Statistica Sinica **24** (2014), 1-23
doi:http://dx.doi.org/10.5705/ss.2011.205

ON LEAST FAVORABLE CONFIGURATIONS FOR STEP-UP-DOWN TESTS

Gilles Blanchard[1], Thorsten Dickhaus[2], Étienne Roquain[3] and Fanny Villers[3]

[1]*Potsdam University,* [2]*Humboldt University and* [3]*UPMC University*

Abstract: This paper investigates an open issue related to false discovery rate (FDR) control of step-up-down (SUD) multiple testing procedures. It has been established that for this type of procedure, under some broad conditions and in an asymptotic sense, the FDR is maximum when the signal strength under the alternative is maximum. In other words, so-called "Dirac uniform configurations" are asymptotically *least favorable* in this setting. It is known that this property also holds in a nonasymptotic sense (for any finite number of hypotheses) for the two extreme versions of SUD procedures, namely step-up and step-down (under additional conditions for the step-down case). It is therefore natural to conjecture that this nonasymptotic *least favorable configuration* property could more generally be true for "intermediate" forms of SUD procedures. We prove that this is not the case. The argument is based on the exact calculations proposed earlier by Roquain and Villers (2011a); we extend them by generalizing Steck's recursion to the case of two populations. Furthermore, we quantify the magnitude of this phenomenon by providing a nonasymptotic upper bound and explicit vanishing rates as a function of the total number of hypotheses.

Key words and phrases: False discovery rate, least favorable configuration, multiple testing, Steck's recursions, step-up-down.

1. Introduction

1.1. Least favorable configurations for multiple testing

In mathematical statistics, so-called least favorable parameter configurations (LFCs) play a pivotal role. For a statistical decision problem over a parameter space Θ with risk $R(\cdot,\cdot)$ and a decision rule δ, an LFC is any element $\theta^*(\delta)$ of Θ that maximizes the risk of δ over Θ. When available, the knowledge of an LFC allows one to obtain a bound on the risk over a possibly large parameter space, including non- or semi-parametric cases where Θ has infinite dimensionality. In practice, LFCs are of interest during the planning phase of an experiment when the aim is to design a procedure δ with an *a priori* guaranteed maximum risk

$$\max_{\theta\in\Theta} R(\theta,\delta) = R(\theta^*(\delta),\delta) = \alpha. \tag{1.1}$$

In particular, if there is a one-parameter family $(\delta_\beta)_\beta$ of candidate decision rules, (1.1) can be used for adequate calibration of $\beta \in I \subset \mathbb{R}$. If the right-hand side (RHS) of (1.1) is invertible w.r.t. β, we can derive a closed form for $\beta(\alpha)$; if not, we can still approximate β by using Monte-Carlo methods simulating the distribution corresponding to the LFC.

While LFC considerations naturally occur in hypothesis testing problems, they are particularly delicate for *multiple* hypothesis testing. The latter issue has been investigated by many authors, see Finner and Roters (2001); Benjamini and Yekutieli (2001); Lehmann and Romano (2005); Finner, Dickhaus, and Roters (2007); Romano and Wolf (2007); Guo and Rao (2008); Somerville and Hemmelmann (2008); Finner, Dickhaus, and Roters (2009); Finner and Gontscharuk (2009); Gontscharuk (2010). In that setting, a family of $m \geq 2$ null hypotheses $H_1, \ldots, H_m$ is to be tested simultaneously under a common statistical model with parameter space Θ, and some type I error criterion is used to account for multiplicity. In applications, it is relevant to determine LFCs over the restricted parameter spaces Θ_{m,m_0} where exactly m_0 out of m of the null hypotheses are true, for a given number m_0 that is fixed when maximizing over $\theta \in \Theta_{m,m_0}$. In the present work, we restrict our attention to multiple testing procedures that depend on the observed data only through a collection of marginal p-values, each associated with an individual null hypothesis. This is a common setting for multiple testing problems in high dimension. Moreover, we consider procedures that reject exactly those null hypotheses having their p-value less than a certain common threshold t^*, which can possibly be data-dependent. We call such procedures threshold-based for short.

LFCs for multiple testing depend crucially on the type I error criterion considered. We first discuss criteria defined through loss functions that only depend on the number of type I errors,

$$R(\theta, \delta) := \mathbb{E}_\theta[\phi(V_m)] , \tag{1.2}$$

where V_m is the number of type I errors of multiple testing procedure δ:

$$V_m = V_m(\theta, \delta) := |\{1 \leq i \leq m : H_i \text{ is true for } \theta \text{ and gets rejected by } \delta\}|. \tag{1.3}$$

Assume that ϕ is a nondecreasing function and that, for threshold-based procedures, t^* is a nonincreasing function in each p-value. When the p-values are jointly independent, it is known that the LFC over Θ_{m,m_0} is a *Dirac-uniform* (DU) distribution, with p-values corresponding to the m_0 true nulls independent uniform variables and the $m - m_0$ p-values under alternatives with point mass 1 at zero. This result is formally recalled in Section S1.2 of the supplement Blanchard et al. (2013). For example, this holds under the above assumptions for the family-wise error rate (FWER), which is the usual type I error concept in traditional multiple hypothesis testing theory.

Over the last two decades, alternative type I error criteria, introduced in applications to genomics, proteomics, neuroimaging, and astronomy, have led to massive multiple testing problems with large systems of hypotheses, see Dudoit and van der Laan (2008); Pantazis et al. (2005); Miller et al. (2001). Here, a less stringent notion of type I error control is needed in order to ensure reasonable power of the corresponding multiple tests. In particular, the false discovery rate (FDR) introduced by Benjamini and Hochberg (1995) has become a standard criterion for type I error control in large-scale multiple testing problems. It does not fall into the class of type I error measures defined by (1.2), and the LFC problem for the FDR criterion turns out to be a challenging issue – even for simple classes of multiple tests and under independence assumptions.

A possible approach to LFCs under the FDR criterion is in an asymptotic sense, when the number m of hypotheses tends to infinity, see Finner, Dickhaus, and Roters (2009) and Gontscharuk (2010). In practice, however, it is desirable to have precise information about LFCs for a fixed number m of hypotheses, in particular for design, planning, and calibration purposes. LFCs for the FDR criterion under *arbitrary* dependencies have also been studied by Lehmann and Romano (2005); Guo and Rao (2008).

1.2. Contributions

We focus on the *nonasymptotic* theory of LFCs under the FDR criterion for so-called step-up-down multiple tests (SUD procedures, for short). These procedures constitute a subclass of threshold-based multiple testing procedures wherein the threshold t^* is obtained by comparing the ordered p-values to a fixed set of critical values, see Tamhane, Liu and Dunnett (1998); Sarkar (2002).

We provide a survey of known LFC results for SUD procedures in specific model classes, in Section 3. New results for LFCs of SUD procedures are derived in Section 4. We establish that the DU configuration is not, generally, a nonasymptotic LFC. Then, since it is known that the DU configuration is the least favorable in an asymptotic sense, we derive precise, nonasymptotic, upper bounds on the difference between the FDR under an arbitrary alternative and under the DU configuration. In particular we analyze, for some specific situations, the rate at which these bounds decay to zero as the number of hypotheses m grows. In Section 5, we give exact formulas for computing the FDR under the so-called two-group *fixed* and *random* mixture models for the p-values. This is a toolbox section for the rest of the paper; the formulas are used to disprove numerically the conjecture studied in Section 4. This section builds on the previous work of Roquain and Villers (2011a,b), which is summarized. A novel addition is an extension of Steck's recursion to compute the joint cumulative distribution function (c.d.f.) of order statistics of two mixed populations, that is used to handle the case of the fixed mixture model. This point had been left open by Roquain and Villers (2011a,b) and is of intrinsic interest, because the computational complexity of existing methods for this kind of function is exponential with k (Glueck et al. (2008)) while Steck's recursion is polynomial. Finally, the exact formulas derived in Section 5 are used to discuss the appropriateness of the FDR criterion in Section 6.2.

2. Mathematical Setting

2.1. Models

Let $\mathcal{F}$ be the set of continuous c.d.f.s from $[0,1]$ into $[0,1]$. We consider a set of $m \geq 2$ null hypotheses $H_1, \ldots, H_m$, and assume the existence of a corresponding collection of tests with associated p-value family $\boldsymbol{p} := (p_i, i \in \{1, \ldots, m\})$, that constitutes the only observed information. Two closely related models for the joint distribution of p-values are common in the literature. The first is the *(two group) fixed mixture model*, denoted $\mathrm{FM}(m, m_0, F)$, with parameters $m \geq 2$, $1 \leq m_0 \leq m$, $F \in \mathcal{F}$. It models $\boldsymbol{p} = (p_i, i \in \{1, \ldots, m\})$ as a family of mutually independent variables, with, for all i,

$$p_i \sim \begin{cases} U(0,1) & \text{if } 1 \leq i \leq m_0, \\ F & \text{if } m_0 + 1 \leq i \leq m, \end{cases}$$

where $U(0,1)$ denotes the uniform distribution on $(0,1)$.

The second model is the *(two group) random mixture model*, denoted by $\mathrm{RM}(m, \pi_0, F)$, with parameters $m \geq 2$, $\pi_0 \in [0,1]$, $F \in \mathcal{F}$. Under this model, m_0 is an (unobserved) binomial random variable $\mathcal{B}(m, \pi_0)$, and $\boldsymbol{p}$ follows the $\mathrm{FM}(m, m_0, F)$ model conditionally on m_0.

Here, the true nulls are assigned to the m_0 (random or not) first coordinates. This is assumed without loss of generality by independence, and because we consider procedures that only depend on the order statistics of the p-values.

Common additional assumptions are that $F(x) \geq x$, for all x, and that F is concave. These are satisfied in the Gaussian location model: $F(t) = \overline{\Phi}(\overline{\Phi}^{-1}(t) - \mu)$, for a given alternative mean $\mu > 0$, where $\overline{\Phi}(z) = \mathbb{P}(Z \geq z)$ for $Z \sim \mathcal{N}(0,1)$. This is the alternative distribution of p-values when testing for $\mu \leq 0$ under a Gaussian location shift model with unit variance. The assumptions are also satisfied for the Dirac δ_0 distribution, as introduced by Finner and Roters (2001). The corresponding distribution in the FM model is called Dirac-uniform (DU) configuration and denoted by $\mathrm{FM}(m, m_0, F \equiv 1)$, or simply $\mathrm{DU}(m, m_0)$. We define similarly $\mathrm{RM}(m, \pi_0, F \equiv 1)$. The DU configuration can be seen as a limit of the Gaussian location model for an alternative mean $\mu = \infty$. It is often considered a natural candidate for an LFC of several global type I error rates, see, e.g., Finner, Dickhaus, and Roters (2007); Romano and Wolf (2007); Somerville and Hemmelmann (2008).

2.2. Procedures

We consider step-up-down (SUD) procedures introduced by Tamhane, Liu and Dunnett (1998), see also Sarkar (2002). Let a *threshold* or *critical value* collection be any nondecreasing sequence $\boldsymbol{t} = (t_k)_{1\le k\le m} \in [0,1]^m$, with $t_0 = 0$.

Definition 1. The *step-up-down procedure* $\mathrm{SUD}_\lambda(\boldsymbol{t})$ of order $\lambda \in \{1,\dots,m\}$ with threshold collection $\boldsymbol{t}$, given a sequence of reordered p-values $p_{(1)} \le p_{(2)} \le \cdots \le p_{(m)}$, with $p_{(0)} = 0$, rejects the ith hypothesis if $p_i \le t^* = t_{\hat{k}}$, with

$$\widehat{k} = \begin{cases} \max\{k \in \{\lambda,\dots,m\} \ : \ \forall k' \in \{\lambda,\dots,k\},\ p_{(k')} \le t_{k'}\} & \text{if } p_{(\lambda)} \le t_\lambda; \\ \max\{k \in \{0,\dots,\lambda\} \ : \ p_{(k)} \le t_k\} & \text{if } p_{(\lambda)} > t_\lambda. \end{cases} \tag{2.1}$$

For convenience, we identify procedures with their rejection sets, e.g., $\mathrm{SUD}_\lambda(\boldsymbol{t}) = \{1 \le i \le m \ : \ p_i \le t_{\hat{k}}\}$. Here $\lambda = 1$ and $\lambda = m$ correspond to the traditional step-down (SD) and step-up (SU) procedures, respectively. An illustration is provided in Figure 1.

A standard choice for $\boldsymbol{t}$ is Simes' (1986) linear critical values $t_k = \alpha k/m$ for a pre-specified level $\alpha \in (0,1)$. The corresponding step-up-down procedure is called the *linear step-up-down* procedure and denoted by LSUD_λ. For $\lambda = 1$ and $\lambda = m$, LSUD_λ is simply written as LSD and LSU, respectively. LSU is the procedure of Benjamini and Hochberg (1995).

More generally, threshold collections are commonly of the form $t_k = \rho(k/m)$ for a function $\rho : [0,1] \to [0,1]$ assumed to satisfy

$$\rho : [0,1] \to [0,1] \text{ is continuous and nondecreasing;} \tag{2.2}$$

$$x \in (0,1] \mapsto \rho(x)/x \text{ is nondecreasing.} \tag{2.3}$$

The function ρ is called the critical value function (and its inverse, the rejection curve, see, e.g., Finner, Dickhaus, and Roters (2009)). Assumptions (2.2) and (2.3) are meaningful in particular for analyzing asymptotics $m \to \infty$, wherein ρ is independent of m. For a fixed m, (2.3) is equivalent to "$k \mapsto t_k/k$ is nondecreasing"; Finner, Gontscharuk, and Dickhaus (2012) call such threshold collections *feasible* critical values.

2.3. False discovery rate and LFCs

Introduced by Benjamini and Hochberg (1995), the FDR of a multiple testing procedure is the averaged ratio of the number of erroneous rejections to the total

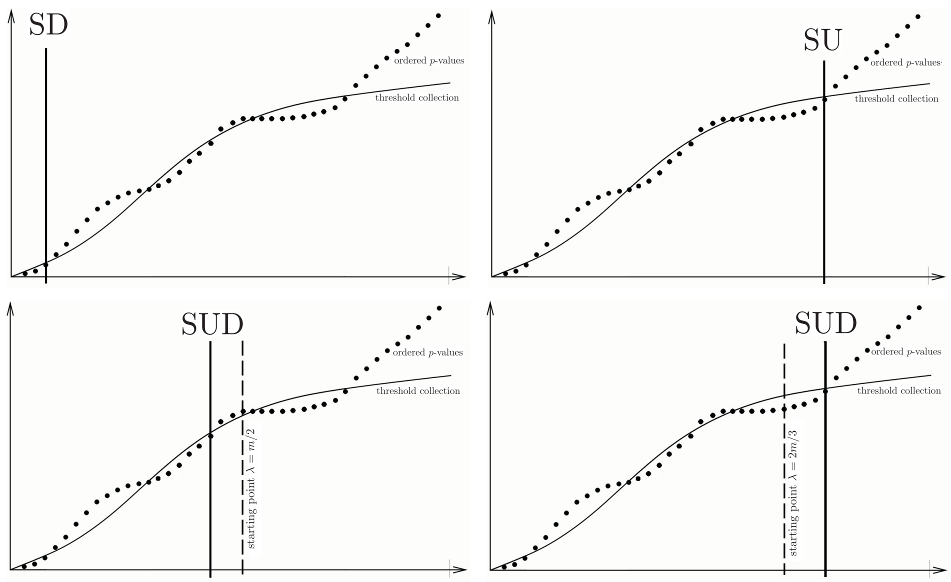

Figure 1. Value of $\widehat{k}$ (vertical solid line), defined by (2.1), for several procedures of the SUD type. The rejection set of the bottom-right SUD procedure (using $\lambda = 2m/3$) coincides with that of the SU procedure for this realization of the p-value family.

number of rejections. In our setting, for P either $\mathrm{FM}(m, m_0, F)$ or $\mathrm{RM}(m, \pi_0, F)$, the FDR of a step-up-down procedure can be written as

$$\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), P) = \mathbb{E}_{\mathbf{p}\sim P}\left[\mathrm{FDP}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, \boldsymbol{p})\right], \tag{2.4}$$

where the false discovery proportion (FDP) is

$$\mathrm{FDP}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, \boldsymbol{p}) = \frac{|\{1 \le i \le m_0 : p_i \le t_{\hat{k}}\}|}{|\{1 \le i \le m : p_i \le t_{\hat{k}}\}| \vee 1}; \tag{2.5}$$

here $|\cdot|$ denotes cardinality and m_0 is fixed or random according as P is $\mathrm{FM}(m, m_0, F)$ or $\mathrm{RM}(m, \pi_0, F)$. We use the short notation $\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F)$ for the quantity $\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), \mathrm{FM}(m, m_0, F))$, resp. $\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), \pi_0, F)$ for the quantity $\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), \mathrm{RM}(m, \pi_0, F))$, or simply $\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), F)$ when the model is unambiguous.

Definition 2. A c.d.f. $F' \in \mathcal{F}$ is a least favorable configuration (LFC) for the FDR of $\mathrm{SUD}_\lambda(\boldsymbol{t})$ in the fixed mixture model with m_0 true hypotheses out of m if $\forall F \in \mathcal{F},\ \mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F) \le \mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F')$.

A similar definition holds for the random mixture model with m hypotheses and proportion π_0 of true hypotheses.

This definition can be restricted to a subclass $\mathcal{G} \subset \mathcal{F}$ (typically, the class of concave c.d.f.s). Finally, if F' is an LFC that is *common* to all values of m_0 for the FM(m, m_0, F) model, then F' is also an LFC in the RM(m, π_0, F) model for any value of π_0 (by integrating over $m_0 \sim \mathcal{B}(m, \pi_0)$).

3. Survey of FDR Comparison Results for SUD Procedures

We briefly survey (and summarize in Figure 2) existing comparison results concerning the FDR of step-up-down procedures under the FM and RM models. They concern inequalities between the FDR of different procedures under the same model, or of the same procedure under different alternatives.

Consider first the problem of the monotonicity of FDR(SUD$_\lambda(\boldsymbol{t})$) in λ (vertical arrows). Provided F is concave, it was established by Theorem 4.1 of Zeisel, Zuk and Domany (2011) that the FDR grows as the rejection set grows. In particular, since $\mathrm{SUD}_\lambda(\boldsymbol{t}) \subseteq \mathrm{SUD}_{\lambda+1}(\boldsymbol{t})$ for any $\lambda \in \{1, \ldots, m-1\}$, we have

$$\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t})) \leq \mathrm{FDR}(\mathrm{SUD}_{\lambda+1}(\boldsymbol{t})), \tag{3.1}$$

both for the FM(m, m_0, F) and RM(m, π_0, F) models. This implies in particular that $\mathrm{FDR}(\mathrm{SD}(\boldsymbol{t})) \leq \mathrm{FDR}(\mathrm{SU}(\boldsymbol{t}))$ for a concave F. Similar inequalities have been obtained under a condition on the threshold collection $\boldsymbol{t}$, rather than on F: Theorem 4.3 of Finner, Dickhaus, and Roters (2009) and Theorem 3.10 of Gontscharuk (2010) establish that, when $k \mapsto t_k/k$ is nondecreasing, for any $\lambda \in \{1, \ldots, m-1\}$,

$$\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t})) \leq \mathrm{FDR}(\mathrm{SU}(\boldsymbol{t})), \tag{3.2}$$

both for the FM(m, m_0, F) and RM(m, π_0, F) models. Though (3.1) and (3.2) suggest that FDR(SU$(\boldsymbol{t})$) is generally larger than FDR(SD$(\boldsymbol{t})$), we show by a counterexample in Section S1.1 of Blanchard et al. (2013) that this is cannot hold in the absence of any assumptions on F or $\boldsymbol{t}$.

We turn to the monotonicity of FDR(SUD$_\lambda(\boldsymbol{t}), F$) in F. For the *step-up*, Theorem 5.3 of Benjamini and Yekutieli (2001) states that $F \leq F'$ implies $\mathrm{FDR}(\mathrm{SU}(\boldsymbol{t}), F) \leq \mathrm{FDR}(\mathrm{SU}(\boldsymbol{t}), F')$ if $k \mapsto t_k/k$ is nondecreasing, with the inequality reversed if $k \mapsto t_k/k$ is nonincreasing. For the *step-down* and under the RM(m, π_0, F) model, Theorem 4.1 of Roquain and Villers (2011a) states that the Dirac-uniform configuration is an LFC under a (complicated) condition on $\boldsymbol{t}$, that is fulfilled in particular by the linear threshold collection $t_k = \alpha k/m$, $\alpha \in (0, 1)$ over the class of concave c.d.f.s. These results thus establish the LFC property of the DU(m, m_0) distribution for SU and SD procedures under appropriate assumptions (F concave and linear threshold family being sufficient).

For SUD procedures that are neither step-up nor step-down (i.e. $\lambda \notin \{1, m\}$), the only comparison result we know of is asymptotic in m. Precisely, combining Theorem 4.3 of Finner, Dickhaus, and Roters (2009) and Lemma 3.7 of Gontscharuk (2010), we obtain the following:

Theorem 1 (Gontscharuk (2010)). *Let $\boldsymbol{t}$ be a threshold collection of the form $t_k = \rho(k/m)$, where ρ satisfies (2.2) and (2.3). Let λ_m be a sequence such that $\lambda_m/m \to \kappa \in [0,1]$, and $m_0(m)$ a sequence such that $m_0(m)/m \to \zeta \in [0,1]$. If $|SUD_{\lambda_m}(\boldsymbol{t})|/m$ converges in probability as $m \to \infty$ under the* $\mathrm{DU}(m, m_0(m))$ *distribution, then we have under the model* $\mathrm{FM}(m, m_0(m), F)$*, for any $F \in \mathcal{F}$:*

$$\limsup_m \left\{FDR(SUD_{\lambda_m}(\boldsymbol{t}), F) - FDR(SUD_{\lambda_m}(\boldsymbol{t}), F \equiv 1)\right\} \leq 0, \qquad (3.3)$$

for all $\zeta \in [0,1]$ if $\kappa > 0$ or for all $\zeta \in [0,1)$ if $\kappa = 0$.

These results leave open the question of (nonasymptotic) LFCs for SUD procedures. Still, as a whole they seem to point towards $\mathrm{DU}(m, m_0)$ as an LFC – at least for a linear threshold family and concave F, thus prompting the following conjecture that motivates the contributions of this paper:

Conjecture 1. *For any $m \geq 2$, the Dirac-uniform configuration is an LFC for the FDR of any linear step-up-down procedure in the* $\mathrm{RM}(m, \pi_0, F)$ *and* $\mathrm{FM}(m, m_0, F)$ *models, at least over the class of concave F.*

4. Analysis of Conjecture 1

4.1. Disproving the conjecture with a numerical counterexample

The exact calculations described in Section 5 allow us the numerical computation of $\mathrm{FDR}(\mathrm{LSUD}_\lambda(\boldsymbol{t}))$, which leads to the following:

Numerical result. *Put $m = 10$, $\alpha = 0.5$, and take $F_0(x) = x$. Then, for any $\lambda \in \{4, 5, 6, 7\}$ we have*

$$FDR(LSUD_\lambda, F) > FDR(LSUD_\lambda, F \equiv 1), \qquad (4.1)$$

under the $\mathrm{FM}(m, m_0, F)$ *model with $m_0 = 7$ and $F = F_0$, and under the* $\mathrm{FM}(m, m_0, F)$ *model with $\pi_0 = 7/10$ and $F = F_0$.*

We display the corresponding values graphically in Figure 3 (top), along with other FDR values computed for the Gaussian alternative c.d.f. $F = \overline{\Phi}(\overline{\Phi}^{-1}(\cdot) - \mu)$, providing additional configurations for which (4.1) holds (e.g., $\mu = 1$ and $\lambda = 5$). The case $F(x) = x$ corresponds to $\mu = 0$, while $F \equiv 1$ corresponds to $\mu = +\infty$. The results for the $\mathrm{FM}(m, m_0, F)$ (top left panel) and $\mathrm{RM}(m, \pi_0, F)$ (top right panel) models are qualitatively the same.

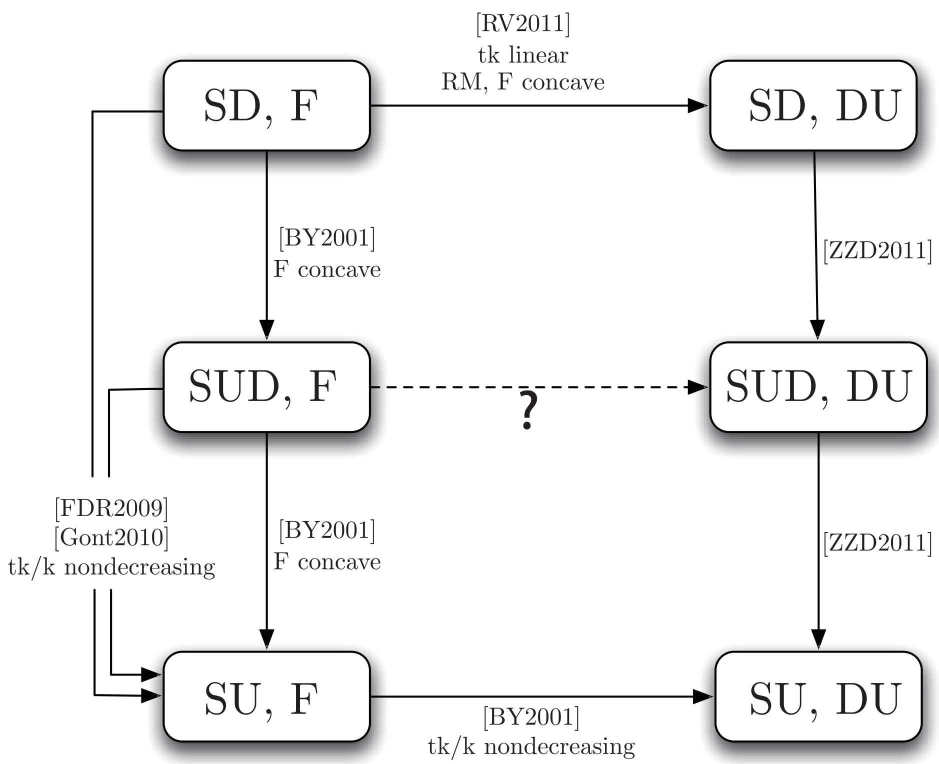

Figure 2. An arrow "A → B" means "FDR(A) $\leq$ FDR(B)". These results hold for the fixed mixture (FM) model except when "RM" is written. The brackets are a shortened reference to the corresponding literature, see main text for more details.

While this disproves Conjecture 1, as expected from the asymptotic analysis of Theorem 1, as m becomes larger, the amplitude of the phenomenon vanishes, see Figure 3 (bottom).The next question is therefore a more precise quantification of the difference of the two sides of (4.1).

Remark 1. The results were double-checked via extensive and independent Monte-Carlo simulations in order to exclude the possibility that the reported phenomenon could be an artifact produced by accumulated rounding errors in the numerical computation.

Remark 2. For counterexample 4.1, we used the identity c.d.f. F_0. By continuity of the exact formulas obtained in Section 5 as a function of the values $F(t_i)$, we conclude that the LFC conjecture is also false over any restricted subclass $\mathcal{F}' \subset \mathcal{F}$ as soon as F_0 is an adherent point of $\mathcal{F}'$ (in the sense of weak convergence of probability measures). Many standard classes have this property, such as the

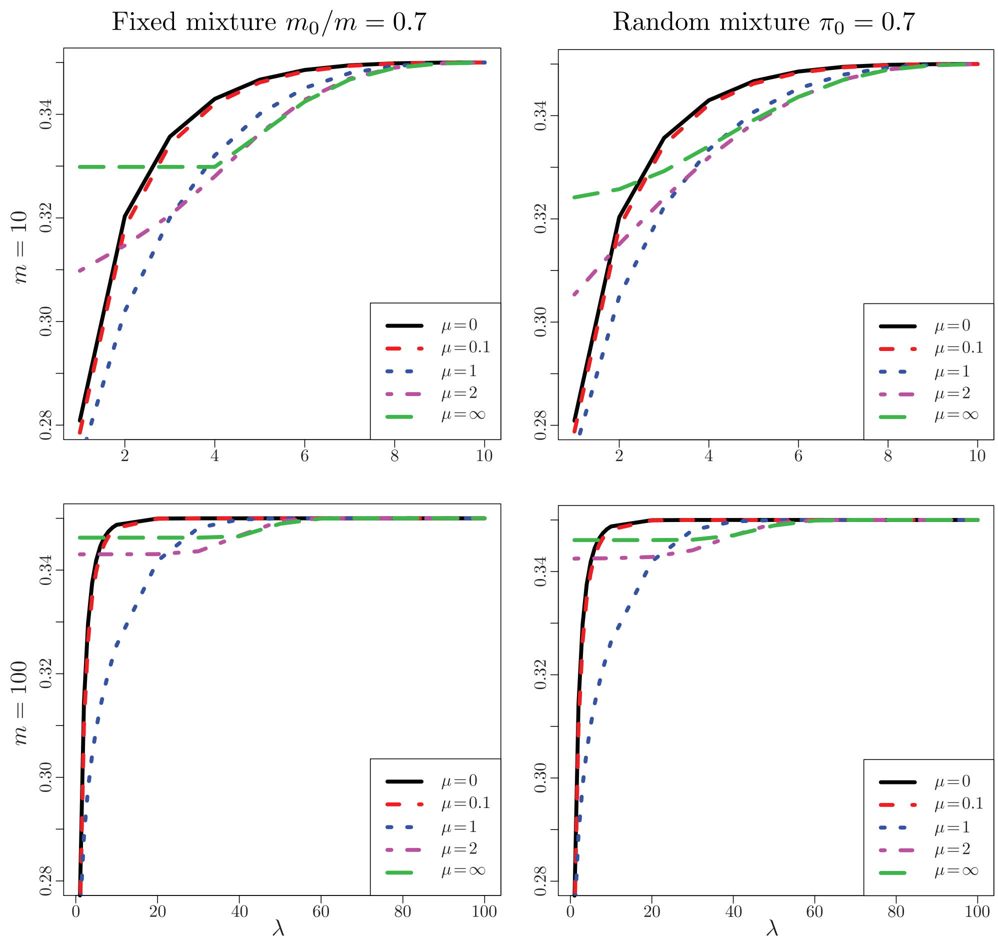

Figure 3. The LFC of LSUD is not always DU. FDR(LSUD_λ) as a function of the order $\lambda \in \{1, \ldots, m\}$. Left: fixed mixture; right: random mixture. The graphs were obtained under a one-sided Gaussian location model with parameter μ. The target FDR level is $\alpha = 0.5$.

one-sided location class $\mathcal{F}' = \{F_\mu \ : \ F_\mu(x) = \overline{D}(\overline{D}^{-1}(x) - \mu), \mu > 0\}$ and the scale class $\mathcal{F}' = \{F_\sigma \ : \ F_\sigma(x) = 2\overline{D}(\overline{D}^{-1}(x/2)/\sigma), \sigma > 1\}$, for which $\overline{D}$ is some upper tail distribution function (see, e.g., Section 2.1 of Neuvial and Roquain (2012) for details).

4.2. Nonasymptotic bound

We now derive a nonasymptotic version of Theorem 1. The main idea is to consider the SUD procedure as a function operating on c.d.f.s, and develop a perturbation analysis when the empirical c.d.f. of the p-values is δ-close to the population c.d.f. (which happens with large probability). In this perspective, we introduce the following notation:

Definition 3. Let $\rho : [0,1] \to [0,1]$ be a continuous and nondecreasing function. For any nondecreasing function $G : [0,1] \to [0,1]$, and $\ell \in [0,1]$, define

$$\mathcal{U}(\ell, G) := \begin{cases} \min\{u \in [\ell,1] \ : \ G(\rho(u)) \le u\} & \text{if } G(\rho(\ell)) \ge \ell; \\ \max\{u \in [0,\ell] \ : \ G(\rho(u)) \ge u\} & \text{if } G(\rho(\ell)) < \ell. \end{cases} \tag{4.2}$$

Here, the infimum and supremum are well-defined, since the considered sets are nonempty. Since G is nondecreasing, $\mathcal{U}(\ell, G)$ is a fixed point of the function $G \circ \rho$ (so that the infimum is indeed a minimum and the supremum, a maximum).

Denote by $\hat{\mathbb{G}}_m(x) := m^{-1} \sum_{i=1}^m \mathbf{1}\{p_i \le x\}$ the empirical c.d.f. of the p-values. The following lemma establishes the close connection between the functional $\mathcal{U}$ and the SUD_λ procedure:

Lemma 1. *For $t_k := \rho(k/m)$, it holds*

$$\mathcal{U}\left(\frac{\lambda}{m}, \hat{\mathbb{G}}_m\right) \le \frac{\widehat{k}}{m} \le \mathcal{U}\left(\frac{\lambda}{m}, (\hat{\mathbb{G}}_m + m^{-1}) \wedge 1\right), \tag{4.3}$$

where $\widehat{k}$ is the number of hypotheses rejected by the SUD_λ procedure (2.1).

The main result of this section is proved in Section 7.2:

Theorem 2. *Let $\rho : [0,1] \to [0,1]$ satisfy* (2.2)–(2.3) *and let $t_k := \rho(k/m)$. For $\zeta, \delta \in (0,1)$ arbitrary constants, let*

$$u_\delta^+ := \mathcal{U}\left(\frac{\lambda}{m}, (G_\zeta^{DU} + \delta) \wedge 1\right) \quad \text{and} \quad u_\delta^- := \mathcal{U}\left(\frac{\lambda}{m}, (G_\zeta^{DU} - \delta) \vee 0\right),$$

where $G_\zeta^{DU}(x) := (1-\zeta) + \zeta x$; and, for any $y \in (0,1)$, let

$$\varepsilon(\delta, m, \zeta, y) := \left(\frac{\rho(u_\delta^+) - \rho(u_\delta^-)}{u_\delta^+}\right)\zeta + \frac{4\zeta}{1-\zeta} e^{-2m(\delta - y - 1/m)_+^2 (1 - y/\zeta)_+}. \tag{4.4}$$

Then, for any $F \in \mathcal{F}$ and $\lambda \in \{1, \ldots, m\}$, in the $\mathrm{FM}(m, m_0, F)$ *model with $0 < m_0 < m$, we have*

$$FDR(SUD_\lambda(\boldsymbol{t}), m_0, F) \le FDR(SUD_\lambda(\boldsymbol{t}), m_0, F \equiv 1) + \varepsilon\left(\delta, m, \frac{m_0}{m}, \frac{1}{m}\right). \tag{4.5}$$

In the $\mathrm{RM}(m, \pi_0, F)$ *model, for $m \ge 3$ we have*

$$FDR(SUD_\lambda(\boldsymbol{t}), \pi_0, F) \le FDR(SUD_\lambda(\boldsymbol{t}), \pi_0, F \equiv 1) + \varepsilon\left(\delta, m, \pi_0, \sqrt{\frac{\log m}{m}}\right) + \frac{2}{m}. \tag{4.6}$$

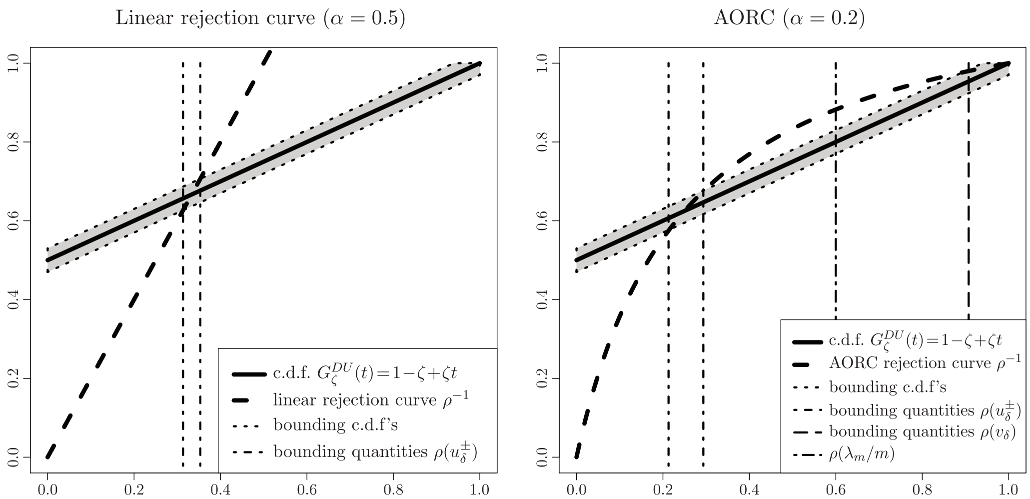

Figure 4. Illustration of u_δ^+ and u_δ^- for the LSUD and the SUD based on the AORC. Here, the horizontal axis is on the "threshold scale" $t = \rho(u)$. $\zeta = 0.5$; $\delta = 0.03$. The area between $(G_\zeta^{DU} - \delta) \vee 0$ and $(G_\zeta^{DU} + \delta) \wedge 1$ is displayed in gray.

The bounds are valid for any $\delta \in (0,1)$, so one can minimize (4.4) over δ to get the sharpest bound. The first term in (4.4) is a bound on the perturbation of the FDR compared to the population case when $\|\mathbb{G}_m - G_\zeta^{DU}\|_\infty \leq \delta$. The second term bounds the probability of the complement of the latter event. As m grows, $\delta = \delta_m$ should be chosen to decrease to zero, while ensuring that the second term remains negligible. We now analyze the behavior of the first term as δ becomes small; and in Section 4.3, we give the resulting rate as $m \to \infty$ and δ_m is chosen appropriately.

As a first example, consider $\rho(x) = \alpha x$ (LSUD, Figure 4, left panel); then $u_\delta^- = (1-\zeta-\delta)/(1-\alpha\zeta) \vee 0$ and $u_\delta^+ = (1-\zeta+\delta)(1-\alpha\zeta) \wedge 1$. Hence, $(\rho(u_\delta^+) - \rho(u_\delta^-))\zeta/u_\delta^+ \leq (2\alpha\zeta\delta)/(1-\zeta+\delta)$. As a result, (4.5) and (4.6) hold by replacing the first term of $\varepsilon(\delta, m, \zeta, y)$ by $(2\alpha\zeta\delta)/(1-\zeta+\delta)$.

Consider now

$$\rho(u) = \frac{\alpha u}{1 - u(1-\alpha)}, \text{ that is, } \rho^{-1}(t) = \frac{t}{\alpha + t(1-\alpha)}, \tag{4.7}$$

with $\zeta > \alpha$. The rejection curve ρ^{-1} (Figure 4 right panel), is the *asymptotically optimal rejection curve* (AORC), introduced by Finner, Dickhaus, and Roters (2009). Here, equation $(G_\zeta^{DU} + \delta)(u) = \rho(u)$ defining u_δ^+ is quadratic with two roots in $[0,1]$; let v_δ be the largest one. If $\lambda/m \geq v_\delta$, we have $u_\delta^+ = v_\delta$ and obtain a singular situation where the bound $\rho(u_\delta^+) - \rho(u_\delta^-)$ remains large for small δ (see Figure 4 for an illustration). This is related to the known fact that

AORC-based SUD procedures become unstable if λ/m is too close to 1. In the "regular" case where $\lambda/m < v_\delta$, u_δ^+ is the smaller root; the two points $\rho(u_\delta^-)$ and $\rho(u_\delta^+)$ are converging to each other as δ becomes small and the bound given in Theorem 2 vanishes. The exact expressions of u_δ^-, u_δ^+, and v_δ in this case can be derived by solving the corresponding quadratic equations. We then have $v_\delta = 1 - \delta\alpha/(\zeta-\alpha) + O(\delta^2)$, $u_\delta^+ = (1-\zeta)/(1-\alpha) + \delta\zeta/(\zeta-\alpha) + O(\delta^2)$, and $u_\delta^- = (1-\zeta)/(1-\alpha) - \delta\zeta/(\zeta-\alpha) + O(\delta^2)$. Since $\rho'((1-\zeta)/(1-\alpha)) = \alpha/\zeta^2$, we have $\rho(u_\delta^+) - \rho(u_\delta^-) = 2\alpha\delta/(\zeta^2-\alpha\zeta) + O(\delta^2)$. Assuming $\lambda/m < v_\delta$, we find that the first term of $\varepsilon(\delta, m, \zeta, y)$ is equivalent to $[2\alpha(1-\alpha)/(\zeta-\alpha)(1-\zeta)]\delta$ as δ tends to zero.

4.3. Convergence rate when m tends to infinity

We use Theorem 2 to obtain an explicit bound on the convergence rate of the LHS expression in (3.3) for specific critical value functions.

Corollary 1. *Let $\alpha \in (0,1)$. Let $t_k := \rho(k/m)$, for either $\rho(x) := \alpha x$ or ρ given by* (4.7) (*AORC*). *Consider the SUD procedure with threshold collection $\boldsymbol{t} = (t_k)_{1\le k\le m}$ and of order $\lambda = \lambda_m$ possibly depending on m. With $(\zeta_m)_m \in (\alpha, 1)$, consider either the* $\mathrm{FM}(m, m_0, F)$ *model with $m_0 = \lfloor \zeta_m m \rfloor$ or the* $\mathrm{RM}(m, \pi_0, F)$ *model with $\pi_0 = \zeta_m$. Assume* $[1/(1-\zeta_m)]\sqrt{(\log m)/m} = o(1)$*; for the AORC case, assume additionally* $\liminf_m \zeta_m > \alpha$ *and* $[1/(1-\lambda_m/m)]\sqrt{(\log m)/m} = o(1)$. *Then for any $F \in \mathcal{F}$,*

$$\left(FDR(SUD_{\lambda_m}(\boldsymbol{t}), F) - FDR(SUD_{\lambda_m}(\boldsymbol{t}), F\equiv 1)\right)_+ = O\Big(\frac{1}{1-\zeta_m}\sqrt{\frac{\log m}{m}}\Big). \quad (4.8)$$

The assumption $\zeta_m > \alpha$ is not restrictive, since when $\zeta_m \le \alpha$, controlling the FDR is a trivial problem: the procedure rejecting all the hypotheses has an FDR (and even an FDP) smaller than $\zeta_m \le \alpha$. While limited for simplicity to the linear and AORC rejection curves, the conclusion of Corollary 1 is more informative than that of Theorem 1. For $\zeta_m = \zeta \in (0,1)$ fixed independently of m, the convergence in (4.8) occurs at a parametric rate, up to a $\log m$ factor. Furthermore, the multiplicative constant in the $O(\cdot)$-notation can be derived explicitly using the arguments from the previous section. Finally, for ζ_m tending to 1 slower than $\sqrt{(\log m)/m}$ ("moderately" sparse case), the bound still converges to zero, though at a slower rate. The assumptions of Corollary 1 are more restrictive than those of Theorem 1 however, excluding in particular the case where ζ_m tends to 1 faster than $\sqrt{(\log m)/m}$ ("highly" sparse case).

5. Exact Formulas

We summarize here some formulas derived by Roquain and Villers (2011a,b) to calculate the joint distribution of the number of false discoveries and the number of discoveries, and additionally contribute a new recursion that makes these formulas fully usable in all considered models. These calculations were used to state the numerical result of Section 4.1, and also provide an exact expression for the FDP distribution.

5.1. A generalization of Steck's recursion

We need as an auxiliary result the multidimensional cumulative distribution function of order statistics in a two population model. Let $k \geq 0$. For any $\boldsymbol{t} = (t_1, \ldots, t_k)$, let

$$\Psi_k(\boldsymbol{t}) = \Psi_k(t_1, \ldots, t_k) := \mathbb{P}\left[U_{(1)} \leq t_1, \ldots, U_{(k)} \leq t_k\right], \tag{5.1}$$

where $(U_i)_{1 \leq i \leq k}$ is a sequence of i.i.d. random variables uniformly distributed on $(0,1)$, and with $\Psi_0(\cdot) \equiv 1$. These functions can be evaluated using Steck's recursion $\Psi_k(\boldsymbol{t}) = (t_k)^k - \sum_{j=0}^{k-2} \binom{k}{j} (t_k - t_{j+1})^{k-j} \Psi_j(t_1, \ldots, t_j)$ (Shorack and Wellner (1986)).

We generalize this to the case of two populations. For $0 \leq k_0 \leq k$ and any threshold collection $\boldsymbol{t} = (t_1, \ldots, t_k)$, let

$$\Psi_{k,k_0,F}(t_1, \ldots, t_k) := \mathbb{P}\left[U_{(1)} \leq t_1, \ldots, U_{(k)} \leq t_k\right], \tag{5.2}$$

where $(U_i)_{1 \leq i \leq k}$ is a sequence of independent random variables, with $(U_i)_{1 \leq i \leq k_0}$ uniformly distributed on $(0,1)$, and $(U_i)_{k_0+1 \leq i \leq k}$ having c.d.f. F. By convention, $\Psi_{0,0,F}(\cdot) = 1$. To our knowledge, existing formulas for computing $\Psi_{k,k_0,F}$ have a complexity exponential with k (Glueck et al. (2008)). We propose a substantially less complex computation:

Proposition 1. *For $0 \leq k_0 \leq k$,*

$$\Psi_{k,k_0,F}(t_1, \ldots, t_k) = (t_k)^{k_0} F(t_k)^{k-k_0} - \sum_{\substack{0 \leq j_0 \leq j \leq k-2 \\ j_0 \leq k_0 \\ j - j_0 \leq k - k_0}} \binom{k_0}{j_0} \binom{k-k_0}{j-j_0}$$
$$\times (t_k - t_{j+1})^{k_0 - j_0} (F(t_k) - F(t_{j+1}))^{k-k_0-j+j_0} \Psi_{j,j_0,F}(t_1, \ldots, t_j). \tag{5.3}$$

The above formula uses $0^0 := 1$. This proposition is proved in Section 7.4. The case $k = k_0$ reduces to the standard Steck's recursion.

5.2. Reformulating the result of Roquain and Villers (2011)

For any threshold collection $\boldsymbol{t} = (t_k)_{1 \leq k \leq m}$, any $F \in \mathcal{F}$, and for any $\pi_0 \in [0,1]$, $0 \leq k \leq m$, $0 \leq j \leq k$, set

$$\mathcal{P}_{m,\pi_0,F}(\boldsymbol{t}, k, j) = \binom{m}{j} \binom{m-j}{k-j} \pi_0^j \pi_1^{k-j} (t_k)^j (F(t_k))^{k-j}$$

$$\times\Psi_{m-k}(1-G(t_m),\ldots,1-G(t_{k+1})); \tag{5.4}$$

$$\widetilde{\mathcal{P}}_{m,\pi_0,F}(\boldsymbol{t},k,j) = \binom{m}{j}\binom{m-j}{k-j}\pi_0^j\pi_1^{k-j}(1-G(t_{k+1}))^{m-k} \times\Psi_{k,j,F}(t_1,\ldots,t_k), \tag{5.5}$$

where $G(t) = \pi_0 t + (1-\pi_0)F(t)$. For any $m_0 \in \{0,\ldots,m\}$, $k \geq 0$, $k \leq m$, $j \leq m_0$, $k-j \leq m-m_0$, set

$$\mathcal{Q}_{m,m_0,F}(\boldsymbol{t},k,j) = \binom{m_0}{j}\binom{m-m_0}{k-j}(t_k)^j(F(t_k))^{k-j} \times\Psi_{m-k,m_0-j,\overline{F}}(1-t_m,\ldots,1-t_{k+1}); \tag{5.6}$$

$$\widetilde{\mathcal{Q}}_{m,m_0,F}(\boldsymbol{t},k,j) = \binom{m_0}{j}\binom{m-m_0}{k-j}(1-t_{k+1})^{m_0-j}(1-F(t_{k+1}))^{m-m_0-k+j} \times\Psi_{k,j,F}(t_1,\ldots,t_k), \tag{5.7}$$

where $\overline{F}(t) = 1-F(1-t)$. The following theorem summarizes some of the results of Roquain and Villers (2011a,b).

Theorem 3 (Roquain and Villers (2011))**.** *Let $R = SUD_\lambda(\boldsymbol{t})$ be a step-up-down procedure of order $\lambda \in \{1,\ldots,m\}$ with a threshold collection $\boldsymbol{t}$. Let $V := R \cap \{1,\ldots,m_0\}$ be the set of false rejections. Then,*

(i) *In the* RM(m,π_0,F) *model, for any $\pi_0 \in [0,1]$, $F \in \mathcal{F}$, $0 \leq k \leq m$, $0 \leq j \leq k$,*

$$\mathbb{P}(|V| = j, |R| = k) = \begin{cases} \mathcal{P}_{m,\pi_0,F}(\boldsymbol{t}\wedge t_\lambda,k,j) & \text{for } k < \lambda, \\ \widetilde{\mathcal{P}}_{m,\pi_0,F}(\boldsymbol{t}\vee t_\lambda,k,j) & \text{for } k \geq \lambda. \end{cases} \tag{5.8}$$

(ii) *In the* FM(m,m_0,F) *model, for any $m_0 \in \{0,\ldots,m\}$, $F \in \mathcal{F}$, $0 \leq k \leq m$, $0 \vee (k-m+m_0) \leq j \leq m_0 \wedge k$,*

$$\mathbb{P}(|V| = j, |R| = k) = \begin{cases} \mathcal{Q}_{m,m_0,F}(\boldsymbol{t}\wedge t_\lambda,k,j) & \text{for } k < \lambda, \\ \widetilde{\mathcal{Q}}_{m,m_0,F}(\boldsymbol{t}\vee t_\lambda,k,j) & \text{for } k \geq \lambda. \end{cases} \tag{5.9}$$

The above formulas in combination with Proposition 1 allow the exact computation of the full joint distribution of $(|V|,|R|)$ in the considered models, therefore also of the distribution of the FDP (which equals $|V|/(|R|\vee 1)$), and of the FDR, its expectation. The computations were found to be numerically tractable up to m of the order of several hundreds.

6. Discussion

6.1. Rejection curve calibration

We discuss some practical consequences of our work. For concreteness, we consider the critical value function

$$\rho_\beta(t) := \frac{t\alpha}{1+\beta-t(1-\alpha)}\,, \tag{6.1}$$

and the associated threshold collection $\boldsymbol{t}_\beta$; α corresponds to the target FDR and β is a tuning parameter. This has been proposed as an *ad hoc* adjustment of the asymptotically optimal rejection curve (4.7) by Finner, Dickhaus, and Roters (2009), further studied by Finner, Gontscharuk, and Dickhaus (2012). The AORC itself, obtained for $\beta = 0$, does not ensure strict FDR control at level α for any finite m. The question is how to choose β as small as possible while ensuring control of the FDR at level α.

In the case of a *step-up* procedure based on this function, it is known that the LFC is a Dirac-Uniform distribution (see Section 3). Therefore, to calibrate β, we can apply the principle delineated at (1.1): find $\beta(\alpha)$ so that

$$\sup_{m_0 \in \{0,\ldots,m\}} \mathrm{FDR}(\mathrm{SU}(\boldsymbol{t}_\beta), m_0, F \equiv 1) = \alpha.$$

The above equation can be solved by numerical search; for each value of β the left-hand side can be computed either by exact computation or by Monte-Carlo approximation. This approach has been advocated by Finner, Dickhaus, and Roters (2009); Finner, Gontscharuk, and Dickhaus (2012) and Gontscharuk (2010), using the exact computations for the FDR of SUD procedures under Dirac-Uniform distributions derived by Dickhaus (2008).

In the case of a more general *step-up-down* procedure, the exact LFC is not known; moreover, we have checked (using a numerical counterexample) that the negative result of Section 4.1 also holds for the critical value function ρ_β. Therefore, the approach delineated above cannot rigorously be applied for exact calibration of β. There is then interest in alternative approaches based on an *upper bound* on the FDR. Here there is the in-depth analysis of Finner, Gontscharuk, and Dickhaus (2012) for SUD procedures, and the elegant result obtained by Gavrilov, Benjamini, and Sarkar (2009), namely, that $\beta \equiv \beta_m = 1/m$ leads to an FDR upper bounded by α for the special case of the step-down procedure ($\lambda = 1$).

6.2. Limitations of FDR as a multiple testing criterion

While the FDR is widely used in practice, one can question how appropriate it is to base the multiple type I error criterion solely on controlling the *expectation* of the FDP. The results of Section 5 may be used to study numerically this issue by exact computation of the point mass function of the FDP under arbitrary configurations for the alternative. Based on this, we investigated the extent to which the distribution of the FDP concentrates around its expectation for a simple Gaussian location model with parameter μ. A graphical representation of this distribution in some specific cases is reported in Blanchard et al. (2013).

We have found that the distribution of the FDP is not concentrated around the corresponding FDR if the effect size μ is close to zero (weak signal) or if the proportion π_0 of true null hypotheses is close to 1 (sparse signal). Thus, even though joint independence of the p-values holds, controlling the FDR alone does not guarantee a small FDP in these cases. Furthermore, when μ and π_0 are fixed with m, such a spread of FDP distribution can arise when adding dependencies between test statistics, see Finner, Dickhaus, and Roters (2007); Delattre and Roquain (2011).

To alleviate such shortcomings, control of the false discovery exceedance (i.e., of the probability that the FDP exceeds a given threshold) has recently been proposed, see Farcomeni (2008) for a review. This brings the question of the corresponding LFC: are Dirac-uniform configurations least favorable for, e.g., $\mathbb{P}(\mathrm{FDP}(\mathrm{LSU}) > x)$? We have found numerically that this is not the case for any x. Hence, finding (possibly approximate) LFCs for the false discovery exceedance remains an open problem.

7. Proofs

7.1. Proof of Lemma 1

Let $\hat{\mathbb{G}}'_m = (\hat{\mathbb{G}}_m + m^{-1}) \wedge 1$. Note that for any $k \in \{1, \dots, m\}$, $p_{(k)} \le t_k$ is equivalent to $\hat{\mathbb{G}}_m(\rho(k/m)) \ge k/m$. We first analyze the case where $p_{(\lambda)} > t_\lambda$, that is, $\hat{\mathbb{G}}_m(\rho(\lambda/m)) < \lambda/m$, and SUD_λ behaves as a step-up, so that:

$$\begin{aligned}\frac{\widehat{k}}{m} &= \max\Big\{\frac{k}{m} \in \{0, \dots, \frac{\lambda}{m}\} \ : \ \hat{\mathbb{G}}_m\Big(\rho\Big(\frac{k}{m}\Big)\Big) \ge \frac{k}{m}\Big\} \\ &= \max\Big\{u \in \Big[0, \frac{\lambda}{m}\Big] \ : \ \hat{\mathbb{G}}_m(\rho(u)) \ge u\Big\} = \mathcal{U}\Big(\frac{\lambda}{m}, \hat{\mathbb{G}}_m\Big),\end{aligned}$$

where the second equality holds because $\mathcal{U}(\ell, \hat{\mathbb{G}}_m)$, being a fixed point of the function $\hat{\mathbb{G}}_m \circ \rho$, belongs to $\{0, 1/m, \dots, m/m\}$. This implies (4.3).

Assume now $p_{(\lambda)} \le t_\lambda$, that is, $\hat{\mathbb{G}}_m(\rho(\lambda/m)) \ge \lambda/m$, and SUD_λ behaves as a step-down. First, suppose $\widehat{k} < m$ holds. Then, on the one hand,

$$\begin{aligned}\frac{\widehat{k}+1}{m} &= \min\Big\{\frac{k}{m} \in \Big\{\frac{\lambda+1}{m}, \dots, \frac{m}{m}\Big\} \ : \ \hat{\mathbb{G}}_m\Big(\rho\Big(\frac{k}{m}\Big)\Big) < \frac{k}{m}\Big\} \\ &= \min\Big\{\frac{k}{m} \in \Big\{\frac{\lambda}{m}, \dots, \frac{m}{m}\Big\} \ : \ \hat{\mathbb{G}}_m\Big(\rho\Big(\frac{k}{m}\Big)\Big) < \frac{k}{m}\Big\} \\ &= \min\Big\{\frac{k}{m} \in \Big\{\frac{\lambda}{m}, \dots, \frac{m}{m}\Big\} \ : \ \hat{\mathbb{G}}'_m\Big(\rho\Big(\frac{k}{m}\Big)\Big) \le \frac{k}{m}\Big\}, \end{aligned} \tag{7.1}$$

because $m\hat{\mathbb{G}}_m(\rho(k/m))$ is an integer. On the other hand, since $\hat{\mathbb{G}}'_m(\rho(\lambda/m)) \ge \lambda/m$ and $m\mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$ is an integer, we have

$$\mathcal{U}\Big(\frac{\lambda}{m}, \hat{\mathbb{G}}'_m\Big) = \min\Big\{u \in [\frac{\lambda}{m}, 1] \ : \ \hat{\mathbb{G}}'_m\Big(\rho(u)\Big) \le u\Big\}$$

$$= \min\Big\{u \in \Big\{\frac{\lambda}{m}, \dots, \frac{m}{m}\Big\} \;:\; \hat{\mathbb{G}}'_m\Big(\rho(u)\Big) \le u\Big\}. \tag{7.2}$$

Combining (7.1) and (7.2), we conclude that $(\widehat{k}+1)/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$, and (4.3) holds. Finally, if $\widehat{k} = m$, then for any $k/m \in \{\lambda/m, \dots, m/m\}$ we have $\hat{\mathbb{G}}_m(\rho(k/m)) \ge k/m$. Hence, for all $k/m \in \{\lambda/m, \dots, (m-1)/m\}$, $\hat{\mathbb{G}}'_m(\rho(k/m)) > k/m$ which entails $\mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m) = 1$. Hence, $\widehat{k}/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$, also implying (4.3).

7.2. Proof of Theorem 2

Consider first the FM(m, m_0, F) model. We establish the following slightly more general inequality: for arbitrary $\delta, \zeta \in (0,1)$, with $\nu_0 = \max_{k \in \{m_0-1, m_0\}} \{|k/m - \zeta|\} \in [0,1]$, we have

$$\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F) \le \mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F \equiv 1) + \frac{m_0}{m\zeta}\varepsilon\,(\delta, m, \zeta, \nu_0)\ . \tag{7.3}$$

Inequality (4.5) is then obtained for $\zeta = m_0/m$ and $\nu_0 = 1/m$.

Preliminaries. With $\hat{\mathbb{G}}_m(x)$ the empirical p-value c.d.f., and $G^{DU}_\zeta(x) := (1-\zeta) + \zeta x$, put $\hat{u} := \hat{k}/m$, where $\hat{k}$ is defined by (2.1). Definition 3 implies that if G, G' are two nondecreasing functions such that $\forall x \in [0,1], G(x) \ge G'(x)$, then $\mathcal{U}(\lambda/m, G) \ge \mathcal{U}(\lambda/m, G')$. Based on the bound (4.3), we deduce that $\forall \delta \in (0,1)$,

$$\Big\{\sup_{x\in[0,1]} |\hat{\mathbb{G}}_m(x) - G^{DU}_\zeta(x)| \le \delta - \frac{1}{m}\Big\} \subset \big\{u^-_\delta \le \hat{u} \le u^+_\delta\big\}\,. \tag{7.4}$$

We now assume the DU(m,k) model. In this case, we have (a.s.) $p_{k'} = 0$ for $k' \ge k+1$, so that $\hat{\mathbb{G}}_m = (m-k)/m + k/m\hat{\mathbb{G}}_k$, implying in turn $G^{DU}_\zeta(x) - \hat{\mathbb{G}}_m(x) = (k/m - \zeta)(1 - \hat{\mathbb{G}}_k(x)) - \zeta(\hat{\mathbb{G}}_k(x) - x)$. For any $\nu > 0$ satisfying $|k/m - \zeta| \le \nu$, we deduce from (7.4) the following relation, which is valid up to a negligible set:

$$\Omega_\delta(k) := \Big\{\sup_{x\in[0,1]} |\hat{\mathbb{G}}_k(x) - x| \le \zeta^{-1}(\delta - \nu - \frac{1}{m})\Big\} \subset \big\{u^-_\delta \le \hat{u} \le u^+_\delta\big\}\,. \tag{7.5}$$

Under the DU(m,k) model, $\hat{\mathbb{G}}_k$ is the empirical c.d.f. of k i.i.d. uniform variables. Using the DKW inequality with Massart's (1990) optimal constant, we get

$$\begin{aligned}\mathbb{P}_{DU(m,k)}[(\Omega_\delta(k))^c] &\le 2\exp\Big\{-\frac{2k\,(\delta-\nu-1/m)_+^2}{\zeta^2}\Big\}\\ &\le 2\exp\Big\{-2m\Big(\delta - \nu - \frac{1}{m}\Big)_+^2(1-\frac{\nu}{\zeta})_+\Big\},\end{aligned} \tag{7.6}$$

because $k/m \ge \zeta - \nu$ and $\zeta \le 1$.

Upper bound. Let $q(x) = \rho(x)/x$ when $x \in (0,1]$ and $q(0) = \lim_{x\to 0^+} \rho(x)/x$ (this limit exists since q is nondecreasing). Applying Theorem 4.3 of Finner, Dickhaus, and Roters (2009), we obtain

$$\begin{aligned}
&\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F)\\
&\quad\leq \frac{m_0}{m}\mathbb{E}_{\mathrm{DU}(m,m_0-1)}[q(\hat{u})]\\
&\quad\leq \frac{m_0}{m}\frac{\rho(u_\delta^+)}{u_\delta^+} + \frac{m_0}{m}\mathbb{P}_{\mathrm{DU}(m,m_0-1)}[(\Omega_\delta(m_0-1))^c],\\
&\quad\leq \frac{m_0}{m}\frac{\rho(u_\delta^+)}{u_\delta^+} + \frac{m_0}{m}2\exp\Big\{-2m\Big(\delta-\nu-\frac{1}{m}\Big)_+^2(1-\frac{\nu}{\zeta})_+\Big\}, \qquad (7.7)
\end{aligned}$$

using (7.5)–(7.6) with $k := m_0 - 1$ and for any $\nu > 0$ satisfying $\nu \geq |[(m_0-1)/m] - \zeta|$.

Lower bound. In the model $\mathrm{DU}(m, m_0)$ with $m_0 < m$, we have $\hat{u} > 0$ a.s. and thus, using (7.5) (with $k := m_0$),

$$\begin{aligned}
\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F\equiv 1) &= \frac{m_0}{m}\mathbb{E}_{\mathrm{DU}(m,m_0)}\Big(\frac{\hat{\mathbb{G}}_{m_0}(\rho(\hat{u}))}{\hat{u}}\Big)\\
&\geq \frac{m_0}{m}\mathbb{E}_{\mathrm{DU}(m,m_0)}\Big(\frac{\hat{\mathbb{G}}_{m_0}(\rho(\hat{u}))}{\hat{u}}\mathbf{1}\{\Omega_\delta(m_0)\}\Big).
\end{aligned}$$

Now, the latter is larger than or equal to

$$\begin{aligned}
&\frac{m_0}{m}\mathbb{E}_{\mathrm{DU}(m,m_0)}\Big(\frac{\hat{\mathbb{G}}_{m_0}(\rho(u_\delta^-))}{u_\delta^+}\mathbf{1}\{\Omega_\delta(m_0)\}\Big)\\
&\qquad\geq \frac{m_0}{m}\frac{\rho(u_\delta^-)}{u_\delta^+} - \frac{m_0}{m}\frac{1}{1-\zeta}\mathbb{P}_{\mathrm{DU}(m,m_0)}\left[(\Omega_\delta(m_0))^c\right],
\end{aligned}$$

because $u_\delta^+ \geq 1-\zeta$. From (7.6), we obtain the lower bound

$$\begin{aligned}
&\mathrm{FDR}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0, F\equiv 1)\\
&\quad\geq \frac{m_0}{m}\frac{\rho(u_\delta^-)}{u_\delta^+} - \frac{m_0}{m}\frac{1}{1-\zeta}2\exp\Big\{-2m\Big(\delta-\nu-\frac{1}{m}\Big)_+^2(1-\frac{\nu}{\zeta})_+\Big\}, \qquad (7.8)
\end{aligned}$$

for any $\nu > 0$ satisfying $\nu \geq |m_0/m - \zeta|$. Finally, (7.7) and (7.8) entail (7.3).

Proof for random mixture model. In the $\mathrm{RM}(m, \pi_0, F)$ model, conditional on m_0, we recover the $\mathrm{FM}(m, m_0, F)$ model and (7.3) holds. Taking $\zeta = \pi_0$, the distribution of m_0 is binomial with parameters (m, ζ). In particular, ν_0 is random. However, we have for any $\gamma \in (0,1)$:

$$\mathbb{E}[\mathrm{FDP}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0)] \leq \mathbb{E}\left[\mathrm{FDP}(\mathrm{SUD}_\lambda(\boldsymbol{t}), m_0)\mathbf{1}\{\nu_0 \leq \gamma\}\right]$$

$$+\mathbb{P}\left(\left|\frac{m_0}{m}-\pi_0\right|>\gamma-\frac{1}{m}\right). \tag{7.9}$$

Using Hoeffding's (1963) inequality, we can write

$$\mathbb{P}\left(\left|\frac{m_0}{m}-\pi_0\right|>\gamma-\frac{1}{m}\right)\leq 2e^{-2m(\gamma-1/m)_+^2}. \tag{7.10}$$

We now combine (7.9) and (7.10) with (4.5), and take $\gamma := \sqrt{\log m/m} \geq \sqrt{\log m/2m}+1/m$ as soon as $m\geq 3$. This finishes the proof.

7.3. Proof of Corollary 1

Consider the FM(m, m_0, F) model with $m_0 < m$ (the case $m_0 = m$ is trivial). Consider $\delta_m \in (2/m, 1)$ satisfying for large m,

$$2\Big(1-\frac{1}{m\zeta_m}\Big)\Big(\delta_m-\frac{2}{m}\Big)^2=\frac{\log m}{m}, \tag{7.11}$$

so that $e^{-2m(\delta_m-2/m)_+^2(1-1/(m\zeta_m))_+} = 1/m$ for large m. Since $\zeta_m > \alpha > 0$ by assumption, we have $\delta_m \asymp \sqrt{(\log m)/m}$. From Theorem 2, it is sufficient to prove that

$$\frac{\rho(u^+_{\delta_m})-\rho(u^-_{\delta_m})}{u^+_{\delta_m}}=O\Big(\frac{\delta_m}{1-\zeta_m}\Big).$$

From the examples of Section 4.2, this holds for the linear critical value function. This also holds for the AORC as soon as $\lambda_m/m < v_{\delta_m}$ for large m and $1/(\zeta_m-\alpha)=O(1)$, which is the case by assumption (the formulas provided at the end of Section 4.2 are also valid when ζ depends on m, because the $O(\cdot)$ are uniform in ζ, for ζ bounded away from α). The proof in the RM(m,π_0,F) model is similar.

7.4. Proof of Proposition 1

We follow the proof of the "regular" Steck's recursion, see Shorack and Wellner (1986) p. 366–369. By using the convention $U_{(0)} = t_0 = 0$ and by considering the smallest j for which $U_{(j+1)} > t_{j+1}$, we can write

$$\begin{aligned}
&\mathbb{P}(U_{(k)}\leq t_k)-\mathbb{P}\left[U_{(1)}\leq t_1,\ldots,U_{(k)}\leq t_k\right]\\
&\qquad=\sum_{j=0}^{k-2}\mathbb{P}(\forall i\leq j, U_{(i)}\leq t_i, U_{(j+1)}>t_{j+1}, U_k\leq t_k)\\
&\qquad=\sum_{j=0}^{k-2}\sum_{X\subset\{1,\ldots,k\},|X|=j}\mathbb{P}(\forall i\leq j, U_{(i)}\leq t_i, \forall i\notin X, t_{j+1}\leq U_i\leq t_k).
\end{aligned}$$

Hence, if $U_{(i:X)}$ denotes the ith smallest member of the set $\{U_i, i \in X\}$, we obtain

$$\begin{aligned}
&(t_k)^{k_0} F(t_k)^{k-k_0} - \Psi_k(t_1, \ldots, t_k) \\
&\quad = \mathbb{P}(U_{(k)} \le t_k) - \mathbb{P}\left[U_{(1)} \le t_1, \ldots, U_{(k)} \le t_k\right] \\
&\quad = \sum_{j=0}^{k-2} \sum_{X \subset \{1,\ldots,k\}, |X|=j} \mathbb{P}(\forall i \le j, U_{(i:X)} \le t_i) \mathbb{P}(\forall i \notin X, t_{j+1} \le U_i \le t_k) \\
&\quad = \sum_{j=0}^{k-2} \sum_{j_0=0}^{j} \sum_{X \subset \{1,\ldots,k\}, |X|=j} \mathbf{1}\{|X \cap \{1, \ldots, k_0\}| = j_0\} \Psi_{j,j_0,F}(t_1, \ldots, t_j) \\
&\qquad \times \mathbb{P}(\forall i \notin X, t_{j+1} \le U_i \le t_k) \\
&\quad = \sum_{\substack{0 \le j_0 \le j \le k-2 \\ j_0 \le k_0 \\ j-j_0 \le k-k_0}} \binom{k_0}{j_0} \binom{k-k_0}{j-j_0} \Psi_{j,j_0,F}(t_1, \ldots, t_j) \\
&\qquad \times (t_k - t_{j+1})^{k_0 - j_0} (F(t_k) - F(t_{j+1}))^{k-k_0-j+j_0}.
\end{aligned}$$

Acknowledgements

This work was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01) and the French ministry of foreign and european affairs (EGIDE - PROCOPE project number 21887 NJ).

References

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

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

Blanchard, G., Dickhaus, T., Roquain, E. and Villers, F. (2013). On least favorable configurations for step-up-down tests. Supplement.

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

Dickhaus, T. (2008). *False Discovery Rate and Asymptotics*, PhD thesis, Heinrich-Heine-Universität Düsseldorf.

Dudoit, S. and van der Laan, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*, Springer Series in Statistics, Springer, New York.

Farcomeni, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion, *Statist. Methods Med. Res.* **17**, 347-388.

Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: asymptotics, *Ann. Statist.* **35**, 1432-1455.

Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve, *Ann. Statist.* **37**, 596-618.

Finner, H. and Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses, *J. Roy. Statist. Soc. Ser. B* **71**, 1031-1048.

Finner, H., Gontscharuk, V. and Dickhaus, T. (2012). False discovery rate control of step-up-down tests with special emphasis on the asymptotically optimal rejection curve, *Scandinavian Journal of Statistics* **39**, 382-397.

Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors, *Biom. J.* **43**, 985-1005.

Gavrilov, Y., Benjamini, Y. and Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence, *Ann. Statist.* **37**, 619-629.

Glueck, D. H., Karimpour-Fard, A., Mandel, J., Hunter, L. and Muller, K. E. (2008). Fast computation by block permanents of cumulative distribution functions of order statistics from several populations, *Commun. Statist. Theory Methods* **37**, 2815-2824.

Gontscharuk, V. (2010). *Asymptotic and Exact Results on FWER and FDR in Multiple Hypothesis Testing*, PhD thesis, Heinrich-Heine-Universität Düsseldorf.

Guo, W. and Rao, M. B. (2008). On control of the false discovery rate under no assumption of dependency, *J. Statist. Plann. Inference* **138**, 3176-3188.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, 13-30.

Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate, *Ann. Statist.* **33**, 1138-1154.

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

Miller, C. J., Genovese, C., Nichol, R. C., Wasserman, L., Connolly, A., Reichart, D., Hopkins, A., Schneider, J. and Moore, A. (2001). Controlling the false-discovery rate in astrophysical data analysis, *The Astronomical Journal* **122**, 3492-3505.

Neuvial, P. and Roquain, E. (2012). On false discovery rate thresholding for classification under sparsity, *Ann. Statist.* **40**, 2572-2600.

Pantazis, D., Nichols, T. E., Baillet, S. and Leahy, R. M. (2005). A comparison of random field theory and permutation methods for statistical analysis of MEG data, *NeuroImage* **25**, 383-394.

Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing, *Ann. Statist.* **35**, 1378-1408.

Roquain, E. and Villers, F. (2011a). Exact calculations for false discovery proportion with application to least favorable configurations, *Ann. Statist.* **39**, 584-612.

Roquain, E. and Villers, F. (2011b). Supplement to "Exact calculations for false discovery proportion with application to least favorable configurations", Ann. Statist. Available online.

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

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance, *Biometrika* **73**, 751-754.

Somerville, P. N. and Hemmelmann, C. (2008). Step-up and step-down procedures controlling the number and proportion of false positives, *Comput. Statist. Data Anal.* **52**, 1323-1334.

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

Zeisel, A., Zuk, O. and Domany, E. (2011). FDR control with adaptive procedures and FDR monotonicity, *Ann. Appl. Statist.* **5**, 943-968.

Universität Potsdam, Institut für Mathematik, Am neuen Palais 10, 14469 Potsdam, Germany.

E-mail: gilles.blanchard@math.uni-potsdam.de

Humboldt-University Berlin, Department of Mathematics, Unter den Linden 6, D-10099 Berlin, Germany.

E-mail: dickhaus@math.hu-berlin.de

Université de Paris 6 (UPMC), Case courrier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France.

E-mail: etienne.roquain@upmc.fr

Université de Paris 6 (UPMC), Case courrier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France.

E-mail: fanny.villers@upmc.fr

(Received August 2011; accepted October 2012)