

Supplement to “Post hoc inference via joint family-wise error rate control”

Gilles Blanchard

*Universität Potsdam, Institut für Mathematik
Karl-Liebknecht-Straße 24-25 14476 Potsdam, Germany
e-mail: gilles.blanchard@math.uni-potsdam.de*

and

Pierre Neuvial

*Institut de Mathématiques de Toulouse;
UMR 5219, Université de Toulouse, CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France
e-mail: pierre.neuvial@math.univ-toulouse.fr*

and

Etienne Roquain

*Sorbonne Université (Université Pierre et Marie Curie), LPSM,
4, Place Jussieu, 75252 Paris cedex 05, France
e-mail: etienne.roquain@upmc.fr*

Abstract: This supplement provides proofs of results of the main paper, as well as additional material: detailed discussion of relation to previous work (higher criticism, closed testing); general properties of templates and reference families; algorithms and numerical experiments.

Contents

S-1 Relation to Goeman and Solari (2011)	3
S-1.1 Closed testing principle and GS approach	3
S-1.2 Closed testing bound as a particular case of JER control	4
S-1.3 Shortcut as using a particular reference family	5
S-2 Properties of linear and balanced procedures	6
S-2.1 Magnitude of $\lambda(\alpha, \mathbb{N}_m)$	6
S-2.2 Effects of step-down algorithm	7
S-2.3 Properties of the balanced template under independence	9
S-3 Relation to higher criticism and detection power optimality	9
S-3.1 Framework	10
S-3.2 Test statistic of the balanced detection procedure	10
S-3.3 Optimality results	11
S-3.4 Numerical experiments for detection power	12

S-4 Two-sample testing	12
S-5 Monte-Carlo approximation for balanced reference family	14
S-6 Proofs	15
S-6.1 Proofs for Section 2	15
S-6.2 Proofs for Section 5	17
S-6.3 Proofs for Section S-2	18
S-6.4 Proofs for Section S-3	20
S-7 Some properties of the Beta distribution	22
References	23

This supplement is organized as follows. In Section S-1, we elaborate on the connection between JER control and the approach to user-agnostic inference via closed testing. Section S-2 gives complements on the properties of linear and balanced templates. In Section S-3, we study the optimality of the linear and balanced reference families in terms of detection power under sparsity assumptions; in doing so we identify connections between the linear reference family and the FDR controlling procedure of Benjamini and Hochberg (1995), and between the balanced reference family and the higher criticism procedure of Donoho and Jin (2004). Section S-4 establishes the validity of assumption (Rand) in permutation-based two-sample multiple testing problems. In Section S-5, we explain how to approximate the balanced thresholds under known dependence using a Monte-Carlo method. Finally, Section S-6 gathers the proofs of the results stated in the main paper and in this supplement.

S-1. Relation to Goeman and Solari (2011)

To the best of our knowledge, the only existing user-agnostic post hoc approach to multiple testing is the method of Goeman and Solari (2011) (GS11 below for short), which inspired for the present work. This method is based on closed testing (Marcus et al., 1976), which relies, in principle, on testing all $2^m - 1$ possible intersections between m hypotheses. In this section, we discuss connections between that work and ours. First, we show that the probabilistic guarantee derived from closed testing in GS11 can be obtained by JER control for the specific choice $\zeta_k = |R_k| - 1$. Then we prove that the shortcut obtained by GS11 for Simes-type local tests can be interpreted as the JER control for a Simes-type template with $\zeta_k = k - 1$, which implies that the post hoc bound derived by this shortcut is identical to the bound we derive from JER control.

S-1.1. Closed testing principle and GS approach

We come back to the general setting of Section 2.1. For any subset $I \subset \mathbb{N}_m$, define the associated intersection hypothesis as $H_{0,I} = \bigcap_{i \in I} H_{0,i}$. Therefore, $H_{0,I}$ is true iff $P \in H_{0,i}$ for all $i \in I$, or equivalently iff $I \subset \mathcal{H}_0(P)$. We will often informally identify the index subset I with the corresponding intersection hypothesis $H_{0,I}$ in the text to simplify statements. Assume that for any index subset I , the intersection null $H_{0,I}$ can be tested by a so-called *local test* $\phi_I(X) \in \{0, 1\}$ of (individual) level α . From the collection of intersection hypotheses that are rejected by their respective local tests, the classical closed testing principle (Marcus et al., 1976) extracts a subcollection¹ \mathfrak{R} of subsets of \mathbb{N}_m , such that rejecting all intersection hypotheses of the collection \mathfrak{R} has now controlled family-wise error rate at level α .

¹The rejected collection \mathfrak{R} is the maximal subcollection of hypotheses $H_{0,I}$ rejected by their local test that is closed by the superset operation on their indices. In other words, the index set I corresponding to intersection hypothesis $H_{0,I}$ is selected to be included in the rejected collection \mathfrak{R} by the closed testing principle iff $\phi_I(X) = 1$, as well as all $\phi_{I'}(X) = 1$ for all $I \subset I'$.

Equivalently, the complementary $\mathfrak{A} = \mathfrak{R}^c$, i.e. the collection of intersection hypotheses not rejected by the closed testing procedure, satisfies for all $P \in \mathcal{P}$:

$$\mathbb{P}_{X \sim P} \left(\forall I \subset \mathcal{H}_0(P), I \in \mathfrak{A} \right) \geq 1 - \alpha. \quad (\text{S-1})$$

From this, GS11 construct a user-agnostic post hoc bound in the following way: on the event (S-1), for any arbitrary $R \subset \{1, \dots, m\}$, $I_0 = R \cap \mathcal{H}_0(P)$ is a subset of $\mathcal{H}_0(P)$ and thus must be included in the collection \mathfrak{A} . This entails that (PH $_{\alpha}$) is satisfied with

$$\bar{V}^{GS}(R) = \max\{|I| : I \in \mathfrak{A}, I \subset R\}. \quad (\text{S-2})$$

S-1.2. Closed testing bound as a particular case of JER control

We now justify that the inequality (S-1) resulting from closed testing can be seen as a JER control. As a result, the GS11 post hoc bound (S-2) can be seen as a particular case of the JER post hoc approach developed here, and more precisely a particular case of the optimal bound V^* defined in (6).

For this, consider $\mathfrak{R} = (R_1, \dots, R_K)$ returned by the closed testing procedure as in Section S-1.1. The closed testing principle implies that \mathfrak{R} is closed by the superset operation, i.e., $I \in \mathfrak{R}$ implies $\forall J \supset I, J \in \mathfrak{R}$. Now let $\zeta(k) := |R_k| - 1$, $1 \leq k \leq K$. Finally, we have

$$\begin{aligned} \left\{ \forall I \subset \mathcal{H}_0(P), I \in \mathfrak{A} \right\}^c &= \left\{ \exists I \subset \mathcal{H}_0(P), I \in \mathfrak{R} \right\} \\ &= \left\{ \mathcal{H}_0(P) \in \mathfrak{R} \right\} \\ &= \left\{ \exists k \in \{1, \dots, K\} : |\mathcal{H}_0(P) \cap R_k| = |R_k| \right\} \\ &= \left\{ \exists k \in \{1, \dots, K\} : |\mathcal{H}_0(P) \cap R_k| > \zeta(k) \right\} \\ &= \mathcal{E}(\mathfrak{R}, \mathcal{H}_0)^c. \end{aligned}$$

Concerning the post hoc bounds, note that

$$\begin{aligned} \mathcal{A}(\mathfrak{R}) &= \{A \subset \mathbb{N}_m : \forall k \in \mathbb{N}_K |R_k \cap A| \leq \zeta_k\} \\ &= \{A \subset \mathbb{N}_m : \forall k \in \mathbb{N}_K |R_k \cap A| \leq |R_k| - 1\} \\ &= \{A \subset \mathbb{N}_m : \forall k \in \mathbb{N}_K R_k \not\subset A\} \\ &= \mathfrak{R}^c = \mathfrak{A}, \end{aligned}$$

where we have used the fact that \mathfrak{R} is closed by superset operation. Hence

$$V_{\mathfrak{R}}^*(R) = \max_{A \in \mathcal{A}(\mathfrak{R})} |R \cap A| = \max_{A \in \mathfrak{A}} |R \cap A| = \max_{A \in \mathfrak{A}, A \subset R} |A| = \bar{V}^{GS}(R),$$

since \mathfrak{A} is closed by subset operation.

S-1.3. Shortcut as using a particular reference family

As noted by GS11, computation of the closed testing output \mathfrak{R} is (in general) not feasible when m is larger than a few dozens. In certain situations, so-called shortcuts can be available for specific choices of the local tests, providing a direct (but possibly approximate and conservative) computation of the closed testing. For instance, it is well-known that Hommel’s step-down procedure is a shortcut for Bonferroni local tests, and Hochberg’s step-up a shortcut for Simes local tests (Huang and Hsu, 2007).

Even if \mathfrak{R} can be computed, we have shown that the calculation of the post hoc bound (S-2) is itself NP-hard in a generic setting. To circumvent this complexity issue, a less time-consuming conservative “shortcut” of the bound (S-2) has been proposed by GS11 for Simes-type local tests:

$$\phi_I(X) = \mathbf{1} \left\{ \exists i \in \{1, \dots, |I|\} : p_{(i:I)} \leq c_i^{|I|} \right\}, \quad I \subset \mathbb{N}_m, \quad (\text{S-3})$$

with the assumptions $c_i^\ell \leq c_i^k$, for $\ell \geq k$ and $c_i^\ell \leq c_j^\ell$, for $i \leq j$. The corresponding bound is defined as:

$$\bar{V}_{\text{Simes}}^{GS}(R) = |R| - (1 + \max\{S_r, 1 \leq r \leq |R|\}) \vee 0, \quad (\text{S-4})$$

where $S_r = \max\{0 \leq s \leq r - 1 : p_{(r:R)} \leq c_{r-s}^m\}$ (with $\max \emptyset = -\infty$). We argue below in Lemma S-1.2 that this bound is in fact equivalent to the post hoc bound $\bar{V}_{\mathfrak{R}}(R)$ defined in (7), for the family $\mathfrak{R} = (R_k, \zeta_k := k - 1)_{1 \leq k \leq m}$ defined by

$$R_k = \{i \in \mathbb{N}_m : p_i \leq c_k^m\}, \quad 1 \leq k \leq m.$$

The next lemma establishes that JER control holds for this family:

Lemma S-1.1. *Assume that the tests $(\phi_I)_{I \subset \mathbb{N}_m}$ form a family of local tests at level α for the considered model, i.e., for any $P \in H_{0,I}$, it holds $\mathbb{P}_{X \sim P}(\phi_I(X) = 1) \leq \alpha$. Then joint control of the k -FWER of R_k at level α , uniformly over $k \in \mathbb{N}_m$, holds; in other words, equation (3) holds for the reference family $\mathfrak{R} = (R_k, \zeta_k := k - 1)_{1 \leq k \leq m}$.*

Proof. For any given distribution P in the model, we have for $I = \mathcal{H}_0 = \mathcal{H}_0(P)$ the local test control

$$\mathbb{P}_{X \sim P} [\exists k \leq m_0 : p_{(k:\mathcal{H}_0)} \leq c_k^{m_0}] \leq \alpha,$$

implying by the monotonicity assumption $c_i^\ell \leq c_i^j$ for $\ell \geq j$:

$$\mathbb{P}_{X \sim P} [\exists k \leq m_0 : p_{(k:\mathcal{H}_0)} \leq c_k^m] \leq \alpha.$$

As we argued in Section 2.4, this is equivalent to $\text{JER}(\mathfrak{R}, P) \leq \alpha$ for the threshold-based reference family $\mathfrak{R} = (R_k, k - 1)_{1 \leq k \leq m}$ using thresholds $t_k := c_k^m$, $k \in \mathbb{N}_m$, see (1). \square

Now, we establish the equivalence of the two bounds:

Lemma S-1.2. For any $R \subset \mathbb{N}_m$, $\bar{V}_{\mathfrak{R}}(R) = \bar{V}_{\text{Simes}}^{GS}(R)$.

Proof. The result comes from

$$\begin{aligned}
& \max\{S_r, 1 \leq r \leq |R|\} \\
&= \max\{s \geq 0 : \exists r \text{ s.t. } 1 \leq r \leq |R| \text{ and } 0 \leq s \leq r-1 \text{ and } p_{(r:R)} \leq c_{r-s}^m\} \\
&= \max\{s \geq 0 : \exists r \text{ s.t. } s+1 \leq r \leq |R| \text{ and } |R_{r-s} \cap R| \geq r\} \\
&= \max\{s \geq 0 : \exists r \leq m \text{ s.t. } s+1 \leq r \leq |R_{r-s} \cap R|\} \\
&= \max\{s \geq 0 : \exists k \leq m \text{ s.t. } 1 \leq k \leq |R_k \cap R| - s\} \\
&= \max\{|R_k \cap R| - k, 1 \leq k \leq m\},
\end{aligned}$$

by letting $k = r - s$. □

A consequence is that using this GS11 shortcut reduces to the post hoc bound studied in this paper. (A remark pointing in that direction is also mentioned at the end of Section 4.2 of GS11.) In particular, for $c_k^m = \alpha k/m$, the reference family reduces to the Simes reference family \mathfrak{R}_0 (17), and the bound $\bar{V}_{\text{Simes}}^{GS}(R)$ has the simple equivalent form given by (18).

S-2. Properties of linear and balanced procedures

S-2.1. Magnitude of $\lambda(\alpha, \mathbb{N}_m)$

Consider the case of known dependence (therefore, with a λ -calibration given by (24)) in the equi-correlated Gaussian one-sided location model for simplicity.

Linear template Let us discuss the magnitude of $\lambda^L(\alpha, \mathbb{N}_m)$. First, for $K = m$ and in the independent case, that is $\Sigma = I_m$, we have $\lambda^L(\alpha, \mathbb{N}_m) = \alpha$ by Theorem 4.1, which means that \mathfrak{R}^L reduces to the Simes reference family \mathfrak{R}^0 (17). Under dependence, Figure S-1 displays $\lambda^L(\alpha, \mathbb{N}_m)$ in the (one-sided) Gaussian ρ -equi-correlated setting, for different values of ρ . The influence of the size K is also illustrated. In a nutshell, we see that the influence of K and ρ is moderate for, say, $\rho \leq 0.2$ (a somewhat realistic range for the dependency strength). The lack of sensitivity with respect to K is not surprising because for the linear template, only the very first k are be important inside the probability of relation (24), as already noted in Section 4.3.

Balanced template Let us discuss the magnitude of $\lambda^B(\alpha, \mathbb{N}_m)$. Since \bar{F} is continuous, each of the $F_k(q_{(k:m)})$ is uniformly distributed on $(0, 1)$, and a simple union bound argument provides the following bounds:

$$\alpha/K \leq \lambda^B(\alpha, \mathbb{N}_m) \leq \alpha. \tag{S-5}$$

Under independence and for $K = m$, the following Lemma provides a more accurate upper bound for m large enough:

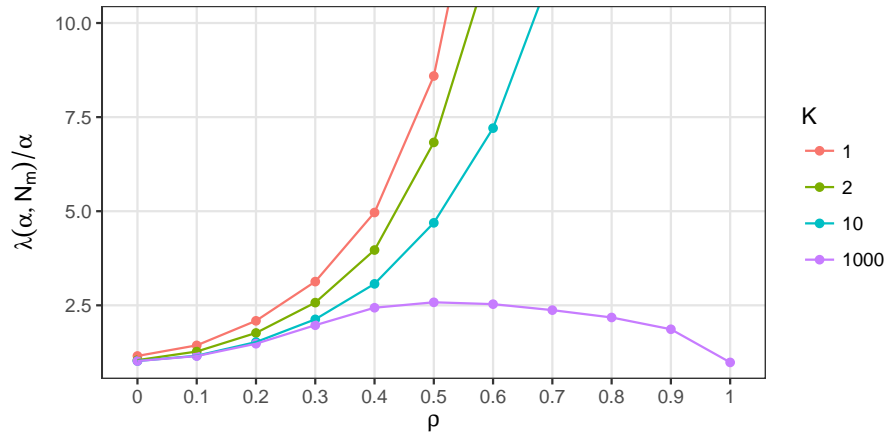


FIG S-1. Influence of the equi-correlation level ρ on the adjustment factor $\lambda^L(\alpha, \mathbb{N}_m)$ for linear template. Different values of K are used. $m = 1,000$; $\alpha = 0.2$; $\pi_0 = 1$. $\lambda^L(\alpha, \mathbb{N}_m)$ was estimated based on $B = 10^4$ Monte-Carlo samples of the joint null distribution.

Lemma S-2.1. In the framework of Proposition S-2.3, consider $\lambda^B(\alpha) = \lambda^B(\alpha, \mathbb{N}_m)$ for $K = m$ (see Section 6.2). Then for m large enough, we have

$$\lambda^B(\alpha, \mathbb{N}_m) \leq 1/(\log m)^{1/4}. \quad (\text{S-6})$$

In particular, $\lambda^B(\alpha)$ tends to zero as m grows to infinity. However, when the size K is kept fixed, say $K = 10$, (S-5) ensures that $\lambda^B(\alpha, \mathbb{N}_m)$ is bounded away from zero. Figure S-2 shows the influence of ρ and K on the value of $\lambda^B(\alpha, \mathbb{N}_m)$ under (one-sided) Gaussian ρ -equi-correlated dependence. Compared to the linear template, we see the sensitivity of $\lambda^B(\alpha, \mathbb{N}_m)$ w.r.t. K and ρ is more substantial. When $\rho = 0$, the value of $\lambda^B(\alpha, \mathbb{N}_m)$ is small for $K = m$ and increases as K becomes smaller, which supports the above theoretical statements. Also, even moderate values of ρ (say, $\rho \leq 0.2$) have a large impact on the value of $\lambda^B(\alpha, \mathbb{N}_m)$.

S-2.2. Effects of step-down algorithm

We focus on the known dependence case, with an independent Gaussian one-sided location model for simplicity. Recall that the rationale behind our proposed step-down algorithm is that, when $\pi_0 = m_0/m$ is smaller than 1, some of the hypotheses will be rejected at each step, which will improve the value of the λ -adjustment by replacing $\lambda(\alpha, \mathbb{N}_m)$ by $\lambda(\alpha, \hat{A})$ for \hat{A} smaller than \mathbb{N}_m . How large is the magnitude of the improvement under independence (and $K = m$)? It turns out that the step-down refinement has a stronger influence for the balanced template than for the linear template.

For the linear reference family, we have $\lambda(\alpha, \mathbb{N}_m) = \alpha$ and the family reduces to the Simes family in this case. From (16), the achieved upper bound (23) on

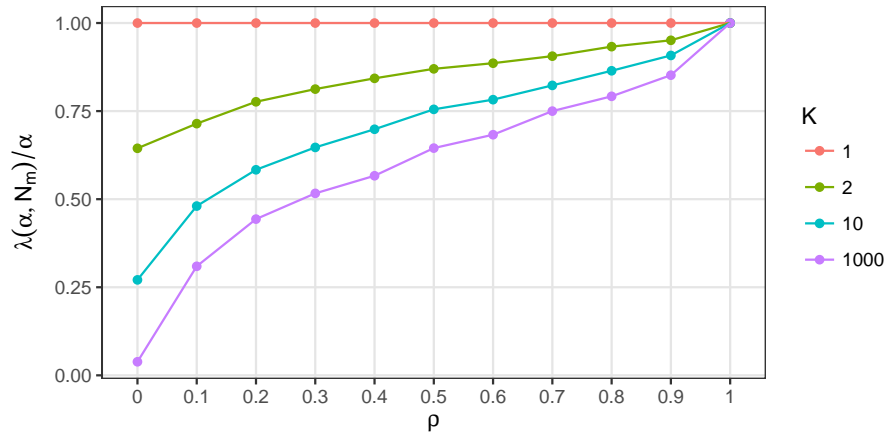


FIG S-2. Influence of the equi-correlation level ρ on the adjustment factor $\lambda^B(\alpha, \mathbb{N}_m)$ for the balanced template. Different values of K are used. $m = 1,000$; $\alpha = 0.2$; $\pi_0 = 1$. $\lambda^B(\alpha, \mathbb{N}_m)$ was estimated based on $B = 10^4$ Monte-Carlo samples of the joint null distribution.

the JER is

$$\mathbb{P}(\exists k \in \{1, \dots, m_0\} : q_{(k:\mathcal{H}_0)} < \alpha k/m) = \pi_0 \alpha.$$

As a consequence, the criterion has a linear dependence w.r.t. π_0 . The best improvement that the step-down algorithm can provide is thus $\lambda = \alpha/\pi_0$.

By contrast, for the balanced reference family, the influence of π_0 is more substantial. The next lemma shows that using the substitute $\lambda^B(\alpha, \mathbb{N}_m)$ instead of $\lambda^B(\alpha, \mathcal{H}_0)$ for the balanced template (in the case $K = m$) results in a JER tending asymptotically to 0 with m if π_0 is bounded away from 1. This justifies the importance of trying to use some kind of adaptive procedure, such as the step-down.

Lemma S-2.2. Consider $\lambda^B(\alpha) = \lambda^B(\alpha, \mathbb{N}_m)$ for $K = m$ and $q_i, i \in \mathbb{N}_m$, i.i.d. $U(0, 1)$ variables. Let $m_0 = \pi_0 m$ for some $\pi_0 \in (0, 1)$ fixed with m . Then, for m large enough,

$$\mathbb{P}\left(\exists k \in \{1, \dots, m_0\} : q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right) \leq \frac{C(\pi_0)}{(\log m)^{1/8}}, \quad (\text{S-7})$$

where $C(\pi_0) = 1 + \frac{64}{(1-\pi_0)^2} \left(1 - e^{-\frac{(1-\pi_0)^2}{32}}\right)^{-1}$.

This shows that the influence of $\pi_0 < 1$ on the achieved JER is substantial and makes the potential improvement of the step-down algorithm all the more important. Of course, the amplitude of this phenomenon decreases as π_0 gets closer to 1, but our numerical experiments suggest that it still exists for cases where $\pi_0 \approx 1$ (sparsity). This is a new feature of step-down type algorithms to the best of our knowledge.

S-2.3. Properties of the balanced template under independence

The following result gathers some of the properties of the balanced template under independence.

Proposition S-2.3. *In the location model (14) under independence, letting $t_k^B(\lambda)$ be the threshold given by (29), we have:*

- (i) for all $k \in \{1, \dots, m\}$ and $\lambda \in [0, 1]$, $t_k^B(\lambda)$ is the λ -quantile of the distribution $\text{Beta}(k, m + 1 - k)$.
- (ii) for any $\lambda \leq 0.5$,

$$t_k^B(\lambda) \leq \frac{k}{m+1};$$

$$t_k^B(\lambda) \geq \frac{k}{m+1} - \left\{ \frac{k}{m+1} \left(1 - \frac{k}{m+1} \right) \right\}^{1/2} \left(\frac{4 \log 1/\lambda}{m} \right)^{1/2}; \quad (\text{S-8})$$

- (iii) for all $m_0 \in \{1, \dots, m\}$ and $k \in \{1, \dots, m_0\}$, for all $\lambda \leq 0.5$,

$$\mathbb{P}(p_{(k:m_0)} \leq t_k^B(\lambda)) \leq \exp\left(-\frac{k}{32} \left(1 - \frac{m_0}{m}\right)^2\right). \quad (\text{S-9})$$

- (iv) for all $k \in \{1, \dots, m\}$ and $\alpha \leq 0.5$,

$$\mathbb{P}\left(p_{(k:m)} \leq \alpha \frac{k}{m}\right) \leq \exp\left(-\frac{k}{4} \left(1 - \alpha - \frac{1}{m+1}\right)^2\right). \quad (\text{S-10})$$

S-3. Relation to higher criticism and detection power optimality

In a nutshell, we show in this section that, as a detection procedure, \mathfrak{R}^B shares some similarity with the calibration of the higher criticism (HC for short) method of [Donoho and Jin \(2004\)](#), DJ04 for short. By contrast, \mathfrak{R}^L (with $K = m$), which is equal to \mathfrak{R}^0 in the setting of this section, is connected to the procedure of [Benjamini and Hochberg \(1995\)](#), BH for short. This induces specific power properties. We evaluate the power of a reference family through its ability of detection of any false null hypothesis:

$$\text{Pow}^*(\mathfrak{R}, P) = \mathbb{P}(\overline{S}_{\mathfrak{R}}(\mathbb{N}_m) \geq 1) = \mathbb{P}(\exists k \in \{1, \dots, K\} : |R_k| \geq k). \quad (\text{S-11})$$

Note that this can be seen as the power of the single test rejecting the null H_0 : “ $\forall i \in \mathbb{N}_m, H_{0,i}$ is true” if there exists $k \in \{1, \dots, m\}$ such that $|R_k| \geq k$. We show that in a special regime, \mathfrak{R}^B is optimal with respect to this criterion, while \mathfrak{R}^L is suboptimal.

Note that the step-down algorithm cannot provide any improvement in terms of detection power: the step-down can potentially make the sets R_k in the reference family larger in comparison to the single-step procedure, but by construction such an improvement can only take place if $|R_1| \geq 1$ in the first place for the single-step procedure (which is the first iteration of the step-down). Hence, we focus on the single-step versions in this section.

S-3.1. Framework

We consider the location model (14) in the Gaussian independent one-sided framework, with the special setting considered in DJ04 where the true/false status of the null hypotheses is randomized with a distribution belonging to some sparse regime. Specifically, we consider the hierarchical model where H_i are i.i.d. $\mathcal{B}(\pi_{1,m})$ and the p -values are independent conditionally on the H_i 's, with

- $p_i(X) | H_i = 0 \sim U(0, 1)$;
- $p_i(X) | H_i = 1$ has for c.d.f. $F_{1,m}(t) = \overline{\Phi}(\overline{\Phi}^{-1}(t) - \mu_m)$.

Hence, overall, the p -values ($p_i, i \in \mathbb{N}_m$) are i.i.d. and of common c.d.f. $G_{1,m}(t) = \pi_{0,m}t + \pi_{1,m}F_{1,m}(t)$, where $\pi_{0,m} = 1 - \pi_{1,m}$. The parameters $\pi_{1,m}, \mu_m$ are taken in the asymptotic range where $\pi_{1,m} = m^{-\beta}$ and $\mu_m = \sqrt{2r \log m}$ for two parameters $\beta \in (1/2, 1)$ and $r \in (0, 1)$.

Let us also recall the optimal asymptotic detection boundary defined by DJ04:

$$\rho^*(\beta) = \begin{cases} \beta - 1/2 & \text{if } \beta \in (1/2, 3/4]; \\ (1 - \sqrt{1 - \beta})^2 & \text{if } \beta \in (3/4, 1). \end{cases} \quad (\text{S-12})$$

This is an optimal detection boundary in the following sense: for $r < \rho^*(\beta)$, any detection procedure will have a risk (type I error rate plus type II error rate) tending to 1, while for $r > \rho^*(\beta)$, there exists a detection procedure that has a risk tending to 0. DJ04 showed that HC achieves this property.

Note that for all $\beta \in (1/2, 3/4]$, the range where $\beta - 1/2 \leq r \leq (1 - \sqrt{1 - \beta})^2$ is usually referred to as sparse/weak, that is, with sparsity and low signal strength. The sparse and weak regime is of interest because DJ04 showed that, in this situation, BH has asymptotically no detection power while HC has full asymptotic power. In particular, this shows that BH does not attain the optimal detection boundary.

S-3.2. Test statistic of the balanced detection procedure

By definition, \mathfrak{R}^B makes a detection if there exists k such that $p_{(k:m)} < t_k^B(\lambda^B(\alpha))$. Furthermore, from (S-5) and (S-8), we have the lower bound

$$t_k^B(\lambda^B(\alpha)) \geq \frac{k}{m+1} - \left\{ \frac{k}{m+1} \left(1 - \frac{k}{m+1} \right) \right\}^{1/2} m^{-1/2} (4 \log(K/\alpha))^{1/2}. \quad (\text{S-13})$$

Hence, \mathfrak{R}^B makes a detection whenever the test statistic

$$\max_{1 \leq k \leq K} \left\{ m^{1/2} \frac{\frac{k}{m+1} - p_{(k:m)}}{\left\{ \frac{k}{m+1} \left(1 - \frac{k}{m+1} \right) \right\}^{1/2}} \right\}$$

exceeds $(4 \log(K/\alpha))^{1/2}$. This is close to the higher criticism procedure of DJ04.

Remark S-3.1. Note that in the definition of Higher Criticism considered in DJ04, the authors have similarly restricted the range of the indices considered to $\{1, \dots, \alpha_0 m\}$, that is, $\alpha_0 m$ plays a similar role to K here. This is useful to tune the power detection ability, as explained in Section S-3.4.

S-3.3. Optimality results

By adapting the proof of DJ04, we can show the following result (proved in Section S-6.4):

Theorem S-3.2. *In the asymptotic setting of Section S-3.1, we have:*

- (i) *Consider any family \mathfrak{R} with thresholds t_k , $1 \leq k \leq m$, that controls the JER at level α in the sense*

$$\mathbb{P}(\exists k \in \{1, \dots, m\} : U_{(k:m)} < t_k) \leq \alpha,$$

for $U_i, i \in \mathbb{N}_m$ i.i.d. uniformly distributed on $(0, 1)$. Then we have

$$\limsup_m \text{Pow}^*(\mathfrak{R}, P) \leq \alpha$$

whenever P is such that $r < \rho^(\beta)$.*

- (ii) *For the balanced family \mathfrak{R}^B with $K = m$, we have $\text{Pow}^*(\mathfrak{R}^B, P) \rightarrow 1$ whenever P is such that $r > \rho^*(\beta)$.*

By contrast, the family \mathfrak{R}^L is sub-optimal, as we show now. This family makes a detection if there exists k such that $p_{(k:m)} < \alpha k/m$, that is, if the Benjamini-Hochberg procedure rejects at least one null hypothesis. The following result is in fact a reformulation of Theorem 1.4 in Donoho and Jin (2004); its proof is given in Section S-6.4 for completeness.

Theorem S-3.3. *Consider the asymptotic setting of Section S-3.1. Then the linear reference family \mathfrak{R}^L satisfies the following:*

- (i) *for $r > (1 - \sqrt{1 - \beta})^2$, $\lim \text{Pow}^*(\mathfrak{R}^L, P) = 1$;*
(ii) *for $r < (1 - \sqrt{1 - \beta})^2$, $\limsup \text{Pow}^*(\mathfrak{R}^L, P) \leq \alpha$.*

Intuitively, the threshold is $\alpha k/m = k/m - (1 - \alpha)k/m$, so the deviation term is not of the correct order. This implies a lack of detection power which makes this procedure miss the optimal boundary.

Let us finally emphasize that the domination of the balanced family/HC w.r.t. the linear family/BH in terms of detection power is less obvious for a moderate value of m , as illustrated in the numerical experiments below where $m = 1,000$. This suggests that the asymptotical regime described in Theorems S-3.2 and S-3.3 is not fully reached for such a value of m (while it seems reached for $m = 10^6$ in DJ04).

S-3.4. Numerical experiments for detection power

We consider the independent case, and we calibrate the parameter $\bar{\mu}$ and π_0 according to the above-defined regime: $\pi_0 = 1 - m^{-\beta}$ and $\bar{\mu} = \sqrt{2r \log m}$, for two parameters β (sparsity) and r (signal strength) taken in the range $\beta \in \{0.5, 0.6, 0.8, 1\}$ and $r \in \{0.05, 0.1, 0.2, 0.5, 1\}$. Note that, however, we do not consider an i.i.d. p -value mixture here, but stick to the framework defined in Section 7. For each setting, we estimate detection power by its empirical counterpart, the proportion \hat{q} of 1,000 simulation runs for which at least one of the subsets R_k of the collection \mathfrak{R} contains more than k elements. Our experiments have been made for a range of values of the target JER level $\alpha \in \{0.01, 0.02, 0.05, 0.10, 0.15, 0.20, 0.25\}$. To summarize the results, we plot in Figure S-3 (top) the empirical detection power \hat{q} as a function of α for each method.

The parameter configurations (β, r) for which the signal is below the asymptotically optimal detection boundary identified by Donoho and Jin (2004) are represented by blue squares in the bottom panel of Figure S-3. As expected from the theory, in such configurations all procedures are powerless, in the sense that the detection power is very close to the JER. Let us focus on the parameter configurations for which detection is asymptotically feasible (green circles and red triangles in the bottom panel of Figure S-3). In such configurations, as expected, K has little influence on detection power for the linear template. For the balanced template, the detection power is substantially higher for $K = 10$ than for $K = m$. This influence of K is consistent with our comments for JER control in Section 7. Overall, the balanced template with $K = m$ has better detection power than the linear template for moderate sparsity ($\beta \in \{0.5, 0.6\}$) and signal ($r \leq 1$). However, for sparser settings ($\beta \in \{0.8, 1\}$) the linear template performs better than the balanced template with $K = m$, and even than the balanced template with $K = 10$ in very sparse scenarios. These numerical results provide a useful complement to the asymptotic statements of the preceding Section; for a finite m , the balanced template/HC is not always superior to the linear template/FWER/FDR. Moreover, in the sparse/weak setting, which is illustrated here by the configurations $\beta = 0.5, r = 0.05$ and $\beta = 0.6, r = 0.1$, the balanced template is only marginally superior to the linear template as a detection procedure; we suspect that the asymptotics of Theorem S-3.2 (and of Donoho and Jin, 2004) are quite slow to kick in and not yet reached for $m = 1000$.

S-4. Two-sample testing

We describe how assumption (Rand) is met in permutation-based two-sample multiple testing problems. This can be seen as a reformulation of Example 5 in Romano and Wolf (2005). Let us consider a two-sample framework where

$$X = (X^{(1)}, \dots, X^{(n_1)}, X^{(n_1+1)}, \dots, X^{(n_1+n_2)}) \in (\mathbb{R}^m)^n$$

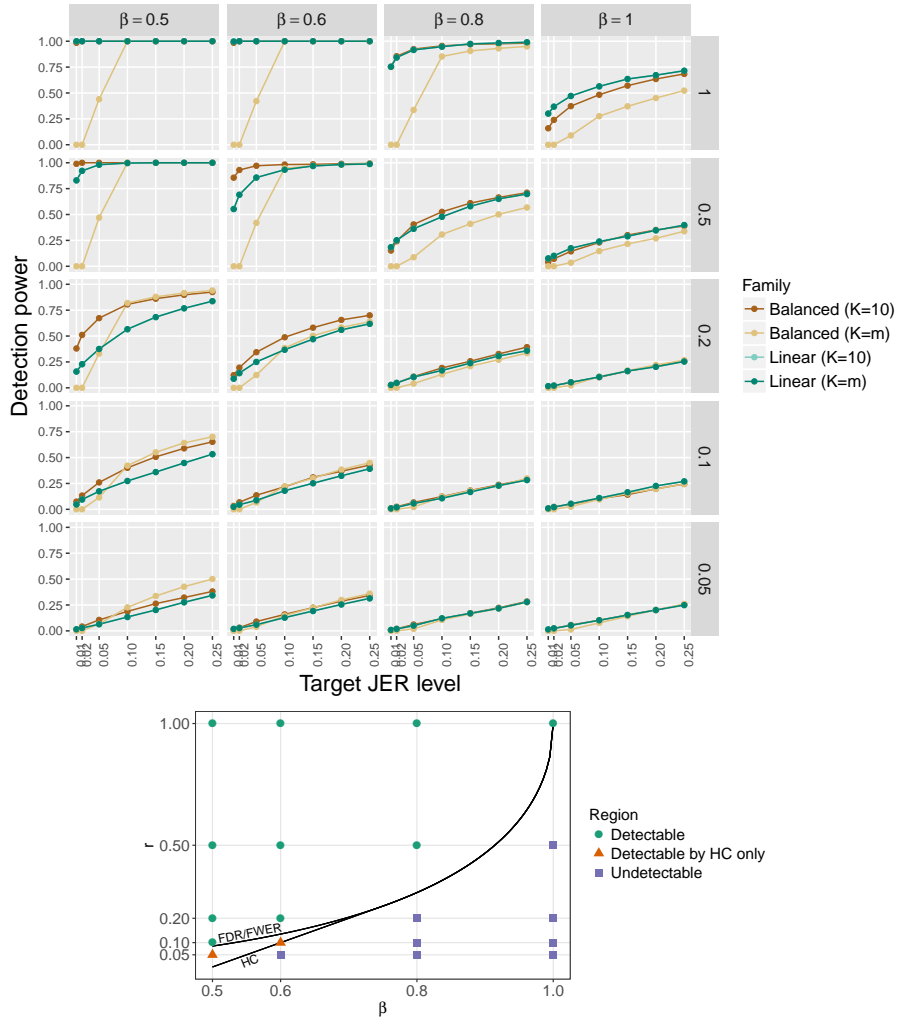


FIG S-3. Top: detection power of JER controlling procedures for independent test statistics in the sparsity range for 4×5 parameter configurations for (β, r) in the sparsity range $[1/2, 1] \times [0, 1]$. Bottom: these 4×5 configurations are positioned with respect to the detection boundaries identified in Donoho and Jin (2004).

is composed of $n = n_1 + n_2$ independent m -dimensional real random vectors with $X^{(j)}$, $1 \leq j \leq n_1$, i.i.d. (case) and $X^{(j)}$, $n_1 + 1 \leq j \leq n$, i.i.d. (control). Then we aim at testing the null hypotheses $H_{0,i} : “\mathcal{D}(X_i^{(1)}) = \mathcal{D}(X_i^{(n_1+1)})”$, simultaneously in $1 \leq i \leq m$, without knowing the dependencies between the coordinates of the $X^{(j)}$'s. Consider any individual p -values $p_i(X)$ function of the line $(X_i^{(j)})_{1 \leq j \leq n}$ (e.g., based on the Mann-Whitney test statistics). Note that $p_{\mathcal{H}_0}(X)$ is thus a measurable function of $(X_i^{(j)})_{i \in \mathcal{H}_0, 1 \leq j \leq n}$. Now, the group \mathcal{G} of the permutation of $\{1, \dots, n\}$ is naturally acting on $\mathcal{X} = (\mathbb{R}^m)^n$ via the permutation of the columns: for all $\sigma \in \mathcal{G}$,

$$\sigma.X = (X^{(\sigma(1))}, \dots, X^{(\sigma(n_1))}, X^{(\sigma(n_1+1))}, \dots, X^{(\sigma(n))}).$$

This entails that $p_{\mathcal{H}_0}(\sigma.X)$ is a measurable function of $(X_i^{(\sigma(j))})_{i \in \mathcal{H}_0, 1 \leq j \leq n}$. As a result, the assumption (Rand) is satisfied as soon as

$$\left((X_i^{(1)})_{i \in \mathcal{H}_0}, \dots, (X_i^{(n)})_{i \in \mathcal{H}_0} \right) \quad (\text{S-14})$$

is an exchangeable vector. However, in the general setting, the latter is not necessarily true, because i.i.d. marginals do not imply an exchangeable joint distribution.

The assumption (Rand) can be covered by making an appropriate additional semi-parametric assumption. Assume that $X^{(1)} \sim Q(\theta^{(1)}, \eta)$ and $X^{(n_1+1)} \sim Q(\theta^{(2)}, \eta)$ for some distribution $Q(\theta, \eta)$ on \mathbb{R}^m only depending on a parameter $\theta \in \mathbb{R}^m$ and on some general nuisance parameter η . Assume the functional $Q(\cdot, \cdot)$ is such that for all $A \subset \mathbb{N}_m$, θ and η , the restriction of $Q(\theta, \eta)$ to the indices of A is only depending on A , $(\theta_i)_{i \in A}$ and η . As a consequence, the null hypotheses can be rewritten as $H_{0,i} : “\theta_i^{(1)} = \theta_i^{(2)}”$, $1 \leq i \leq m$, and it is now clear that the vector (S-14) is i.i.d. and thus exchangeable. A typical instance for such a functional $Q(\cdot, \cdot)$ is given by $Q(\theta, \eta) = \mathcal{N}(\theta, \eta)$ where η is some (unknown) covariance matrix.

S-5. Monte-Carlo approximation for balanced reference family

We consider the balanced reference family $\mathfrak{R}^{B, sd}$ given in Section 6.2. We explain here in detail the construction of the reference thresholds $t_k^B(\lambda^B(\alpha, \widehat{A}))$, $1 \leq k \leq K$ in the case of known dependence, using a Monte-Carlo approximation.

1. Draw $q^{(1)}, \dots, q^{(B)}$ i.i.d. according to the distribution ν_m (on $[0, 1]^m$), and define the matrix

$$M_0 = \begin{pmatrix} q_1^{(1)} & q_2^{(1)} & \dots & q_m^{(1)} \\ q_1^{(2)} & q_2^{(2)} & \dots & q_m^{(2)} \\ \vdots & \vdots & & \vdots \\ q_1^{(B)} & q_2^{(B)} & \dots & q_m^{(B)} \end{pmatrix};$$

2. Define for all $A \subset \{1, \dots, m\}$ (denoting $a := |A|$), the matrix

$$M(A) = \begin{pmatrix} q_{(1:A)}^{(1)} & q_{(2:A)}^{(1)} & \cdots & q_{(a:A)}^{(1)} \\ q_{(1:A)}^{(2)} & q_{(2:A)}^{(2)} & \cdots & q_{(a:A)}^{(2)} \\ \vdots & \vdots & & \vdots \\ q_{(1:A)}^{(B)} & q_{(2:A)}^{(B)} & \cdots & q_{(a:A)}^{(B)} \end{pmatrix}$$

by ordering the rows of the submatrix of M_0 whose row indices are in A .

3. Consider $M(A)$ for $A = \mathbb{N}_m$ and approximate $F_k(x)$ by $\tilde{F}_k(x) = B^{-1} \sum_{b=1}^B \mathbb{1} \left\{ q_{(k:m)}^{(b)} \leq x \right\}$.

For each λ , approximate $t_k^B(\lambda)$ by $\tilde{t}_k^B(\lambda)$, defined as the λ -quantile of the sample

$$\left(q_{(k:m)}^{(1)}, \dots, q_{(k:m)}^{(B)} \right).$$

4. Consider the matrix of “ranks”

$$Z(A) = \begin{pmatrix} Z_{1,1}(A) & Z_{1,2}(A) & \cdots & Z_{1,a}(A) \\ Z_{2,1}(A) & Z_{2,2}(A) & \cdots & Z_{2,a}(A) \\ \vdots & \vdots & & \vdots \\ Z_{B,1}(A) & Z_{B,2}(A) & \cdots & Z_{B,a}(A) \end{pmatrix}$$

where we let $Z_{b,k}(A) = \tilde{F}_k(q_{(k:A)}^{(b)})$, for $1 \leq k \leq K \wedge a$ and $1 \leq b \leq B$.

5. Build the vector

$$U = \left(\min_{1 \leq k \leq K \wedge a} \{Z_{1,k}(A)\}, \dots, \min_{1 \leq k \leq K \wedge a} \{Z_{B,k}(A)\} \right),$$

by taking the minimum within each line of $Z(A)$. Approximate now $\lambda^B(\alpha, A)$ by $\tilde{\lambda}^B(\alpha, A) = U_{(\lceil \alpha B \rceil)}$, i.e., the α empirical quantile of the sample $U = (U_1, \dots, U_B)$.

6. Use Algorithm 1 with \tilde{t}_k^B and $\tilde{\lambda}^B$ instead of t_k^B and λ^B , respectively, to obtain \tilde{A} .

Note that in the above construction, it is only required to calculate the first $K \wedge a$ elements instead of the first a elements.

Remark S-5.1. Although JER control by λ -calibration for the balanced reference family under unknown dependence is not guaranteed by Proposition 5.8, we have also implemented this λ -calibration. Formally, the construction is identical, using $q_k^{(b)} = p_k(g_b \cdot X)$ for $b = 1 \dots B$ and $k \in \mathbb{N}_m$. In this case, we have $\tilde{F}_k = F_k$, $\tilde{t}_k = t_k$, and $\tilde{\lambda}^B = \lambda^B$.

S-6. Proofs

S-6.1. Proofs for Section 2

Proof of Proposition 2.1. The result is more natural to prove in terms of true positives ($|R \cap \mathcal{H}_1|$) that false positives ($|R \cap \mathcal{H}_0|$), so we introduce the notation

$S^*(R) = |R| - V^*(R)$ and $\bar{S}(R) := |R| - \bar{V}(R)$. Let us equivalently establish

$$\mathcal{E}(\mathfrak{R}, \mathcal{H}_0) = \{\forall R \subset \mathbb{N}_m, |R \cap \mathcal{H}_1| \geq \bar{S}(R)\} = \{\bar{S}(\mathcal{H}_0) = 0\} \quad (\text{S-15})$$

$$= \{\forall R \subset \mathbb{N}_m, |R \cap \mathcal{H}_1| \geq S^*(R)\} = \{S^*(\mathcal{H}_0) = 0\}. \quad (\text{S-16})$$

We show a circular inclusion of the events in (S-15), starting with $\mathcal{E}(\mathfrak{R}, \mathcal{H}_0)$. First, for $1 \leq k \leq K$, $|R_k \cap \mathcal{H}_0| \leq \zeta_k$, implies that for any R ,

$$|R \cap \mathcal{H}_1| + \zeta_k \geq |R \cap \mathcal{H}_1| + |R_k \cap \mathcal{H}_0| \geq |R \cap R_k|,$$

which entails $|R \cap \mathcal{H}_1| \geq \bar{S}(R)$ by taking a maximum over all possible values of k . Secondly, if $\forall R, |R \cap \mathcal{H}_1| \geq \bar{S}(R)$, then for $R = \mathcal{H}_0$, we obtain $\bar{S}(\mathcal{H}_0) = 0$. Finally, $\bar{S}(\mathcal{H}_0) = 0$ implies that for all k , $|\mathcal{H}_0 \cap R_k| - \zeta_k \leq 0$.

Similarly for (S-16), the event $\mathcal{E}(\mathfrak{R}, \mathcal{H}_0)$ is by definition equivalent to $\mathcal{H}_0 \in \mathcal{A}(\mathfrak{R})$. This implies for any R : $|R \cap \mathcal{H}_1| = |R \setminus \mathcal{H}_0| \geq S^*(R)$. Specializing for $R = \mathcal{H}_0$ as above, this entails $S^*(\mathcal{H}_0) = 0$. Finally, the latter event implies in turn that there must exist $A \in \mathcal{A}(\mathfrak{R})$ with $\mathcal{H}_0 \subset A$, but since any subset of an element of $\mathcal{A}(\mathfrak{R})$ also belongs to $\mathcal{A}(\mathfrak{R})$, we conclude $\mathcal{H}_0 \in \mathcal{A}(\mathfrak{R})$. \square

Proof of Proposition 2.3. We prove that the specific subproblem of computing $V_{\mathfrak{R}}^*(R)$ under the following restrictions is already NP-hard:

- $|R_k| = 2$ for all k ;
- $\zeta_k = 1$ for all k ;
- $R = \mathbb{N}_m$.

Namely, we can formally embed as an instance of this setting the well-known NP-complete problem of finding a maximal independent set of vertices in an arbitrary graph G , in the following way: let K be the number of edges in the graph; construct the family of sets by associating to each edge e of G the set R_e containing the two vertices it joins, and $\zeta_e = 1$. Then elements of $\mathcal{A}(\mathfrak{R})$ are exactly the subsets of independent vertices of G , that is, the subsets that do not contain a pair of vertices connected by an edge. Taking $R = \mathbb{N}_m$, computing $V^*(R) = \max_{A \in \mathcal{A}(\mathfrak{R})} |A|$ is then equivalent to finding the maximal size of an independent vertex set in G . \square

Proof of Proposition 2.4. Obviously, $\tilde{\zeta}_k \leq \zeta_k$ and thus $\bar{V}_{\tilde{\mathfrak{R}}}(R) \leq \bar{V}_{\mathfrak{R}}(R)$. Let us prove the reverse inequality:

$$\begin{aligned} \bar{V}_{\tilde{\mathfrak{R}}}(R) &= \min_{k \in \{1, \dots, K\}} \left(|R \setminus R_k| + \min_{j \in \{1, \dots, K\}} (|R_k \setminus R_j| + \zeta_j) \wedge |R_k| \right) \wedge |R| \\ &= \min_{j, k \in \{1, \dots, K\}} (|R \setminus R_k| + |R_k \setminus R_j| + \zeta_j) \wedge |R| \\ &\geq \min_{j \in \{1, \dots, K\}} (|R \setminus R_j| + \zeta_j) \wedge |R|, \end{aligned}$$

where we used $|E \setminus F| + |F \setminus G| \geq |E \setminus G|$. \square

Proof of Proposition 2.5. For convenience, we recall the notation

$$\mathcal{A}(\mathfrak{R}) := \{A \subset \mathbb{N}_m : \forall k = 1, \dots, K, |R_k \cap A| \leq \zeta_k\}$$

in the definition of V^* . Let $R \subset \mathbb{N}_m$; it is straightforward to check that $V_{\mathfrak{R}}^*(R) \leq \bar{V}_{\mathfrak{R}}(R)$, since V^* is optimal; in fact for all $A \in \mathcal{A}$ and $k \in \{1, \dots, K\}$, we have $|R \cap A| \leq |R \cap A \cap R_k| + |R \cap A \cap R_k^c| \leq (\zeta_k + |R \cap R_k^c|) \wedge |R|$. We now prove the reverse inequality, by showing that there exists a set $A \in \mathcal{A}(\mathfrak{R})$ such that $A \subset R$ and $|A| \geq \bar{V}_{\mathfrak{R}}(R)$. For this, let $\tilde{\zeta}_k$ be defined as in (10) applied to the family $(R_k \cap R, \zeta_k)$, $1 \leq k \leq K$. Formally,

$$\tilde{\zeta}_k = \min_{1 \leq j \leq K} \{ |(R_k \cap R) \setminus (R_j \cap R)| + \zeta_j \} \wedge |R_k \cap R|, \quad 1 \leq k \leq K,$$

which means that (12) is satisfied and in particular

$$\tilde{\zeta}_k - \tilde{\zeta}_{k-1} \leq |(R_k \cap R) \setminus (R_{k-1} \cap R)|, \quad 1 \leq k \leq K,$$

with the conventions $\tilde{\zeta}_0 = 0$ and $R_0 = \emptyset$. Now construct a set A by picking $\tilde{\zeta}_k - \tilde{\zeta}_{k-1}$ elements in each $(R_k \cap R) \setminus (R_{k-1} \cap R)$ for $1 \leq k \leq K$ (which is possible by the latter display) and add the elements of $R \setminus (R_K \cap R)$. We now check that A satisfies the constraints ensuring $A \in \mathcal{A}(\mathfrak{R})$, using the nestedness assumption and the fact that $A \subset R$ by construction, for all $k \in \{1 \dots K\}$,

$$|R_k \cap A| = |R_k \cap R \cap A| = \sum_{j=1}^k |(R_j \cap R \cap A) \setminus (R_{j-1} \cap R \cap A)| = \tilde{\zeta}_k \leq \zeta_k.$$

Moreover, by Proposition 2.4, $\bar{V}_{\mathfrak{R}}(R) \leq |R \setminus (R_K \cap R)| + \tilde{\zeta}_K$. Therefore, $\bar{V}_{\mathfrak{R}}(R) \leq |R \setminus (R_K \cap R)| + |R_K \cap R \cap A| = |A|$, and the result is proved. \square

S-6.2. Proofs for Section 5

Proof of Proposition 5.5. Consider the event Ω for which

$$\forall k \in \{1, \dots, K\}, \quad p_{(k:\mathcal{H}_0)}(X) \geq t_k(\lambda(\alpha, \mathcal{H}_0)), \quad (\text{S-17})$$

which occurs with probability at least $1 - \alpha$ by (24). Now, since $t_1(\lambda(\alpha, \cdot))$ is a non-increasing function on the subsets of \mathbb{N}_m , we have on Ω , for all $j \geq 0$,

$$\mathcal{H}_0 \subset A^{(j-1)} \Rightarrow p_{(1:\mathcal{H}_0)}(X) \geq t_1(\lambda(\alpha, A^{(j-1)})) \Rightarrow \mathcal{H}_0 \subset A^{(j)},$$

and thus $\mathcal{H}_0 \subset \hat{A}$, which itself entails

$$\forall k \in \{1, \dots, K\}, \quad p_{(k:\mathcal{H}_0)}(X) \geq t_k(\lambda(\alpha, \hat{A})).$$

Since Ω is of probability at least $1 - \alpha$, the result is proved. \square

Second proof of Theorem 5.8. Let $\mathcal{G}' = (g_1, g_2, \dots, g_B)$. We denote in this proof $\lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')$ instead of $\lambda(\alpha, \mathcal{H}_0)$ to underline the dependence of this functional w.r.t. the data X and the subset $\mathcal{G}' = (g_1, g_2, \dots, g_B)$. By Section A, it is sufficient to prove

$$\mathbb{P}\left(\Psi(X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')\right) \leq \alpha. \quad (\text{S-18})$$

We use here an elegant technique recently proposed by [Hemerik and Goeman \(2017\)](#). Consider an independent variable $U \in \{1, \dots, B\}$ uniformly distributed. We easily check that, for any $j \in \{1, \dots, B\}$, $\mathcal{G}' \cdot g_j^{-1}$ has the same distribution as \mathcal{G}' (up to permutation of the elements), which entails that $\mathcal{G}' \cdot g_U^{-1}$ has the same distribution as \mathcal{G}' (up to permutation of the elements). Now, since the functional $\lambda(\alpha, X, \mathcal{H}_0, \cdot)$ is invariant by permutation, $\lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}' \cdot g_U^{-1})$ has the same distribution as $\lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')$ (conditionally on X). As a consequence, the LHS of (S-18) is equal to

$$\mathbb{P}\left(\Psi(X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}' \cdot g_U^{-1})\right) = \mathbb{P}\left(\Psi(g_U \cdot X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')\right),$$

where we used that, by [\(Rand\)](#),

$$\begin{aligned} & (p_{\mathcal{H}_0}(X), p_{\mathcal{H}_0}(g_1 \cdot g_U^{-1} \cdot X), p_{\mathcal{H}_0}(g_2 \cdot g_U^{-1} \cdot X), \dots, p_{\mathcal{H}_0}(g_B \cdot g_U^{-1} \cdot X)) \\ & \sim (p_{\mathcal{H}_0}(g_U \cdot X), p_{\mathcal{H}_0}(g_1 \cdot X), p_{\mathcal{H}_0}(g_2 \cdot X), \dots, p_{\mathcal{H}_0}(g_B \cdot X)), \end{aligned}$$

because it is true conditionally on \mathcal{G}' and U . Now, the result follows because

$$\begin{aligned} & \mathbb{P}\left(\Psi(g_U \cdot X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}') \mid X, \mathcal{G}'\right) \\ & = B^{-1} \sum_{j=1}^B \mathbb{1}\{\Psi(g_j \cdot X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')\} \leq \alpha \end{aligned}$$

by definition of $\lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}')$. □

S-6.3. Proofs for Section S-2

Proof of Lemma S-2.1. Let $\lambda_0 = 1/(\log m)^{1/4}$ and consider U_1, \dots, U_m i.i.d. $\sim U(0, 1)$. By definition of $\lambda^B(\alpha)$, it is sufficient to prove that for m large enough, the probability $\mathbb{P}(\exists k \in \{1, \dots, m\} : U_{(k:m)} < t_k(\lambda_0))$ is larger than α .

For this, use the lower bound (S-8) to write for a large m ,

$$\begin{aligned}
& \mathbb{P}(\exists k \in \{1, \dots, m\} : U_{(k:m)} < t_k(\lambda_0)) \\
&= \mathbb{P}(\exists k \in \{1, \dots, m\} : U_{(k:m)} \leq t_k(\lambda_0)) \\
&\geq \mathbb{P}\left(\exists k \in \{1, \dots, m\} : U_{(k:m)} \leq \right. \\
&\quad \left. \frac{k}{m+1} - \left\{ \frac{k}{m+1} \left(1 - \frac{k}{m+1}\right) \right\}^{1/2} m^{-1/2} (4 \log(1/\lambda_0))^{1/2} \right) \\
&= \mathbb{P}\left(Z_m \geq (4 \log(1/\lambda_0))^{1/2}\right),
\end{aligned}$$

where we let

$$Z_m = \max_{1 \leq k \leq m} \left\{ \frac{m^{1/2}}{\left\{ \frac{k}{m+1} \left(1 - \frac{k}{m+1}\right) \right\}^{1/2}} \left(\frac{k}{m+1} - U_{(k:m)} \right) \right\}.$$

Since $(4 \log(1/\lambda_0))^{1/2} = (\log \log m)^{1/2}$, we conclude by applying Lemma S-7.2. \square

Proof of Lemma S-2.2. Let $N \in \{1, \dots, m_0 - 1\}$ be some integer to be chosen later. By a union bound argument, we have

$$\begin{aligned}
& \mathbb{P}\left(\exists k \in \{1, \dots, m_0\} : q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right) \\
&\leq \sum_{k=1}^N \mathbb{P}\left(q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right) + \sum_{k=N+1}^{m_0} \mathbb{P}\left(q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right).
\end{aligned}$$

For the first term, since $q_{(k:m)}$ is stochastically smaller than $q_{(k:m_0)}$, we have

$$\sum_{k=1}^N \mathbb{P}\left(q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right) \leq N \lambda^B(\alpha) \leq N / (\log m)^{1/4},$$

by Lemma S-2.1. For the second term, by (S-9) ($\lambda^B(\alpha)$ begin smaller than 0.5 for large enough m by S-6) and letting $r = e^{-\frac{(1-\pi_0)^2}{32}}$, we have for large enough m :

$$\begin{aligned}
& \sum_{k=N+1}^{m_0} \mathbb{P}\left(q_{(k:m_0)} < t_k(\lambda^B(\alpha))\right) \\
&\leq \sum_{k=N+1}^{m_0} r^k \leq (1-r)^{-1} r^N \leq (1-r)^{-1} \frac{32}{N(1-\pi_0)^2},
\end{aligned}$$

because $e^{-u} \leq 1/u$ for all $u > 0$. Choosing $N = \lfloor (\log m)^{1/8} \rfloor$ yields the desired result. \square

Proof of Proposition S-2.3. Item (i) just follows from the definition. The proof of item (ii) is straightforward from a classical bound for the Beta distribution, see relation (S-21) and Lemma S-7.3 in Section S-7. For item (iii), we use item (ii) and $m/(m+1) \geq 1/2$ and $m_0/(m_0+1) \geq 1/2$, to write

$$\begin{aligned} & \mathbb{P} \left[p_{(k:m_0)} \leq t_k(\lambda) \right] \\ & \leq \mathbb{P} \left[p_{(k:m_0)} \leq \frac{k}{m+1} \right] \\ & = \mathbb{P} \left[m_0^{1/2} \left(p_{(k:m_0)} - \frac{k}{m_0+1} \right) \leq -(k)^{1/2} (km_0)^{1/2} \left(\frac{1}{m_0+1} - \frac{1}{m+1} \right) \right] \\ & \leq \mathbb{P} \left[m_0^{1/2} \left(p_{(k:m_0)} - \frac{k}{m_0+1} \right) \leq - \left\{ \frac{k}{m_0+1} \right\}^{1/2} (k/2)^{1/2} \left(\frac{m-m_0}{m+1} \right) \right] \\ & \leq \mathbb{P} \left[m_0^{1/2} \left(p_{(k:m_0)} - \frac{k}{m_0+1} \right) \leq - \left\{ \frac{k}{m_0+1} \right\}^{1/2} (k/2)^{1/2} \left(1 - \frac{m_0}{m} \right) / 2 \right], \end{aligned}$$

and we conclude by using (S-21). For (iv), the reasoning is similar, using

$$\begin{aligned} & \mathbb{P} \left[p_{(k:m)} \leq \alpha \frac{k}{m} \right] \\ & = \mathbb{P} \left[m^{1/2} \left(p_{(k:m)} - \frac{k}{m+1} \right) \leq - \left\{ \frac{k}{m} \right\}^{1/2} \sqrt{k} \left(\left(1 - \frac{1}{m+1} \right) - \alpha \right) \right]. \end{aligned}$$

□

S-6.4. Proofs for Section S-3

Proof of Theorem S-3.2. For proving (i), we note that any family \mathfrak{R} with thresholds t_k , $1 \leq k \leq m$, controlling the JER at level α induces a test $\varphi = \mathbb{1} \{ \exists k : p_{(k)} < t_k \}$ of level α of $H_0 : “\mu_m = 0”$ (i.e., $p_i, i \in \mathbb{N}_m$ are all i.i.d. uniform) against $H_1 : “\mu_m = \sqrt{2r \log m}”$. Hence, it will have less power than the likelihood ratio test (LRT) of level α . Now, as claimed in Section 1.1 of [Donoho and Jin \(2004\)](#) (itself referring to [Ingster, 1999](#)), the null hypothesis and the alternative hypothesis merge asymptotically whenever $r < \rho^*(\beta)$. Hence, the asymptotic power of the LRT is less than α .

Now consider the balanced family \mathfrak{R}_α^B and prove (ii). Write t_k for $t_k^B(\lambda^B(\alpha))$ for simplicity. The basic inequality for our proof is the following: for any $k \in \{1, \dots, m\}$,

$$\text{Pow}^*(\mathfrak{R}_\alpha^B, P) \geq \mathbb{P} (U_{(k:m)} \leq \pi_{0,m} t_k + \pi_{1,m} F_{1,m}(t_k)). \quad (\text{S-19})$$

From (S-13), now write for any $k \in \{1, \dots, m\}$,

$$\begin{aligned}
\text{Pow}^*(\mathfrak{R}_\alpha^B, P) &\geq \mathbb{P}\left(U_{(k:m)} \leq \frac{k}{m+1}\right. \\
&\quad \left. - \left\{\frac{k}{m+1} \left(1 - \frac{k}{m+1}\right)\right\}^{1/2} \left(\frac{4 \log(m/\alpha)}{m}\right)^{1/2}\right. \\
&\quad \left. + \pi_{1,m} (F_{1,m}(t_k) - t_k)\right) \\
&\geq \mathbb{P}\left(\max_{1 \leq k \leq m} \left\{m^{1/2} \frac{U_{(k:m)} - \frac{k}{m+1}}{\left\{\frac{k}{m+1} \left(1 - \frac{k}{m+1}\right)\right\}^{1/2}}\right\}\right. \\
&\quad \left.\leq - (4 \log(m/\alpha))^{1/2} + \pi_{1,m} \frac{F_{1,m}(t_k) - t_k}{k^{1/2}/(m+1)}\right), \tag{S-20}
\end{aligned}$$

because $k^{1/2}/(m+1) \geq \left\{\frac{k}{m+1} \left(1 - \frac{k}{m+1}\right)\right\}^{1/2} m^{-1/2}$. Let $r_k > 0$ such that $\bar{\Phi}^{-1}(t_k) = \sqrt{2r_k \log m}$. Then, choosing $k = \lfloor m^{1-q} \log m \rfloor$ for some $q \in (0, 1)$, we have $r_k \rightarrow q$ as m tends to infinity. To see this, first note that $\lambda^B(\alpha) \rightarrow 0$ by Lemma S-2.1, and that $\lambda^B(\alpha) \geq \alpha/m$ by (S-5). Therefore, (S-8) in Proposition S-2.3 entails that $t_k \sim k/m$. Then, recalling that $\bar{\Phi}^{-1}(u) \sim \sqrt{2 \log 1/u}$ as $u \rightarrow 0$, our choice of r_k yields $r_k \rightarrow q$. Furthermore, denoting by ϕ the density of the standard gaussian distribution, we have for $q > r$:

$$\begin{aligned}
\pi_{1,m} F_{1,m}(t_k) &= m^{-\beta} \bar{\Phi} \left(\bar{\Phi}^{-1}(t_k) - \sqrt{2r \log m} \right) \\
&= m^{-\beta} \bar{\Phi} \left(r_k^{1/2} - r^{1/2} \right) \sqrt{2 \log m} \\
&\sim D m^{-\beta} \phi \left((q^{1/2} - r^{1/2}) \sqrt{2 \log m} \right) / \sqrt{\log m} \\
&\sim D' m^{-\beta - (q^{1/2} - r^{1/2})^2} / \sqrt{\log m},
\end{aligned}$$

for some constants $D, D' > 0$. This entails

$$\pi_{1,m} \frac{F_{1,m}(t_k) - t_k}{k^{1/2}/(m+1)} \sim D' m^{\frac{1+q}{2} - \beta - (q^{1/2} - r^{1/2})^2} / \log m,$$

Let $f(q) = \frac{1+q}{2} - \beta - (q^{1/2} - r^{1/2})^2$. Since $f(q_0) > 0$ for $q_0 = (4r) \wedge 1$ and $r > \rho^*(\beta)$, by continuity of f there exists $q \in (0, 1)$ such that $f(q) > 0$. Now, (ii) comes from (S-20) and the fact that the sequence of random variables

$$\max_{1 \leq k \leq m} \left\{ m^{1/2} \frac{U_{(k:m)} - \frac{k}{m+1}}{\left\{\frac{k}{m+1} \left(1 - \frac{k}{m+1}\right)\right\}^{1/2}} \right\} / (\log \log m)^{1/2}$$

is tight (see Lemma S-7.2).

□

Proof of Theorem S-3.3. Let us first prove (i). For any $k \in \{1, \dots, m\}$,

$$\text{Pow}^*(\mathfrak{R}_\alpha^0, P) \geq \mathbb{P} \left(U_{(k:m)} \leq \frac{k}{m} (\pi_{0,m}\alpha + m\pi_{1,m}F_{1,m}(\alpha k/m)/k) \right)$$

Let $r_k > 0$ such that $\bar{\Phi}^{-1}(\alpha k/m) = \sqrt{2r_k \log m}$, so that when $k = \lfloor \log m \rfloor$, $r_k \rightarrow 1$ as m tends to infinity. Then we have for some universal constant $D > 0$,

$$m\pi_{1,m}F_{1,m}(\alpha k/m)/k \sim Dm^{1-\beta-(1-r^{1/2})^2}/(\log m)^{3/2},$$

and thus the latter tends to infinity. Hence, for any $M > 0$, for m large enough, we have

$$\text{Pow}^*(\mathfrak{R}_\alpha^0, P) \geq \mathbb{P} \left(U_{(k:m)} \leq M \frac{k}{m} \right).$$

Then (i) is proved because $mU_{(k:m)}/k$ tends to 1 in probability. Now, let us show (ii). We have

$$\begin{aligned} \text{Pow}^*(\mathfrak{R}_\alpha^0, P) &= \mathbb{P} \left(\exists k \in \{1, \dots, m\} : U_{(k:m)} \leq \alpha \frac{k}{m} \left(\pi_{0,m} + \pi_{1,m} \frac{F_{1,m}(\alpha k/m)}{\alpha k/m} \right) \right) \\ &\leq \mathbb{P} \left(\exists k \in \{1, \dots, m\} : U_{(k:m)} \leq \alpha \frac{k}{m} \left(\pi_{0,m} + \pi_{1,m} \frac{F_{1,m}(\alpha/m)}{\alpha/m} \right) \right) \end{aligned}$$

because $F_{1,m}(x)/x$ is decreasing. Now, we have

$$\pi_{1,m} \frac{F_{1,m}(\alpha/m)}{\alpha/m} \sim Dm^{1-\beta-(1-r^{1/2})^2}/(\log m)^{1/2},$$

for some universal constant $D > 0$, and thus the latter tends to zero as soon as $r < (1 - \sqrt{1 - \beta})^2$. Hence, for any $\varepsilon \in (0, 1)$, for m large enough,

$$\text{Pow}^*(\mathfrak{R}_\alpha^0, P) \leq \mathbb{P} \left(\exists k \in \{1, \dots, m\} : U_{(k:m)} \leq \alpha \frac{k}{m} (1 + \varepsilon) \right) \leq \alpha(1 + \varepsilon),$$

by applying the Simes inequality. The result comes by making ε tends to zero. \square

S-7. Some properties of the Beta distribution

We recall the following result (Shorack and Wellner, 1986, p.454-455):

Lemma S-7.1. *for U_1, \dots, U_n i.i.d. $\sim U(0, 1)$, any $\ell \in \{1, \dots, n\}$ and $x \geq 1$, we have*

$$\mathbb{P} \left(n^{1/2} \left(U_{(\ell:n)} - \frac{\ell}{n+1} \right) \leq - \left\{ \frac{\ell}{n+1} \left(1 - \frac{\ell}{n+1} \right) \right\}^{1/2} x \right) \leq e^{-x^2/4} \quad (\text{S-21})$$

Here is another lemma, which is a consequence of (24) in (Shorack and Wellner, 1986, p.601):

Lemma S-7.2. *Let U_1, U_2, \dots i.i.d. $\sim U(0, 1)$ and consider*

$$Z_n = \max_{1 \leq \ell \leq n} \left\{ n^{1/2} \frac{\frac{\ell}{n+1} - U_{(\ell:n)}}{\left\{ \frac{\ell}{n+1} \left(1 - \frac{\ell}{n+1} \right) \right\}^{1/2}} \right\},$$

then we have, as n grows to infinity,

$$\mathbb{P}((\log \log n)^{1/2} \leq Z_n \leq 2(\log \log n)^{1/2}) \rightarrow 1 \quad (\text{S-22})$$

Lemma S-7.3. *Let U_1, U_2, \dots i.i.d. $\sim U(0, 1)$, then, for all $m \geq 2$, for all $k \in \{1, \dots, m\}$,*

$$\mathbb{P}(U_{(k:m)} \leq (k+1)/m) \geq 0.5. \quad (\text{S-23})$$

Proof. We can assume $k \leq m-1$. Now, by considering $Z \sim \mathcal{B}(m, (k+1)/m)$, we have

$$\mathbb{P}(U_{(k:m)} \leq (k+1)/m) = \mathbb{P}(Z \geq k) = \mathbb{P}(Z \geq (k+1) - 1) \geq 0.5,$$

where we used that for any binomial distribution, the median and the mean are at a distance at most 1 (see, e.g., Kaas and Buhrman, 1980). \square

References

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.
- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994.
- Goeman, J. J. and Solari, A. (2011). Multiple testing for exploratory research. *Statist. Sci.*, 26(4):584–597.
- Hemerik, J. and Goeman, J. J. (2017). False discovery proportion estimation by permutations: confidence for significance analysis of microarrays. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Huang, Y. and Hsu, J. C. (2007). Hochberg’s step-up method: cutting corners off Holm’s step-down method. *Biometrika*, 94(4):965–975.
- Ingster, Y. I. (1999). Minimax detection of a signal for l^n -balls. *Math. Methods Statist.*, 7(4):401–428.
- Kaas, R. and Buhrman, J. (1980). Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18.
- Marcus, R., Peritz, E., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660.
- Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.*, 100(469):94–108.

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