New procedures for discrete tests with proven false discovery rate control

Sebastian Döhler, Guillermo Durand, Etienne Roquain

New procedures for discrete tests with proven false discovery rate control

Sebastian Döhler

Darmstadt University of Applied Sciences
D-64295 Darmstadt, Germany
e-mail: sebastian.doehler@h-da.de

and

Guillermo Durand

Sorbonne Universités, UPMC
4, Place Jussieu, 75252 Paris cedex 05, France
e-mail: guillermo.durand@upmc.fr

and

Etienne Roquain

Sorbonne Universités, UPMC
4, Place Jussieu, 75252 Paris cedex 05, France
e-mail: etienne.roquain@upmc.fr

Abstract: We consider the problem of testing simultaneously many null hypotheses when the test statistics have a discrete distribution. We present new modifications of the Benjamini-Hochberg procedure that incorporate the discrete structure of the data in an appropriate way. These new procedures are theoretically proved to control the false discovery rate (FDR) for any fixed number of null hypotheses. A strong point of our FDR controlling methodology is that it allows to incorporate at once the discreteness and the quantity of signal of the data (so called "π_0-adaptation"). Finally, the power advantage of the new methods is demonstrated by using both numerical experiments and real data sets.

Keywords and phrases: false discovery rate, discrete hypothesis testing, type I error rate control, adaptive procedure, step-up algorithm, step-down algorithm.

1. Introduction

Multiple testing procedures are now routinely used to find significant items in massive and complex data. An important focus has been given to method controlling the false discovery rate (FDR) because this scalable type I error rate "survives" to high dimension. Since the original procedure of Benjamini and Hochberg (1995), much efforts have been undertaken to design FDR controlling procedures that adapt to various underlying structures of the data, as the quantity of signal, the signal strength and the dependencies, among others.

In this work, we deal with adaptation to the discreteness structure, which is faced in various applications where data are collected under the form of counts, as clinical trials, genome-wide association studies (GWAS) or next generation sequencing (NGS). A well known fact is that using discrete test statistics can generate a severe power loss, already at the stage of the single tests. A consequence is that using "blindly" the BH procedure with discrete p-values will control the FDR in a too conservative manner.

In the literature, building multiple testing procedures that take into account the discreteness of the test statistics has a long history, that can be traced back to Tukey and Mantel (1980): some null hypotheses can be *a priori* excluded from the study, because the corresponding tests are unable to produce sufficiently small p-values. This results in a multiplicity reduction that should increase the power. While this idea has been exploited in Tarone (1990) and Westfall and Wolfinger

(1997) for family-wise error rate, an attempt has been made for FDR later in Gilbert (2005). More recently, Heyse (2011) has proposed a more powerful solution, relying on the following averaged cumulative distribution function (c.d.f.):

$$\overline{F}(t) = \frac{1}{m}\sum_{i=1}^{m} F_i(t), \ \ t \in [0,1], \tag{1}$$

where each F_i corresponds to the c.d.f. of the i-th test p-value. To illustrate the potential benefit of using $\overline{F}$, Figure 1 displays this function for the pharmacovigilance data from Heller and Gur (2011) (see section 5 for more details). The critical values of the Heyse procedure can be obtained by inverting $\overline{F}$ at the values $\alpha k/m, 1 \leq k \leq m$. Thus, the smaller the $\overline{F}$-values, the larger the critical values. Here, Heyse critical values improve the BH critical values roughly by a factor 3, thereby yielding a potentially strong rejection enhancement. Furthermore, since the functions F_i's are known, so is $\overline{F}$. Hence, the user has a good prior idea of the improvements reachable by this discrete approach. Unfortunately, the down-side of the Heyse procedure is that it does not rigorously control the FDR in general, as shown by[1] Heller and Gur (2011) (simulations) and Döhler (2016) (formal proof).

Meanwhile, solutions of other natures have been explored by modifying directly the p-values, either by randomization (see Habiger (2015) and references therein), or by shrinking them to build so-called *midP-values* (see Heller and Gur (2011) and references therein). Other work use the discreteness to provide a better estimate of the FDR, see, e.g., Pounds and Cheng (2006), or to make groups and p-value weighting, see Chen and Doerge (2015b).

Overall, although many new procedures have been produced in literature, only few of them are proved to achieve a rigorous FDR control under standard conditions, especially in the finite sample case. To the best of our knowledge, we can only refer to the discretized version of the procedure of Benjamini and Liu (1999) introduced by Heller and Gur (2011) and to the asymptotic work of Ferreira (2007). The aim of this paper is to fill the gap by proposing new procedures that achieve both theoretical validity and good practical performances.

In this paper, we introduce procedures relying on the following modifications of $\overline{F}$ function:

$$\overline{F}_{\mathrm{SU}}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{F_i\,(t)}{1-F_i\,(\tau_m)}; \ \ \overline{F}_{\mathrm{SD}}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{F_i\,(t)}{1-F_i\,(t)}, \ \ t \in [0,1],$$

(an appropriate choice of τ_m is given in section 3.2). Figure 1 displays these functions and show they are very close to the original $\overline{F}$ for small values of t (in a more general manner, note that $\overline{F}(t) \leq \overline{F}_{\mathrm{SD}}(t) \leq \overline{F}(t)/(1-t)$ a soon as $F_i(t) \leq t$ for all i). The main advantages of these modifications is that the corresponding procedures actually provide proven FDR control at the desired level under standard conditions. Furthermore, we show that they have "adaptive" counterparts that are uniform improvements maintaining the FDR control, see Sections 3 and 4. Here, "adaptive" means that the derived critical values can be designed in a way that "implicitly estimate" the overall proportion of true null hypotheses. To explore in detail the performance of these procedures, practical experiments are done in Sections 5 and 6 with true and simulated data.

2. Preliminaries

2.1. General model

Let us observe a random variable X, defined on a probabilistic space and valued in an observation space $(\mathcal{X}, \mathfrak{X})$. We consider a set $\mathcal{P}$ of possible distributions for the distribution of X and we denote the true one by P. We assume that m null hypotheses $H_{0,i}$, $1 \leq i \leq m$, are available

[1] suggestion sd: examples given in Heller and Gur (2011) and Döhler (2016) .

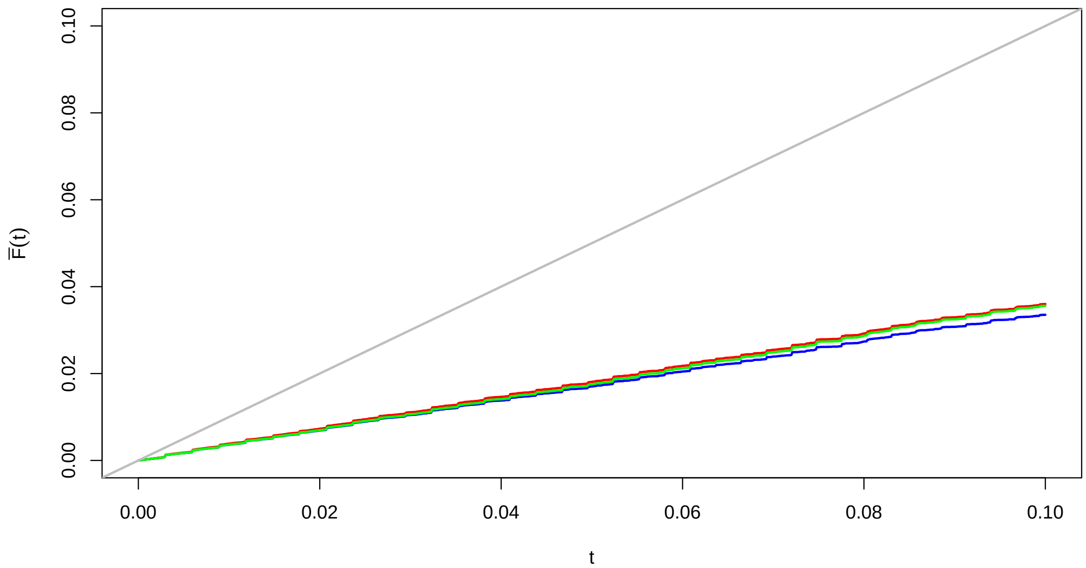

Fig 1. *Plots of variants of $\overline{F}$ for the pharmacovigilance data. The grey line corresponds to the uniform case, the discrete variants are represented by blue (for $\overline{F}$), green (for $\overline{F}_{\mathrm{SD}}$) and red (for $\overline{F}_{\mathrm{SU}}$) lines.*

for P and we denote the corresponding set of true null hypotheses by $\mathcal{H}_0(P) = \{1 \leq i \leq m : H_{0,i}$ is satisfied by $P\}$. We also denote by $m_0(P) = |\mathcal{H}_0(P)|$ the number of true nulls.

We assume that the user has at hand a set of p-values to test each null, that is, a set of random variables $\{p_i(X), 1 \leq i \leq m\}$, valued in $[0,1]$. Throughout the paper, we also make the important (but classical) assumption that the p-values $p_i(X)$, $1 \leq i \leq m$, are mutually independent.

Now, we denote $\mathcal{F} = \{F_i, 1 \leq i \leq m\}$, where for each $i \in \{1, \dots, m\}$, the function

$$F_i(t) = \sup_{P \in \mathcal{P}\,:\, i \in \mathcal{H}_0(P)} \mathbf{P}_{X\sim P}(p_i(X) \leq t),\ \ t \in [0,1],\ \ 1 \leq i \leq m$$

is assumed to be *known*. Note that we necessarily have $F_i(\cdot)$ non decreasing, $F_i(t) \in [0,1]$, $F_i(1) = 1$ and we add the technical condition $F_i(0) = 0$. Loosely, each F_i corresponds to the cumulative distribution of p_i under the null. Above, we have put a supremum to cover the case where the null hypothesis is composite: in that situation, each F_i is adjusted according to the least favorable configuration within the null $H_{0,i}$.

Here are some conditions on $\mathcal{F}$ that will be useful to compare some of the studied procedures (these conditions are not assumed in our results unless explicitly mentioned):

$$F_i(t) \leq t,\ \ t \in [0,1],\ \ 1 \leq i \leq m, \tag{2}$$
$$F_i(t) = t,\ \ t \in [0,1],\ \ 1 \leq i \leq m \tag{3}$$

Condition (2) ensures that the p-values have marginals stochastically lower-bounded by a uniform variable under the null. This is the classical setting which is used in most of the work dealing with FDR controlling theory, see, e.g., Benjamini and Hochberg (1995). Condition (3) is more restrictive : whenever each null hypothesis is a singleton, it implies that the p-values have uniform marginals under the null.

2.2. Discrete and continuous modelling

In this paper, we will assume that we are in one of the two following situations :

- Continuous case: for all $i \in \{1, \dots, m\}$, F_i is continuous. In that case, we let $\mathcal{A}_i = [0, 1]$, $1 \le i \le m$ and $\mathcal{A} = \cup_{i=1}^m \mathcal{A}_i = [0, 1]$.
- Discrete case: each p-value p_i (both under the null and alternative) takes values in some finite set $\mathcal{A}_i = \{a_{i,k}, 0 \le k \le K_i\}$, where $K_i \ge 0$ and $(a_{i,k})_{0 \le k \le K_i} \in [0, 1]^{K_i+1}$ is an increasing sequence (with $a_{i,0} = 0$, $a_{i,K_i} = 1$). We denote $\mathcal{A} = \cup_{i=1}^m \mathcal{A}_i$ the overall p-value support.

The continuous setting typically comes for situations where the p-values are calibrated from test statistics having a continuous distribution under the null. In this situation, (3) is often satisfied. The discrete setting typically arises in situations where the p-values are calibrated from test statistics having a finitely supported distribution under the null. In this situation, (3) is typically violated and we relax it under the weaker form:

$$F_i(t) = t, \ \ t \in \mathcal{A}_i, \ \ 1 \le i \le m. \tag{4}$$

In the discrete framework, let us underline that while (4) will typically hold, the equality $F_i(t) = t$, $t \in \mathcal{A}$ will fail in general because $\mathcal{A}$ contains points of $\mathcal{A}_j$ for $j \ne i$. For the function $\overline{F}$ defined by (1), this entails that $\overline{F}(t)$ will be smaller than t in general (see Figure 1), which is exactly the property that we want to exploit in this paper.

To illustrate the above framework, we provide below two simple examples (for more advanced examples, see for instance Chen and Doerge (2015b)).

Example 2.1 (Gaussian testing). Observe $X = (X_i)_{1 \le i \le m}$ with independent coordinates and marginals $X_i \sim \mathcal{N}(\mu_i, 1)$, $\mu_i \in \mathbb{R}$ is the parameter of interest, $1 \le i \le m$. In that situation, a possible hypothesis testing problem is to consider the nulls $H_{0,i}$: "$\mu_i \le 0$" against $H_{1,i}$: "$\mu_i > 0$". Then $p_i(X) = 1 - \Phi(X_i)$, $1 \le i \le m$, is a family of p-values satisfying (3) (where Φ denotes the c.d.f. of a standard Gaussian variable).

Example 2.2 (Binomial testing). Observe $X = (X_i)_{1 \le i \le m}$ with independent coordinates and marginals $X_i \sim \mathcal{B}(n_i, \theta_i)$, where $n_i \ge 1$ is known and $\theta_i \in (0, 1)$ is the parameter of interest, $1 \le i \le m$. In that situation, a possible hypothesis testing problem is to consider the nulls $H_{0,i}$: "$\theta_i \le 1/2$" against $H_{1,i}$: "$\theta_i > 1/2$". Then $p_i(X) = T_i(X_i)$, $1 \le i \le m$, define a family of p-values where $T_i(x) = 2^{-n_i} \sum_{j=0}^{x} \binom{n_i}{j}$ is the upper-tail distribution function of a binomial distribution of parameters $(n_i, 1/2)$. The support of the p-values under the null and alternative is covered by letting $K_i = n_i + 1$ and $a_{i,k} = 2^{-n_i} \sum_{j=0}^{k-1} \binom{n_i}{j}$, $1 \le k \le K_i$. We merely check in that case that (3) is violated while (2) and (4) hold.

2.3. Step-wise procedures

First define a critical value sequence as any nondecreasing sequence $\tau = (\tau_k)_{1 \le k \le m} \in [0, 1]^m$ (with $\tau_0 = 0$ by convention).

The *step-up procedure* of critical value sequence τ, denoted by $\mathbf{SU}(\tau)$, rejects the i-th hypothesis if $p_i \le \tau_{\widehat{k}}$, with

$$\widehat{k} = \max\{k \in \{0, 1, ..., m\} \ : \ p_{(k)} \le \tau_k\},$$

where $p_{(1)} \le p_{(2)} \le ... \le p_{(m)}$ denote the ordered p-values (with the convention $p_{(0)} = 0$).

The *step-down procedure* of critical value sequence τ, denoted by $\mathbf{SD}(\mathbf{t})$, rejects the i-th hypothesis if $p_i \le \tau_{\widetilde{k}}$, with

$$\widetilde{k} = \max\{k \in \{0, 1, ..., m\} \ : \ \forall k' \le k, \ p_{(k')} \le \tau_{k'}\}.$$

It is straightforward to check that, for the same set of critical values, the step-up version rejects always more hypotheses than the step-down version. More comments and illustrations on step-wise procedures can be found in Blanchard et al. (2014) and Dickhaus (2014), among others.

2.4. False discovery rate

We measure the quantity of false positives of a step-up (resp. step-down) procedure by using the false discovery rate (FDR), introduced and popularized by Benjamini and Hochberg (1995), which is defined as the averaged proportion of errors among the rejected hypotheses. More formally, for some procedure R rejecting the i-th hypothesis if $p_i \leq \hat{t}(X)$ (for some threshold $\hat{t}(X)$),

$$\mathrm{FDR}(R, P) = \mathbf{E}\left[\frac{\sum_{i\in\mathcal{H}_0(P)} \mathbf{1}\{p_i \leq \hat{t}(X)\}}{1 \vee \sum_{i=1}^m \mathbf{1}\{p_i \leq \hat{t}(X)\}}\right], \ P \in \mathcal{P}. \tag{5}$$

The aim of this work is to propose procedures that control the FDR at a prescribed level α and that incorporate the knowledge of the F_i's in a way that increases the number of discoveries.

3. Methods

3.1. Existing methods

Here are some existing procedures, that we use as benchmarks in our analysis.

- [BH]: the seminal procedure proposed in Benjamini and Hochberg (1995), corresponding to the step-up procedure $\mathbf{SU}(\tau)$, with critical values $\tau_k = \alpha k/m$, $1 \leq k \leq m$;
- [BR-λ]: an adaptive version of BH procedure that was proposed in Blanchard and Roquain (2009), corresponding to the step-up procedure $\mathbf{SU}(\tau)$, with critical values

$$\tau_k = \left((1-\lambda)\frac{\alpha k}{m-k+1}\right) \wedge \lambda, \ 1 \leq k \leq m; \tag{6}$$

- [GBS]: an adaptive version of BH procedure that has been proposed in Gavrilov et al. (2009), corresponding to the step-down procedure $\mathbf{SD}(\tau)$, with critical values

$$\tau_k = \frac{\alpha k}{m-(1-\alpha)k+1}, \ 1 \leq k \leq m; \tag{7}$$

- [Heyse]: the step-up procedure $\mathbf{SU}(\tau)$ using critical values given by

$$\tau_k = \max\{t \in \mathcal{A} \,:\, \overline{F}(t) \leq \alpha k/m\}, \ 1 \leq k \leq m; \tag{8}$$

  where $\overline{F}$ is defined by (1). This procedure was proposed in Heyse (2011).

The rationale behind the critical values of [BR-λ] and [GBS] is that they are intended to mimic the oracle critical values $\tau_k = \alpha k/m_0(P)$, $1 \leq k \leq m$, which are less conservative than those of [BH] when $m_0(P)/m$ is not close to 1, see, e.g., Benjamini et al. (2006); Blanchard and Roquain (2009) for more details on adaptive procedures.

Let us now comment [Heyse]. First, in the continuous setting where (2) holds, $\overline{F}(t) \leq t$, $t \in [0,1]$, and thus the critical values given by (8) satisfy $\tau_k \geq \alpha k/m$, $1 \leq k \leq m$, which means that [Heyse] is always less conservative (or equal) than [BH]. When (3) additionally holds, $\overline{F}(t) = t$, $t \in [0,1]$, and the two critical value sequences are the same. Second, in the discrete setting where (2) holds, $\mathcal{A}$ is finite and τ_k is not provided anymore to be above $\alpha k/m$. However, [Heyse] is also less conservative (or equal) than [BH] in this case, as stated in the following result (proved in Section A).

Lemma 3.1. *Consider the model of Section 2.1 assuming (2), both in the continuous and discrete setting described in Section 2.2. Then the set of nulls rejected by [Heyse] is larger than the one of [BH] (almost surely). Furthermore, under (4), these two rejection sets are equal (almost surely) if $F_i = F_j$ for all $i \neq j$.*

The equality case of Lemma 3.1 was provided in Proposition 2.3 of Heller and Gur (2011), who presented it as a limitation of Heyse procedure in the discrete case. However, we advocate that the condition $F_i = F_j$ for all $i \neq j$ is a somehow extreme configuration which is rarely met in practice (in the discrete case). More typically, the F_i's have an heterogeneous structure implying that $\overline{F}(t)$ is smaller than t (see Figure 1). This entails that [Heyse] can substantially improve [BH] (see Figure 2).

Here, our point is that, while [Heyse] incorporates the knowledge of the F_i's in an appropriate way (see also Remark 3.2 below), it is not correctly well calibrated for a rigorous FDR control: as shown in Heller and Gur (2011); Döhler (2016), it fails to control the FDR in general. We propose suitable modifications of [Heyse] in the next sections.

Remark 3.2 (Empirical Bayes point of view on the Heyse procedure). We note here that [Heyse] corresponds to a suitable empirical Bayes procedure. Let us consider the "binomial example" of Section 2.2, but assume now that the counts $n_1, \dots, n_m$ are observed from a sample $N_1, \dots, N_m$ i.i.d. of a priori distribution ν. Unconditionally, the p-values p_i, $i \in \mathcal{H}_0$, are thus i.i.d. with c.d.f. $\bar{F}_0 = \sum_{n\geq 0} \nu(\{n\}) F_{0,n}$, where $F_{0,n}$ is the c.d.f. jumping at each $x_{k,n} = 2^{-n}\sum_{j=0}^{k-1}\binom{n}{j}$ with $F_{0,n}(x_{k,n}) = x_{k,n}$, $1 \leq k \leq n+1$. This suggests to choose the step-up procedure with critical values $\tau_k = \max\{t \,:\, \bar{F}_0(t) \leq \alpha k/m\}$. Following an empirical Bayes approach, the prior ν can be estimated by $\hat{\nu}(\{n\}) = m^{-1}\sum_{i=1}^m \mathbf{1}_{\{N_i=n\}}$, which gives rise to the estimator of $\bar{F}_0$ given by $\hat{\bar{F}}_0 = \sum_{n\geq 0}\hat{\nu}(\{n\})F_{0,n} = m^{-1}\sum_{i=1}^m F_{0,N_i}$, which is equal to $\overline{F}$ given by (1). Hence, the corresponding (empirical Bayes) step-up procedure reduces to [Heyse].

3.2. Two new methods

We now present two procedures that aim at correcting [Heyse] :

- [DBH-SU]: the step-up procedure $\mathbf{SU}(\tau)$ using the critical values defined in the following way

$$\begin{aligned}\tau_m &= \max\left\{t \in \mathcal{A} \,:\, \frac{1}{m}\sum_{i=1}^m \frac{F_i(t)}{1-F_i(t)} \leq \alpha\right\}\\ \tau_k &= \max\left\{t \in \mathcal{A} \,:\, t \leq \tau_m,\, \frac{1}{m}\sum_{i=1}^m \frac{F_i(t)}{1-F_i(\tau_m)} \leq \alpha k/m\right\},\ 1 \leq k \leq m-1.\end{aligned} \tag{9}$$

- [DBH-SD]: the step-down procedure $\mathbf{SD}(\tau)$ using the critical values defined in the following way :

$$\tau_k = \max\left\{t \in \mathcal{A} \,:\, \frac{1}{m}\sum_{i=1}^m \frac{F_i(t)}{1-F_i(t)} \leq \alpha k/m\right\},\ 1 \leq k \leq m. \tag{10}$$

[DBH-SU] can be seen as a correction of [Heyse]: the correction term in the critical values (9) lies in the additional denominator $1 - F_i(\tau_m)$. A consequence is that [DBH-SU] can be more conservative than [BH]. However, the amplitude of this phenomenon is always light, as the next lemma shows (proved in Section A).

Lemma 3.3. *Under the conditions of Lemma 3.1, the set of nulls rejected by [DBH-SU] is larger than the one of [BH] taken at level $\alpha/(1+\alpha)$ (almost surely).*

At this point, it is useful to ask whether we can build a procedure that incorporate the F_i's while being a uniform improvement of [BH]. We have found a solution (under some mild conditions), called [RBH], but the improvement brought by the F_i's information is less substantial than for [DBH-SU], so we have chosen to not report [RBH] in the main stream of the paper. We refer the reader to Appendix B.1 for more details.

Now, we briefly discuss [DBH-SD]. The following result can be established:

Lemma 3.4. *Under the conditions of Lemma 3.1, the set of nulls rejected by [DBH-SD] is (almost surely) larger than the one of the step-down procedure with critical values* $(\alpha k/m)/(1+\alpha k/m)$*,* $1 \leq k \leq m$*.*

Finally note that [DBH-SD] uses critical values always larger than those of [DBH-SU], but uses a step-down algorithm instead of a step-up algorithm. Hence, [DBH-SD] and [DBH-SU] are not comparable in general.

3.3. Adaptive versions

In this section, we define adaptive versions of [DBH-SU] and [DBH-SD] in the following way:

- [A-DBH-SU]: the step-up procedure $\mathbf{SU}(\tau)$ using the critical values defined in the following way

$$\tau_m = \max\left\{t \in \mathcal{A} \,:\, m^{-1}\sum_{i=1}^{m} \frac{F_i\left(t\right)}{1-F_i\left(t\right)} \leq \alpha\right\}$$
$$\tau_k = \max\left\{t \in \mathcal{A} \,:\, t \leq \tau_m,\, \left(\frac{F\left(t\right)}{1-F\left(\tau_m\right)}\right)_{(1)} + \cdots + \left(\frac{F\left(t\right)}{1-F\left(\tau_m\right)}\right)_{(m-k+1)} \leq \alpha k\right\},\ 1 \leq k \leq m-1, \tag{11}$$

  where each $\left(\frac{F(t)}{1-F(\tau_m)}\right)_{(j)}$ denotes the j-th largest elements of the set $\left\{\frac{F_i(t)}{1-F_i(\tau_m)}, 1 \leq i \leq m\right\}$.
- [A-DBH-SD]: the step-down procedure $\mathbf{SD}(\tau)$ using the critical values defined in the following way :

$$\tau_k = \max\left\{t \in \mathcal{A} \,:\, \left(\frac{F\left(t\right)}{1-F\left(t\right)}\right)_{(1)} + \cdots + \left(\frac{F\left(t\right)}{1-F\left(t\right)}\right)_{(m-k+1)} \leq \alpha k\right\},\ 1 \leq k \leq m, \tag{12}$$

  where each $\left(\frac{F(t)}{1-F(t)}\right)_{(j)}$ denotes the j-th largest elements of the set $\left\{\frac{F_i(t)}{1-F_i(t)}, 1 \leq i \leq m\right\}$.

Note that the critical values of [A-DBH-SU] and [A-DBH-SD] are clearly larger than or equal to those of their non-adaptive counterparts [DBH-SU] and [DBH-SD], respectively. This means that using adaptive versions is always less conservative. The only cost is computational: deriving (11) and (12) can be time-consuming because the ordering should be done for each value of t.

Additionally, the following result holds (proved in Section A).

Lemma 3.5. *Under the conditions of Lemma 3.1, the following holds:*

(i) the set of nulls rejected by [A-DBH-SU] is larger than the one of [BR-λ] (almost surely), where λ is taken equal to the largest critical values of [A-DBH-SU] (or [DBH-SU]);
(ii) the set of nulls rejected by [A-DBH-SD] is larger than the one of [GBS] (almost surely);

The above lemma ensures that the user can incorporate the knowledge of the F_i's in adaptive procedures with a "no loss" guarantee with respect to [BR] and [GBS]. This is a somehow striking fact, coming loosely from a "fortunate marriage" between the proof technics of discreteness theory and adaptation theory.

4. FDR controlling results

In this section, we provide results showing that the new proposed methods appropriately control the FDR. The proofs are all deferred to Appendix C.

Theorem 4.1 (Step-up). *In the model of section 2.1, consider any critical values τ_k, $1 \leq k \leq m$, satisfying*

$$\min\left(\sum_{i=1}^{m} \max_{1\leq k\leq m} \frac{F_i(\tau_k)}{k}, \max_{1\leq k\leq m} \max_{\substack{X\subset\{1,\ldots,m\}\\|X|=m-k+1}} \left(\frac{1}{k}\sum_{i\in X}\frac{F_i\left(\tau_k\right)}{1-F_i\left(\tau_m\right)}\right)\right) \leq \alpha,$$

Then, we have for all $P \in \mathcal{P}$, $\mathrm{FDR}(\mathbf{SU}(\tau), P) \leq \alpha$.

Theorem 4.1 generalizes the original result of Benjamini and Hochberg (1995) and Theorem 9 of Blanchard and Roquain (2009) (for the choice $\lambda = \alpha/(1+\alpha)$).

Theorem 4.2 (Step-down). *In the model of section 2.1, consider any critical values τ_k, $1 \leq k \leq m$, satisfying*

$$\max_{1\leq k\leq m} \max_{\substack{X\subset\{1,\ldots,m\}\\|X|=m-k+1}} \left(\frac{1}{k}\sum_{i\in X}\frac{F_i\left(\tau_k\right)}{1-F_i\left(\tau_k\right)}\right) \leq \alpha.$$

Then, we have for all $P \in \mathcal{P}$, $\mathrm{FDR}(\mathbf{SD}(\tau), P) \leq \alpha$.

Theorem 4.2 extends Theorem 1.1 of Gavrilov et al. (2009). Also, our proof is much simpler than the original proof of Gavrilov et al. (2009), see Appendix C.2.

Corollary 4.3. *In the model of section 2.1, the procedures [DBH-SU]; [DBH-SD]; [A-DBH-SU]; [A-DBH-SD] all control the FDR at level α.*

5. Empirical data

To illustrate the performance of FDR-controlling procedures for discrete data, we analyse two benchmark data sets which have also been used in previous publications. In what follows, our main goal is to compare the performance of the new procedures [DBH-SU], [A-DBH-SU] and [A-DBH-SD] to the classical [BH] procedure. As a further benchmark we also include [Heyse] in the analysis. All analyses were performed using the R language for statistical computing (R Core Team, 2016).

5.1. *Pharmacovigilance data*

This data set is derived from a database for reporting, investigating and monitoring adverse drug reactions due to the Medicines and Healthcare products Regulatory Agency in the United Kingdom. It contains the number of reported cases of amnesia as well as the total number of adverse events reported for each of the $m = 2446$ drugs in the database. For more details we refer to Heller and Gur (2011) and to the accompanying R-package 'discreteMTP' (Heller et al., 2012), which also contains the data. Heller and Gur (2011) investigate the association between reports of amnesia and suspected drugs by performing for each drug a Fisher's exact test (one-sided) for testing association between the drug and amnesia while adjusting for multiplicity by using several (discrete) FDR procedures.

5.2. *Next generation sequencing data*

We also revisit the next generation sequencing (NGS) count data analysed by Chen and Doerge (2015b), to which we also refer for more details. More specifically, we reanalyse the methylation data set for cytosines of Arabidopsis in Lister et al. (2008) which is part of the R-package 'fdrDiscreteNull' (Chen and Doerge, 2015a). This data set contains the counts for a biological entity under two different biological conditions or treatments. Following Chen and Doerge (2015b), $m = 7421$ genes whose treatment-wise total counts are positive but row-total counts are no greater than 100 are analysed using the exact binomial test, see Chen and Doerge (2015b).

5.3. Results

Table 1 summarizes the number of discoveries for the pharmacovigilance and NGS data when using the respective FDR procedures at level $\alpha = 0.05$.

TABLE 1
Number of rejections (discoveries) for the pharmacovigilance and Arabidopsis methylation data.

Data set	[BH]	[DBH-SU]	[Heyse]	[A-DBH-SU]	[A-DBH-SD]
Pharmacovigilance	24	27	27	27	27
Arabidopsis methylation	2097	2358	2379	2446	2453

Compared to the classical FDR controlling procedures, the new procedures are able to detect three additional candidates linking amnesia and drugs in the pharmacovigilance data. Note also that for this data, they reject the same number of hypotheses as [Heyse], even though [Heyse] is not correctly calibrated for FDR control. For the Arabidopsis data, the new procedures improve considerably on [BH]. Moreover, there is a clear separation between the adaptive and non-adaptive procedures.

Figure 2 illustrates graphically the data and the critical constants of the involved multiple testing procedures. In particular, the benefit of taking discreteness into account becomes more

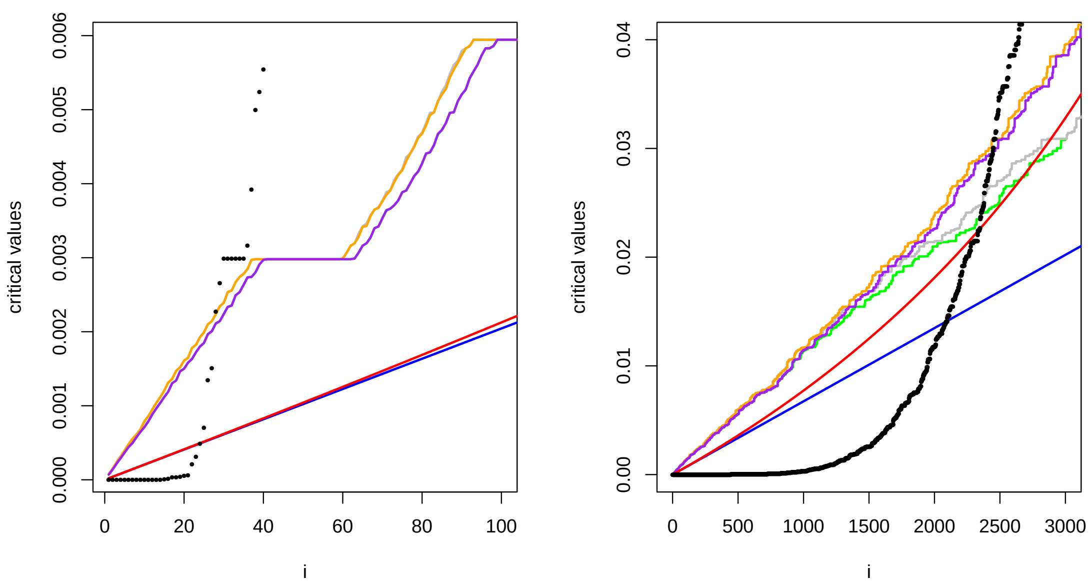

FIG 2. *Critical constants and sorted p-values (represented by black dots) for the pharmacovigilance (left panel) and Arabidopsis methylation data (right panel). The [BH], [GBS], [DBH-SU], [A-DBH-SU], [A-DBH-SD] and [Heyse] critical constants are represented respectively by blue, red, green, purple, orange and grey solid lines.*

apparent: for the pharmacovigilance data, the discrete critical values are considerably (by a factor of $2.5 - 3.5$) larger than their respective classical counterparts. This leads to more powerful procedures. For the NGS data, we can observe quite clearly that the [DBH-SU] critical constants are dominated by the [A-DBH-SU] constants, as explained in section 3. This leads to roughly 100 additional rejections. Again, the discrete critical values are considerably larger than their respective classical counterparts. In 3.2 we mentioned that the correction factor $1 - F_i(\tau_m)$, introduced for guaranteeing FDR control of [DBH-SU], may lead to a procedure which is more conservative than [BH]. However, Figure 2 shows that – at least for the data sets considered here – this risk is by far compensated by the benefit of taking discreteness adequately into account.

6. Simulation study

We now investigate the power of the procedures from the previous section in a simulation study similar to those described in Gilbert (2005), Heller and Gur (2011) and Döhler (2016). Again, we focus on comparing the performance of the new discrete procedures to [BH].

6.1. Simulated Scenarios

We simulate a two-sample problem in which a vector of m independent binary responses ("adverse events") is observed for each subject in two groups, where each group consists of $N = 25$ subjects. Then, the goal is to simultaneously test the m null hypotheses H_{0i} : "$p_{1i} = p_{2i}$", $i = 1, \ldots, m$, where p_{1i} and p_{2i} are the success probabilities for the ith binary response in group 1 and 2, respectively. We take $m = 800, 2000$ where $m = m_1 + m_2 + m_3$ and data are generated so that the response is $Bernoulli(0.01)$ at m_1 positions for both groups, $Bernoulli(0.10)$ at m_2 positions for both groups and $Bernoulli(0.10)$ at m_3 positions for group 1 and $Bernoulli(q)$ at m_3 positions for group 2 where $q = 0.15, 0.25, 0.4$ represents weak, moderate and strong effects respectively. The null hypothesis is true for the m_1 and m_2 positions while the alternative hypothesis is true for the m_3 positions. We also take different configurations for the proportion of false null hypotheses, m_3 is set to be 10%, 30% and 80% of the value of m, which represents small, intermediate and large proportion of effects (the proportion of true nulls π_0 is 0.9, 0.7, 0.2, respectively). Then, m_1 is set to be 20%, 50% and 80% of the number of true nulls (that is, $m - m_3$) and m_2 is taken accordingly as $m - m_1 - m_3$.

For each of the 54 possible parameter configurations specified by m, m_3, m_1 and q, 10000 Monte Carlo trials are performed, that is, 10000 data sets are generated and for each data set, an unadjusted two-sided p-value from Fisher's exact test is computed for each of the m positions, and the multiple testing procedures mentioned above are applied at level $\alpha = 0.05$. The power of each procedure was estimated as the fraction of the m_3 false null hypotheses that were rejected, averaged over the 10000 simulations. For random number generation the R-function *rbinom* was used. The two-sided p-values from Fisher's exact test were computed using the R-function *fisher.test*.

6.2. Results

Table 2 displays the (average) power of the five procedures under investigation. For weak and moderate effects, i.e. $q = 0.15$ and $q = 0.25$, none of the procedure possesses relevant power. For strong effects, the results are summarized in Figure 3. (Since the power of the discrete procedures is slightly increasing in m_1 for fixed m_3 and q, we present – in order to avoid over-optimism – the configuration with smallest m_1).

The results are consistent with the findings of the previous section: The new discrete procedures are considerably more powerful than their classical counterparts and perform roughly similarly for small and intermediate proportions of alternatives. When the proportion of alternatives is large, the benefit of using adaptive procedures – especially the [A-DBH-SU] procedure – is clearly visible in Figure 3.

Loosely, this experiment supports the fact the new proposed procedures have power performance of the "same order" as [Heyse]. Their advantage is thus that they have an additional theoretical guarantee.

Acknowledgements

This work has been supported by the French grant ANR-16-CE40-0019.

Table 2
Average power of FDR procedures for $N = 25$ (see text).

m	m_3	m_1	q	[BH]	[Heyse]	[DBH-SU]	[A-DBH-SU]	[A-DBH-SD]
800	80	144	0.15	0.0000	0.0004	0.0003	0.0003	0.0004
		144	0.25	0.0004	0.0197	0.0177	0.0177	0.0135
		144	0.4	0.0803	0.4425	0.4247	0.4247	0.4130
		360	0.15	0.0000	0.0007	0.0006	0.0006	0.0007
		360	0.25	0.0004	0.0244	0.0209	0.0209	0.0153
		360	0.4	0.0803	0.4529	0.4509	0.4509	0.4487
		576	0.15	0.0000	0.0009	0.0007	0.0007	0.0008
		576	0.25	0.0004	0.0343	0.0259	0.0259	0.0231
		576	0.4	0.0803	0.5367	0.4741	0.4741	0.4999
	240	112	0.15	0.0000	0.0003	0.0003	0.0003	0.0002
		112	0.25	0.0005	0.0276	0.0249	0.0249	0.0157
		112	0.4	0.2148	0.5365	0.5012	0.5012	0.4951
		280	0.15	0.0000	0.0003	0.0003	0.0003	0.0002
		280	0.25	0.0005	0.0315	0.0272	0.0272	0.0175
		280	0.4	0.2147	0.5758	0.5536	0.5536	0.5495
		448	0.15	0.0000	0.0005	0.0003	0.0003	0.0004
		448	0.25	0.0005	0.0372	0.0308	0.0308	0.0207
		448	0.4	0.2145	0.5920	0.5741	0.5741	0.5775
	640	32	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		32	0.25	0.0010	0.0378	0.0341	0.0341	0.0174
		32	0.4	0.4243	0.6174	0.5955	0.6828	0.6621
		80	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		80	0.25	0.0010	0.0388	0.0347	0.0347	0.0179
		80	0.4	0.4242	0.6282	0.6128	0.6841	0.6638
		128	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		128	0.25	0.0010	0.0400	0.0354	0.0354	0.0183
		128	0.4	0.4240	0.6353	0.6265	0.6854	0.6656
2000	200	360	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		360	0.25	0.0001	0.0156	0.0142	0.0142	0.0100
		360	0.4	0.0730	0.4486	0.4317	0.4317	0.4197
		900	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		900	0.25	0.0001	0.0192	0.0166	0.0166	0.0125
		900	0.4	0.0730	0.4517	0.4511	0.4511	0.4509
		1440	0.15	0.0000	0.0003	0.0002	0.0002	0.0002
		1440	0.25	0.0001	0.0286	0.0211	0.0211	0.0165
		1440	0.4	0.0730	0.5402	0.4684	0.4684	0.4984
	600	280	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		280	0.25	0.0001	0.0239	0.0213	0.0213	0.0115
		280	0.4	0.2058	0.5350	0.4988	0.4988	0.4960
		700	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		700	0.25	0.0001	0.0290	0.0239	0.0239	0.0132
		700	0.4	0.2058	0.5750	0.5590	0.5590	0.5516
		1120	0.15	0.0000	0.0002	0.0002	0.0002	0.0002
		1120	0.25	0.0001	0.0350	0.0283	0.0283	0.0157
		1120	0.4	0.2057	0.5908	0.5739	0.5739	0.5761
	1600	80	0.15	0.0000	0.0001	0.0001	0.0001	0.0000
		80	0.25	0.0003	0.0379	0.0342	0.0342	0.0126
		80	0.4	0.4223	0.6196	0.5928	0.6860	0.6591
		200	0.15	0.0000	0.0001	0.0001	0.0001	0.0000
		200	0.25	0.0003	0.0387	0.0351	0.0351	0.0131
		200	0.4	0.4222	0.6281	0.6152	0.6869	0.6602
		320	0.15	0.0000	0.0001	0.0001	0.0001	0.0000
		320	0.25	0.0003	0.0396	0.0360	0.0360	0.0137
		320	0.4	0.4220	0.6327	0.6278	0.6877	0.6617

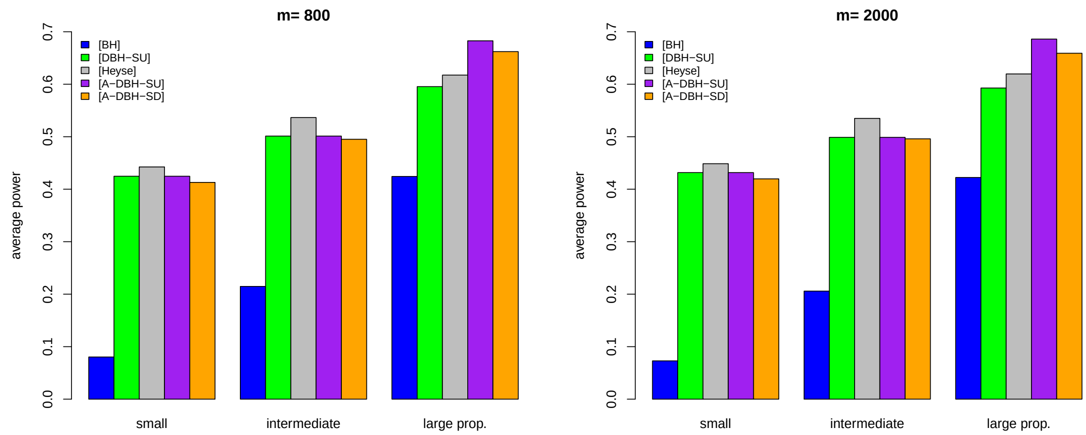

Fig 3. *Average power for the [BH] and discrete procedures in the simulation study. The left panel presents results for* $m = 800$*, the right panel for* $m = 2000$.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series 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 Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *J. Statist. Plann. Inference*, 82(1-2):163–170.

Blanchard, G., Dickhaus, T., Roquain, E., and Villers, F. (2014). On least favorable configurations for step-up-down tests. *Statist. Sinica*, 24(1):1–23.

Blanchard, G. and Roquain, E. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871.

Chen, X. and Doerge, R. (2015a). *fdrDiscreteNull: False Discovery Rate Procedure Under Discrete Null Distributions*. R package version 1.0.

Chen, X. and Doerge, R. (2015b). A weighted fdr procedure under discrete and heterogeneous null distributions. *arXiv:1502.00973*.

Dickhaus, T. (2014). *Simultaneous statistical inference*. Springer, Heidelberg. With applications in the life sciences.

Döhler, S. (2016). A discrete modification of the Benjamini-Yekutieli procedure. *ArXiv e-prints*.

Ferreira, J. A. (2007). The Benjamini-Hochberg method in the case of discrete test statistics. *Int. J. Biostat.*, 3:Art. 11, 18.

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.

Gilbert, P. (2005). A modified false discovery rate multiple-comparisons procedure for discrete data, applied to human immunodeficiency virus genetics. *Journal of the Royal Statistical Society. Series C*, 54(1):143–158.

Habiger, J. D. (2015). Multiple test functions and adjusted p-values for test statistics with discrete distributions. *J. Statist. Plann. Inference*, 167:1–13.

Heller, R. and Gur, H. (2011). False discovery rate controlling procedures for discrete tests. *ArXiv e-prints*.

Heller, R., Gur, H., and Yaacoby, S. (2012). *discreteMTP: Multiple testing procedures for discrete test statistics*. R package version 0.1-2.

Heyse, J. F. (2011). A false discovery rate procedure for categorical data. In *Recent Advances in*

Bio- statistics: False Discovery Rates, Survival Analysis, and Related Topics, pages 43–58.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, 133(3):523–536.

Mantel, N. (1980). A biometrics invited paper. assessing laboratory evidence for neoplastic activity. *Biometrics*, 36(3):381–399.

Pounds, S. and Cheng, C. (2006). Robust estimation of the false discovery rate. *Bioinformatics*, 22(16):1979–1987.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Tarone, R. E. (1990). A modified bonferroni method for discrete data. *Biometrics*, 46(2):515–522.

Westfall, P. and Wolfinger, R. (1997). Multiple tests with discrete distributions. *The American Statistician*, 51(1):3–8.

Appendix A: Proofs for lemmas comparing procedures

The lemmas presented here rely on the fact that, there is almost surely no p-value in $[0,1]\backslash\mathcal{A}$ (both in the continuous and discrete cases). All symbols "$=$" or "$\subset$" are intended to be valid almost surely in this section.

A result which will be extensively used in the proofs of this section is the following one : for p-values valued in the set $\mathcal{A}$, then the step-up procedure with critical values τ_k, $1 \leq k \leq m$, has the same rejection set as the step-up procedure with critical values $\xi_k = \max\{t \in \mathcal{A} \,:\, t \leq \tau_k\}$, $1 \leq k \leq m$. This fact comes from the simple following observation : for all k,

$$\begin{aligned}\{1 \leq i \leq m \,:\, p_i \leq \tau_k\} &= \{1 \leq i \leq m \,:\, p_i \in \mathcal{A},\, p_i \leq \tau_k\} \\ &= \{1 \leq i \leq m \,:\, p_i \in \mathcal{A},\, p_i \leq \xi_k\} = \{1 \leq i \leq m \,:\, p_i \leq \xi_k\}.\end{aligned}$$

The ξ_k's are called the "effective" critical values of $\mathbf{SD}(\tau)$ or $\mathbf{SU}(\tau)$ in the sequel.

A.1. Proof of Lemma 3.1

The effective critical values of the BH procedure are given by $\xi_k = \max\{t \in \mathcal{A} \,:\, t \leq \alpha k/m\}$, $1 \leq k \leq m$. If (2) holds, then $\overline{F}(t) \leq t$ and each ξ_k is clearly smaller than the k-th critical values of [Heyse]. This implies that the rejection set of [Heyse] is larger than the one of [BH]. Conversely, under (4) and if $F_i = F_j = \overline{F}$ for all $i \neq j$, we always have $\overline{F}(t) = F_i(t) = t$ for $t \in \mathcal{A}$. This implies that the ξ_k's are the critical values of [Heyse] and shows the reversed inclusion.

A.2. Proof of Lemma 3.3

Let τ_k, $1 \leq k \leq m$, be the critical values of [DBH-SU]. Let $\xi_k = \max\left\{t \in \mathcal{A} \,:\, t \leq \frac{\alpha}{1+\alpha}\frac{k}{m}\right\}$ be the effective critical values of the [BH] procedure at level $\alpha/(1+\alpha)$. Now, for all $t \in [0,1]$, we have by (2),

$$\overline{F}_{\mathrm{SU}}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1-F_i(\tau_m)} \leq \frac{t}{m}\sum_{i=1}^{m}\frac{1}{1-F_i(\tau_m)} = t\cdot(1+\overline{F}_{\mathrm{SU}}(\tau_m)) \leq t\cdot(1+\alpha), \tag{13}$$

where the last inequality follows from the definition of τ_m. Thus we have $\overline{F}_{\mathrm{SU}}(\xi_m) \leq \alpha$, which in turn implies $\xi_m \leq \tau_m$. Additionally, the bound (13) yields for $1 \leq k < m$

$$\begin{aligned}\tau_k &= \max\left\{t \in \mathcal{A} \,:\, t \leq \tau_m,\, \overline{F}_{\mathrm{SU}}(t) \leq \alpha k/m\right\} \\ &\geq \max\{t \in \mathcal{A} \,:\, t \leq \tau_m,\, t(1+\alpha) \leq \alpha k/m\} \\ &= \max\{t \in \mathcal{A} \,:\, t(1+\alpha) \leq \alpha k/m\} \\ &= \xi_k,\end{aligned}$$

where we used that $\xi_m \leq \tau_m$. The result follows.

A.3. Proof of Lemma 3.4

The proof is analogue to the proof of Lemma 3.3 and is left to the reader.

A.4. Proof of Lemma 3.5

Let us first focus on the case (i) and denote by τ_k, $1 \leq k \leq m$, the critical values of [A-DBH-SU]. From (2), we have for $1 \leq k \leq m-1$,

$$\begin{aligned}\tau_k &\geq \max\{t \in \mathcal{A} \,:\, t \leq \tau_m,\, t \leq \alpha k(1-\tau_m)/(m-k+1)\}\\ &= \max\left\{t \in \mathcal{A} \,:\, t \leq \left((1-\tau_m)\frac{\alpha k}{m-k+1}\right) \wedge \tau_m\right\},\end{aligned}$$

which correspond to the effective critical values of [BR-λ] with $\lambda = \tau_m$. Now consider the case (ii) and denote again by τ_k, $1 \leq k \leq m$, the critical values of [A-DBH-SD]. From (2), we have for $1 \leq k \leq m$,

$$\tau_k \geq \max\{t \in \mathcal{A} \,:\, (m-k+1)t/(1-t) \leq \alpha k\} = \max\{t \in \mathcal{A} \,:\, t \leq \alpha k/(m-k(1-\alpha)+1)\}$$

which correspond to the effective critical values of [GBS]. This implies the result.

Appendix B: Additional materials

B.1. Procedure [RBH]

The procedure [RBH] is defined as the step-up procedure using the critical values $\tau_k = \lambda_\alpha k/m$, $1 \leq k \leq m$, where $\lambda_\alpha = \max\{\lambda \in [0,1] \,:\, \Psi(\lambda_\alpha) \leq \alpha\}$ for

$$\Psi(\lambda) = \min\left(\lambda, \max_{1\leq k\leq m}\left(\frac{1}{k}\sum_{i=1}^{m}\frac{F_i\left(\lambda k/m\right)}{1-F_i\left(\lambda\right)}\right)\right).$$

The following result is straightforward from Theorem 4.1.

Corollary B.1. *In the model of section 2.1 with the additional assumption (2), we have* $\forall P \in \mathcal{P}$, $\mathrm{FDR}(\mathrm{RBH}, P) \leq \alpha$.

Moreover, if α is such that the equality $\Psi(\lambda_\alpha) = \alpha$ holds true, then $\lambda_\alpha \geq \Psi(\lambda_\alpha) = \alpha$ and [RBH] always dominates [BH] in terms of critical values and thus also of rejection set.

B.2. Lemma for step-up procedures

It is straightforward to check that

$$\widehat{k} = \max\left\{k \in \{0,1,...,m\} \,:\, \sum_{i=1}^{m}\mathbf{1}\{p_i \leq \tau_k\} \geq k\right\},$$

which is sometimes more handy, because this definition avoids to rely explicitly on the order statistics of the p-values.

Let us introduce the following modifications of $\mathbf{SU}(\tau)$:

- $\mathbf{SU}'(\tau) = \mathbf{SU}(\tau')$ the step-up with m critical values defined by $(\tau'_1, \ldots, \tau'_m) = (\tau_2, \ldots, \tau_m, \tau_m)$;
- for some given index $i \in \{1, \ldots, m\}$, $\mathbf{SU}'^{-i}(\tau) = \mathbf{SU}(\tau'^{-i})$ the step-up with $m-1$ critical values defined by $(\tau'^{-i}_1, \ldots, \tau'^{-i}_{m-1}) = (\tau_2, \ldots, \tau_m)$ and restricted to the p-values of the set $\{p_j, j \neq i\}$.

The following lemma holds

Lemma B.2. *For all $i \in \{1,\dots,m\}$, the following assertions are equivalent: (i) $p_i \leq \tau_{\hat{k}}$; (ii) $p_i \leq \tau_{\hat{k}'^{-i}+1}$; (iii) $\widehat{k}'^{-i}+1=\widehat{k}$, where $\widehat{k}'^{-i}$ denotes the number of rejected hypotheses of the procedure $\mathbf{SU}'^{-i}(\tau)$. Moreover, we have $\{p_i > \tau_m\} \subset \{\widehat{k}' = \widehat{k}'^{-i}\}$, where $\widehat{k}'$ denotes the number of rejected hypotheses of the procedure $\mathbf{SU}'(\tau)$.*

Proof. It is not difficult to check that $\widehat{k}'^{-i} \geq \widehat{k}-1$ always holds: this comes from the inequality

$$\widehat{k}-1 = \sum_{j=1}^m \mathbf{1}\{p_j \leq \tau_{\hat{k}}\} - 1 \leq \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\hat{k}}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\hat{k}-1}\},$$

because $\tau'^{-i}_{\ell-1} = \tau_\ell$ for $\ell \in \{2,\dots,m\}$ (note that we can assume without loss of generality $\widehat{k} \geq 1$ here). This means that (i) implies (ii). Now, when $p_i \leq \tau_{\hat{k}'^{-i}+1}$, we have

$$\hat{k}'^{-i} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\hat{k}'^{-i}}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\hat{k}'^{-i}+1}\} = \sum_{j=1}^m \mathbf{1}\{p_j \leq \tau_{\hat{k}'^{-i}+1}\} - 1$$

which implies $\hat{k}'^{-i}+1 \leq \sum_{j=1}^m \mathbf{1}\{p_j \leq \tau_{\hat{k}'^{-i}+1}\}$ and thus $\widehat{k}'^{-i}+1 \leq \widehat{k}$. Since, again, $\widehat{k}'^{-i} \geq \widehat{k}-1$ always holds, we have $\widehat{k}'^{-i}+1=\widehat{k}$. Hence, (ii) implies (iii). Now, if $\widehat{k}'^{-i}+1=\widehat{k}$, we have

$$\begin{aligned}\mathbf{1}\{p_i \leq \tau_{\hat{k}}\} &= \sum_{j=1}^m \mathbf{1}\{p_j \leq \tau_{\hat{k}}\} - \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\hat{k}}\} = \widehat{k} - \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\hat{k}'^{-i}+1}\} \\ &= \widehat{k} - \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\hat{k}'^{-i}}\} = \widehat{k} - \widehat{k}'^{-i} = 1,\end{aligned}$$

by definition of τ'^{-i}, which gives that (iii) implies (i). Now, to prove the last statement, we first note that $\widehat{k}' \geq \widehat{k}'^{-i}$ always holds. Furthermore, if $p_i > \tau_m$ let us prove $\widehat{k}' \leq \widehat{k}'^{-i}$. First, $\widehat{k}'=m$ is impossible because p_i is above τ_m and thus p_i cannot be rejected by $\mathbf{SU}'(\tau)$. Hence, $\widehat{k}' \leq m-1$ and thus $\tau'^{-i}_{\hat{k}'}$ is well defined. Now, since $p_i > \tau_m$, we obtain

$$\sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\hat{k}'}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'_{\hat{k}'}\} = \sum_{j=1}^m \mathbf{1}\{p_j \leq \tau'_{\hat{k}'}\} = \widehat{k}',$$

which implies $\widehat{k}' \leq \widehat{k}'^{-i}$ by definition of $\mathbf{SU}'^{-i}(\tau)$. □

B.3. Lemma for step-down procedures

It is straightforward to check that

$$\widetilde{k} = \max\left\{k \in \{0,1,...,m\} \;:\; \forall k' \leq k,\; \sum_{i=1}^m \mathbf{1}\{p_i \leq \tau_{k'}\} \geq k'\right\}.$$

Let us introduce the following modifications of $\mathbf{SD}(\tau)$:

- for some given index $i \in \{1,\dots,m\}$, $\mathbf{SD}^{-i}(\tau) = \mathbf{SD}(\tau^{-i})$ the step-down procedure with $m-1$ critical values defined by $(\tau_1^{-i},\dots,\tau_{m-1}^{-i}) = (\tau_1,\dots,\tau_{m-1})$ and restricted to the p-values of the set $\{p_j, j\neq i\}$.
- for some given index $i \in \{1,\dots,m\}$, $\mathbf{SD}'^{-i}(\tau) = \mathbf{SD}(\tau'^{-i})$ the step-down procedure with $m-1$ critical values defined by $(\tau_1'^{-i},\dots,\tau_{m-1}'^{-i}) = (\tau_2,\dots,\tau_m)$ and restricted to the p-values of the set $\{p_j, j\neq i\}$.

The following lemma holds

Lemma B.3. *For all $i \in \{1,\dots,m\}$, the following assertions are equivalent: (i) $p_i \leq \tau_{\tilde{k}}$; (ii) $p_i \leq \tau_{\tilde{k}+1}$; (iii) $p_i \leq \tau_{\tilde{k}^{-i}+1}$; (iv) $\widetilde{k}'^{-i}+1=\widetilde{k}$, where $\widetilde{k}^{-i}$ is the number of rejections of $\mathbf{SD}^{-i}(\tau)$ and $\widetilde{k}'^{-i}$ is the number of rejections of $\mathbf{SD}'^{-i}(\tau)$. Moreover, we have $\{p_i > \tau_{\tilde{k}^{-i}+1}\} \subset \{\widetilde{k} = \widetilde{k}^{-i}\}$.*

Proof. First, we check that $\widetilde{k}'^{-i}+1 \geq \widetilde{k}$ always holds. Since $\sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\tilde{k}'^{-i}+1}\} < \widetilde{k}'^{-i}+1$, we have

$$\sum_{j=1}^{m} \mathbf{1}\{p_j \leq \tau_{\tilde{k}'^{-i}+2}\} \leq 1 + \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\tilde{k}'^{-i}+1}\} < \widetilde{k}'^{-i}+2,$$

which gives $\widetilde{k} < \widetilde{k}'^{-i}+2$ by definition of $\widetilde{k}$ and thus $\widetilde{k} \leq \widetilde{k}'^{-i}+1$. Now, if $p_i \leq \tau_{\tilde{k}}$, we have

$$\sum_{j\neq i} \mathbf{1}\{p_j \leq \tau'^{-i}_{\tilde{k}}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\tilde{k}+1}\} = \sum_{j=1}^{m} \mathbf{1}\{p_j \leq \tau_{\tilde{k}+1}\} - 1 < \widetilde{k}+1-1,$$

so that $\widetilde{k} > \widetilde{k}'^{-i}$ and thus $\widetilde{k} \geq \widetilde{k}'^{-i}+1$. This proves that (i) implies (iv). Next, if $p_i > \tau_{\tilde{k}^{-i}+1}$, then

$$\sum_{j=1}^{m} \mathbf{1}\{p_j \leq \tau_{\tilde{k}^{-i}+1}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau_{\tilde{k}^{-i}+1}\} = \sum_{j\neq i} \mathbf{1}\{p_j \leq \tau^{-i}_{\tilde{k}^{-i}+1}\} < \widetilde{k}^{-i}+1,$$

which entails $\widetilde{k} < \widetilde{k}^{-i}+1$ and thus $\widetilde{k} \leq \widetilde{k}^{-i}$. This proves $\widetilde{k} \neq \widetilde{k}'^{-i}+1$. Hence, (iv) implies (iii). The fact that (iii) implies (ii) is obvious because $\widetilde{k} \geq \widetilde{k}^{-i}$ always holds. Finally, we merely check that $\widetilde{k}$ is such that

$$\widetilde{k} = \sum_{j=1}^{m} \mathbf{1}\{p_j \leq \tau_{\tilde{k}}\} = \sum_{j=1}^{m} \mathbf{1}\{p_j \leq \tau_{\tilde{k}+1}\},$$

which means that the set of p-values rejected at threshold $\tau_{\tilde{k}}$ is the same as the set of p-values rejected at threshold $\tau_{\tilde{k}+1}$. This gives that (ii) implies (i). For the last assertion, it has been proved in the above reasoning while showing that (iv) implies (iii). □

Appendix C: Proofs of controlling results

C.1. Proof of Theorem 4.1

By using Lemma B.2 (ii) and (iii) and independence, we easily obtain

$$\mathrm{FDR}(\mathbf{SU}(\tau)) = \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{\mathbf{1}\{p_i \leq \tau_{\hat{k}}\}}{\widehat{k}}\right) = \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{\mathbf{1}\{p_i \leq \tau_{\hat{k}'^{-i}+1}\}}{\widehat{k}'^{-i}+1}\right) \leq \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'^{-i}+1}\right)}{\widehat{k}'^{-i}+1}\right),$$

where the last expectation is taken only with respect to $(p_j, j\neq i)$. Now, on the one hand,

$$\sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'^{-i}+1}\right)}{\widehat{k}'^{-i}+1}\right) \leq \sum_{i=1}^{m} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'^{-i}+1}\right)}{\widehat{k}'^{-i}+1}\right) \leq \sum_{i=1}^{m} \max_{1\leq k\leq m} \frac{F_i(\tau_k)}{k}.$$

Next, on the other hand, by using again the independence,

$$\begin{aligned}\sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'^{-i}+1}\right)}{\widehat{k}'^{-i}+1}\right) &\leq \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'^{-i}+1}\right)}{1-F_i\left(\tau_m\right)} \frac{\mathbf{1}\{p_i > \tau_m\}}{\widehat{k}'^{-i}+1}\right)\\ &= \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\hat{k}'+1}\right)}{1-F_i\left(\tau_m\right)} \frac{\mathbf{1}\{p_i > \tau_m\}}{\widehat{k}'+1} \mathbf{1}\{\widehat{k}'+1 \leq m\}\right),\end{aligned}$$

where the latter equality comes from the last assertion of Lemma B.2. Now, since $\tau_{\hat{k}'+1} \leq \tau_m$, we have that the last display is smaller or equal to

$$\mathbf{E}\left(\sum_{i=1}^{m} \frac{F_i\left(\tau_{\hat{k}'+1}\right)}{1-F_i\left(\tau_m\right)} \frac{\mathbf{1}\{p_i > \tau_{\hat{k}'+1}\}}{\widehat{k}'+1} \mathbf{1}\{\widehat{k}'+1 \leq m\}\right)$$
$$\leq \max_{0\leq k\leq m-1} \max_{X\subset\{1,\dots,m\}, |X|=m-k} \sum_{i\in X} \frac{F_i\left(\tau_{k+1}\right)}{1-F_i\left(\tau_m\right)} \frac{1}{k+1},$$

by taking the maximum over all the possible realizations of the set $X = \{1 \leq i \leq m \,:\, p_i > \tau_{\hat{k}'+1}\} = \{1 \leq i \leq m \,:\, p_i > \tau'_{\hat{k}'}\}$ which is the index set corresponding to the non-rejected null hypotheses of $\mathbf{SU}(\tau')$ (the latter being by definition of cardinality $m - \hat{k}'$). This concludes the proof.

C.2. Proof of Theorem 4.2

It is similar to the step-up case, but relies now on Lemma B.3 (*iii*) and (*iv*) (and still independence):

$$\mathrm{FDR}(\mathbf{SD}(\tau)) = \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{\mathbf{1}\{p_i \leq \tau_{\tilde{k}}\}}{\tilde{k}}\right) = \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{\mathbf{1}\{p_i \leq \tau_{\tilde{k}^{-i}+1}\}}{\tilde{k}'^{-i}+1}\right) \leq \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\tilde{k}^{-i}+1}\right)}{\tilde{k}^{-i}+1}\right).$$

By using independence and the last assertion of Lemma B.3, we obtain

$$\sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\tilde{k}^{-i}+1}\right)}{\tilde{k}^{-i}+1}\right) \leq \sum_{i\in\mathcal{H}_0} \mathbf{E}\left(\frac{F_i\left(\tau_{\tilde{k}^{-i}+1}\right)}{1-F_i\left(\tau_{\tilde{k}^{-i}+1}\right)} \frac{\mathbf{1}\{p_i > \tau_{\tilde{k}^{-i}+1}\}}{\tilde{k}^{-i}+1}\right)$$
$$\leq \mathbf{E}\left(\sum_{i=1}^{m} \frac{F_i\left(\tau_{\tilde{k}+1}\right)}{1-F_i\left(\tau_{\tilde{k}+1}\right)} \frac{\mathbf{1}\{p_i > \tau_{\tilde{k}+1}\}}{\tilde{k}+1} \mathbf{1}\{\tilde{k}+1 \leq m\}\right)$$
$$\leq \mathbf{E}\left(\max_{0\leq k\leq m-1} \max_{X\subset\{1,\dots,m\}, |X|=m-k} \sum_{i\in X} \frac{F_i\left(\tau_{k+1}\right)}{1-F_i\left(\tau_{k+1}\right)} \frac{1}{k+1}\right),$$

because $\{1 \leq i \leq m : p_i > \tau_{\tilde{k}+1}\}$ is equal to $\{1 \leq i \leq m : p_i > \tau_{\tilde{k}}\}$, which is the set of non-rejected hypotheses of $\mathbf{SD}(\tau)$. Since $\mathbf{SD}(\tau)$ rejects exactly $\tilde{k}$ hypotheses, the proof is completed.