



HAL
open science

On k-FWE-based critical values for controlling the false discovery proportion under dependence

Sylvain Delattre, Etienne Roquain

► **To cite this version:**

Sylvain Delattre, Etienne Roquain. On k-FWE-based critical values for controlling the false discovery proportion under dependence. 2013. hal-00905060v1

HAL Id: hal-00905060

<https://hal.science/hal-00905060v1>

Preprint submitted on 15 Nov 2013 (v1), last revised 3 Jun 2015 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On k -FWE-based critical values for controlling the false discovery proportion under dependence

Sylvain Delattre

Université Paris Diderot, LPMA,
e-mail: sylvain.delattre@univ-paris-diderot.fr

and

Etienne Roquain

UPMC Université Paris 6, LPMA,
e-mail: etienne.roquain@upmc.fr

Abstract: The false discovery proportion (FDP) is a convenient way to account for false positives in an high dimensional setting where a large number of tests are performed simultaneously. The Benjamini-Hochberg procedure is now widely used and is known to control the expectation of the FDP, called the false discovery rate (FDR). However, when the individual tests are correlated, controlling the FDR can be unsuitable to ensure that the actually achieved FDP is close (or below) the targeted level. This rises the question of controlling the quantiles of the distribution of the FDP, which is a challenging question that has received a growing attention in the recent literature. This paper elaborates upon the general principle let down by Romano and Wolf (2007) in [Romano and Wolf \(2007\)](#) (RW) that builds FDP controlling procedures from k -family-wise error rate (k -FWE) controlling procedures, while incorporating known dependencies in an appropriate manner. This method is revisited as follows : first, choose a device to upper-bound the k -FWE, for all k . Second, build the corresponding critical values, possibly adaptively to the number m_0 of true null hypotheses. Third, use these critical values into a step-wise procedure (either step-down or step-up). The goal of the paper is to study the obtained FDP when using this methodology. Our first result provides sample finite bounds, while our second result is asymptotic in the number m of hypotheses. Overall, this paper can be seen as a validation of RW's paradigm for controlling the FDP under dependence.

AMS 2000 subject classifications: Primary 62H15; secondary 60F17.

Keywords and phrases: multiple testing, false discovery rate, pairwise correlation, Gaussian multivariate distribution, functional central limit theorem.

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 have correlations. Pertaining to this class of high-dimensional problems, a common issue is to make simultaneously a huge number of 0/1 decisions, possibly correlated, and with a valid control on the overall amount of false discoveries (items declared to be 1 wrongly). A convenient way to account for false discoveries in high dimension 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 so-called ‘‘Benjamini and Hochberg procedure’’ (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 dependencies, see, e.g., [Delattre and Roquain \(2011\)](#); [Guo et al. \(2013\)](#); [Korn et al. \(2004\)](#), which makes the use of the BH procedure questionable.

To illustrate deeper this phenomenon, Figure 1 displays the distribution of the FDP of the (oracle) BH procedure in the classical one-sided Gaussian multiple testing framework, when the test statistics are all ρ -equicorrelated. As ρ increases, the distribution of the FDP get less concentrated and turns out to be drastically skewed for $\rho = 0.1$ (in particular it falls outside the Gaussian regime). Clearly, in that case, the mean fails to describe accurately the overall behavior of the FDP distribution. In particular, although the mean of the FDP is below $\alpha = 0.2$, the true value of FDP is not ensured to be small in that 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 $\mathbb{P}(\text{FDP} > \alpha) \leq \zeta$. While taking $\zeta = 1/2$ 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: due to the increasingly heavy upper-tail, while the 95%-quantile gets substantially larger, the median gets slightly smaller. This suggests that the (oracle) 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 controlling the $(1 - \zeta)$ -quantile of the FDP is essential in the presence of dependence.

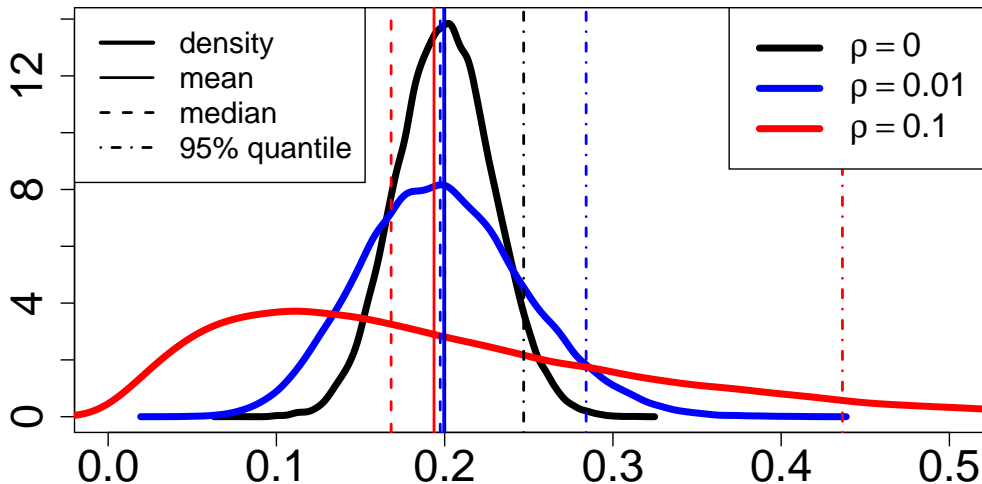


FIG 1. *Density of the false discovery proportion of the (oracle) BH procedure when increasing the dependence. $m = 1000$, $m_0 = 800$, 5×10^4 simulations, Gaussian one-sided equicorrelated model.*

1.2. Multiple testing framework

We observe a random variable X whose distribution belongs to some set \mathcal{P} . For $m \geq 2$, 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 $\mathcal{P}_H \subset \mathcal{P}$ (assumed to be non-empty) 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

$$\forall H \in \{0, 1\}^m \text{ with } H_i = 0, \text{ we have } p_i(X) \sim U(0, 1) \text{ when } X \sim P \text{ with } P \in \mathcal{P}_H.$$

In this paper, a leading example is the *Gaussian multivariate framework*, for which $X = H\mu + Y$, where $H\mu = (H_i\mu_i)_{1 \leq i \leq m}$ for $H \in \{0, 1\}^m$, $\mu \in (\mathbb{R}_+ \setminus \{0\})^m$ and Y is a m -dimensional centered Gaussian vector with covariance matrix Γ such that $\Gamma_{i,i} = 1$. We focus on the one-sided multiple testing problem of “ $\mathbb{E}X_i = 0$ ” versus “ $\mathbb{E}X_i > 0$ ”, which is equivalent to test “ $H_i = 0$ ” against “ $H_i = 1$ ”. The p -values are given by $p_i(X) = \Phi(X_i)$. Importantly, in this paper, the matrix Γ is sometimes assumed to be known. This is in accordance with practical situations where we test whether an outcome is associated to multiple covariates (e.g., genome-wide association studies), see Section 2 in [Fan et al. \(2012\)](#). A special case of interest is the equi-correlated case:

$$\Gamma_{i,j} = \rho, \text{ for all } i \neq j, \text{ where } \rho \in [-(m-1)^{-1}, 1], \quad (\rho\text{-equi})$$

which is extensively used throughout the paper.

When (ρ -equi) does not hold, the joint null 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 that 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 propose to add to our model random effects, that make H random.

More formally, we distinguish between the three 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

$$\Theta^F = \{(H, P) : H \in \{0, 1\}^m, P \in \mathcal{P}_H\}.$$

- 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 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 $\theta = (m_0, (P_H)_H)$, to be chosen in the set

$$\Theta^U = \{(m_0, (P_H)_H) : m_0 \in \{0, 1, \dots, m\}, P_H \in \mathcal{P}_H \text{ for all } H\}.$$

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

- Bernoulli mixture model: contrary to the “Uniform mixture model”, the distribution of H is $\mathcal{B}(1 - \pi_0)^{\otimes m}$, for some parameter $\pi_0 \in [0, 1]$. Note that it can be obtained from an

uniform model by taking in addition $m_0 \sim \mathcal{B}(m, \pi_0)$. The parameters of the model are given by $\theta = (\pi_0, (P_H)_H)$ to be chosen in

$$\Theta^B = \{(\pi_0, (P_H)_H) : \pi_0 \in [0, 1], P_H \in \mathcal{P}_H \text{ for all } H\}.$$

Also, we denote $\pi_1 = 1 - \pi_0$.

While the fixed mixture model is the most commonly used model for multiple testing, the Bernoulli mixture model has become somewhat standard since its introduction by [Efron et al. \(2001\)](#), see, e.g., [Genovese and Wasserman \(2004\)](#); [Roquain and Villers \(2011\)](#); [Storey \(2003\)](#), because it is mathematically more convenient while it incorporates a meaningful *a priori* on the trueness/falseness of the null hypotheses. The uniform mixture model, which is new to our knowledge, removes part of the random effects of the Bernoulli mixture to keep m_0 deterministic. While it still derives some benefit from the random mixture, it is additionally convenient for the adaptation to m_0 , as we will see later on. With some abuse, we denote $m_0(\theta)$ (or m_0 when not ambiguous) the number of true null in the fixed/uniform mixture models.

1.3. Type I error rates

First, for $t \in [0, 1]$, denote $V_m(t) = \sum_{i=1}^m (1 - H_i) \mathbf{1}\{p_i(X) \leq t\}$ and $R_m(t) = \sum_{i=1}^m \mathbf{1}\{p_i(X) \leq t\}$ 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 is defined as the probability that more than k true nulls have a p -value smaller than \hat{t}_m , see, e.g., [Hommel and Hoffman \(1988\)](#); [Lehmann and Romano \(2005\)](#). Formally, for $\theta \in \Theta$ (in one of the models defined in Section 1.2 and Θ being the corresponding parameter space)

$$k\text{-FWER}(\hat{t}_m) = \mathbb{P}_\theta(V_m(\hat{t}_m) \geq k). \quad (1)$$

Note that when $k = 1$, the latter corresponds to the traditional FWER. From (1), providing $k\text{-FWER}(\hat{t}_m) \leq \zeta$ (for all $\theta \in \Theta$ and some $\zeta \in (0, 1)$), ensures that, with probability at least $1 - \zeta$, less than $k - 1$ false discoveries are made by the thresholding procedure \hat{t}_m . However, the meaning of k in this criterion may be seen as troublesome: while ensuring at most $k - 1 = 5$ false discoveries seems fair for more than 100 discoveries (say), it is certainly unacceptable when less than 10 discoveries (say) are made. Hence, it seems appropriate to relate k to the overall number of discoveries of the procedure. To this end, for some threshold $t \in [0, 1]$, define the false discovery proportion at threshold t as follows:

$$\text{FDP}_m(t) = \frac{V_m(t)}{R_m(t) \vee 1}. \quad (2)$$

The quantity $\text{FDP}_m(t)$ is not observable because it depends on the unknown process $V(t)$. Furthermore, $\text{FDP}_m(t)$ is random, so that $\text{FDP}_m(t) \leq \alpha$ can hold only with high probability. Controlling the FDP via a threshold $t = \hat{t}_{m, \alpha, \zeta}$ (potentially depending on the data) thus corresponds to the following probabilistic bound:

$$\forall \theta \in \Theta, \mathbb{P}_\theta(\text{FDP}_m(\hat{t}_{m, \alpha, \zeta}) \leq \alpha) \geq 1 - \zeta, \quad (3)$$

for some pre-specified values $\alpha, \zeta \in (0, 1)$. As mentioned before, (3) corresponds to upper-bound the $(1 - \zeta)$ -quantile of the distribution of $\text{FDP}_m(\widehat{t}_{m,\alpha,\zeta})$ by α . Since $\text{FDP}_m(t) > \alpha$ is equivalent to $V_m(t) \geq \lfloor \alpha R_m(t) \rfloor + 1$, note that FDP control and k -FWER control are intrinsically linked.

From an historical point of view, the introduction of the FDP goes back to Eklund in the 1960's (as reported in Seeger (1968)), that has presented the FDP as a solution to the ‘‘mass-significance problem’’. Much later, the seminal paper of Benjamini and Hochberg 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.

1.4. Step-up and step-down procedures

Consider a nondecreasing sequence $(\tau_\ell)_{1 \leq \ell \leq m}$ of nonnegative values (critical values), with the convention $\tau_0 = 0$. The corresponding step-up (resp. step-down) procedure is defined as rejecting the p -values smaller than $\tau_{\widehat{\ell}}$, where $\widehat{\ell}$ is defined by either of the two following quantities:

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

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

Let us also recall the so-called ‘‘switching relation’’, establishing that $p_{(\ell)} \leq \tau_\ell$ is equivalent to $R(\tau_\ell) \geq \ell$. This entails that $R(\tau_{\widehat{\ell}}) = \widehat{\ell}$ both in the step-up and step-down cases.

1.5. Proposed approach and relation to previous literature

The problem of finding multiple testing procedures ensuring (3) has received a growing attention in the last decades, see for instance Chi and Tan (2008); Dudoit and van der Laan (2008); Guo et al. (2013); Lehmann and Romano (2005); Romano and Shaikh (2006a,b); Romano and Wolf (2007); Roquain (2011); Roquain and Villers (2011). To investigate this challenging issue, existing work either proposed FDP controlling procedures that are too conservative (see Guo et al. (2013); Lehmann and Romano (2005); Romano and Shaikh (2006a,b)), either established the FDP control under very restrictive assumptions (e.g., independence in Guo and Romano (2007), alternative p -values all equal to 0 in Romano and Wolf (2007)). The present paper is an attempt to fill the gap between these two extreme situations by providing theoretical supports for the general method exposed in RW Romano and Wolf (2007). However, strictly speaking, contrary to RW who used resampling based technics, our approach is guided by the case where the dependencies are known and Gaussian multivariate. This setup, already considered in Fan et al. (2012); Troendle (2000) for instance, allows exact (or alternatively Monte-Carlo) calculations and we find it appropriate to study accurately the property of a method capturing the dependencies like RW's method.

RW's method, itself generalizing the approach of Korn et al. (2004), is to build an FDP controlling procedure from k -FWER controlling procedures at level γ (rejecting, say, R_k hypotheses), by a step-down algorithm starting from $k = 1$ and stopping the first time that $R_k < k/\alpha - 1$. In Section 2, we reformulate this method by defining k -FWE-based critical values, as coming from a ‘‘bounding device’’, that can be chosen in a way that adapts to m_0 and/or to the dependence structure. This gives rise to several critical values families,

including the well-known Lehmann-Romano critical values [Lehmann and Romano \(2005\)](#) and those found in [Guo and Romano \(2007\)](#). The derived critical values are then used into a step-down or step-up procedure. Sections 3 and 4 present our major contributions w.r.t. RW's work: finite sample bounds and asymptotic control for the FDP of the derived procedures. While our finite sample bounds are shown to improve the state-of-the-art bounds [Guo et al. \(2013\)](#); [Romano and Shaikh \(2006a,b\)](#), the asymptotic results hold when the *number m of hypotheses tends to infinity* and rely on a type of weak dependence assumption between the individual p -values (functional central limit theorem). This is markedly different from the setting of RW, because their control is established when the number of observations (sample size) tends to infinity, while the number m of hypotheses being fixed. The latter corresponds to the case where all the alternative p -values are (asymptotically) equal to 0. Since this so-called ‘‘Dirac Uniform’’ configuration (see [Finner et al. \(2007\)](#)) is not proved to be least favorable, this is a seriously stringent assumption.

Next, let us underline that previous work studying the asymptotic FDP distribution all rely on the use of the functional delta method [Delattre and Roquain \(2011\)](#); [Delattre and Roquain \(2012\)](#); [Farcomeni \(2006\)](#); [Genovese and Wasserman \(2004\)](#); [Neuvial \(2008\)](#). This tool, that relies on the tedious and complex calculations of derivatives [Neuvial \(2008\)](#), is advantageously not required here.

2. Building k -FWE-based critical values

2.1. Heuristic

Starting from arbitrary critical values $(\tau_\ell)_{1 \leq \ell \leq m}$, and by taking an indice $\hat{\ell}$ such that $R(\tau_{\hat{\ell}}) = \hat{\ell}$, we have

$$\begin{aligned} \mathbb{P}_\theta(\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha) &= \mathbb{P}_\theta(V_m(\tau_{\hat{\ell}}) > \alpha R(\tau_{\hat{\ell}})) \\ &= \mathbb{P}_\theta(V_m(\tau_{\hat{\ell}}) \geq \lfloor \alpha \hat{\ell} \rfloor + 1). \end{aligned} \quad (4)$$

Hence, by taking τ_ℓ such that $\mathbb{P}_\theta(V_m(\tau_\ell) \geq \lfloor \alpha \ell \rfloor + 1) = (\lfloor \alpha \ell \rfloor + 1)\text{-FWER}(\tau_\ell) \leq \zeta$ for all ℓ , we should get that (4) is below ζ . However, the above reasoning does not rigorously establish (3) (with $\hat{t}_{m,\alpha,\zeta} = \tau_{\hat{\ell}}$), because it implicitly assumed that $\hat{\ell}$ is determinist. Nevertheless, this heuristic is the starting point for building the critical values.

2.2. Bounding device

Let us consider as model 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 $u \in \{k, \dots, m\}$, which is non-increasing in k , with $B_m^0(0, k, u) = 0$ and such that

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

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

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

$$\tilde{B}_m(t, k, \ell) = \sup_{k \leq u \leq m - \ell + k} \{B_m^0(t, k, u)\}; \quad (\text{Bound-adapt})$$

which are additionally assumed to be non-decreasing and left-continuous in t . While $\bar{B}_m(t, k)$ is obviously non-increasing in k because $B_m^0(t, k, u)$ is, the fact that $\tilde{B}_m(t, k, \ell)$ is non-increasing in k is an additional assumption that we make throughout the paper.

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 non adaptive (resp. adaptive, oracle) k -FWE-based critical values associated to the bounding function B_m^0 are respectively defined as follows:*

$$\bar{\tau}_\ell = \max \{t \in [0, 1] : \bar{B}_m(t, \lfloor \alpha \ell \rfloor + 1) \leq \zeta\}, \quad 1 \leq \ell \leq m; \quad (5)$$

$$\tilde{\tau}_\ell = \max \{t \in [0, 1] : \tilde{B}_m(t, \lfloor \alpha \ell \rfloor + 1, \ell) \leq \zeta\}, \quad 1 \leq \ell \leq m; \quad (6)$$

$$\tau_\ell^0 = \max \{t \in [0, 1] : B_m^0(t, \lfloor \alpha \ell \rfloor + 1, m_0) \leq \zeta\}, \quad 1 \leq \ell \leq m. \quad (7)$$

The critical values $\tilde{\tau}_\ell$, $\ell = 1, \dots, m$, are said *adaptive* because they implicitly (over-) estimate m_0 by $m - \ell + \lfloor \alpha \ell \rfloor + 1$. This way to adapt to π_0 , that produces procedures often referred to as “one-stage” (by contrast with “two-stage”, see [Benjamini et al. \(2006\)](#); [Blanchard and Roquain \(2009\)](#); [Sarkar \(2008a\)](#)), has been used in [Finner et al. \(2009\)](#); [Gavrilov et al. \(2009\)](#) and has been proved to be asymptotically optimal in a specific sense (AORC), see [Finner et al. \(2009\)](#). Finally, it is worth to check that $\bar{\tau}_m \leq \tilde{\tau}_m < 1$ (this comes from $B_m^0(1, k, u) = 1$ for all $u \geq k$) 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. Some resulting critical values are displayed in [Figure 2](#) and [3](#).

Markov By Markov’s inequality, we have

$$\mathbb{P}_\theta(V_m(t) \geq k) \leq \frac{\mathbb{E}_\theta(V_m(t))}{k} = \frac{m_0 t}{k} =: B_m^0(t, k, m_0).$$

This gives rise to the critical values

$$\bar{\tau}_\ell = \frac{\zeta(\lfloor \alpha \ell \rfloor + 1)}{m}; \quad \tilde{\tau}_\ell = \frac{\zeta(\lfloor \alpha \ell \rfloor + 1)}{m - \ell + \lfloor \alpha \ell \rfloor + 1}, \quad 1 \leq \ell \leq m. \quad (8)$$

The adaptive critical values $(\tilde{\tau}_\ell)_{1 \leq \ell \leq m}$ are those proposed by Lehmann and Romano in [Lehmann and Romano \(2005\)](#). Note that these critical values do not adapt to the underlying dependence structure of the p -values.

Pairwise By using Markov’s inequality “at the second order”, we obtain

$$\mathbb{P}_\theta(V_m(t) \geq k) \leq \mathbb{P}_\theta((V_m(t))^2 \geq k^2) \leq \frac{m_0 t + m_0(m_0 - 1)F_\theta^{(2)}(t)}{k^2},$$

where $F_\theta^{(2)}$ denotes the function:

$$F_\theta^{(2)}(t) = \frac{1}{m_0(m_0 - 1)} \mathbb{E}_\theta \left(\sum_{i \neq j} (1 - H_i)(1 - H_j) \mathbf{1}\{p_i(X) \leq t, p_j(X) \leq t\} \right). \quad (9)$$

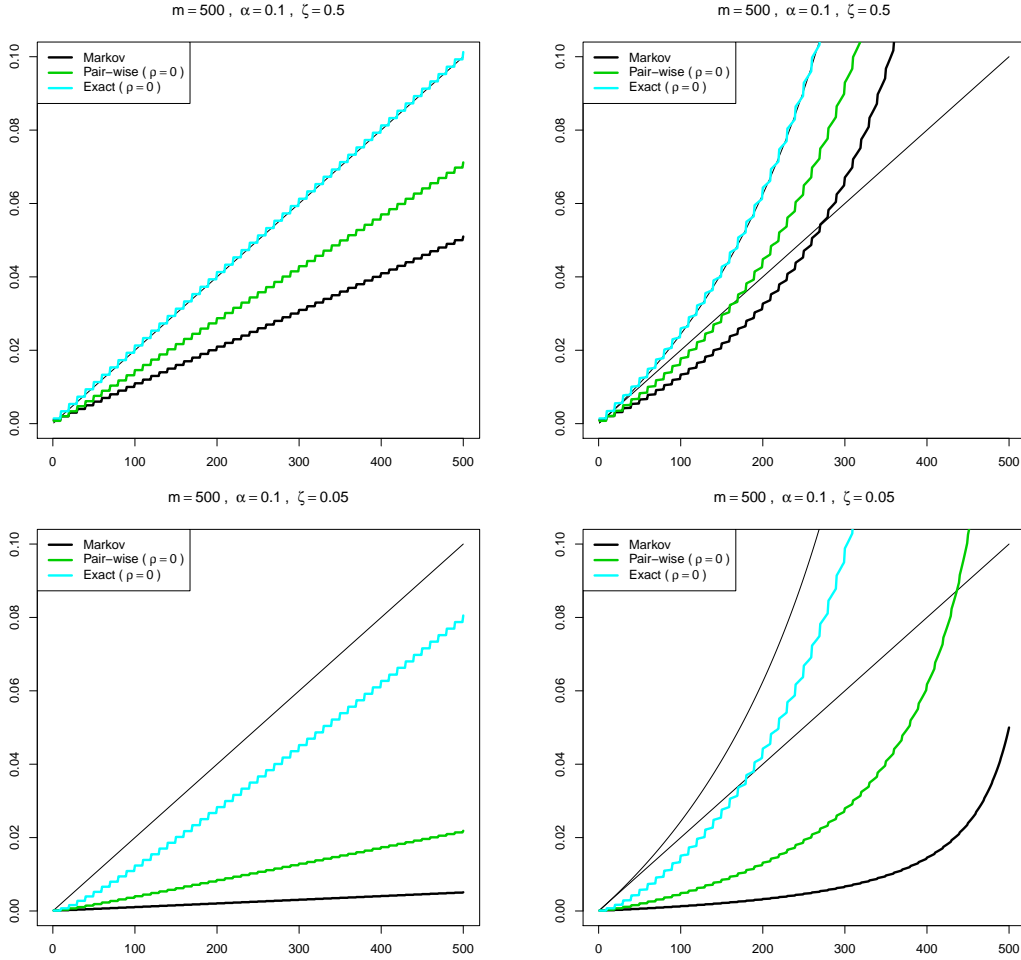


FIG 2. k -FWE-based critical values. Left: $\bar{\tau}_k$'s (non adaptive); Right: $\tilde{\tau}_k$'s (adaptive). Exact calculations of critical values under Gaussian independence ($\rho = 0$). For comparison, the solid thin black line denotes either the BH critical values $t_k = \alpha k/m$ (non adaptive) or the AORC critical values $t_k = \alpha k/(m - (1 - \alpha)k)$ defined in [Finner et al. \(2009\)](#) (adaptive).

Hence, we can take

$$B_m^0(t, k, u) = k^{-2} \left(ut + u(u - 1) \sup \left\{ F_\theta^{(2)}(t), \theta \in \Theta, m_0(\theta) = u \right\} \right). \quad (10)$$

As a first instance, assuming

$$p_i, i \in \mathcal{H}_0, \text{ are pairwise independent (cond. on } H \text{ in model } \Theta^U), \quad (\text{Pw-indep})$$

we get $F_\theta^{(2)}(t) = t^2$ and thus $B_m^0(t, k, u) = (ut + u(u - 1)t^2)/k^2$, which leads for instance to

$$\tilde{\tau}_\ell = \left(\frac{\zeta([\alpha\ell] + 1)^2}{(m - \ell + [\alpha\ell])(m - \ell + [\alpha\ell] + 1)} + \frac{1}{4(m - \ell + [\alpha\ell])^2} \right)^{1/2} - \frac{1}{2(m - \ell + [\alpha\ell])}.$$

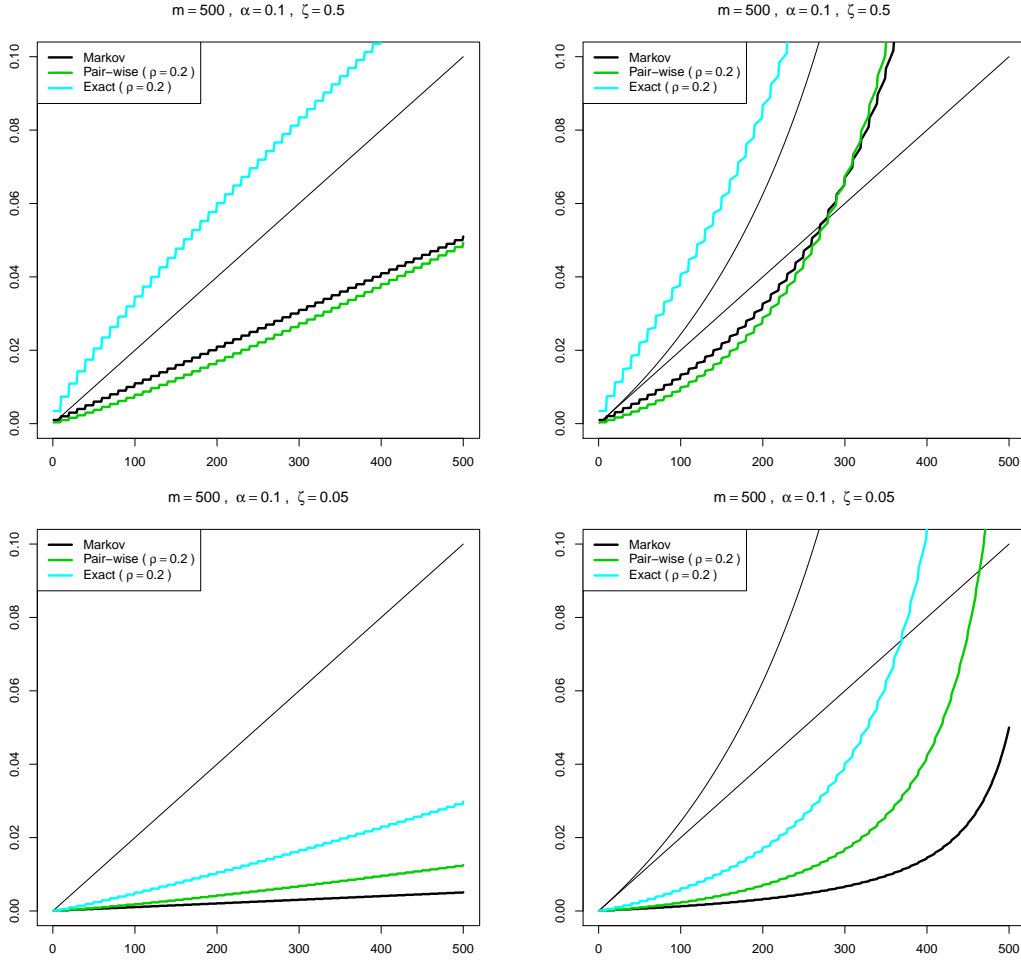


FIG 3. Same as Figure 3 under Assumption (ρ -equi) with $\rho = 0.2$.

As a second illustration, in the case (ρ -equi) with $\rho \geq 0$, we easily get

$$F_{\theta}^{(2)}(t) = \int_0^t \Phi \left(\frac{\Phi^{-1}(t) - \rho \Phi^{-1}(x)}{\sqrt{1 - \rho^2}} \right) dx, \quad (11)$$

which can lead to critical values (adaptive or not) by a numerical inversion. Third, in the Gaussian case with known covariance matrix Γ (but not necessarily equi-correlated), we easily get that $\bar{B}_m(t, k) = B_m^0(t, k, m) = (mt + m(m-1)F_{\Gamma}^{(2)}(t)) / k^2$ where

$$F_{\Gamma}^{(2)}(t) = \frac{1}{m(m-1)} \sum_{1 \leq i \neq j \leq m} \int_0^t \Phi \left(\frac{\Phi^{-1}(t) - \Gamma_{i,j} \Phi^{-1}(x)}{\sqrt{1 - \Gamma_{i,j}^2}} \right) dx. \quad (12)$$

The adaptive case is more delicate because the supremum in (10) lies over the set of all the matrices of the form $(\Gamma_{i,j})_{i,j \in A}$, for $A \subset \{1, \dots, m\}$, $|A| = u$, which can be a huge set when

$u < m$. For this reason, it can be more convenient to consider the *uniform* mixture model (rather than the fixed mixture model) for which $B_m^0(t, k, u) = \left(ut + u(u-1)F_\Gamma^{(2)}(t) \right) / k^2$, where $F_\Gamma^{(2)}$ is given by (12). This leads to adaptive critical values by a numerical inversion. All these critical values take into account the pairwise dependence structure of the data.

Exact In some cases, closed-formulas can be derived for the RHS of (Bound). First, assuming

$$p_i, i \in \mathcal{H}_0, \text{ are mutually independent (cond. on } H \text{ in model } \Theta^U), \quad (\text{Indep})$$

the distribution of $V(t)$ 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 have been introduced in Guo and Romano (2007). Second, the following exact formula can be used under (ρ -equi) when $\rho \geq 0$:

$$B_m^0(t, k, u) = \int_0^1 \sum_{j=k}^u \binom{u}{j} (F_\rho(t, x))^j (1 - F_\rho(t, x))^{u-j} dx, \quad (13)$$

where $F_\rho(t, x) = \Phi \left((\Phi^{-1}(t) - \sqrt{\rho} \Phi^{-1}(x)) / (\sqrt{1-\rho}) \right)$. This comes from the well known decomposition $X_i \sim H_i \mu_i + \sqrt{\rho} U + \sqrt{1-\rho} \xi_i$, where U and ξ_i are all i.i.d. $\mathcal{N}(0, 1)$. Third, in the Gaussian case where Γ is known, the quantities in the RHS of (Bound) can be computed by a Monte-Carlo method as follows: on the one hand, in the non adaptive case, we can upper-bound $V_m(t)$ by the full null process

$$V'_m(t) = m^{-1} \sum_{i=1}^m \mathbf{1} \{ \Phi(Y_i) \leq t \} \quad (14)$$

whose distribution is known since $Y \sim \mathcal{N}(0, \Gamma)$ is. On the other hand, in the adaptive case, the original null process $V_m(t)$ can be easily generated in the uniform model Θ^U in the case $m_0(\theta) = u$, for an arbitrary u . This leads to non-adaptive and adaptive critical values that incorporate any pre-specified covariance matrix Γ .

3. Finite sample results

In this section, we provide upper bounds for the probability that the FDP of a step-down/step-up procedure exceeds α . First, this section gathers part of the most prominent state-of-the-art bounds in an unified way, second we provide improvements when the covariance dependence is known. These results hold for any critical values, but are particularly interesting when used with k -FWE-based critical values.

3.1. A first unifying bound

Proposition 3.1. *For any critical values $(\tau_\ell)_{1 \leq \ell \leq m}$, consider either the corresponding step-down procedure (SD) or the corresponding step-up procedure (SU), with rejection number $\tilde{\ell}$. Then the following holds, both in the fixed model ($\Theta = \Theta^F$) and the uniform model ($\Theta = \Theta^U$): for all $\theta \in \Theta$,*

$$\mathbb{P}_\theta \left(\text{FDP}_m(\tau_{\tilde{\ell}}) > \alpha \right) \leq \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{P}_\theta \left(V_m(\tau_\ell) \geq d(\ell, m, m_0), \tilde{\ell} = \ell \right), \quad (15)$$

where $b_\alpha(m_0)$, $d(\ell, m, m_0)$ and $\tilde{\ell}$ are taken as follows:

- (i) *Step-down case:* for all $u \in \{0, \dots, m\}$, $b_\alpha(u) = (\lfloor (m-u)/(1-\alpha) \rfloor + 1) \wedge (\lceil u/\alpha \rceil - 1) \wedge m$;
 $d(\ell, m, u) = \lfloor \alpha \ell \rfloor + 1$; $\tilde{\ell} = \hat{\ell}^{(1)}$, where $\hat{\ell}^{(1)} = \min\{\ell \in \{1, \dots, m\} : S_m(\tau_\ell) < (1-\alpha)\ell\}$
 (with the convention $\min \emptyset = m+1$) and by denoting $S_m(t) = R_m(t) - V_m(t)$ the number
 of true discoveries at threshold t .
- (ii) *Step-up case:* for all $u \in \{0, \dots, m\}$, $b_\alpha(u) = (\lceil u/\alpha \rceil - 1) \wedge m$; $d(\ell, m, u) = (\lfloor \alpha \ell \rfloor + 1) \vee$
 $(\ell - m + u)$; $\tilde{\ell} = \hat{\ell}$.

Moreover, in the step-up case, (15) is an equality.

Proposition 3.1 (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 3.1 (ii) is essentially based on Romano and Shaikh (2006b) and we provide a short proof for it in Section 6.1.

Now, let us apply Proposition 3.1 in the step-down case, for critical values $(\tau_\ell)_{1 \leq \ell \leq m}$ that are k -FWE-based (associated to some bounding device B_m^0 and either adaptive or not). From (15), the FDP control (3) holds for $\hat{t}_{m,\alpha,\zeta} = \tau_{\hat{\ell}}$ when the RHS of (15) is below ζ . This holds in several particular cases:

- (a) Dirac-Uniform configuration (established in Romano and Wolf (2007)): when all the p -values under the alternative are equal to zero. In that case, $\hat{\ell}^{(1)}$ is deterministic and is equal to $\ell^* = (\lfloor (m - m_0)/(1 - \alpha) \rfloor + 1) \wedge m$. Hence, the RHS of (15) is $\mathbb{P}_\theta(V_m(\tau_{\ell^*}) \geq \lfloor \alpha \ell^* \rfloor + 1) \leq B_m^0(\tau_{\ell^*}, \lfloor \alpha \ell^* \rfloor + 1, m_0)$, which is below ζ (because $m - \ell^* + \lfloor \alpha \ell^* \rfloor + 1 \geq m_0$).
- (b) Independence (established in Guo and Romano (2007) with the exact device): when for all θ , the p -values $(p_i, i : H_i = 0)$ are independent of the p -values $(p_i, i : H_i = 1)$. This implies that $\hat{\ell}^{(1)}$ is independent of the process $V_m(\cdot)$ and thus that the probability $\mathbb{P}_\theta(V_m(\tau_\ell) \geq \lfloor \alpha \ell \rfloor + 1 \mid \hat{\ell}^{(1)} = \ell)$ is below ζ for all ℓ .
- (c) when for all $\theta \in \Theta$, $\lfloor \alpha b_\alpha(m_0(\theta)) \rfloor = 0$ (e.g., $m_0(\theta) \in \{1, m\}$ or $\lfloor \alpha m \rfloor = 0$): the RHS of (15) is bounded by $\mathbb{P}_\theta(V_m(\tau_{b_\alpha(m_0)}) \geq \lfloor \alpha b_\alpha(m_0) \rfloor + 1)$ and thus is below ζ .
- (d) under (ρ -equi) when $\rho = 1$: since in that case $V_m(t) = m_0 \mathbf{1}\{p_1 \leq t\}$ we have $\tau_\ell \leq \zeta$ for all $\ell \leq b_\alpha(m_0)$. Next, $\mathbf{1}\{V_m(\tau_\ell) \geq \lfloor \alpha \ell \rfloor + 1\} \leq \mathbf{1}\{p_1 \leq \tau_\ell\} \leq \mathbf{1}\{p_1 \leq \tau_{b_\alpha(m_0)}\}$ for all $\ell \leq b_\alpha(m_0)$, which implies the control.

Additionally to the above results, numerical experiments indicate that the FDP control should hold under (ρ -equi) for many values of ρ and μ , see Section 5. However, we were not able to provide a valid proof of the latter statement.

Now, in the step-up case, Proposition 3.1 (ii) implies the following bound when applied with k -FWE-based critical values:

Corollary 3.2. *Consider some bounding device B_m^0 and the associated k -FWE-based critical values $(\tau_\ell)_{1 \leq \ell \leq m}$, being either adaptive or not and computed either in the fixed mixture model or in the uniform mixture model. Then the corresponding step-up procedure (SU) is such that*

$$\mathbb{P}(\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha) \leq \mathbb{P}(V_m(\nu_k^0) \geq \hat{k}) = \mathbb{P}(q_{(\hat{k})} \leq \nu_k^0), \quad (16)$$

for $\hat{k} = V_m(\tau_{\hat{\ell}})$, where $q_{(1)} \leq \dots \leq q_{(m_0)}$ denotes the ordered p -values under the null and where

$$\nu_k^0 = \max\{t \in [0, 1] : B_m^0(t, k, m_0) \leq \zeta\}. \quad (17)$$

The proof is provided in Section 6.2. It relies on the same idea than the proof of Theorem 3.1 in Guo et al. (2013). Unfortunately, by contrast to the step-down case, using the k -FWE-based critical values into a step-up procedure does not lead in general to a valid finite sample FDP control. This is shown in the simulations of Section 5. As proved by Theorem 3.1 of Guo et al. (2013), however, a valid FDP control holds when using the Lehmann-Romano critical values (8) and when Simes' inequality is valid. Namely, by considering the B_m^0 associated to Markov's inequality, that is, $B_m^0(t, k, m_0) = m_0 t/k$ or equivalently $\nu_k^0 = \zeta k/m_0$, inequality (16) entails a finite sample FDP control when the celebrated Simes' inequality applies (see Simes (1986)),

$$\mathbb{P} \left(\bigcup_{k=1}^{m_0} \{q_{(k)} \leq \zeta k/m_0\} \right) \leq \zeta. \quad (18)$$

The latter holds under positive dependence conditions, see Sarkar (2008b) for instance. As an illustration, this is the case when $\rho \geq 0$ under Assumptions (ρ -equi) (see also Benjamini and Yekutieli (2001)).

3.2. A second unifying bound

When the FDP control cannot be established, an alternative approach is to upper-bound the RHS of (15) by a computable quantity. The next proposition is to be proved in Section 6.3.

Proposition 3.3. *In the setting of Proposition 3.1, assume moreover that there exists a family of random variables $(Z_{\ell, \ell'})_{1 \leq \ell, \ell' \leq m}$ satisfying: for all ℓ, ℓ' ,*

$$\mathbf{1} \{V_m(\tau_\ell) \geq d(\ell', m, m_0)\} \leq Z_{\ell, \ell'} \text{ a.s.} \quad (19)$$

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

$$\mathbb{P}_\theta (\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha) \leq \sum_{\ell=1}^{b_\alpha(m_0)} (\mathbb{E}_\theta (Z_{\ell, \ell-1}) - \mathbb{E}_\theta (Z_{\ell-1, \ell-1})) \wedge (\mathbb{E}_\theta (Z_{\ell, \ell}) - \mathbb{E}_\theta (Z_{\ell-1, \ell})), \quad (20)$$

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

Starting from arbitrary critical values, Proposition 3.3 provides conservative upper-bounds from inequalities of the type (19). For instance, choosing $Z_{\ell, \ell'} = V_m(\tau_\ell)/d(\ell', m, m_0)$, $Z_{\ell, \ell'} = (V_m(\tau_\ell))^2/(d(\ell', m, m_0))^2$ and $Z_{\ell, \ell'} = \mathbf{1} \{V_m(\tau_\ell) \geq d(\ell', m, m_0)\}$ leads to upperbounding (20) by

$$\max_{1 \leq u \leq m} \left\{ u \sum_{\ell=1}^{b_\alpha(u)} \frac{\tau_\ell - \tau_{\ell-1}}{d(\ell, m, u)} \right\}; \quad (21)$$

$$\max_{1 \leq u \leq m} \left\{ \sum_{\ell=1}^{b_\alpha(u)} \left(u(\tau_\ell - \tau_{\ell-1}) + u(u-1) \max_{\substack{\theta \in \Theta \\ m_0(\theta)=u}} \{F_\theta^{(2)}(\tau_\ell) - F_\theta^{(2)}(\tau_{\ell-1})\} \right) (d(\ell, m, u))^{-2} \right\}; \quad (22)$$

$$\max_{1 \leq u \leq m} \left\{ \sum_{\ell=1}^{b_\alpha(u)} \max_{\substack{\theta \in \Theta \\ m_0(\theta)=u}} \{(\mathbb{P}_\theta (V_m(\tau_\ell) \geq d(\ell-1, m, u)) - \mathbb{P}_\theta (V_m(\tau_{\ell-1}) \geq d(\ell-1, m, u))) \wedge (\mathbb{P}_\theta (V_m(\tau_\ell) \geq d(\ell, m, u)) - \mathbb{P}_\theta (V_m(\tau_{\ell-1}) \geq d(\ell, m, u)))\} \right\}, \quad (23)$$

where $F_{\theta}^{(2)}$ is defined by (9). While the upper-bound (21) has been established in Romano and Shaikh (2006a) in the step-down case and Romano and Shaikh (2006b) in the step-up case, Proposition 3.3 allows to derive a sharper upper-bound (23) when additional informations are available concerning the joint null-distribution. The new bound also improves the recent result of Guo et al. (2013), which combines the Markov and pairwise approaches, see Figure 4. Remember that the step-down and step-up procedures using Markov-based (i.e., Lehmann-Romano) critical values (8) provably control the FDP in the (nonnegative) equi-correlated setting. From the left column of Figure 4, the new proposed (“Exact”) bound shows that, if the equi-correlation ρ is known and large enough, an improvement over the Lehmann-Romano procedure is possible by incorporating the dependence structure (the bound is below ζ). This is, to the best of our knowledge, the first time that such improvement is proved to be possible. By contrast, the bound proposed in Guo et al. (2013) is always above the targeted level ζ . Nevertheless, to be completely fair, we should notice that their bound was established under slightly weaker assumptions (pairwise Gaussian rather than multivariate Gaussian).

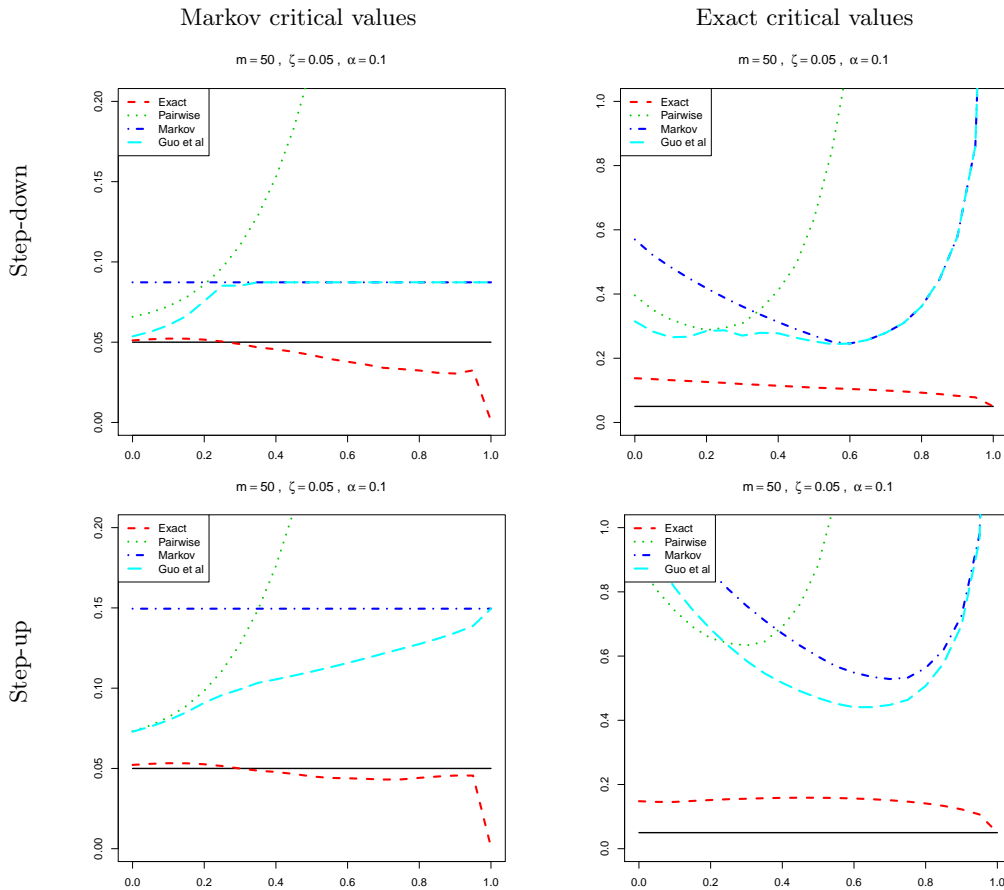


FIG 4. Bounds derived from (21) (Markov), (22) (Pairwise) and (23) (Exact), displayed in function of ρ , the equicorrelation between the Gaussian test statistics (while assuming $(\rho\text{-equi})$). Two k -FWE-based adaptive critical values are used as initial critical values: Markov (see (8), first column) or exact (see (13), second column). The step-down (first row) and step-up (second row) algorithms are considered.

Finally, when using the Exact-based critical values (right column in Figure 4), however, there is no room left for such improvement: the “Exact” bounds are above ζ . Since simulations

indicate that the FDP control is close to be achieved in these cases (see Section 5), these bounds are still not totally satisfactory to understand the subtle behavior of k -FWE-based approaches. The next section complements this view by studying an asymptotic framework (w.r.t. m), for which ζ is a valid asymptotic bound (in the step-up case).

4. Asymptotic results

The present section validates the use of k -FWE-based critical values into a step-up procedure for controlling the FDP as the number of hypotheses m tends to infinity. Roughly speaking, the essence of the argumentation is as follows: when $\hat{\ell}/m$ converges in probability to some deterministic quantities, then the fluctuations of $\hat{\ell}$ asymptotically disappear in (4), and thus this probability is bounded by ζ from (Bound). This argument is rigorously formalized in this section.

4.1. Setting and assumptions

We consider the convergence of probabilities of the type (4) when m is tending to infinity. Hence, we should formally consider a *sequence* of models $(\Theta^{(m)}, m \geq 1)$ and a sequence of parameters $(\theta^{(m)}, m \geq 1)$ with $\theta^{(m)} \in \Theta^{(m)}$ for all $m \geq 1$. The latter sequence is assumed to be fixed once for all throughout this section.

Roughly speaking, we make two kinds of assumptions to get our results: first, the bounding device should have some “appropriate” convergence properties (typically the case for the Markov, pairwise and exact devices, as we will see in Section 4.2). Second, and maybe more importantly, we should make *distributional* assumptions, that is, we should make assumptions on the sequence $(\theta^{(m)}, m \geq 1)$. The latter are given and discussed below. Also, with some abuses, we denote here $m_0(m)$ the number of true nulls (possibly random in the Bernoulli mixture model).

The first assumption on $(\theta^{(m)})_m$ relies on the behavior of the (rescaled) process $V_m(\cdot)$.

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

For instance, if the following technical condition holds

$$m_0(m)/m \rightarrow \pi_0, \text{ where } \pi_0 \in (0, 1) \text{ (in probability),} \quad (24)$$

Assumption (FLT) holds (with $r_m = \sqrt{m}$) when the p -values $(p_i, H_i = 0)$ are i.i.d. by Donsker Theorem. More generally, dependencies satisfying “mixing” conditions also lead to Donsker-type theorems and thus to (FLT), see, e.g., [Dedecker and Prieur \(2007\)](#); [Doukhan et al. \(2010\)](#) or [Farcomeni \(2007\)](#). Recently, some efforts have been undertaken to consider others type of dependence, not necessarily locally structured, see [Bardet and Surgailis \(2013\)](#); [Soulier \(2001\)](#) and [Delattre and Roquain \(2012\)](#). The latter can be more suitable to model high dimensional data, see [Friguet et al. \(2009\)](#).

While providing an exhaustive review of settings encompassing Assumption (FLT) exceeds the scope of the paper, let us discuss this assumption in the multivariate Gaussian case. In that case, condition (FLT) only relies on the behavior of the $m_0(m) \times m_0(m)$ submatrix $(\Gamma_{i,j})_{i,j:H_i=0,H_j=0}$ as m grows. Since the positions i for which $H_i = 0$ are unknown, it is certainly more desirable to assume

$$\text{(FLT) holds for the full null process } V'_m(t) \text{ defined by (14),} \quad \text{(FLT')}$$

which involves the whole matrix Γ (known) and not an unknown portion of it. Lemma 7.3 in Section 7.5 shows that, in the Bernoulli random mixture model, Assumption (FLT') entails Assumption (FLT). Hence, to the price of adding a random effect on m_0 , (FLT') can be used to derive sufficient conditions on (the sequence) $\Gamma = \Gamma^{(m)}$ to ensure (FLT). For instance, either of the two following conditions is sufficient:

- Stationary: $\Gamma_{i,j}^{(m)} = r(|i-j|)$, where r is a nonnegative function, fixed with m , with $r(0) = 1$, that satisfies either $\sum_{n \geq 1} r(n) < +\infty$ (short-range Csörgő and Mielniczuk (1996)) or $r(n) = n^{-D}L(n)$, $0 < D < 1$, L being a function slowly varying at infinity (long-range Dehling and Taqqu (1989));
- “Vanish-second-order” Delattre and Roquain (2012): for $\bar{\rho}_m = \frac{1}{m(m-1)} \sum_{i \neq j} \Gamma_{i,j}^{(m)}$, $r_m = (m^{-1} + |\bar{\rho}_m|)^{-1/2}$,

$$\frac{r_m^2}{m^2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)} \right)^2 \rightarrow 0; \quad \frac{r_m^{4+\varepsilon_0}}{m^2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)} \right)^4 \rightarrow 0, \quad \text{for some } \varepsilon_0 > 0.$$

As a particular case, under equi-correlation (ρ -equi) with parameter $\rho = \rho_m$, the latter is equivalent to $\rho_m \rightarrow 0$, see also Delattre and Roquain (2011).

The second type of assumption on $(\theta^{(m)})_m$ concerns the process $R_m(\cdot)$.

$$\left\{ \begin{array}{l} \text{The process } \widehat{G}_m(t) = R_m(t)/m, t \in [0, 1], \text{ is such that } \|\widehat{G}_m - G\|_\infty = o_P(1), \text{ for} \\ \text{some } G : [0, 1] \rightarrow [0, 1] \text{ continuous.} \end{array} \right. \quad \text{(Consist)}$$

Condition (Consist) is a “law of large number”-type assumption (by contrast to (FLT) which is of the “central limit”-type). In the Gaussian multivariate setting, (Consist) is satisfied under (24) and if the following conditions hold (see Lemma A.1):

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

$$m^{-2} \sum_{i,j=1}^m (\Gamma_{i,j})^2 \rightarrow 0. \quad \text{(LLNdep)}$$

Here, $G(t) = \pi_0 t + \pi_1 F_1(t)$ and $F_1(t) = \int_0^\infty \Phi(\Phi^{-1}(t) - \beta) d\nu(\beta)$. Note that H can be random in Condition (Conv-alt): in that case, the weak convergence means that the distance between the two distributions tends to zero in probability. Condition (Conv-alt) extends the usual situation where the p -value distribution is fixed with m under the alternative (as assumed in Delattre and Roquain (2011); Delattre and Roquain (2012); Genovese and Wasserman (2004);

Neuivial (2008)). Similarly to above, (Conv-alt) depends on the unknown H , which can be seen as a limitation, because it is unknown. To circumvent this problem, we can use that, in the Bernoulli model, (Conv-alt) is implied by

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

as stated by Lemma 7.3 in Section 7.5. As an illustration, when $\mu_i = f(i/m)$ for some continuous function f on $[0, 1]$ which is positive on $(0, 1]$, then (Conv-alt') holds by taking ν equal to the image distribution of the Lebesgue measure on $[0, 1]$ via f .

Finally, let us underline that under (ρ -equi), assumptions (LLNdep) and (FLT) are both satisfied whenever $\rho = \rho_m \rightarrow 0$ (see Delattre and Roquain (2011)).

4.2. A general result and applications

In this section, we consider a sequence of parameters $(\theta^{(m)}, m \geq 1)$ (in the fixed, uniform or Bernoulli model), some bounding device B_m^0 and the associated k -FWE-based critical values $(\tau_\ell)_{1 \leq \ell \leq m}$ as in Definition 2.1, being either adaptive (6) or not (5). Let us underline that when $(\theta^{(m)}, m \geq 1)$ is taken in the Bernoulli model, inequality (Bound), on which the bounding device relies on, is assumed to hold either in the fixed or in the uniform model.

Theorem 4.1. *Consider the step-up procedure associated to the κ -kapped critical values $\tau'_\ell = \tau_\ell \wedge \kappa$, $\ell = 1, \dots, m$ with rejection number $\widehat{\ell}$, for some $\kappa \in (0, 1)$. Assume that $\theta^{(m)}$ satisfies (FLT) and assume*

$$\tau'_\ell \text{ converges in probability to } t^*, \text{ for some } t^* \in (0, 1). \quad (\text{Conv-threshold})$$

Then we have

$$\limsup_m \left\{ \mathbb{P}_{\theta^{(m)}} \left(\text{FDP}_m \left(\tau'_\ell \right) > \alpha \right) \right\} \leq \zeta. \quad (25)$$

Theorem 4.1 is proved in Section 7.1. Note that the FDP control is established for the κ -kapped critical values and not the original critical values. The smaller κ , the more restrictive the kapping. We will see that choosing $\kappa \in (0, 1)$, arbitrary close to 1, is often possible so that this condition is not very restrictive. Next, Theorem 4.1 introduces the novel Assumption (Conv-threshold). The following result identifies sufficient conditions to ensure it (see Section 7.2 for a proof).

Proposition 4.2. *Denote by f_m the rejection curve associated to the critical values τ_ℓ on $[0, \kappa]$, that is, for $t \in [0, \kappa]$,*

$$f_m(t) = m^{-1} \times \min\{\ell \in \{0, \dots, m+1\} : \tau_\ell \geq t\}, \quad (26)$$

(with the convention $\tau_0 = 0, \tau_{m+1} = 1$). Assume that $\theta^{(m)}$ satisfies (Consist) for some function G and assume the following for the rejection curve f_m :

$$f_m(t) \text{ converges to } f_\infty(t) \text{ for all } t \in [0, \kappa] \text{ where } f_\infty : [0, \kappa] \rightarrow [0, 1] \text{ is continuous with } f_\infty(t) > 0 \text{ for } t \in (0, \kappa]. \quad (27)$$

Let

$$t^* = \sup\{t \in [0, \kappa] : G(t) \geq f_\infty(t)\} \quad (28)$$

and assume further

$$t^* > 0 \text{ and } \{G(t^*) = f_\infty(t^*) \Rightarrow G(u_p) > f_\infty(u_p) \text{ for all } p \text{ for some } u_p \uparrow t^*\}. \quad (\text{StableLastCross})$$

Then the step-up procedure associated to the κ -capped critical values $\tau'_\ell = \tau_\ell \wedge \kappa$, $\ell = 1, \dots, m$, satisfies **(Conv-threshold)**.

Let us discuss the conditions introduced by Proposition 4.2. As mentioned in Section 4.1, a typical situation for which **(Consist)** holds is the Gaussian setting satisfying **(24)**-**(Conv-alt)**-**(LLNdep)**. Under these assumptions, we easily check that condition **(27)** holds for the three types of device defined in Section 2.3, see Lemma 7.1 in Section 7.3. Figure 5 displays the resulting limiting functions

$$f_\infty(t) = \frac{t}{\alpha\lambda_1 + (1-\alpha)t\lambda_2} \wedge 1, \quad (29)$$

where λ_1 and λ_2 are given in each case in Table 1. As for **(StableLastCross)**, it is implied for instance by the two conditions

$$\begin{aligned} G(t) = f_\infty(t) \text{ has at most one solution on } (0, \kappa]; & \quad (\text{Unique}) \\ \lim_{t \rightarrow 0^+} G(t)/f_\infty(t) \in (1, +\infty]; & \quad (\text{Noncritical}) \end{aligned}$$

Assumption **(Unique)** and **(Noncritical)** hold for instance for G given by **(Conv-alt)** and f_∞ given by **(29)** with $\lambda_2 = 0$ (non adaptive case). In the adaptive case, that is $\lambda_2 = 1$, **(Noncritical)** holds but we can find very specific parameter values for which f_∞ has several intersection points with G (e.g., for $\lambda_1 = 1$, $\alpha = 0.009$, $\pi_0 = 0.01$, $\beta \equiv 2.7$). In that case, κ can always be chosen small enough to ensure **(Unique)**.

In conclusion, we are able to state the following corollaries.

Corollary 4.3. *In the Gaussian setting, consider the **non-adaptive** k -FWE-based critical values τ_ℓ , $\ell = 1, \dots, m$, computed in the uniform mixture model, and associated to the Markov, pairwise or exact device. Then, in the Bernoulli mixture model, the asymptotic FDP control **(25)** holds for the step-up procedure with critical values τ_ℓ , $\ell = 1, \dots, m$, for a covariance matrix $\Gamma^{(m)}$ satisfying **(FLT')** and alternative means μ_i , $i = 1, \dots, m$, satisfying **(Conv-alt')**. Moreover, the same conclusion holds for the **adaptive** case, except that the asymptotic FDP control holds for the κ -capped (adaptive) critical values, where $\kappa \in (0, 1)$ is taken small enough to ensure that $G(t) = \pi_0 t + \pi_1 \int_0^\infty \Phi(\Phi^{-1}(t) - \beta) d\nu(\beta)$ satisfies **(Unique)**.*

Corollary 4.4. *In a non-necessarily Gaussian setting, consider the Lehmann-Romano critical values **(8)** (adaptive or not). Then, for the corresponding step-up procedure, the asymptotic FDP control **(25)** holds in a fixed mixture model satisfying Assumptions **(FLT)**, **(Consist)** and under condition **(StableLastCross)**, taken with $\kappa = \zeta\alpha$ (non adaptive) or $\kappa = \zeta$ (adaptive).*

Note that the FDP control of Lehmann-Romano step-up procedure stated in Corollary 4.4 is markedly different from the state-of-the-art finite sample control of Guo et al. (2013): while it is asymptotic, it does not relies on Simes' inequality (see **(18)**), but rather on a central limit type assumption for the dependencies. Hence, it can be seen as “relaxing” the assumption of positively regressively dependent p -values, which is often used to ensure Simes' inequality.

Device	Markov	Pairwise	Exact	Adaptive	No	Yes
λ_1	ζ	$\zeta^{1/2}$	1	λ_2	0	1

TABLE 1

Values of λ_1 and λ_2 to be used in (29) for each set of critical values.

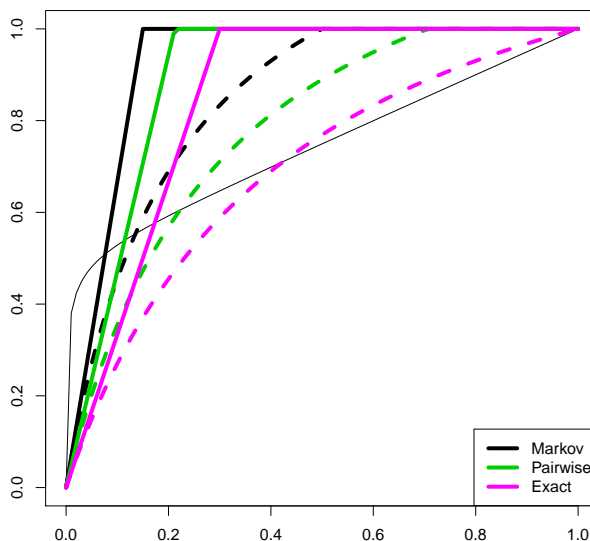


FIG 5. $f_\infty(t)$ given by (29) in function of t , for the different bounding devices and $\alpha = 0.3$, $\zeta = 0.5$. Non adaptive (solid lines) ; adaptive (dashed). The thin solid black line is $G(t) = \pi_0 t + (1 - \pi_0)\Phi(\Phi^{-1}(t) - \beta)$ in the one-sided Gaussian case for $\pi_0 = 0.5$ and alternative means all equal to $\beta = 3$.

Remark 4.5. The threshold of the exact- k -FWE-based step-up procedure asymptotically coincides with the limit threshold of the original BH procedure (or AORC procedure in the adaptive case). This is due to our “weak dependence” type assumptions. However, please note that it does not imply that they are asymptotically equivalent in terms of FDP distribution: while it is well known that the BH/AORC procedures do not control the FDP asymptotically (even under independence, see [Neuivial \(2008\)](#)), our analysis shows that this is the case for the k -FWE-based step-up procedures.

5. Numerical experiments

This section complements Sections 3 and 4 by evaluating the probability that the FDP exceeds α for a fixed m with numerical experiments.

We consider the Gaussian multivariate framework defined in Section 1.2, where $\mu_i = \beta$ for all i (all alternative means are equal) and the random set $\{i : H_i = 0\}$ is uniformly distributed among the subsets of $\{1, \dots, m_0\}$ of cardinal m_0 (uniform mixture model). This section presents the result in the case where the matrix Γ is assumed to be ρ -equicorrelated, see (ρ -equi). The case of a 3-factor model provides similar results and is postponed to the supplement [Delattre and Roquain \(2013\)](#), see Figures S-2 and S-3 therein. Figure 6 displays $\mathbb{P}(\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha)$, computed with 3×10^5 simulations, for a procedure coming from a Markov, pairwise or exact device, being either adaptive (solid) or not (dashed) and being either step-

down (SD, lines 1 and 2 on Figure 6) or step-up (SU, lines 3 and 4 on Figure 6). The parameter values/ranges are as follows: $\alpha = 0.1$, $m = 100$, $\rho \in \{0.1, 0.5\}$, $\zeta \in \{0.05, 0.5\}$, $m_0 \in \{50, 90\}$, $\beta \in \{0.01, 1, 1.1, \dots, 4.9, 5\}$. Remember that Theorem 4.1 only provides the (asymptotic) FDP control when ρ is small ($\rho = \rho_m \rightarrow 0$), so certainly not when $\rho = 0.5$. Also, for an arbitrary $\rho \geq 0$, it is known that the FDP control holds for the Markov device (because Simes' inequality holds here, see Section 3.1).

First, Figure 6 shows that the pairwise and Markov devices are conservative (and sometimes very conservative) in the sense that the actually achieved exceedance probability is much smaller than the targeted level ζ . By contrast, the exact device is much more accurate and often has a probability very close to ζ in its adaptive form. This seems fair since the exact device uses the full joint distribution of the null p -values. Second, it is worth to note that the adaptive procedures improve over their non-adaptive counterparts, especially when m_0/m is small ($= 0.5$ here) and β is large (say, $\beta \geq 2$). Here, some ‘‘signal’’ (in quantity and strength) is needed to make the adaptation effective. This corroborates previous studies designing adaptive procedures, see, e.g., Benjamini et al. (2006); Blanchard and Roquain (2009).

Third, while for *step-up* procedures the probability can exceeds ζ (e.g., $\pi_0 = 0.9$, $\zeta = 0.05$, $\rho = 0.1$, $\beta = 3$), this is not the case for the *step-down* procedures. Moreover, this observation is also true in the 3-factor model. This intriguing fact might suggest that a formal FDP controlling result exists. To investigate deeper this issue, we reproduced the experiment with a *perfect* adaptation step, that is, by taking τ_ℓ^0 (see (7)) instead of $\tilde{\tau}_\ell$ (see (6)). The result, displayed in the supplement in Figure S-1, shows that the FDP control is not maintained. This shows that the FDP control appearing on Figure 6 is a very subtle phenomenon that might be hard to capture formally (because a proof valid for $\tilde{\tau}_\ell$ should fail for τ_ℓ^0).

Finally, note that we only study here the type I error control, not the power. The reason is that we find it clear from the critical values (see Figures 2 and 3) that the exact device leads to more discoveries (and thus more power) than Markov/pairwise devices.

6. Proofs for finite sample results

6.1. Proof of Proposition 3.1 (ii)

Since $\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha$ implies $\lfloor \alpha \hat{\ell} \rfloor + 1 \leq m_0$, we have $\hat{\ell} \leq b_\alpha(m_0)$. Also, $\hat{\ell} = R_m(\tau_{\hat{\ell}}) \leq m_1 + V_m(\tau_{\hat{\ell}})$, which implies $V_m(\tau_{\hat{\ell}}) \geq \hat{\ell} - m_1$. This implies (15) in case (ii).

6.2. Proof of Corollary 3.2

Let $\hat{k} = V_m(\tau_{\hat{\ell}})$ and note that $\hat{k} \leq m_0$ and $\{\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha\} = \{\hat{k} \geq \lfloor \alpha \hat{\ell} \rfloor + 1\}$. First, in the non-adaptive case, we have by definition of \bar{B}_m , for all t and $k \leq m_0$,

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

Hence, we have by definition of the (non-adaptive) critical values, whenever $\hat{k} \geq \lfloor \alpha \hat{\ell} \rfloor + 1$,

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

which is larger than or equal to $B_m^0(\tau_{\hat{\ell}}, \hat{k}, m_0)$. Hence, we obtain

$$\{\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha\} \subset \{B_m^0(\tau_{\hat{\ell}}, \hat{k}, m_0) \leq \zeta\} \subset \{\tau_{\hat{\ell}} \leq \nu_{\hat{k}}^0\},$$

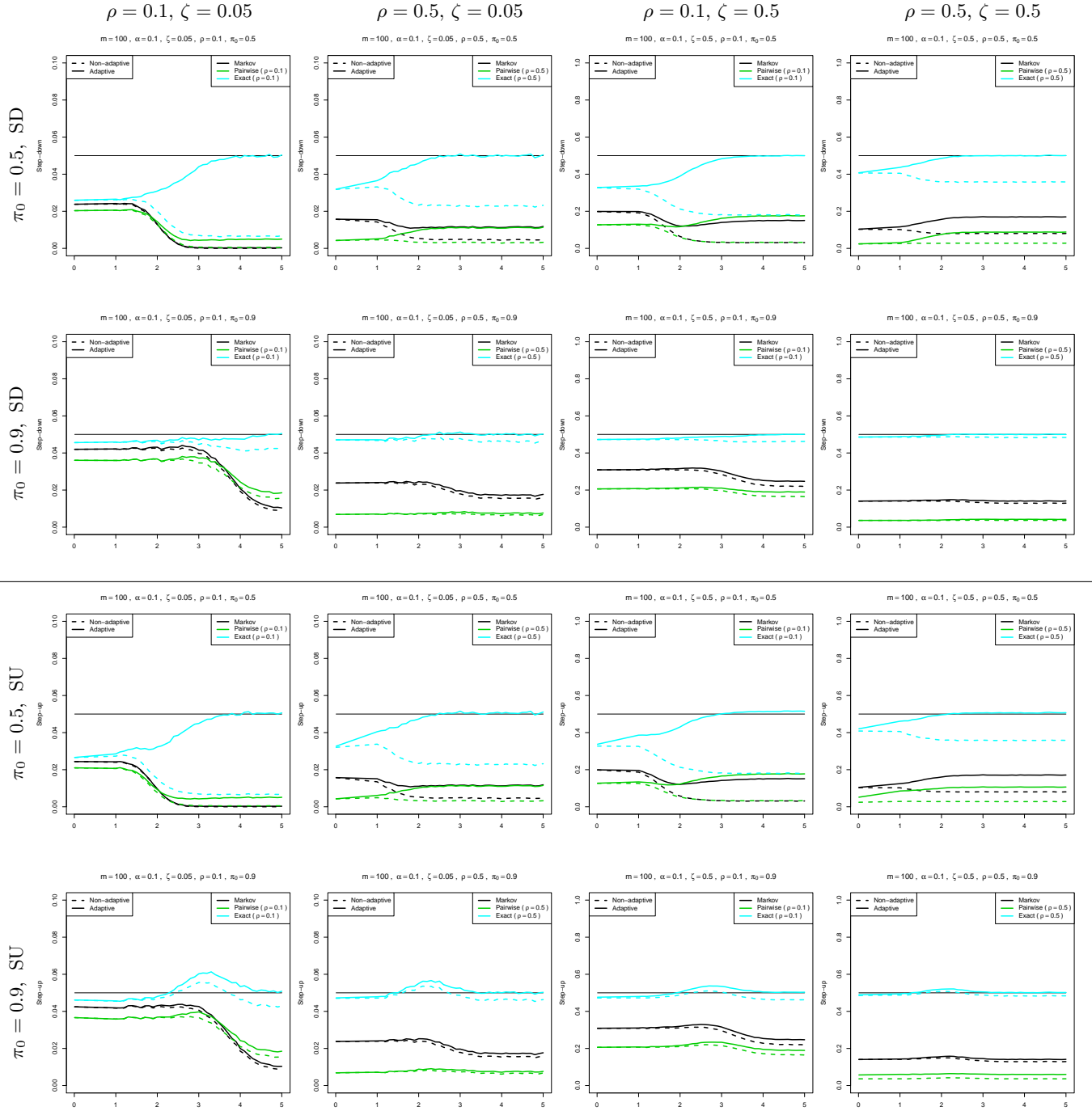


FIG 6. FDP exceedance probability $\mathbb{P}(\text{FDP}_m(\tau_{\hat{\ell}}) > \alpha)$ in function of β for several model parameters and Γ satisfying (ρ -equi), see text.

and thus (16). Second, in the adaptive case, we use that $m_0 \leq m - \hat{\ell} + \hat{k}$. Thus, we have for all t ,

$$\tilde{B}_m(t, \hat{k}, \hat{\ell}) = \sup_{\hat{k} \leq u \leq m - \hat{\ell} + \hat{k}} \left\{ B_m^0(t, \hat{k}, u) \right\} \geq B_m^0(t, \hat{k}, m_0).$$

Hence, whenever $\widehat{k} \geq \lfloor \alpha \widehat{\ell} \rfloor + 1$, we have

$$\zeta \geq \widetilde{B}_m(\tau_{\widehat{\ell}}, \lfloor \alpha \widehat{\ell} \rfloor + 1, \widehat{\ell}) \geq \widetilde{B}_m(\tau_{\widehat{\ell}}, \widehat{k}, \widehat{\ell}) \geq B_m^0(\tau_{\widehat{\ell}}, \widehat{k}, m_0). \quad (30)$$

Hence, $\tau_{\widehat{\ell}} \leq \nu_{\widehat{k}}^0$ and the proof is finished.

6.3. Proof of Proposition 3.3

From (15), we derive

$$\begin{aligned} \mathbb{P}_\theta (\text{FDP}_m(\tau_{\widehat{\ell}}) > \alpha) &\leq \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{P}_\theta \left(V_m(\tau_\ell) \geq d(\ell, m, m_0), \widetilde{\ell} = \ell \right) \\ &\leq \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta \left(Z_{\ell, \ell} \mathbf{1} \{ \widetilde{\ell} = \ell \} \right). \end{aligned}$$

Now, the RHS of the previous display is equal to

$$\begin{aligned} &\sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta \left(Z_{\ell, \ell} \mathbf{1} \{ \widetilde{\ell} \geq \ell \} \right) - \sum_{\ell=1}^{b_\alpha(m_0)-1} \mathbb{E}_\theta \left(Z_{\ell, \ell} \mathbf{1} \{ \widetilde{\ell} \geq \ell + 1 \} \right) \\ &= \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta \left(Z_{\ell, \ell} \mathbf{1} \{ \widetilde{\ell} \geq \ell \} \right) - \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta \left(Z_{\ell-1, \ell-1} \mathbf{1} \{ \widetilde{\ell} \geq \ell \} \right) \\ &= \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta \left((Z_{\ell, \ell} - Z_{\ell-1, \ell-1}) \mathbf{1} \{ \widetilde{\ell} \geq \ell \} \right) \\ &\leq \sum_{\ell=1}^{b_\alpha(m_0)} \mathbb{E}_\theta (Z_{\ell, \ell-1} - Z_{\ell-1, \ell-1}) \wedge \mathbb{E}_\theta (Z_{\ell, \ell} - Z_{\ell-1, \ell}), \end{aligned}$$

which proves the result.

7. Proofs for asymptotic results

7.1. Proof of Theorem 4.1

Denote τ'_ℓ by \widehat{t} for short. Let us consider $q_m^0(t)$ the $(1-\zeta)$ -quantile of the distribution of $V_m(t)$, for all $t \in [0, 1]$. We have by (Bound) and (30), $\mathbb{P}(\text{FDP}_m(\widehat{t}) > \alpha) \leq \mathbb{P}(V_m(\widehat{t}) > q_m^0(\widehat{t}))$. Let us now consider the process

$$U_m(t) = r_m(V_m(t)/m - q_m^0(t)/m) = Z_m(t) - r_m(q_m^0(t)/m - (m_0/m)t)$$

and consider a compact set $K \subset (0, 1)$ containing t^* . Applying Lemma A.3, we have that the function sequence

$$r_m(q_m^0(t)/m - (m_0/m)t)_{t \in K}$$

converges uniformly to $(q_\zeta(t))_{t \in K}$, where $q_\zeta(t)$ is the $(1-\zeta)$ -quantile of $Z(t)$. This entails that $U_m(\cdot)$ is a C-tight process with

$$(U_m(t))_{t \in K} \rightsquigarrow (Z(t) - q_\zeta(t))_{t \in K}.$$

Next for any $\varepsilon > 0$ and $\eta > 0$, by using (**Conv-threshold**), we have

$$\begin{aligned} \mathbb{P}(\text{FDP}_m(\hat{t}) > \alpha) &\leq \mathbb{P}(\text{FDP}_m(\hat{t}) > \alpha, |\hat{t} - t^*| \leq \varepsilon) + \mathbb{P}(\text{FDP}_m(\hat{t}) > \alpha, |\hat{t} - t^*| > \varepsilon) \\ &\leq \mathbb{P}(U_m(\hat{t}) > 0, |\hat{t} - t^*| \leq \varepsilon) + o(1) \\ &\leq \mathbb{P}(\{U_m(\hat{t}) - U_m(t^*)\} + U_m(t^*) > 0, |\hat{t} - t^*| \leq \varepsilon) + o(1) \\ &\leq \mathbb{P}\left(\sup_{s \in K \cap [t^* - \varepsilon, t^* + \varepsilon]} \{U_m(s) - U_m(t^*)\} > \eta\right) + \mathbb{P}(U_m(t^*) > -\eta) + o(1). \end{aligned}$$

By letting $m \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, we get

$$\limsup_m \{\mathbb{P}(\text{FDP}_m(\hat{t}) > \alpha)\} \leq \mathbb{P}(Z(t^*) > q_C(t^*) - \eta)$$

Since the c.d.f. of $Z(t^*)$ is continuous, we obtain the result by making η converging to zero in the last display.

7.2. Proof for Proposition 4.2

First observe that since f_m is a nondecreasing function and since f_∞ is continuous, then the convergence of f_m to f_∞ is uniform on $[0, \kappa]$. Also, f_∞ is a nondecreasing function. By Lemma A.2, we have $\tau'_\ell = \tilde{t}$, where

$$\tilde{t} = \sup\{t \in [0, 1] : \widehat{\mathbb{G}}_m(t) \geq f_m(t)\}. \quad (31)$$

Note that $\tilde{t} \leq \kappa$. For some realization such that $\sup_{t \in [0, 1]} |\widehat{\mathbb{G}}_m(t) - G(t)| \rightarrow 0$, assume that \tilde{t} converges to some $t \in [0, \kappa]$ and let us prove that t is equal to t^* given by (28).

First, since $\widehat{\mathbb{G}}_m(\tilde{t}) \geq f_m(\tilde{t})$, we have $G(t) \geq f_\infty(t)$ and thus $t \leq t^*$. Let us prove $t \geq t^*$. First, if $G(t^*) \neq f_\infty(t^*)$, that is, $G(t^*) > f_\infty(t^*)$, we have $\widehat{\mathbb{G}}_m(t^*) \geq f_m(t^*)$ for m large enough and thus $t \geq t^*$. Second, if $G(t^*) = f_\infty(t^*)$, we have by (**StableLastCross**), that $G(u_p) > f_\infty(u_p)$ for all p , for some $u_p \uparrow t^*$. This entails, for all p , for m large enough, $\widehat{\mathbb{G}}_m(u_p) > f_m(u_p)$ and thus $t \geq u_p$. Hence, $t \geq t^*$ by making p tends to infinity.

Finally, by using a subsequence argument, this implies that on the event $\{\sup_{t \in [0, 1]} |\widehat{\mathbb{G}}_m(t) - G(t)| \rightarrow 0\}$, we have that \tilde{t} converges to t^* . This gives the result because of Assumption (**Consist**).

7.3. Computing f_∞ for Markov, Pairwise and Exact devices

Lemma 7.1. *Condition (27) holds with $\kappa = 1$ for each rejection curve f_m corresponding to the devices defined in Section 2.3: Markov under no assumption ; Pairwise and Exact in the Gaussian setting satisfying (**LLNdep**) and in the uniform model. Additionally, the limiting rejection curves are given by (29).*

To prove Lemma 7.1, we use Lemma 7.2 in Section 7.4. First, for Markov device, it is clear that the assumption of Lemma 7.2 holds with $g(u) = \zeta\alpha u$ (non-adaptive) or $g(u) = \zeta\alpha u/(1 - u(1 - \alpha))$ (adaptive). This gives the result.

Second, consider the pairwise device in the Gaussian case, in the uniform model, with the assumption (**LLNdep**). Remember that, from Section 2.3, we have

$$B_m(t, \ell_m) = \frac{\bar{m}_0(\ell_m)t + \bar{m}_0(\ell_m)(\bar{m}_0(\ell_m) - 1)F_\Gamma^{(2)}(t)}{(\lfloor \alpha \ell_m \rfloor + 1)^2},$$

with $B_m(t, \ell_m) = \overline{B}_m(t, \lfloor \alpha \ell_m \rfloor + 1)$, $\overline{m}_0(\ell_m) = m$ (non adaptive) or $B_m(t, \ell_m) = \widetilde{B}_m(t, \lfloor \alpha \ell_m \rfloor + 1, \ell_m)$, $\overline{m}_0(\ell_m) = m - \ell_m + \lfloor \alpha \ell_m \rfloor + 1$ (adaptive). From (LLNdep), we have $\Gamma_{I,J}$ that converges in probability to 0 (I and J being two indexes independent and uniformly distributed in $\{1, \dots, m\}$). Hence, by the continuous mapping theorem, for all t ,

$$\int_0^t \Phi \left(\frac{\Phi^{-1}(t) - \Gamma_{I,J} \Phi^{-1}(x)}{\sqrt{1 - \Gamma_{I,J}^2}} \right) dx$$

converges in probability to t^2 and thus $F_\Gamma^{(2)}(t) \rightarrow t^2$. This entails that whenever $\ell_m/m \rightarrow u$, $B_m(t, \ell_m)$ converges to $(t/(\alpha u))^2$ (non adaptive) or $(t(1 - (1 - \alpha)u)/(\alpha u))^2$ (adaptive). Hence, by applying Lemma 7.2, we obtain condition (27).

Third, consider the exact device under the same assumptions. The exact device is such that

$$B_m(t, \ell_m) = \mathbb{P}_{m_0 = \overline{m}_0(\ell_m)}(V_m(t) \geq \lfloor \alpha \ell_m \rfloor + 1),$$

with $\overline{m}_0(\ell_m)$ defined as above. By assuming (LLNdep), we have that $V_m(t)/\overline{m}_0(\ell_m)$ converges in probability to t . Letting $g(u) = \alpha u$ (non adaptive) or $g(u) = \alpha u/(1 - (1 - \alpha)u)$ (adaptive), this implies that whenever $\ell_m/m \rightarrow u$ and $t \neq g(u)$, $B_m(t, \ell_m)$ converges to $\mathbf{1}\{t \geq g(u)\}$. Hence, Lemma 7.2 can be applied, which gives that condition (27) holds.

7.4. A useful result to prove the convergence (27)

Lemma 7.2. *Let us consider some bounding device B_m^0 as in (Bound) and the corresponding critical values τ_ℓ , $\ell = 1, \dots, m$, being either adaptive or not. Assume that there exists a function $g : [0, 1] \rightarrow [0, g(1)]$ that is one to one with $g(0) = 0$, and such that for all sequence $\ell_m \in \{1, \dots, m\}$ with ℓ_m/m converging to some $u \in [0, 1]$, we have $\forall t \in [0, 1]$,*

- $\lim_m B_m(t, \ell_m) < \zeta$ for all $t < g(u)$;
- $\lim_m B_m(t, \ell_m) > \zeta$ for all $t > g(u)$;

where $B_m(t, \ell_m)$ denotes either $\overline{B}_m(t, \lfloor \alpha \ell_m \rfloor + 1)$ (non adaptive case) or $\widetilde{B}_m(t, \lfloor \alpha \ell_m \rfloor + 1, \ell_m)$ (adaptive case). Then, f_m defined by (26) (with $\kappa = 1$), satisfies that for any $t \in [0, 1]$, $f_m(t) \rightarrow f_\infty(t)$, with $f_\infty(t) = g^{-1}(t)$ for $t \in [0, g(1)]$ and $f_\infty(t) = 1$ for $t \in (g(1), 1]$.

From an intuitive point of view, in Lemma 7.2, a natural candidate for $g(u)$ is the value of t solving the equation $\lim_m B_m(t, \ell_m) = \zeta$.

We now prove Lemma 7.2. First, let us prove the following assertion:

$$\text{for any sequence } \ell_m/m \rightarrow u \in [0, 1], \text{ we have } \tau_{\ell_m} \text{ converges to } g(u). \quad (32)$$

By assumption, for all $\varepsilon > 0$, $\lim_m B_m(g(u) - \varepsilon, \ell_m) < \zeta$. This implies that, for large m , $B_m(g(u) - \varepsilon, \ell_m) < \zeta$ and thus $g(u) - \varepsilon \leq \tau_{\ell_m}$ by definition of τ_{ℓ_m} . Similarly, since $\lim_m B_m(g(u) + \varepsilon, \ell_m) > \zeta$, $g(u) + \varepsilon \geq \tau_{\ell_m}$ for large m . This provides that τ_{ℓ_m} converges to $g(u)$.

Let us now prove Lemma 7.2. Note that, by (32), we have $\tau_1 \rightarrow g(0) = 0$ and $\tau_m \rightarrow g(1)$. Also, $f_m(0) = 0$ and $f_m(1) = 1 + 1/m \rightarrow 1$ (because $\tau_m < 1$). Take $t \in (0, 1)$ and consider any limit point $u \in [0, 1]$ of $f_m(t)$. The result is proved if we show that $u = f_\infty(t)$. First, if $t \in (g(1), 1]$, then $t > \tau_m$ for m large enough and thus $f_m(t) = 1 + 1/m$ which entails

$f_m(t) \rightarrow 1 = f_\infty(t)$. Second, if $t \in (0, g(1))$, we have for m large enough that $t \in [\tau_1, \tau_m]$ (because $\tau_1 \rightarrow 0 < t$ and $\tau_m \rightarrow g(1) > t$). Hence, by definition of f_m , we have

$$\tau_{(mf_m(t)-1)\vee 1} \leq t \leq \tau_{(mf_m(t))\wedge m}.$$

Applying (32) twice with $\ell_m = (mf_m(t)-1)\vee 1$ and $\ell_m = (mf_m(t))\wedge m$ (along the appropriate subsequence), we obtain that $\tau_{(mf_m(t)-1)\vee 1}$ and $\tau_{(mf_m(t))\wedge m}$ both converge to $g(u)$. Hence, $g(u) = t$, which gives $u = g^{-1}(t) = f_\infty(t)$. Third, if $t = g(1)$, since f_m is nondecreasing, $f_m(g(1) - \varepsilon) \leq f_m(g(1))$. Hence, by making m tends to infinity, we have $f_\infty(g(1) - \varepsilon) \leq \liminf_m f_m(g(1)) \leq 1$. Now making ε tends to 0, we get $f_m(g(1)) \rightarrow f_\infty(g(1)) = 1$. This concludes the proof.

7.5. Assumptions (Conv-alt) and (FLT) in Bernoulli mixture model

Lemma 7.3. *In the Bernoulli random mixture model with $\pi_0 \in (0, 1)$ and in the multivariate Gaussian setting, we have the following:*

- Assumption (Conv-alt') entails Assumption (Conv-alt);
- Assumption (FLT') (with a rate r_m s.t. $r_m m^{-1/2}$ has a limit in $[0, \infty]$) entails Assumption (FLT).

Proof. Assume that the H_i 's are i.i.d. Then for any continuous bounded function Ψ ,

$$m^{-1} \sum_{i=1}^m H_i \Psi(\mu_i) = \pi_1 m^{-1} \sum_{i=1}^m \Psi(\mu_i) + o_P(1) = \pi_1 \int \Psi d\nu + o_P(1),$$

by using Assumption (Conv-alt). This implies Assumption (Conv-alt') because $m_1(m)/m$ converges almost surely to π_1 .

Let us now turn to the second result and assume (FLT') (we denote the rate by r'_m and the limit by Z') and observe that, defining $r_m = \min(r'_m, \sqrt{m})$,

$$r_m(V_m(t)/m - (m_0/m)t) = (r_m/\sqrt{m}) (\xi_t^m - t\xi_1^m) + \pi_0(r_m/r'_m) r'_m(V'_m(t)/m - t) \quad (33)$$

with

$$\xi_t^m = m^{-1/2} \sum_{i=1}^m (1 - H_i - \pi_0) \mathbf{1} \{ \Phi(Y_i) \leq t \}.$$

Since (FLT') entails (LLNdep) in the Gaussian case, and up to consider a subsequence, we can assume that, a.s.,

$$\sup_{t \in [0,1]} \left| m^{-1} \sum_{i=1}^m \mathbf{1} \{ \Phi(Y_i) \leq t \} - t \right| \rightarrow 0.$$

Next, we merely check that the process $(\xi_t^m)_{t \in [0,1]}$ is a martingale (conditionally on $(Y_i)_i$) with quadratic variation

$$m^{-1} \sum_{i=1}^m (1 - H_i - \pi_0)^2 \mathbf{1} \{ \Phi(Y_i) \leq t \} = \pi_0 \pi_1 m^{-1} \sum_{i=1}^m \mathbf{1} \{ \Phi(Y_i) \leq t \} + o_p(1).$$

Since the latter converges to $\pi_0 \pi_1 t$ (a.s.), Theorem VIII.3.11 in [Jacod and Shiryaev \(2003\)](#) shows that $(\xi_t^m)_{t \in [0,1]}$ converges in distribution in the Skorokhod space to $\sqrt{\pi_0 \pi_1} B_t$ (conditionally on $(Y_i)_i$), where B_t is a Brownian motion. Hence, unconditionally on Y , we have

that the couples $(\xi_t^m - t\xi_1^m, r'_m(V'_m(t)/m - t))_{t \in [a,b]}$ converge in distribution to $(\sqrt{\pi_0\pi_1}(B_t - tB_1), Z'(t))_{t \in [a,b]}$ where the processes B and Z' are independent. Finally, by using (33), we get (FLT) with the rate r_m and a limit $Z(t) = \theta_1\sqrt{\pi_0\pi_1}(B_t - tB_1) + \pi_0\theta_2Z'(t)$ with θ_1, θ_2 two nonnegative numbers such that $\theta_1 + \theta_2 > 0$. \square

Appendix A: Auxiliary results

Lemma A.1. *In the Gaussian setting, conditions (24), (Conv-alt) and (LLNdep) imply (Consist).*

Proof. From (LLNdep) and (24), Proposition 2.1 in Delattre and Roquain (2012) (for instance) ensures that, for all t , $V_m(t)/m$ converges in probability to $\pi_0 t$. Similarly, letting for i with $H_i = 1$, $p'_i = \Phi(\Phi^{-1}(p_i) - \mu_i) = \Phi(Y_i)$. We have that for all t , $m^{-1} \sum_{i=1}^m H_i \mathbf{1}\{p'_i \leq t\}$ converges to $\pi_1 t$. Let us now consider the case of $S_m(t)/m = m^{-1} \sum_{i=1}^m H_i \mathbf{1}\{p'_i \leq \Phi(\Phi^{-1}(t) - \mu_i)\}$. For this, we define the uniform random index $I \sim m^{-1} \sum_{i=1}^m H_i \delta_i$. By (Conv-alt), we have $\mu_I \rightsquigarrow \nu$. Moreover, it is not difficult to see that, up to consider a subsequence, a.s., the joint convergence $(p'_I, \mu_I) \rightsquigarrow U(0, 1) \otimes \nu$ holds. Hence, $S_m(t)/m$ is converging to $\pi_1 F_1(t)$. \square

Lemma A.2. *Let us consider a step-up procedure with critical values $(\tau_k)_{1 \leq k \leq m}$ and a rejection threshold $\tau_{\hat{\ell}}$, where $\hat{\ell}$ is defined by (SU). Consider the function f_m defined by (26) and \tilde{t} be defined by (31). Then the supremum into (31) is a maximum, that is, $\widehat{\mathbb{G}}_m(\tilde{t}) \geq f_m(\tilde{t})$. Furthermore, $\tilde{t} = \tau_{\tilde{k}}$.*

Proof. The first point holds because f_m is left-continuous. Next, since $f_m(t) \leq k/m$ is equivalent to $\tau_k \geq t$, we have

$$\tau_k = \max\{t \in [0, 1] : f_m(t) \leq k/m\}.$$

Hence, $\widehat{\mathbb{G}}_m(\tau_{\tilde{k}}) \geq \tilde{k}/m \geq f_m(\tau_{\tilde{k}})$ and $\tau_{\tilde{k}} \leq \tilde{t}$. Conversely, let us prove $\tilde{t} \leq \tau_{\tilde{k}}$. Since $f_m(\tilde{t}) \leq 1$, we have $\tilde{t} \leq \tau_m$. We can thus define

$$\tilde{k} = \min\{k \in \{0, \dots, m\} : \tilde{t} \leq \tau_k\}.$$

By definition of \tilde{k} , the result is shown if we prove $\widehat{\mathbb{G}}_m(\tau_{\tilde{k}}) \geq \tilde{k}/m$. Assume $\tilde{k} \geq 1$ (otherwise the result is trivial). By definition of \tilde{k} , $\tau_{\tilde{k}-1} < \tilde{t}$, which in turn implies $f_m(\tilde{t}) > (\tilde{k}-1)/m$ (by definition of $\tau_{\tilde{k}-1}$). This gives $\widehat{\mathbb{G}}_m(\tilde{t}) \geq f_m(\tilde{t}) > (\tilde{k}-1)/m$. Hence, $\widehat{\mathbb{G}}_m(\tau_{\tilde{k}}) \geq \widehat{\mathbb{G}}_m(\tilde{t}) \geq \tilde{k}/m$, which concludes the proof. \square

Lemma A.3. *Let $\gamma \in (0, 1)$ and V_m be a process valued in $D([0, 1])$. Assume that for some compact $K \subset [0, 1]$, the following convergence holds*

$$(V_m(t))_{t \in K} \rightsquigarrow (V(t))_{t \in K},$$

where $V \in C(K)$, a.s., and where for all $t \in K$, $V(t)$ as a continuous increasing c.d.f. Let

$$q_m(t) = \min\{x : \mathbb{P}(V_m(t) \leq x) \geq \gamma\}.$$

Then the function sequence q_m converge uniformly on K to the function

$$q(t) = \min\{x : \mathbb{P}(V(t) \leq x) \geq \gamma\}.$$

Proof. Consider an arbitrary sequence $t_m \in K$ tending to $t \in K$ and show that $q_m(t_m)$ tends to $q(t)$. Up to consider a subsequence assume that $q_m(t_m)$ is tending to some q^* . Then we have

$$\gamma \leq \lim_m \mathbb{P}(V_m(t_m) \leq q_m(t_m)) = \mathbb{P}(V(t) \leq q^*),$$

because $V_m(t_m) = \{V_m(t_m) - V_m(t)\} + V_m(t) \xrightarrow{P} V(t)$ and the c.d.f. of $V(t)$ is continuous. Similarly, for all $\varepsilon > 0$,

$$\gamma \geq \lim_m \mathbb{P}(V_m(t_m) \leq q_m(t_m) - \varepsilon) = \mathbb{P}(V(t) \leq q^* - \varepsilon).$$

By making ε decreases to zero, we get $\mathbb{P}(V(t) \leq q^*) = \gamma$ which entails $q^* = q(t)$ because the e.d.f. of $V(t)$ is one to one. □

Acknowledgments

We are very grateful to Yoav Benjamini and Pierre Neuvial for helpful discussions. The second author was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01).

References

- Bardet, J.-M. and Surgailis, D. (2013). Moment bounds and central limit theorems for Gaussian subordinated arrays. *J. Multivariate Anal.*, 114:457–473.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.
- Blanchard, G. and Roquain, E. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871.
- Chi, Z. and Tan, Z. (2008). Positive false discovery proportions: intrinsic bounds and adaptive control. *Statist. Sinica*, 18(3):837–860.
- Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields*, 104(1):15–25.
- Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.*, 117(1):121–142.
- Dehling, H. and Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U -statistics. *Ann. Statist.*, 17(4):1767–1783.
- Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.*, 81(1):111–115.
- Delattre, S. and Roquain, E. (2012). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *ArXiv e-prints*.
- Delattre, S. and Roquain, E. (2013). Supplement to: “On k -FWE-based critical values for controlling the false discovery proportion under dependence”. Technical report.

- Doukhan, P., Lang, G., Surgailis, D., and Teyssière, G., editors (2010). *Dependence in probability and statistics*, volume 200 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Dudoit, S. and van der Laan, M. J. (2008). *Multiple testing procedures with applications to genomics*. Springer Series in Statistics. Springer, New York.
- Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160.
- Fan, J., Han, X., and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association*, 107(499):1019–1035.
- Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Stat. Methods Appl.*, 15(1):43–73.
- Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.*, 34(2):275–297.
- Finner, H., Dickhaus, T., and Roters, M. (2007). Dependency and false discovery rate: asymptotics. *Ann. Statist.*, 35(4):1432–1455.
- Finner, H., Dickhaus, T., and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618.
- Friguet, C., Kloareg, M., and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist. Assoc.*, 104(488):1406–1415.
- Gavrilov, Y., Benjamini, Y., and Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.*, 37(2):619–629.
- Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061.
- Guo, W., Li, H., and Sarkar, S. K. (2013). Further results on controlling the false discovery proportion. *Ann. Statist.* Under revision. Available at <http://web.njit.edu/wguo/research.html>.
- 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:Art. 3, 35 pp. (electronic).
- Hommel, G. and Hoffman, T. (1988). Controlled uncertainty. In *Multiple Hypotheses Testing*, pages 154–161. Springer, Heidelberg.
- Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition.
- Kim, K. I. and van de Wiel, M. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9(1):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(2):379–398.
- Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.*, 33:1138–1154.
- Neuval, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.*, 2:1065–1110.
- Perone Pacifico, M., Genovese, C., Verdinelli, I., and Wasserman, L. (2004). False discovery control for random fields. *J. Amer. Statist. Assoc.*, 99(468):1002–1014.
- Romano, J. P. and Shaikh, A. M. (2006a). On stepdown control of the false discovery proportion. volume 49 of *Lecture Notes-Monograph Series*, pages 33–50.

- Romano, J. P. and Shaikh, A. M. (2006b). Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.*, 34(4):1850–1873.
- Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.*, 35(4):1378–1408.
- Roquain, E. (2011). Type I error rate control for testing many hypotheses: a survey with proofs. *J. Soc. Fr. Stat.*, 152(2):3–38.
- Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, 39(1):584–612.
- Sarkar, S. K. (2008a). On methods controlling the false discovery rate. *Sankhya, Ser. A*, 70:135–168.
- Sarkar, S. K. (2008b). On the Simes inequality and its generalization. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, volume 1 of *Inst. Math. Stat. Collect.*, pages 231–242. Inst. Math. Statist., Beachwood, OH.
- Seeger, P. (1968). A note on a method for the analysis of significances en masse. *Technometrics*, 10(3):586–593.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.
- Soulier, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statist. Probab. Lett.*, 54(2):193–203.
- Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q -value. *Ann. Statist.*, 31(6):2013–2035.
- Troendle, J. F. (2000). Stepwise normal theory multiple test procedures controlling the false discovery rate. *J. Statist. Plann. Inference*, 84(1-2):139–158.