > \zeta + 10^{-3}$. Admittedly, the FDP control is just slightly violated. Nevertheless, this gives numerical evidence of the following fact.

FACT 3.4. RW's heuristic does not always provide a valid FDP control for finite m in its oracle step-down form in model (Gauss-ρ-equi).

Note that the case of the *non-oracle* adaptive version is studied with extensive simulations in Section S-6 in Supplementary Material, the conclusion is similar. Fact 3.4 is interesting from a theoretical point of view: it annihilates any hope of finding a general finite sample proof of FDP control in the step-down case, even under a very simple form of positive dependence.

3.2. *Existing modifications*. Facts 3.3 and 3.4 indicate that to obtain a provable finite sample control, it is appropriate to slightly decrease the initial k-FWE-based critical values $(\tau_\ell)_{1 \leq \ell \leq m}$. Interestingly, several existing procedures that provably control the FDP can be reinterpreted as modifications of the τ_ℓ's. In the literature, we have identified the following principles that provide a control of the FDP under general dependence:

- The "diminution" principle Guo, He and Sarkar (2014), Romano and Shaikh (2006a, 2006b): first, establish a rigorous upper-bound for $\mathbb{P}(\mathrm{FDP} > \alpha)$ for a step-down or step-up procedure with arbitrary critical values $(c_\ell(x))_{1 \leq \ell \leq m}$ depending on a single parameter x. Second, adjust x to make the bound smaller than ζ. As an illustration, Romano and Shaikh (2006a, 2006b) have proposed the following bound that can be rewritten as follows (see Section 7.2 for a proof):

$$C^{\mathrm{RS}}(x) = \max_{1\le u\le m}\left\{u\sum_{\ell=1}^{b_\alpha(u)}\frac{c_\ell(x)-c_{\ell-1}(x)}{d(\ell,m,u)}\right\};\tag{21}$$

where for all $u,\ell\in\{0,\dots,m\}$, we let

$$b_\alpha(u) = \begin{cases}\big(\lfloor (m-u)/(1-\alpha)\rfloor+1\big)\wedge\big(\lceil u/\alpha\rceil-1\big)\wedge m, & \\ & \text{(step-down)},\\ \big(\lceil u/\alpha\rceil-1\big)\wedge m, & \text{(step-up)},\end{cases}\tag{22}$$

$$d(\ell,m,u) = \begin{cases}\lfloor\alpha\ell\rfloor+1, & \text{(step-down)},\\ \big(\lfloor\alpha\ell\rfloor+1\big)\vee(\ell-m+u), & \text{(step-up)}.\end{cases}\tag{23}$$

This bound does not incorporate the dependence. Finally, let us mention that the diminution principle has been recently followed by using much more sophisticated bounds that incorporate the pairwise dependence; see Theorems 3.7 and 3.8 in Guo, He and Sarkar (2014).

- The "augmentation" principle Farcomeni (2009), van der Laan, Dudoit and Pollard (2004): consider the 1-FWE controlling procedure at level ζ rejecting the null hypotheses corresponding to the set $\mathcal{R}^{(1)}=\{1\le i\le m : p_i\le\tau_1(\zeta)\}$, denote $\ell^{(1)}$ the number of rejections of $\mathcal{R}^{(1)}$ and
$$\widetilde{\ell}^{\mathrm{Aug}} = \big\lfloor \ell^{(1)}/(1-\alpha)\big\rfloor\wedge m.$$
Then the "augmented" procedure rejects the nulls associated to the $\widetilde{\ell}^{\mathrm{Aug}}$ smallest p-values. This procedure can incorporate the dependence if $\mathcal{R}^{(1)}$ is appropriately chosen.
- The "simultaneous" k-FWE control proposed in Genovese and Wasserman (2006): consider critical values $(\tau_\ell(\zeta/m))_{1\le\ell\le m}$ (with ζ divided by m), and let
$$\widetilde{\ell}^{\mathrm{sim}} = \left\lfloor\frac{\max\{R(\tau_\ell(\zeta/m))-\lfloor\alpha\ell\rfloor : \ell\le R(\tau_\ell(\zeta/m)),\ell\ge 0\}}{1-\alpha}\right\rfloor\wedge m.$$
Then the "simultaneous" procedure rejects the nulls corresponding to the $\widetilde{\ell}^{\mathrm{sim}}$ smallest p-values. Again, this procedure is able to incorporate the dependence if the τ_ℓ's are suitably built.

3.3. *Two new modifications*. This section presents new results that can be seen as modifications of k-FWE based procedures that ensure finite sample FDP control. Both modifications incorporate the dependence between the p-values. Furthermore, the numerical experiments of Section 5 show that they are more powerful than the state-of-the-art procedures described in Section 3.2.

A first modification. The first result follows the "diminution" principle. For any arbitrary critical values $(c_\ell(x))_{1\le\ell\le m}$ (depending on a variable x), let $C^{\mathrm{ex}}(x)$ be

$$\begin{aligned}\max_{1\le u\le m}\Bigg\{\sum_{\ell=1}^{b_\alpha(u)}\max_{\substack{\theta\in\Theta\\ m_0(\theta)=u}}&\big\{\big(\mathbb{P}_\theta\big(V_m\big(c_\ell(x)\big)\ge d(\ell-1,m,u)\big)\\
&\quad-\mathbb{P}_\theta\big(V_m\big(c_{\ell-1}(x)\big)\ge d(\ell-1,m,u)\big)\big)\\
&\quad\wedge\big(\mathbb{P}_\theta\big(V_m\big(c_\ell(x)\big)\ge d(\ell,m,u)\big)\\
&\qquad-\mathbb{P}_\theta\big(V_m\big(c_{\ell-1}(x)\big)\ge d(\ell,m,u)\big)\big)\big\}\Bigg\},\end{aligned}\tag{24}$$

where $b_\alpha(u)$ and $d(\ell,m,u)$ are given by (22) and (23), respectively. The following result is established in Section 7.2.

THEOREM 3.5. *Let us consider either the fixed model* ($\Theta=\Theta^F$) *or the uniform model* ($\Theta=\Theta^U$) *and any family of critical values* $(c_\ell(x))_{1\le\ell\le m}$, $x\ge 0$, *such that* $c_m(0)=0$. *Consider some* $x^\star\ge 0$ *satisfying* $C^{\mathrm{ex}}(x^\star)\wedge C^{\mathrm{RS}}(x^\star)\le\zeta$, *where* $C^{\mathrm{ex}}(\cdot)$ *is defined by* (24) *and* $C^{\mathrm{RS}}(\cdot)$ *by* (21). *Let* $\hat\ell$ *be the number of rejections of the step-down* (SD) [*resp*., *step-up* (SU)] *algorithm associated to the critical values* $(c_\ell(x^\star))_{1\le\ell\le m}$. *Then the FDP control* (6) *holds*, *with* $\hat t_m=\tau_{\hat\ell}$.

Theorem 3.5 can be applied with any starting critical values $(c_\ell(x))_{1\le\ell\le m}$. A choice in accordance with RW's heuristic is $c_\ell(x)=x\tilde\tau_\ell$, $\ell\in\{1,\dots,m\}$, $x\ge 0$, where $(\tilde\tau_\ell)_{1\le\ell\le m}$ are the adaptive k-FWE based critical values (9) for some appropriate bounding device. Next, while Theorem 3.5 does not require any assumption on the dependence, it implicitly assumes that the function $C^{\mathrm{ex}}(\cdot)$ is known or easily computable. This is the case, for instance, in the model (facmod) because we have

$$\begin{aligned}C^{\mathrm{ex}}(x)=\max_{1\le u\le m}\Bigg\{\sum_{\ell=1}^{b_\alpha(u)}&\big(B_m^0\big(c_\ell(x),d(\ell-1,m,u),u\big)\\
&\quad-B_m^0\big(c_{\ell-1}(x),d(\ell-1,m,u),u\big)\big)\\
&\quad\wedge\big(B_m^0\big(c_\ell(x),d(\ell,m,u),u\big)\\
&\qquad-B_m^0\big(c_{\ell-1}(x),d(\ell,m,u),u\big)\big)\Bigg\},\end{aligned}\tag{25}$$

where $B_m^0(t,k,u)$ is the exact bounding device defined by (16). A second illustration is the Gaussian case where Γ is known but arbitrary and where the model is $\Theta=\Theta^U$. In this situation, $C^{\mathrm{ex}}(x)$ in (24) can be approximated by Monte Carlo calculations. Finally, let us underline that Theorem 3.5 provides FDP control even if the incorporated dependence is not positive.

A second modification. The second result presented in this section relies on the K-Markov device $B_m^0(t,k,u)$ given by (15) (for some integer $K \geq 1$). It specifically uses the two following assumptions (here $\Theta = \Theta^F$ only):

(Exch-$\mathcal{H}_0$) for all $\theta \in \Theta$, for any permutation σ of $\{1,\dots,m\}$ with $\sigma(i)=i$ for all $i \notin \mathcal{H}_0(\theta)$, the distribution of $(p_{\sigma(i)})_{1\le i\le m}$ is equal to the one of $(p_i)_{1\le i\le m}$;

(Posdep) for all $\theta \in \Theta$, for any measurable nondecreasing set $D \subset [0,1]^m$ and subset $X \subset \mathcal{H}_0(\theta)$, $x \in [0,1] \mapsto \mathbb{P}_\theta\big((p_i)_{1\le i\le m} \in D \mid \forall i \in X, p_i \le x\big)$ is nondecreasing.

In (Posdep), a set $D \subset [0,1]^m$ is said nondecreasing if for any $x, y \in [0,1]^m$ such that $x \in D$, the inequality $x \le y$ (holding component-wise) entails $y \in D$. Condition (Posdep) induces a form of positive dependence between the p-values. It is stronger than the condition of positive dependence ensuring FDR control for the BH procedure, for which the conditioning holds w.r.t. only one element; see Benjamini and Yekutieli (2001), Blanchard, Delattre and Roquain (2014), Blanchard and Roquain (2008). However, assumption (Posdep) is satisfied as soon as the p-value family is multivariate totally positive of order 2 (MTP2); see Sarkar (1969). We refer to Section 3 of Karlin and Rinott (1981) for several examples of MTP2 models, which thus satisfy assumption (Posdep). More explicit examples will be provided at the end of the section.

Now, assuming (Exch-$\mathcal{H}_0$), we consider for $\ell \in \{1,\dots,m\}$,

$$\tau_\ell^{\text{new}} = \begin{cases} \tilde{\tau}_\ell\big(\lambda\zeta, m(\ell)\big), & \text{if } \ell \ge \ell_K, \\ \left(\dfrac{(1-\lambda)\zeta(\lfloor \alpha\ell \rfloor + 1)}{m(\ell)}\right) \wedge \tilde{\tau}_{\ell_K}(\lambda\zeta, m), & \text{if } \ell < \ell_K, \end{cases} \tag{26}$$

where $\ell_K = \lceil (K-1)/\alpha \rceil$, $\lambda \in [0,1]$ is some tuning parameter. Also, $\tilde{\tau}_\ell(\zeta,u)$ denotes the value of t obtained by solving the equation

$$\begin{aligned} &\sup_{\substack{\theta\in\Theta \\ m_0(\theta)=u}} \Big\{P_\theta\Big(\max_{i\in X_0}\{p_i\} \le t\Big)\Big\} \\ &\quad = \zeta \frac{(\lfloor \alpha\ell \rfloor + 1)(\lfloor \alpha\ell \rfloor + 1 - 1)\cdots(\lfloor \alpha\ell \rfloor + 1 - K + 1)}{u(u-1)\cdots(u-K+1)}, \end{aligned} \tag{27}$$

where X_0 denotes any subset of $\mathcal{H}_0$ of cardinal K. The following result holds; see Section 7.4 for a proof.

THEOREM 3.6. *In the fixed model $\Theta = \Theta^F$, let $\hat{\ell}$ be the number of rejections of the step-up (SU) algorithm associated to the critical values $(\tau_\ell^{\text{new}})_{1\le\ell\le m}$ given by (26). Then the finite sample FDP control (6) holds for $\hat{t}_m = \tau_{\hat{\ell}}^{\text{new}}$ under assumptions (Posdep) and (Exch-$\mathcal{H}_0$).*

The proof of Theorem 3.6 is given in Section 7.4. It shares some similarity with the proofs developed in Sarkar (2007) in the FDR case. When $K = 1$ and $\lambda = 1$, the critical values $(\tau_\ell^{\text{new}})_{1\le\ell\le m}$ are the Lehmann–Romano critical values (13), and thus Theorem 3.6 is in accordance with Proposition 3.1(ii) and Theorem 3.1 of Guo, He and Sarkar (2014) because Simes's inequality is valid in that case. The originality of Theorem 3.6 lies in the case $K > 1$ that allows us to incorporate the dependence in an FDP controlling procedure. Below, some examples are provided in the one-sided location models (3):

(i) When the noise Y is Gaussian multivariate, assumption (Exch-$\mathcal{H}_0$) imposes equicorrelation between the p-values $(p_i, i \in \mathcal{H}_0)$, say ρ-equicorrelation with $\rho \in [0, 1)$. In this case, equation (27) can be solved by using $\mathbb{P}_\theta(\forall i \in X_0, p_i \le t) = \mathbb{E}[(F_0(t, W))^K]$, where F_0 is given by (F_0-Gauss-ρ-equi) and $W \sim \mathcal{N}(0, 1)$. Furthermore, the p-value family is MTP2 if and only if $-\Gamma^{-1}$ has nonnegative off-diagonal elements; see, for example, Rinott and Scarsini (2006). For instance, both assumptions are satisfied if Γ is ρ-equi-correlated (Gauss-ρ-equi). Additional examples can be provided with matrices Γ such that $(\Gamma_{i,j})_{i\in\mathcal{H}_0, j\in\mathcal{H}_0}$ is ρ-equi-correlated while $-\Gamma^{-1}$ has nonnegative off-diagonal elements.

(ii) Consider (facmod) in the particular case where $X_i = \mu_i + c_i W + \zeta_i - (a/b)w_0$, where $c_1 \sim \gamma(a, b)$, W is a positive random variable ($w_0 = \mathbb{E}W$) and ζ_1 is centered with a log-concave density. In this case, the p-value family is MTP2 by Proposition 3.7 and 3.9 of Karlin and Rinott (1980), which entails (Posdep). Assumption (Exch-$\mathcal{H}_0$) also clearly holds, and the LHS of (27) is $\mathbb{E}[(F_0(t, W))^K]$, where F_0 is defined by (F_0-facmod).

REMARK 3.7. Assumption (Posdep) is, strictly speaking, weaker than MTP2 property. For instance, (Posdep) is satisfied in the Gaussian case where $\Gamma_{i,j} \ge 0$ for $i \in \mathcal{H}_1$ and $j \in \mathcal{H}_0$ and $\Gamma_{i,j} = 1$ for $i, j \in \mathcal{H}_0$.

4. Asymptotic results. The goal of this section is to study RW's heuristic from an asymptotic point of view.

4.1. *Setting and assumptions.* In this section, the FDP control under study is asymptotic: we search $\hat{t}_m$ such that

$$\limsup_m \{\mathbb{P}_{\theta^{(m)}}(\mathrm{FDP}_m(\hat{t}_m) > \alpha)\} \le \zeta. \tag{28}$$

This requires us to consider a *sequence* of models $(\Theta^{(m)}, m \ge 1)$ (fixed mixture models here) and a sequence of parameters $(\theta^{(m)}, m \ge 1)$ with $\theta^{(m)} \in \Theta^{(m)}$ for all $m \ge 1$. The latter sequence is assumed to be fixed once for all throughout this section. Moreover, we will assume throughout this study the following common assumption:

$$m_0(\theta^{(m)})/m \to \pi_0 \qquad \text{where } \pi_0 \in (0, 1). \tag{29}$$

In particular, any sparse situation where $m_0(\theta^{(m)})/m \to 1$ is excluded. Also, under (29), we let $\pi_1 = 1 - \pi_0 \in (0, 1)$.

Useful assumptions on $(\theta^{(m)}, m \geq 1)$ are the following weak dependence assumptions on the processes $\widehat{\mathbb{G}}_m(t) = R_m(t)/m$ and $\widehat{\mathbb{G}}_{0,m}(t) = V_m(t)/m_0$, $t \in [0, 1]$:

(weakdep) $\quad \|\widehat{\mathbb{G}}_m - G\|_\infty = o_P(1) \quad$ for some continuous $G : [0, 1] \to [0, 1]$;

(weakdep0) $\quad \|\widehat{\mathbb{G}}_{0,m} - I\|_\infty = o_P(1) \quad$ for $I(t) = t$, $t \in [0, 1]$.

These weak dependence conditions are widely used in the context of multiple testing; see, for example, Gontscharuk and Finner (2013), Storey, Taylor and Siegmund (2004) and the stronger condition (FLT) further on. In the particular one-sided Gaussian multivariate setting, these conditions have been studied in Delattre and Roquain (2015a), Fan, Han and Gu (2012), Schwartzman and Lin (2011) (among others). Lemma S-3.1 of the Supplementary Material states that assumptions (weakdep) and (weakdep0) are satisfied with $G(t) = \pi_0 t + \pi_1 F_1(t)$ and $F_1(t) = \int_0^\infty \bar{\Phi}(\bar{\Phi}^{-1}(t) - \beta)\, d\nu(\beta)$ under (29) if the following conditions hold:

(Conv-alt)
$$(m_1)^{-1} \sum_{i=1}^{m} H_i \delta_{\mu_i} \xrightarrow{\text{weak}} \nu$$
for a distribution ν on $\mathbb{R}^+$ with $\nu(\{0\}) = 0$,

(weakdepGauss)
$$m^{-2} \sum_{i,j=1}^{m} (\Gamma_{i,j})^2 \to 0.$$

Also, let us underline that under (Gauss-ρ-equi), assumption (weakdepGauss) is satisfied whenever $\rho = \rho_m \to 0$.

Finally, we also explore in this section *strong dependence*, through the factor model (facmod). This includes (Gauss-ρ-equi) for a parameter $\rho \in (0, 1)$ taken fixed with m.

4.2. *The BH procedure and FDP control.* Let us go back to Figure 1. When $\rho = 0$, even if the BH procedure is only intended to control the expectation of the FDP at level α, the 95% quantile of the FDP is still close to α. This comes from the concentration of the FDP of the BH procedure around $\pi_0 \alpha < \alpha = 0.2$ as m grows to infinity. It is well known that this quantile converges to $\pi_0 \alpha$ as m grows to infinity, so that the limit in (28) is equal to zero; see, for example, Neuvial (2008). In other words, the FDP concentration combined with the slight amount of conservativeness due to $\pi_0 < 1$ "prevents" the FDP from exceeding α. The consequence is simple: the BH procedure controls the FDP asymptotically in the sense of (28) under independence. As a matter of fact, the latter also holds under weak dependence.

LEMMA 4.1. *Consider the BH procedure, that is, the step-up procedure* (SU) *associated to the linear critical values* $\tau_\ell = \alpha\ell/m$, $1 \le \ell \le m$. *Assume that* $(\theta^{(m)}, m \ge 1)$ *satisfies* (29), (weakdep), (weakdep0) *and further assume that* G *satisfies the following property*:

(Exists) *there exists* $t \in (0, 1)$, *such that* $G(t) > t/\alpha$.

Then we have $\mathbb{P}_{\theta^{(m)}}(\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha) \to 0$.

Although this result seems new, its proof, provided in Section 8.1, can certainly be considered as standard; see, for example, Finner, Dickhaus and Roters (2007), Genovese and Wasserman (2004). Also, while Lemma 4.1 does not require any Gaussian assumption in general, all the assumptions of Lemma 4.1 are satisfied under (29), (Conv-alt) and (weakdepGauss).

In the literature, even under independence, it is common to exclude the BH procedure while studying (28). For instance, Proposition 4.1 in Chi and Tan (2008) shows that the "oracle" version of the BH procedure, that is, the step-up procedure with critical values $\alpha\ell/m_0$, $\ell \in \{1, \dots, m\}$, has a FDP exceeding α with a probability tending to $1/2$. Since the oracle BH procedure is often considered to be better than the original BH procedure, it is thus tempting to exclude the BH procedure when studying an FDP control of type (28). Lemma 4.1 shows that, perhaps surprisingly, this is a mistake: BH procedure is interesting when providing (28) and does not suffer from the same drawback as the oracle BH procedure.

By contrast, if the dependence is not weak, the BH procedure can fail to control the FDP as m is tending to infinity. Under (Gauss-ρ-equi) and when the p-values under the alternative are zero (Dirac uniform configuration), this fact has been formally established in Theorem 2.1 of Finner, Dickhaus and Roters (2007), by showing that the limit of the FDP of the BH procedure is not deterministic anymore and hence can exceed α with a positive probability, which is obviously not related to ζ (because BH critical values do not depend on ζ).

4.3. *RW's heuristic under weak dependence.* The two results provided in this section both validate the use of RW's heuristic under weak dependence. Since the BH procedure is valid in this case, they are of limited interest in practice. Nevertheless, we believe that they suitably complement our overview on RW's heuristic. The first result is proved in Section 8.2 via technics similar to those used for proving Lemma 4.1.

THEOREM 4.2. *Consider the one-sided location model* (3) *with the full null process* $V'_m(\cdot)$ *being defined by* (17). *Assume that the* (*nonadaptive*) *exact bounding device is such that for all* t, k $\bar{B}_m(t, k) = \mathbb{P}_\theta(V'_m(t) \ge k)$ *for the parameter* $\theta = \theta^{(m)}$ *at hand. Consider the critical values* $\bar{\tau}_\ell$, $1 \le \ell \le m$, *derived from* $\bar{B}_m$ *as in* (8), *and consider the corresponding step-up procedure* (SU). *Assume that* $(\theta^{(m)}, m \ge 1)$ *satisfies* (29), (weakdep), (weakdep0), (Exists) *and that* $V'_m(t)/m$ *converges in* *probability to t for any $t \in [0, 1]$ (i.e., weak dependence for the full null process). Then we have $\mathbb{P}_{\theta^{(m)}}(\mathrm{FDP}_m(\bar{\tau}_{\hat{\ell}}) > \alpha) \to 0$.*

The above result shows that RW's procedure used with the exact bounding device turns out to have an asymptotic exceedance probability of zero under weak dependence, likewise the BH procedure. Again, this is due to the convergence of the FDP toward $\pi_0\alpha < \alpha$. Hence, perhaps disappointingly, ζ plays no role in the limit, which indicates that using the simpler BH procedure seems more appropriate in this case.

Nevertheless, when m_0 is known, an interesting point is that, while the oracle BH procedure fails to control the FDP (as discussed in the above section), oracle RW's method maintains the FDP control. To show this, we need to slightly strengthen the assumption (weakdep0) by assuming that the following central limit theorem holds for the (rescaled) process $V_m(\cdot)$:

(FLT) $\begin{cases} \text{There is a rate } r_m \to \infty \text{ such that the process } Z_m(t) = r_m\big(V_m(t)/m - (m_0(m)/m)t\big) \text{ satisfies, for any } K = [a,b] \subset (0,1), \text{ the convergence } \big(Z_m(t)\big)_{t\in K} \rightsquigarrow \big(Z(t)\big)_{t\in K} \text{ (for the Skorokhod topology), for a process } \big(Z(t)\big)_{t\in K} \text{ with continuous paths and such that the random variable } Z(t) \text{ has a continuous increasing c.d.f. for all } t \in K. \end{cases}$

For instance, under (29), assumption (FLT) holds when the p-values $(p_i, H_i = 0)$ are i.i.d. by Donsker's theorem. More generally, dependencies satisfying "mixing" conditions can also lead to (FLT); see, for example, Dedecker and Prieur (2007), Doukhan et al. (2010) or Farcomeni (2007). Recently, some efforts have been undertaken to consider other types of dependence, not necessarily locally structured; see Bardet and Surgailis (2013), Soulier (2001). In the case of a Gaussian multivariate structure, explicit sufficient conditions on Γ are provided in Delattre and Roquain (2015a). The following result is proved in Section 8.3.

THEOREM 4.3. *Assume that the exact oracle bounding device is such that for all t, k $B^0_m(t,k,m_0) = \mathbb{P}_\theta(V_m(t) \geq k)$ for the parameter $\theta = \theta^{(m)}$ at hand. Consider the critical values τ^0_ℓ, $1 \leq \ell \leq m$, derived from B^0_m as in (10), and consider the corresponding step-up procedure (SU). Assume that $(\theta^{(m)}, m \geq 1)$ satisfies (29) with $\pi_0 > \alpha$, (weakdep), (FLT) and further assume that G satisfies the following property*:

(Unique) $\quad G(t) = \pi_0 t/\alpha$ *has at most one solution on* $(0, 1)$;

(NonCritical) $$\lim_{t\to 0^+} G(t)/t \in (\pi_0/\alpha, +\infty].$$

Then we have $\mathbb{P}_{\theta^{(m)}}(\mathrm{FDP}_m(\tau^0_{\hat{\ell}}) > \alpha) \to \zeta$.

Roughly speaking, the essence of the argumentation is as follows: when $\hat{t}$ converges in probability to some deterministic quantities, then the fluctuations of $\hat{\ell}/m$ asymptotically disappear in probability (7), and thus the latter is equal to ζ by definition of the oracle exact bounding device. Note that a similar reasoning has been made at the end of Section 7 in Genovese and Wasserman (2006). Here, we derive sufficient conditions that make this informal argument rigorous.

Markedly, in Theorem 4.3, the limit of the probability is exactly ζ; hence there is no loss in the level of RW's method. However, since m_0 is often unknown (and seems hard to estimate at a rate faster than r_m), the interest of this result remains mainly theoretical. Finally note that (Unique) and (NonCritical) are classical conditions when studying asymptotic properties of step-up procedures; see, for example, Chi (2007), Genovese and Wasserman (2002), Neuvial (2008).

4.4. *RW's heuristic under strong dependence.* Here, we study the asymptotic properties of RW's method under strong dependence by focusing on models of the type (facmod). A crucial assumption is as follows:

(Posdep-facmod) for any $t \in (0, 1)$, the function $w \mapsto F_0(t, w)$ is increasing,

where F_0 is given by (F_0-facmod). The latter assumption is a form of positive dependence which is specific to (facmod): it roughly means that the variable W disturbs each p-value distribution in an "unidirectional" manner. For instance, (Posdep-facmod) is satisfied if $\mathbb{P}(c_1 \geq 0) = 1$, $\mathbb{P}(c_1 > 0) > 0$ and ξ_1 has a distribution function which is continuous increasing on $\mathbb{R}$. As an illustration, (Posdep-facmod) is satisfied under (Gauss-ρ-equi) ($\rho > 0$) but not necessarily under (alt-ρ-equi).

Asymptotic view of RW's heuristic. Under appropriate assumptions, the exact device (16) is such that, for a sequence k_m with $k_m/m \to \kappa$,

$$\bar{B}_m(t, k_m) = \mathbb{P}(N_m/m \geq k_m/m) \to \mathbb{P}_W\big(F_0(t, W) \geq \kappa\big), \tag{30}$$

where F_0 is defined by (F_0-facmod), and N_m follows a binomial distribution of parameters m and $F_0(t, W)$, conditionally on W. By taking $\kappa = F_0(t, q_\zeta)$ where q_ζ is such that $\mathbb{P}(W \geq q_\zeta) \leq \zeta$, the probability on the RHS of (30) is smaller than or equal to $\mathbb{P}_W(F_0(t, W) \geq F_0(t, q_\zeta)) \leq \mathbb{P}_W(W \geq q_\zeta) \leq \zeta$, provided that (Posdep-facmod) holds. Now, RW's heuristic (taken in an asymptotic sense) leads to the following equation for the critical values:

$$F_0(\tau_\ell, q_\zeta) = \alpha\ell/m, \qquad 1 \leq \ell \leq m. \tag{31}$$

Under independence, this gives the BH critical values. In the equi-correlated case (Gauss-ρ-equi) (with $\rho \geq 0$), this yields the critical values (2) mentioned in the Introduction of the paper.

A modification based on DKW's concentration inequality. The following result shows that a simple modification of (31) provides FDP control (see Section 8.4 for a proof).

THEOREM 4.4. *Let* $\lambda \in (0, 1)$. *In a model* (facmod) *satisfying* (Posdep-facmod), *consider any critical values* τ_ℓ, $1 \le \ell \le m$, *satisfying*

$$F_0(\tau_\ell, q_{\zeta(1-\lambda)}) \le \left(\frac{\alpha\ell}{m} - \left(\frac{-\log(\lambda\zeta/2)}{2m}\right)^{1/2}\right)_+, \qquad 1 \le \ell \le m, \tag{32}$$

where for any $x \in (0, 1)$, $q_x \in \mathbb{R}$ *is such that* $\mathbb{P}(W \ge q_x) \le x$ *and* F_0 *is defined by* (F_0-facmod). *Consider the step-up procedure* (SU) *associated to the critical values* τ_ℓ, $\ell = 1, \dots, m$. *Then it controls the FDP*; *that is*, (6) *holds with* $\hat{t} = \tau_{\hat{\ell}}$.

While this result is nonasymptotic, it is intended to be used for large values of m in order to reduce the influence of the remainder terms. However, even for large values of m, (32) imposes to set the first critical values to zero, which may be undesirable. The next section presents conditions allowing to drop these annoying remainder terms.

An asymptotic validation of RW's heuristic. Here, we present situations for which the raw critical values (31) can be used to get an asymptotic FDP control. We consider the following additional distribution assumptions on c_1, W and ξ_1 in (facmod):

(i) c_1 is a random variable with a finite support in $\mathbb{R}^+$;
(ii) the distribution function of W is continuous;
(iii) the function $x \in \mathbb{R} \mapsto \bar{F}_\xi(x) = \mathbb{P}(\xi_1 \ge x)$ is continuous increasing and is such that, for all $y \in \mathbb{R}$, as $x \to +\infty$,

$$\frac{\bar{F}_\xi(x-y)}{\bar{F}_\xi(x)} \to \begin{cases} +\infty, & \text{if } y > 0, \\ 0, & \text{if } y < 0. \end{cases} \tag{33}$$

When ξ_1 has a log-concave density, condition (33) can be reformulated in terms of a density ratio; see, for example, the relations in Section S-5 of Neuvial and Roquain (2012). For instance, a simple class of distributions satisfying (33) are the so-called *Subbotin* distributions, for which the density of ξ_1 is given by $e^{-|x|^\gamma/\gamma}$ (up to a constant), for some parameter $\gamma > 1$; see Section 5 of Neuvial (2013) for more details on this.

The following result is proved in Section 8.5.

THEOREM 4.5. *Consider the one-sided testing problem* (3) *with all alternative means equal to some* $\beta > 0$, *and assume that* $\theta^{(m)}$ *satisfies* (29). *Consider* (facmod) *satisfying the assumptions* (i), (ii) *and* (iii) *above. Consider the step-up procedure* (SU) *associated to the critical values* τ_ℓ, $\ell = 1, \dots, m$ *satisfying* (31). *Then the asymptotic FDP control* (28) *holds with* $\hat{t}_m = \tau_{\hat{\ell}}$.

As a first illustration, conditions (i), (ii) and (iii) of Theorem 4.5 are satisfied in the case (Gauss-ρ-equi) for some $\rho \in [0, 1)$. Hence, a direct corollary is that the step-up procedure with the critical values given by (2) controls the FDP asymptotically. Furthermore, in Section S-2 of the Supplementary Material, we complement this result by proving that this FDP control is maintained if the value of ρ in (2) is replaced by any estimator $\hat{\rho}_m$ provided that

$$(\log m)(\hat{\rho}_m - \rho)^2 = o_P(1). \tag{34}$$

As a second illustration, consider a model (facmod) where c_1 is uniform on $\{k/r, 0 \le k \le r\}$, and ξ_1 is γ-Subbotin (for some $\gamma > 1$). Here, W can be any random variable with continuous distribution function. For this particular dependence structure, Theorem 4.5 establishes the asymptotic FDP control of the step-up procedure with critical values τ_ℓ given by the equation

$$(r+1)^{-1} \sum_{k=0}^{r} \bar{D}_\gamma\big(\bar{F}^{-1}(\tau_\ell) - q_\zeta k/r\big) = \alpha \ell/m, \qquad 1 \le \ell \le m,$$

where $\bar{D}_\gamma$ denotes the upper-tail function of a γ-Subbotin distribution.

5. Numerical experiments.

This section evaluates the power of the procedures considered in Section 3 with a proven FDP control. The power is evaluated by using the standard false nondiscovery rate (FNR), defined as the expected ratio of errors among the accepted null hypotheses. Table 1 summarizes the procedures that have been considered. The simulation are made in model (Gauss-ρ-equi) where the alternative means μ_i are all equal to some parameter β.

TABLE 1
Procedures used in Figures 4 and 5; see Sections 3.2 and 3.3 for more details. All the procedures are step-up; "e.b.d." means "exact bounding device"

Procedures not using the value of ρ	
[Bonf]	the raw Bonferroni procedure
[LR]	Lehmann Romano's procedure (13)
[AugBonf]	augmentation with $\tau_1 = \zeta/m$
[SimLR]	simultaneous k-FWE with (13)
[DimMarkovLR]	diminution with (21) and $c_\ell(x) = x\tilde{\tau}_\ell$ coming from (13)

Procedures incorporating the value of ρ	
[AugEx]	augmentation with τ_1 coming from e.b.d.
[SimEx]	simultaneous k-FWE with $\tilde{\tau}_\ell$ coming from e.b.d.
[Split1/2]	new procedure (26) with $\lambda = 1/2$ and $K = 2$
[Split0.95]	new procedure (26) with $\lambda = 0.95$ and $K = 2$
[RWExact]	nonmodified k-FWE with $\tilde{\tau}_\ell$ coming from e.b.d.
[DimExEx]	new diminution with (25) and $c_\ell(x) = x\tilde{\tau}_\ell$ coming from e.b.d.
[DimGuoLR]	diminution following Theorem 3.8 of Guo, He and Sarkar (2014) with $c_\ell(x) = x\tilde{\tau}_\ell$ coming from (13)

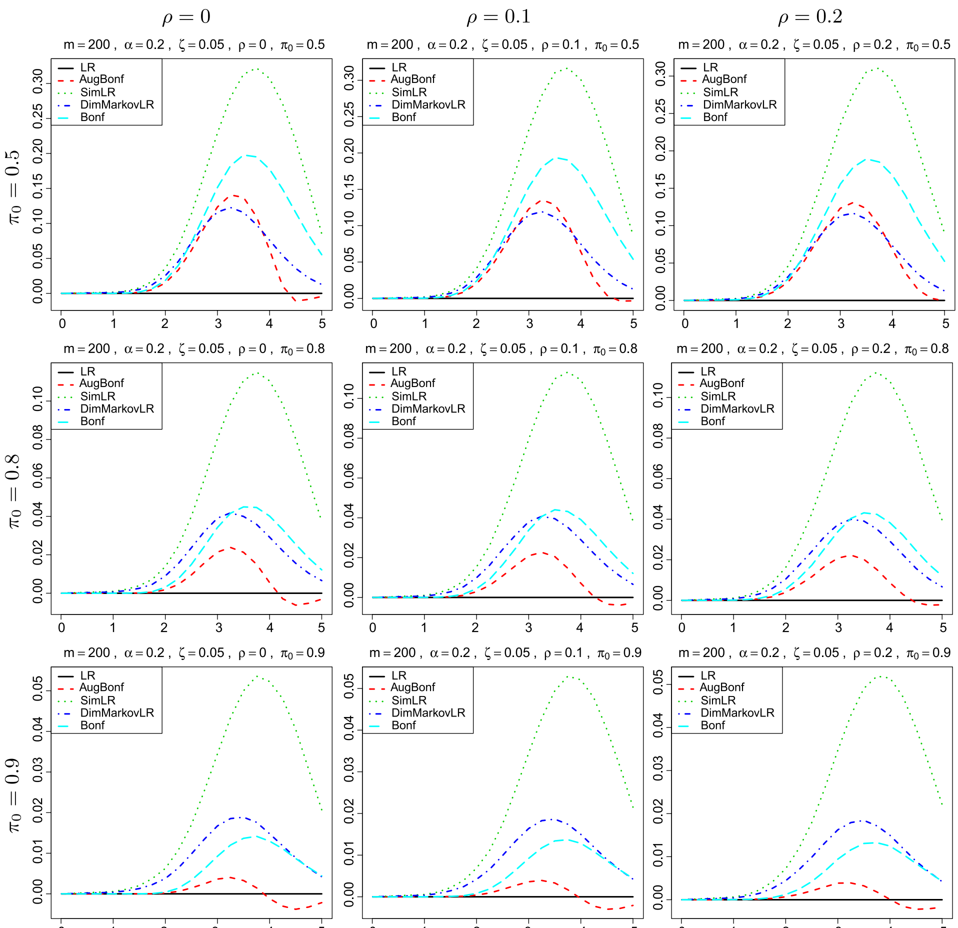

FIG. 4. *Relative FNR to the Lehmann–Romano procedure in function of β; see text and Table 1. Procedures not using the value of ρ.*

Figure 4 displays the power of procedures that do not incorporate the value of ρ; see Table 1. Note that, according to Proposition 3.1(ii), [LR] controls the FDP because Simes's inequality is valid here. Hence it does not use the true value of ρ but uses nevertheless an assumption on the dependence structure. This is not the case of [AugBonf], [SimLR], [DimMarkovLR] and [Bonf] that control the FDP for any dependence structure. As one can expect, [LR] essentially dominates the other procedures. Also, while [AugBonf] comes in second position, [SimLR] is even worst than [Bonf] and should be avoided here.

Now, while incorporating the value of ρ, we will loosely say that a procedure is *admissible* if it performs better than [LR] at least for *a reasonable amount* of parameter configurations. Figure 5 displays the power of procedures incorporating the value of ρ (except [LR] and [Bonf] that we have added only for comparison);

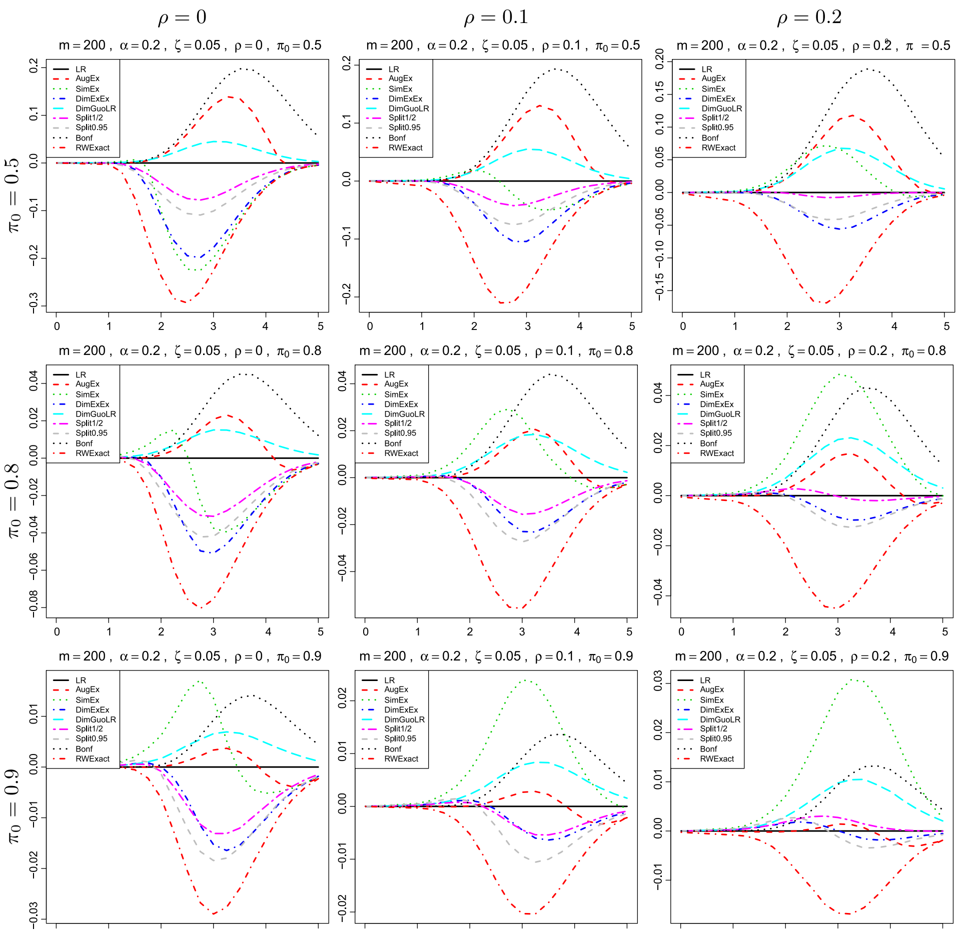

FIG. 5. *Relative FNR to the Lehmann–Romano procedure in function of β; see text and Table 1. Procedures using the value of ρ (except [LR] and [Bonf]).*

see Table 1. Note that, except [RWExact], all the procedures have a proven FDP control, so that the power comparison is fair. First, [DimGuoLR] is *not admissible*, which indicates that the interest of the bounds found in Guo, He and Sarkar (2014) are mainly theoretical in our setting. Second, [AugEx] only improves [LR] in a very small region, which shows that, as one can expect, providing 1-FWE control for controlling the FDP is too conservative in general. As for [SimEx], things are more balanced: when $\rho = 0$, it improves [LR] when many rejections are possible (π_0 not large or β large) but does worst otherwise. We think that this is due to the nature of the [LR] critical values, which are design to perform well when only few nulls are expected to be rejected. When ρ is larger, however, [SimEx] quickly deteriorates. An explanation is that the simultaneity in [SimEx] is obtained via an union bound, which is conservative when the dependence is strong. Finally, our new procedures [Split1/2], [Split0.95] and [DimExEx] seem to be all *admissible* in these simulations and substantially outperform the other procedures. Also, none of the three procedures uniformly dominates the others. For instance, taking $\lambda = 1/2$ rather than $\lambda = 0.95$ is better when less rejections are expected, but worst otherwise, while [DimExEx] seems often better than [Split1/2].

Let us mention that additional simulations have been done in Section S-5 of the Supplementary Material. We briefly report some comments here. First, the large value $\rho = 0.5$ has been tried. This deteriorates the relative performance of all the procedures (except maybe [AugEx]), and in particular of the K-Markov based procedure because the distribution of the maximum between null p-values get closer to the uniform. Second, simulations have been performed in model (alt-ρ-equi) for $a = 0.5$. In this model, the positive dependence property is lost. Hence while [DimExEx] still provably controls the FDP, this is not anymore the case of [LR], [Split1/2] and [Split0.95]. However, it is interesting to note that the FDP control seems to be maintained in the simulations; see Figure S-4 in the Supplementary Material. As for the power, the conclusions are qualitatively the same as in model (Gauss-ρ-equi).

6. Conclusion and discussion. This paper investigated the FDP control in the case where the dependence is partly/fully incorporated, by using an extension of RW's heuristic. We provided two new approaches that offer finite sample control: the first one (Theorem 3.5) followed the diminution principle and can be used as soon as the joint distribution of the null p-values can be computed. The second one (Theorem 3.6) offered a finite sample control under a particular type of positive dependence (Posdep) and exchangeability. Next, an important part of our work concerned the asymptotic FDP control: while we established that RW's heuristic is valid under weak dependence (Theorems 4.3 and 4.2), we noticed that the interest of the latter has to be balanced with the fact that the simple BH procedure can be used in this case (Lemma 4.1). Then, still based on RW's heuristic, we proposed new critical values that provide asymptotic control under model (facmod) (Theorems 4.4 and 4.5). Markedly, while it still relies on a positive dependence assumption (Posdep-facmod), this condition has a much simpler form than (Posdep).

Our leading example is related to one-sided testing, so we can legitimately ask whether our results can be extended to two-sided testing, that is, when $p_i = 2\bar{\Phi}(|X_i|)$ (by using the notation of Section 1.3). In the model (facmod) with $\xi_1 \sim -\xi_1$, the bounding device calculations done in Section 2.3 can be clearly generalized to the two-sided case by replacing $F_0(t, w)$ by $F_0^{(2)}(t, w) = F_0(t/2, w) + F_0(t/2, -w)$. Hence we can define new critical values coming from the corresponding exact bounding device and combine it with the diminution principle presented in Theorem 3.5. However, the other results of the paper cannot be directly generalized to the two-sided case because $F_0^{(2)}(t, w)$ may be not increasing w.r.t. w.

While this paper solved some issues, it opened several directions of research. For instance, is the asymptotic FDP control of Theorem 4.5 still true when using the original critical values of RW's method rather than their asymptotic counterpart? We believe that this issue intrinsically relies on the Poisson asymptotic regime, which was (essentially) not considered here in our asymptotic FDP controlling results. Finally, a crucial, but probably very challenging issue is the validity of RW's approach in the case of permutation tests with an arbitrary and unknown dependence structure.

7. Proofs for finite sample results.

7.1. *An unifying bound.*

PROPOSITION 7.1. *For any critical values $(\tau_\ell)_{1\le\ell\le m}$, consider either the corresponding step-down* (SD) *or step-up* (SU) *procedure, with rejection number $\hat{\ell}$. Then the following holds, both in the fixed model ($\Theta = \Theta^F$) and the uniform model ($\Theta = \Theta^U$): for all $\theta \in \Theta$,*

$$\mathbb{P}_\theta\big(\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha\big) \le \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{P}_\theta\big(V_m(\tau_\ell) \ge d(\ell, m, m_0), \tilde{\ell} = \ell\big), \tag{35}$$

where $b_\alpha(m_0)$, $d(\ell, m, m_0)$ are defined in (22) *and* (23), *respectively, and $\tilde{\ell}$ is taken as follows*:

(i) *Step-down case*: $\tilde{\ell} = \hat{\ell}^{(1)}$, *where* $\hat{\ell}^{(1)} = \min\{\ell \in \{1, \dots, m\} : S_m(\tau_\ell) < (1-\alpha)\ell\}$ (*with the convention* $\min\varnothing = m+1$) *and by denoting $S_m(t) = R_m(t) - V_m(t)$ the number of true discoveries at threshold t.*

(ii) *Step-up case*: $\tilde{\ell} = \hat{\ell}$.

Moreover, in the step-up case, (35) *is an equality.*

Proposition 7.1(i) is a reformulation of Theorem 5.2 in Roquain (2011) in our framework and is based on ideas presented in the proofs of Lehmann and Romano (2005), Romano and Wolf (2007). Proposition 7.1(ii) is essentially based on Romano and Shaikh (2006b), and we provide a short proof below.

PROOF OF PROPOSITION 7.1. Since $\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha$ implies $\lfloor \alpha\hat{\ell}\rfloor + 1 \le m_0$, we have $\hat{\ell} \le b_\alpha(m_0)$. Also, $\hat{\ell} = R_m(\tau_{\hat{\ell}}) \le m_1 + V_m(\tau_{\hat{\ell}})$, which implies $V_m(\tau_{\hat{\ell}}) \ge \hat{\ell} - m_1$. This implies (35) in case (ii). $\square$

7.2. *A new bound.*

PROPOSITION 7.2. *In the setting of Proposition 7.1, assume moreover that there exists a family of random variables $(Z_{\ell,\ell'})_{1\le \ell,\ell'\le m}$ satisfying: for all ℓ,ℓ',*

$$\mathbf{1}\{V_m(\tau_\ell)\ge d(\ell',m,m_0)\}\le Z_{\ell,\ell'} \qquad a.s. \tag{36}$$

and, a.s., $Z_{\ell,\ell'}$ is nondecreasing in ℓ and nonincreasing in ℓ'. Then for all $\theta\in\Theta$,

$$\begin{aligned}&\mathbb{P}_\theta(\mathrm{FDP}_m(\tau_{\hat{\ell}})>\alpha)\\ &\quad\le \sum_{\ell=1}^{b_\alpha(m_0)}(\mathbb{E}_\theta(Z_{\ell,\ell-1})-\mathbb{E}_\theta(Z_{\ell-1,\ell-1}))\wedge(\mathbb{E}_\theta(Z_{\ell,\ell})-\mathbb{E}_\theta(Z_{\ell-1,\ell})),\end{aligned} \tag{37}$$

by letting $Z_{0,\ell'}=0$ and $Z_{\ell,0}=1$ for $\ell'\ge 0$, $\ell\ge 1$.

Applied with $Z_{\ell,\ell'}=V_m(\tau_\ell)/d(\ell',m,m_0)$, Proposition 7.2 establishes the Romano–Shaikh bound (21). Applied with $Z_{\ell,\ell'}=\mathbf{1}\{V_m(\tau_\ell)\ge d(\ell',m,m_0)\}$, Proposition 7.2 entails Theorem 3.5.

PROOF OF PROPOSITION 7.2. From (35), we derive

$$\begin{aligned}\mathbb{P}_\theta(\mathrm{FDP}_m(\tau_{\hat{\ell}})>\alpha)&\le\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{P}_\theta(V_m(\tau_\ell)\ge d(\ell,m,m_0),\tilde{\ell}=\ell)\\ &\le\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{E}_\theta(Z_{\ell,\ell}\mathbf{1}\{\tilde{\ell}=\ell\}).\end{aligned}$$

Now, the RHS of the previous display is equal to

$$\begin{aligned}&\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{E}_\theta(Z_{\ell,\ell}\mathbf{1}\{\tilde{\ell}\ge\ell\})-\sum_{\ell=1}^{b_\alpha(m_0)-1}\mathbb{E}_\theta(Z_{\ell,\ell}\mathbf{1}\{\tilde{\ell}\ge\ell+1\})\\ &\quad=\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{E}_\theta(Z_{\ell,\ell}\mathbf{1}\{\tilde{\ell}\ge\ell\})-\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{E}_\theta(Z_{\ell-1,\ell-1}\mathbf{1}\{\tilde{\ell}\ge\ell\})\\ &\quad=\sum_{\ell=1}^{b_\alpha(m_0)}\mathbb{E}_\theta((Z_{\ell,\ell}-Z_{\ell-1,\ell-1})\mathbf{1}\{\tilde{\ell}\ge\ell\}).\end{aligned}$$

Now, since $Z_{\ell,\ell'}$ is nonincreasing w.r.t. ℓ', the quantity $Z_{\ell,\ell}-Z_{\ell-1,\ell-1}$ is below $(Z_{\ell,\ell-1}-Z_{\ell-1,\ell-1})\wedge(Z_{\ell,\ell}-Z_{\ell-1,\ell})$, and the latter is nonnegative because $Z_{\ell,\ell'}$ is nondecreasing w.r.t. ℓ. This entails the result. $\square$

7.3. *Proof of Proposition* 3.2. Let $\hat{k} = V_m(\tau_{\hat{\ell}})$, and note that $\hat{k} \le m_0$ and $\{\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha\} = \{\hat{k} \ge \lfloor \alpha\hat{\ell} \rfloor + 1\}$. First, in the nonadaptive case, we have by definition of $\bar{B}_m$, for all t and $k \le m_0$,

$$\bar{B}_m(t,k) = \sup_{0 \le u \le m} \{B_m^0(t,k,u)\} \ge B_m^0(t,k,m_0).$$

Hence, we have by definition of the (nonadaptive) critical values,

$$\zeta \ge \bar{B}_m\big(\tau_{\hat{\ell}}, \lfloor \alpha\hat{\ell} \rfloor + 1\big) \ge B_m^0\big(\tau_{\hat{\ell}}, \lfloor \alpha\hat{\ell} \rfloor + 1, m_0\big),$$

which is larger than or equal to $B_m^0(\tau_{\hat{\ell}}, \hat{k}, m_0)$ whenever $\hat{k} \ge \lfloor \alpha\hat{\ell} \rfloor + 1$. Hence we obtain

$$\{\mathrm{FDP}_m(\tau_{\hat{\ell}}) > \alpha\} \subset \{B_m^0(\tau_{\hat{\ell}}, \hat{k}, m_0) \le \zeta, \hat{k} \ge 1\} \subset \{\tau_{\hat{\ell}} \le \nu_{\hat{k}}^0, \hat{k} \ge 1\},$$

and thus (19) holds. Second, in the adaptive case, we use that $m_0 \le m - R_m(\tau_{\hat{\ell}}) + V_m(\tau_{\hat{\ell}}) = m - \hat{\ell} + \hat{k}$. Thus whenever $\hat{k} \ge \lfloor \alpha\hat{\ell} \rfloor + 1$, we have for all t,

$$\begin{aligned} \zeta &\ge \tilde{B}_m\big(\tau_{\hat{\ell}}, \lfloor \alpha\hat{\ell} \rfloor + 1, \hat{\ell}\big) = \sup_{\lfloor \alpha\hat{\ell} \rfloor + 1 \le k' \le \hat{\ell}} \Big\{ \sup_{0 \le u \le m - \hat{\ell} + k'} B_m^0\big(\tau_{\hat{\ell}}, k', u\big) \Big\} \\ &\ge \sup_{0 \le u \le m - \hat{\ell} + \hat{k}} B_m^0(\tau_{\hat{\ell}}, \hat{k}, u) \ge B_m^0(\tau_{\hat{\ell}}, \hat{k}, m_0). \end{aligned} \tag{38}$$

Hence, this implies $\tau_{\hat{\ell}} \le \nu_{\hat{k}}^0$ and $\hat{k} \ge 1$, and the proof is complete.

7.4. *Proof of Theorem* 3.6. First observe that the critical values (26) can be obtained by modifying the K-Markov Bounding device $B_m^0(t,k,u)$ defined by (15) as follows:

$$(B_m^0)'(t,k,u) = \begin{cases} B_m^0(t,k,u)/\lambda, & \text{if } k \ge K, \\ \dfrac{ut}{(1-\lambda)k} \vee \big(B_m^0(t,K,m)/\lambda\big), & \text{if } k < K, \end{cases}$$

(the second bounding value being infinite when $\lambda = 1$). Note that the associated adaptive bounding device (Bound-adapt) is equal to $(B_m^0)'(t,k,m-\ell+k)$ and thus gives rise to the adaptive critical values (26). By using Proposition 3.2 and by letting $\widehat{k} = V_m(\tau_{\hat{\ell}}^{\mathrm{new}})$, we get

$$\mathbb{P}_\theta\big(\mathrm{FDP}_m(\tau_{\hat{\ell}}^{\mathrm{new}}) > \alpha\big) = \sum_{k=1}^{m_0} \mathbb{P}_\theta\big(V_m(\nu_k^0) \ge k, \hat{k} = k\big),$$

where $\nu_k^0 = \max\{t \in [0,1] : B_m^0(t,k,m_0) \le \lambda\zeta\}$ for all $k \ge K$ and where $\nu_k^0 = \max\{t \in [0,1] : (B_m^0(t,K,m)/\lambda) \vee (m_0 t/(k(1-\lambda))) \le \lambda\zeta\}$ for all $k < K$. It follows that the above display is smaller than or equal to $T_1 + T_2$, where we let

$$T_1 = \sum_{k=K}^{m_0} \mathbb{P}_\theta\big(V_m(\nu_k^0) \ge k, \hat{k} = k\big); \qquad T_2 = \sum_{k=1}^{(K-1)\wedge m_0} \mathbb{P}_\theta\big(V_m(\nu_k^0) \ge k, \hat{k} = k\big),$$

with by convention $T_1 = 0$ when $K > m_0$. By (14), (Exch-$\mathcal{H}_0$), and since $\hat{k}$ is permutation invariant (as a function of the p-values), we obtain

$$T_1 \le \sum_{k=K}^{m_0} \frac{1}{\binom{k}{K}} \sum_{X \subset \mathcal{H}_0 : |X|=K} \mathbb{P}_\theta\Big(\hat{k} = k, \max_{i \in X}\{p_i\} \le \nu_k^0\Big)$$
$$= \sum_{k=K}^{m_0} \frac{m_0(m_0-1)\cdots(m_0-K+1)}{k(k-1)\cdots(k-K+1)} \mathbb{P}_\theta\Big(\hat{k} = k, \max_{1 \le i \le K}\{q_i\} \le \nu_k^0\Big),$$

where $q_1, \dots, q_{m_0}$ denotes the p-values under the null, that is, the p-values of the set $\{p_i, i \in \mathcal{H}_0\}$. Next, by using that $B_m^0(\nu_k^0, k, m_0) \le \lambda\zeta$ for $k \ge K$ and (15), we get

$$T_1 \le \lambda\zeta \sum_{k=K}^{m_0} \mathbb{P}_\theta\Big(\hat{k} = k | \max_{1 \le i \le K}\{q_i\} \le \nu_k^0\Big)$$
$$\le \lambda\zeta \sum_{k=K}^{m_0} \Big\{\mathbb{P}_\theta\Big(\hat{k} \le k | \max_{1 \le i \le K}\{q_i\} \le \nu_k^0\Big) - \mathbb{P}_\theta\Big(\hat{k} \le k-1 | \max_{1 \le i \le K}\{q_i\} \le \nu_k^0\Big)\Big\}.$$

Now, since the p-value subset of $[0, 1]^m$ defined by the relation $\hat{k} \le k-1$ is nondecreasing, assumption (Posdep) ensures

$$T_1 \le \lambda\zeta \sum_{k=K}^{m_0} \Big\{\mathbb{P}_\theta\Big(\hat{k} \le k | \max_{1 \le i \le K}\{q_i\} \le \nu_k^0\Big) - \mathbb{P}_\theta\Big(\hat{k} \le k-1 | \max_{1 \le i \le K}\{q_i\} \le \nu_{k-1}^0\Big)\Big\},$$

which is below $\lambda\zeta$ because the sum is telescopic.

Now, for T_2, we use the same type of reasoning with $\mathbf{1}\{V_m(\nu_k^0) \ge k\} \le \frac{1}{k}\sum_{i=1}^{m_0} \mathbf{1}\{q_i \le \nu_k^0\}$ and $m_0\nu_k^0 \le (1-\lambda)\zeta k$ for $k < K$,

$$T_2 \le \sum_{i=1}^{m_0} \sum_{k=1}^{(K-1)\wedge m_0} \frac{1}{k} \mathbb{P}_\theta(\hat{k} = k, q_i \le \nu_k^0)$$
$$\le (1-\lambda)\zeta m_0^{-1} \sum_{i=1}^{m_0} \sum_{k=1}^{(K-1)\wedge m_0} \mathbb{P}_\theta(\hat{k} = k | q_i \le \nu_k^0)$$
$$\le (1-\lambda)\zeta m_0^{-1} \sum_{i=1}^{m_0} \sum_{k=1}^{(K-1)\wedge m_0} \mathbb{P}_\theta(\hat{k} \le k | q_i \le \nu_k^0) - \mathbb{P}_\theta(\hat{k} \le k-1 | q_i \le \nu_k^0)$$
$$\le (1-\lambda)\zeta m_0^{-1} \sum_{i=1}^{m_0} \sum_{k=1}^{(K-1)\wedge m_0} \{\mathbb{P}_\theta(\hat{k} \le k | q_i \le \nu_k^0) - \mathbb{P}_\theta(\hat{k} \le k-1 | q_i \le \nu_{k-1}^0)\},$$

by using again assumption (Posdep). Finally, the last display is below $(1-\lambda)\zeta$, because the sum is telescopic. This completes the proof.

8. Proofs for asymptotic results. In this section, the following well-known lemma will be extensively used.

LEMMA 8.1. *Let $\hat{\ell}$ be the number of rejections of the step-up* (SU) *algorithm associated to some critical values $(\tau_\ell)_{1\le\ell\le m}$. Consider the function f_m defined by*

$$f_m(t) = m^{-1} \times \min\{\ell \in \{0, \dots, m+1\} : \tau_\ell \ge t\}, \tag{39}$$

with the conventions $\tau_0 = 0$, $\tau_{m+1} = 1$. Let $\hat{t}$ be defined by

$$\hat{t} = \sup\{t \in [0,1] : \widehat{\mathbb{G}}_m(t) \ge f_m(t)\}. \tag{40}$$

Then the supremum into (40) *is a maximum, that is, $\widehat{\mathbb{G}}_m(\hat{t}) \ge f_m(\hat{t})$. Furthermore, $\hat{t} = \tau_{\hat{\ell}}$.*

8.1. *Proof of Lemma* 4.1. Actually, we prove the result for a more general class of procedures, where $\hat{t} = \tau_{\hat{\ell}}$ is obtained by (40) for a sequence of functions $f_m = \hat{f}_m$ (possibly random) which is uniformly close to $f_\infty(t) = t/\alpha$ on every compact of $(0, \alpha]$, that is,

$$\sup_{b\le t\le\alpha} |\hat{f}_m(t) - t/\alpha| \to 0 \qquad \text{a.s. for all } b \in (0, \alpha). \tag{41}$$

Note that $\hat{f}_m = f_\infty$ gives the BH procedure by Lemma 8.1. Next, since $R_m(\hat{t}) \ge m\hat{f}_m(\hat{t})$,

$$\begin{aligned}
&\mathbb{P}(\mathrm{FDP}_m(\hat{t}) > \alpha) \\
&\quad \le \mathbb{P}(V_m(\hat{t})/m > \alpha \hat{f}_m(\hat{t})) \\
&\quad = \mathbb{P}((m_0/m)(\widehat{\mathbb{G}}_{0,m}(\hat{t}) - \hat{t}) - \alpha(\hat{f}_m(\hat{t}) - \hat{t}/\alpha) > (1 - m_0/m)\hat{t}) \\
&\quad \le \mathbb{P}\Big((m_0/m)\|\widehat{\mathbb{G}}_{0,m} - I\|_\infty + \alpha \sup_{t^\star\le t\le\alpha} |\hat{f}_m(t) - t/\alpha| > (1 - m_0/m)t^\star\Big) \\
&\qquad + \mathbb{P}(\hat{t} \le t^\star),
\end{aligned}$$

for some $t^\star > 0$ satisfying $G(t^\star) > t^\star/\alpha$ [which exists by (Exists)]. By (29) and (weakdep0), it is sufficient to check that $\mathbb{P}(\hat{t} \le t^\star)$ tends to zero. For this, we use (weakdep) that ensures

$$\begin{aligned}
\mathbb{P}(\hat{t} > t^\star) &\ge \mathbb{P}(\widehat{\mathbb{G}}_m(t^\star) > \hat{f}_m(t^\star)) \\
&\ge \mathbb{P}(G(t^\star) > f_\infty(t^\star) + |G(t^\star) - \widehat{\mathbb{G}}_m(t^\star)| + |\hat{f}_m(t^\star) - f_\infty(t^\star)|) \to 1,
\end{aligned}$$

which completes the proof.

8.2. *Proof of Theorem* 4.2. By the proof of Lemma 4.1, it is sufficient to show that $f_m(t)$ defined by (39) is such that $f_m(t) \to t/\alpha$ for all $t \in [0, 1]$. This is an easy consequence of the fact that, since $V'_m(t)/m$ converges in probability to t, for any sequence $(\ell_m)_m$ with ℓ_m/m converging to some u, $\bar{B}_m(t, \lfloor\alpha\ell_m\rfloor + 1)$ converges to 1 if $t > \alpha u$ and 0 if $t < \alpha u$.

8.3. *Proof of Theorem* 4.3. First, by assumption (weakdep), we can assume that the convergence $\sup_{t\in[0,1]}|\widehat{\mathbb{G}}_m(t)-G(t)|\to 0$ is almost sure. Next, let us prove

$$\hat{t} \text{ converges a.s. to } t^\star\in(0,1), \tag{42}$$

where $t^\star=\sup\{t\in[0,1]:G(t)\geq \pi_0 t/\alpha\}$. First, $t^\star$ lies in $(0,1)$ by (NonCritical) and because $\pi_0>\alpha$. Then, by Lemma 8.1, we have $\hat{t}=\sup\{t\in[0,1]:\widehat{\mathbb{G}}_m(t)\geq f_m(t)\}$ where $f_m(t)$ is given (39). As in proof of Theorem 4.2, we easily check that for all $t\in[0,1]$, $f_m(t)$ converges to $\pi_0 t/\alpha$. As a result, since f_m is a nondecreasing function, the convergence of $f_m(t)$ to $\pi_0 t/\alpha$ is uniform on $[0,1]$. Now, to establish (42), it is sufficient to show that if $\hat{t}$ converges to some $t\in[0,1]$ along a subsequence, then we have $t=t^\star$. First, since $\widehat{\mathbb{G}}_m(\hat{t})\geq f_m(\hat{t})$, we have $G(t)\geq \pi_0 t/\alpha$ and thus $t\leq t^\star$. Let us prove $t\geq t^\star$. We have by (Unique) and (NonCritical) that $G(u_p)>f_\infty(u_p)$ for all p, for some $u_p\uparrow t^\star$. This yields, for all p and m large enough, $\widehat{\mathbb{G}}_m(u_p)>f_m(u_p)$ and thus $t\geq u_p$. Hence, $t\geq t^\star$ by making p tends to infinity. This proves (42).

Now, we have $\mathbb{P}(\mathrm{FDP}_m(\hat{t})>\alpha)=\mathbb{P}(V_m(\hat{t})>\alpha\hat{\ell})=\mathbb{P}(Z_m(\hat{t})>\Upsilon_m)$, by letting $\Upsilon_m=r_m(\alpha\hat{\ell}/m-\tau^0_{\hat{\ell}}m_0/m)$. By assumption (FLT), we have that $Z_m(\hat{t})$ converges in distribution to $Z(t^\star)$. Let $q_m(t)$ denotes the $(1-\zeta)$-quantile of $Z_m(t)$. From Lemma S-3.2, we have that the function sequence $q_m(t)$ converges uniformly to $q_\zeta(t)$ for t in any compact of $(0,1)$, where $q_\zeta(t)$ denotes the $(1-\zeta)$-quantile of $Z(t)$. From above, the proof is complete if we show

$$\Upsilon_m \text{ converges a.s. to } q_\zeta(t^\star). \tag{43}$$

Let us prove (43). By definition of B^0_m, we have $\mathbb{P}(V_m(\tau^0_\ell)>\alpha\ell)\leq\zeta<\mathbb{P}(V_m((\tau^0_\ell+\varepsilon/r_m)\wedge 1)>\alpha\ell)$, for all $\varepsilon>0$. Note that the latter uses that $m_0>\alpha m$ (for m large enough). This shows that for all $\ell\in\{1,\dots,m\}$, $q_m(\tau^0_\ell)\leq r_m(\alpha\ell/m-\tau^0_\ell m_0/m)\leq q_m((\tau^0_\ell+\varepsilon/r_m)\wedge 1)+\varepsilon$. Hence, applying this relation to $\ell=\hat{\ell}$, we get that for all $\varepsilon>0$, a.s., $q_\zeta(t^\star)\leq\liminf_m\Upsilon_m\leq\limsup_m\Upsilon_m\leq q_\zeta(t^\star)+\varepsilon$. Then (43) is derived by making ε tend to zero.

8.4. *Proof of Theorem* 4.4. We have

$$\begin{aligned}\mathbb{P}\big(\mathrm{FDP}_m(\hat{t})>\alpha\big)&\leq\mathbb{P}\big((m_0/m)\widehat{\mathbb{G}}_{0,m}(\hat{t})>\alpha\hat{\ell}/m,\hat{t}>0\big)\\ &\leq\mathbb{P}\bigg(\big\|\widehat{\mathbb{G}}_{0,m}(\cdot)-F_0(\cdot,W)\big\|_\infty>\Big(\frac{-\log(\lambda\zeta/2)}{2m_0}\Big)^{1/2}\bigg)\\ &\qquad+\mathbb{P}\big(F_0(\hat{t},W)>F_0(\hat{t},q_{\zeta(1-\lambda)}),\hat{t}\in(0,1)\big).\end{aligned} \tag{44}$$

Now, conditionally on W, the p_i's are i.i.d. of distribution function $F_0(\cdot,W)$. Hence, by applying the Dvoretzky–Kiefer–Wolfowitz inequality with the tight constant [see Massart (1990)], we get that the first term in the previous display is smaller than $\lambda\zeta$, which in turn implies that (44) is smaller than $\lambda\zeta+\mathbb{P}(W\geq q_{\zeta(1-\lambda)})\leq\zeta$ by (Posdep-facmod).

8.5. *Proof of Theorem* 4.5. First note that since $\bar{F}_\xi$ is continuous and increasing, so is $\bar{F}$ and $t \in [0,1] \mapsto F_0(t,w)$, for all w. Hence (31) defines the τ_ℓ's in an unique manner. Next, if $\mathbb{P}(c_1 > 0) = 0$, then $F_0(t,w) = t$ for all w, and thus the considered procedure is the BH procedure, which controls the FDP asymptotically by Lemma 4.1. Hence we can assume that $\mathbb{P}(c_1 > 0) > 0$. Let us denote the support of c_1 by $\{v_1, \dots, v_r\}$, for $r \geq 1$, $v_i \geq 0$, $v_i \neq v_j$ for $i \neq j$. We thus have that at least one v_i is positive. In particular, assumption (Posdep-facmod) holds.

Then, by using the Skorokhod representation theorem, up to consider a subsequence, we can assume that $(\hat{t}, W)$ is almost surely converging to some (T, W) (on appropriate subspaces). Denote $\kappa_\zeta = \max_{1\leq i\leq r}\{v_i q_\zeta\}$, $\kappa_W = \max_{1\leq i\leq r}\{v_i W\}$, and let us establish

$$T > 0 \qquad \text{a.s. if } \kappa_W + \beta > \kappa_\zeta. \tag{45}$$

For this, note that by Lemma 8.1, $\hat{t}$ is obtained by (40) with $f_m(t) = F_0(t, q_\zeta)/\alpha$, which gives $F_0(\hat{t}, q_\zeta) = \max\{t' \in [0,1] | \widehat{\mathbb{G}}'_m(t') \geq t'/\alpha\}$, where $\widehat{\mathbb{G}}'_m(t') = m^{-1}\sum_{i=1}^m \mathbf{1}\{F_0(p_i, q_\zeta) \leq t'\}$. Now observe that there exists a constant $D \in (0,1)$ such that for all u,

$$D\bar{F}_\xi\big(\bar{F}^{-1}(u) - \kappa_\zeta\big) \leq F_0(u, q_\zeta) \leq \bar{F}_\xi\big(\bar{F}^{-1}(u) - \kappa_\zeta\big). \tag{46}$$

It follows that $\widehat{\mathbb{G}}'_m(t')$ is lower-bounded by $m^{-1}\sum_{i=1}^m H_i \mathbf{1}\{\bar{F}_\xi(\bar{F}^{-1}(p_i) - \kappa_\zeta) \leq t'\}$, which by the law of large numbers [because (c_i, ξ_i) are i.i.d.] converges a.s. toward

$$\pi_1 \mathbb{P}\big(\bar{F}_\xi\big(\bar{F}_\xi^{-1}(t') + \kappa_\zeta - c_1 W - \beta\big) | W\big) \geq D' \bar{F}_\xi\big(\bar{F}_\xi^{-1}(t') + \kappa_\zeta - \kappa_W - \beta\big),$$

where D' is some positive constant. Assume now $\kappa_W + \beta > \kappa_\zeta$. By (33), the slope of $\bar{F}_\xi(\bar{F}_\xi^{-1}(t') + \kappa_\zeta - \kappa_W - \beta)$ is infinite in 0. Hence, for m large enough we have $F_0(\hat{t}, q_\zeta) > t'_0$, where t'_0 denotes any $t' \in (0,1)$ such that $D'\bar{F}_\xi(\bar{F}_\xi^{-1}(t') + \kappa_\zeta - \kappa_W - \beta) > t'/\alpha$. As a result, $T > 0$ and (45) is proved.

Now, we establish

$$\text{For all } \varepsilon > 0, \exists t_\varepsilon(W) \in (0,1), \text{ s.t.} \qquad \mathbb{P}\big(\hat{\ell} \geq 1, \hat{t} \leq t_\varepsilon(W) | W\big) \leq \varepsilon\alpha/D \qquad \text{if } \kappa_W + \beta < \kappa_\zeta. \tag{47}$$

By the LHS of (46), we obtain that $\widehat{\mathbb{G}}'_m(t')$ is upper-bounded by

$$\begin{aligned} m^{-1} &\sum_{i=1}^m \mathbf{1}\{\bar{F}_\xi(\beta + c_i W + \xi_i - \kappa_\zeta) \leq t'/D\} \\ &\leq m^{-1} \sum_{i=1}^m \mathbf{1}\{q_i \leq \bar{F}_\xi\big(\bar{F}_\xi^{-1}(t'/D) + \kappa_\zeta - \kappa_W - \beta\big)\}, \end{aligned}$$

where we let $q_i = \bar{F}_\xi(\xi_i)$, for $1 \leq i \leq m$, which are i.i.d. uniform. Now assume $\kappa_\zeta - \kappa_W - \beta > 0$, and take any $\varepsilon > 0$. By (33), there exists $t'_\varepsilon(W) \in (0,1)$ such that $\forall t' \in (0, t'_\varepsilon(W)]$, we have $\bar{F}_\xi(\bar{F}_\xi^{-1}(t'/D) + \kappa_\zeta - \kappa_W - \beta) \le \varepsilon t'/D$. Then we have

$$\mathbb{P}(\hat{\ell} \ge 1, \alpha\hat{\ell}/m \le t'_\varepsilon(W) | W) \le \mathbb{P}(\exists \ell \in \{1, \dots, m\} : q_{(\ell)} \le \varepsilon D^{-1}\alpha\ell/m | W),$$

which is below $\varepsilon D^{-1}\alpha$ by using Simes's inequality; see, for example, (18). This provides (47) by taking $t_\varepsilon(W) \in (0, 1)$ such that $F_0(t_\varepsilon(W), q_\zeta) = t'_\varepsilon(W)$.

The last argument is that when $T > 0$ a.s., we have $F_0(T, q_\zeta) > 0$ a.s. and thus

$$\mathrm{FDP}(\hat{t}) = \frac{m_0}{m}\frac{\widehat{\mathbb{G}}_{0,m}(\hat{t})}{\widehat{\mathbb{G}}_m(\hat{t})} = \frac{m_0}{m}\alpha\frac{\widehat{\mathbb{G}}_{0,m}(\hat{t})}{F_0(\hat{t}, q_\zeta)} \to \pi_0\alpha\frac{F_0(T, W)}{F_0(T, q_\zeta)}. \tag{48}$$

Now, by combining (45), (47) and (48), we obtain

$$\begin{aligned}
&\limsup_m \mathbb{P}(\mathrm{FDP}_m(\hat{t}) > \alpha) \\
&\quad\le \mathbb{E}\Big(\limsup_m \mathbf{1}\{\mathrm{FDP}_m(\hat{t}) > \alpha, \kappa_W + \beta > \kappa_\zeta\}\Big) \\
&\qquad + \limsup_m \mathbb{P}(\mathrm{FDP}_m(\hat{t}) > \alpha, \kappa_W + \beta < \kappa_\zeta, \hat{t} \le t_\varepsilon(W)) \\
&\qquad + \mathbb{E}\Big(\limsup_m \mathbf{1}\{\mathrm{FDP}_m(\hat{t}) > \alpha, \kappa_W + \beta < \kappa_\zeta, \hat{t} > t_\varepsilon(W)\}\Big) \\
&\quad\le \mathbb{P}\left(\pi_0\alpha\frac{F_0(T, W)}{F_0(T, q_\zeta)} \ge \alpha, \kappa_W + \beta > \kappa_\zeta, T > 0\right) + \varepsilon\alpha D^{-1} \\
&\qquad + \mathbb{P}\left(\pi_0\alpha\frac{F_0(T, W)}{F_0(T, q_\zeta)} \ge \alpha, \kappa_W + \beta < \kappa_\zeta, T \ge t_\varepsilon(W)\right).
\end{aligned}$$

Also note that $T < 1$ a.s. on the two above events, because $\pi_0\alpha < \alpha$. Hence we get

$$\limsup_m \mathbb{P}(\mathrm{FDP}_m(\hat{t}) > \alpha) \le \mathbb{P}(F_0(T, W) > F_0(T, q_\zeta), T \in (0, 1)) + \varepsilon\alpha D^{-1},$$

and the result comes from (Posdep-facmod) and by letting ε tends to zero.

Acknowledgments. We acknowledge an Associate Editor and three referees for their relevant suggestions that have substantially improved the quality of the manuscript. We are very grateful to Yoav Benjamini, Gilles Blanchard and Pierre Neuvial for helpful discussions. We also thank Bertrand Michel and Jacques Portes for their support during the numerical experiments.

SUPPLEMENTARY MATERIAL

Supplement to "New procedures controlling the false discovery proportion via Romano–Wolf's heuristic" (DOI: 10.1214/14-AOS1302SUPP; .pdf). The supplement presents additional materials for the paper; see Delattre and Roquain (2015b).

REFERENCES

BARDET, J.-M. and SURGAILIS, D. (2013). Moment bounds and central limit theorems for Gaussian subordinated arrays. *J. Multivariate Anal.* **114** 457–473. MR2993899

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

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

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

BLANCHARD, G., DELATTRE, S. and ROQUAIN, E. (2014). Testing over a continuum of null hypotheses with false discovery rate control. *Bernoulli* **20** 304–333. MR3160584

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

BLANCHARD, G. and ROQUAIN, É. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2871. MR2579914

BLANCHARD, G., DICKHAUS, T., ROQUAIN, É. and VILLERS, F. (2014). On least favorable configurations for step-up-down tests. *Statist. Sinica* **24** 1–23. MR3184590

CHI, Z. (2007). On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **35** 1409–1431. MR2351091

CHI, Z. and TAN, Z. (2008). Positive false discovery proportions: Intrinsic bounds and adaptive control. *Statist. Sinica* **18** 837–860. MR2440397

DEDECKER, J. and PRIEUR, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117** 121–142. MR2287106

DELATTRE, S. and ROQUAIN, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.* **81** 111–115. MR2740072

DELATTRE, S. and ROQUAIN, E. (2015a). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *Bernoulli*. To appear.

DELATTRE, S. and ROQUAIN, E. (2015b). Supplement to "New procedures controlling the false discovery proportion via Romano–Wolf's heuristic." DOI:10.1214/14-AOS1302SUPP.

DOUKHAN, P., LANG, G., SURGAILIS, D. and TEYSSIÈRE, G., eds. (2010). *Dependence in Probability and Statistics. Lecture Notes in Statistics* **200**. Springer, Berlin. MR2741808

DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. Springer, New York. MR2373771

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

FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887

FARCOMENI, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Stat.* **34** 275–297. MR2346640

FARCOMENI, A. (2009). Generalized augmentation to control the false discovery exceedance in multiple testing. *Scand. J. Stat.* **36** 501–517. MR2549707

FINNER, H., DICKHAUS, T. and ROTERS, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist.* **35** 1432–1455. MR2351092

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

FRIGUET, C., KLOAREG, M. and CAUSEUR, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist. Assoc.* **104** 1406–1415. MR2750571

GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517. MR1924303

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

GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101** 1408–1417. MR2279468

GONTSCHARUK, V. and FINNER, H. (2013). Asymptotic FDR control under weak dependence: A counterexample. *Statist. Probab. Lett.* **83** 1888–1893. MR3069893

GUO, W., HE, L. and SARKAR, S. K. (2014). Further results on controlling the false discovery proportion. *Ann. Statist.* **42** 1070–1101. MR3210996

GUO, W. and ROMANO, J. (2007). A generalized Sidak–Holm procedure and control of generalized error rates under independence. *Stat. Appl. Genet. Mol. Biol.* **6** 35 pp. (electronic). MR2306938

HOMMEL, G. and HOFFMAN, T. (1988). Controlled uncertainty. In *Multiple Hypotheses Testing* 154–161. Springer, Heidelberg.

KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. MR0599685

KARLIN, S. and RINOTT, Y. (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Ann. Statist.* **9** 1035–1049. MR0628759

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

KORN, E. L., TROENDLE, J. F., MCSHANE, L. M. and SIMON, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371

LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723.

LEHMANN, E. L. and ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

NEUVIAL, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2** 1065–1110. MR2460858

NEUVIAL, P. (2013). Asymptotic results on adaptive false discovery rate controlling procedures based on kernel estimators. *J. Mach. Learn. Res.* **14** 1423–1459. MR3081930

NEUVIAL, P. and ROQUAIN, E. (2012). On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **40** 2572–2600. MR3097613

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

RINOTT, Y. and SCARSINI, M. (2006). Total positivity order and the normal distribution. *J. Multivariate Anal.* **97** 1251–1261. MR2276157

ROMANO, J. P. and SHAIKH, A. M. (2006a). On stepdown control of the false discovery proportion. In *Optimality. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **49** 33–50. IMS, Beachwood, OH. MR2337829

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

ROMANO, J. P. and WOLF, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. MR2351090

ROQUAIN, E. (2011). Type I error rate control for testing many hypotheses: A survey with proofs. *J. SFdS* **152** 3–38. MR2821220

ROQUAIN, E. and VILLERS, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612. MR2797857

SARKAR, T. K. (1969). Some lower bounds of reliability. Ph.D. thesis, Stanford Univ., ProQuest LLC, Ann Arbor, MI. MR2618992

SARKAR, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* **35** 2405–2420. MR2382652

SARKAR, S. K. (2008). On methods controlling the false discovery rate. *Sankhyā* **70** 135–168. MR2551809

SCHWARTZMAN, A. and LIN, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika* **98** 199–214. MR2804220

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

SOULIER, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statist. Probab. Lett.* **54** 193–203. MR1858634

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766

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

UNIV. PARIS DIDEROT
SORBONNE PARIS CITÉ
LPMA, UMR 7599, F-75205, PARIS
FRANCE
E-MAIL: sylvain.delattre@univ-paris-diderot.fr

SORBONNE UNIVERSITÉS
UPMC UNIV. PARIS 6
UMR 7599, LPMA, F-75005, PARIS
FRANCE
E-MAIL: etienne.roquain@upmc.fr