Submitted to the Annals of Statistics

SUPPLEMENT TO "POST HOC CONFIDENCE BOUNDS ON FALSE POSITIVES USING REFERENCE FAMILIES"

By Gilles Blanchard, Pierre Neuvial and Etienne Roquain

This supplement provides proofs of results of the main paper, as well as additional material: detailed discussion of relation to previous work (augmentation procedures of van der Laan et al. (2004), inversion method of Genovese and Wasserman, 2006, closed testing of Goeman and Solari, 2011, higher criticism of Donoho and Jin, 2004); complements on JER control based on Simes and Hommel inequalities; general properties of templates and reference families; algorithmic details concerning Monte-Carlo and permutation-based calibration and additional numerical experiments.

CONTENTS

This supplement is organized as follows. In Section S-1, we elaborate on the connection between JER control and previous work, more specifically the confidence envelope approach of Genovese and Wasserman (2004, 2006), Meinshausen and Bühlmann (2005), Meinshausen (2006) as well as the closed testing approach of Goeman and Solari (2011) to user-agnostic inference. We provide complements on JER control based on Simes and Hommel inequalities in Sections S-2, and on the properties of linear and balanced templates in Section S-3. In Section S-4, 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-5 establishes the validity of assumption (Rand) in permutation-based two-sample multiple testing problems. In Section S-6, we explain how to approximate the balanced thresholds under known dependence using a Monte-Carlo method. Finally, Section S-7 gathers the proofs of the results stated in the main paper and in this supplement.

S-1. Relation to previous work on confidence envelopes and user-agnostic bounds. The present work was inspired in particular by the seminal works of Genovese and Wasserman (2006) (GW06 below for short) and Goeman and Solari (2011) (GS11 below for short), both of which provide confidence bounds on the number of false positives $|\mathcal{H}_0 \cap R|$ uniformly over any rejection region R, and thus a user-agnostic control of the number (or proportion) of false discoveries. Such bounds are called "confidence envelopes" by GW06 and "post hoc bounds" by GS11. As stated in the introduction, related ideas where used by Meinshausen and Bühlmann (2005), Meinshausen (2006) when the coverage was restricted to selection sets R that are p-value level sets, that is of the form $R = \{i : p_i \leq t\}$, for some $t \in [0, 1]$. In this section, we elucidate the relation to these works, first between each other, then to ours.

To give an overview, following the terminology used by GW06 we first need to introduce a distinction between

- "Inversion procedures", based on family of local tests for all intersection hypotheses, which also form the basis over which GS11 is built;
- "Augmentation procedures", based on the control of the k-family-wise error rate on a specific set R_k, and an "interpolation" from this control to all possible rejection sets.

In a nutshell, we argue that the JER approach advocated in the present paper suppresses this distinction and encompasses both approaches as particular cases. More precisely, the inversion procedure can be seen as a particular case of the bound V^*, and the bound $\overline{V}$ can be seen as an extension of the augmentation principle.

We organize this section as follows. We start with a comparison with augmentation procedures and approaches based on uniform confidence bounds on the distribution function of null p-values in Section S-1.1. We then turn to the comparison to inversion procedures based on local intersection tests, which are presented in detail in Section S-1.2. There we remark that the closed testing approach of GS11 is essentially equivalent to GW06's inversion procedure[1]. In the following sections, we discuss connections between these approaches and ours. In Section S-1.3, we show that the probabilistic guarantee derived from the inversion procedure of GW06 or closed testing in GS11 can be equivalently obtained by JER control of a specific form combined with the optimal bound V^*. Finally, in Section S-1.4, we prove that the shortcut bound obtained by GS11 for certain Simes-type local tests can be obtained equivalently by our approach via JER control using a Simes-type template with $\zeta_k = k-1$.

S-1.1. *Augmentation procedures and confidence envelopes on p-value levels sets.* The "augmentation procedure" of Genovese and Wasserman (2006, Theorems 1,2,3) is based on a single reference set R_k with controlled k-family-wise error rate (extending an idea due to van der Laan et al. 2004, initially considered for $k=1$). These authors then propose the bound (called confidence envelope) $V^{\mathrm{aug}}(R) = \max(|R \setminus R_k| + (k-1), |R|)$, which is a particular case of our bound $\overline{V}(R)$ given by (8), when the reference family only consists of one element $(R_k, \zeta_k = k-1)$. Obviously, with a larger reference family enjoying joint error rate control, our bound $\overline{V}(R)$ is obtained by taking the minimum of all possible such augmentation bounds over elements of the reference family.

The term confidence envelope was also used in relation to bounds based on a uniform upper confidence bound over the empirical distribution function of null p-values (Genovese and Wasserman, 2004; Meinshausen and Bühlmann, 2005; Meinshausen, 2006) (also called "quantile bounding function" in the two latter references). The principle is the following: let us denote p-value level sets as $\widetilde{R}_t = \{i : p_i \leq t\}$ and assume we have at hand a bound $B(t)$ such that

$$\mathbb{P}\left(\forall t \in [0,1] \,:\, |\mathcal{H}_0 \cap \widetilde{R}_t| \leq B(t)\right) \geq 1-\alpha.$$

[1]We express our thanks to Aaditya Ramdas for first pointing out this fact to us.

Such a bound can be obtained under the independence assumption by classical inequalities for i.i.d. uniform variables (Genovese and Wasserman, 2004), or by permutation approaches in the case of general dependence (Meinshausen and Bühlmann, 2005; Meinshausen, 2006). Meinshausen (2006) establishes that a simultaneous bound over the false discoveries all $(\widetilde{R}_t)s$ is given by

$$\overline{V}^{\mathrm{Mein}}(\widetilde{R}_t) = |\widetilde{R}_t| - \max_{0\leq\tau\leq t}(|\widetilde{R}_\tau| - B(\tau)).$$

This can be interpreted in our setting as follows. We consider the reference family $(\widetilde{R}_\tau, \zeta_\tau = B(\tau))_{0\leq\tau\leq 1}$ (it is formally an uncountable family, but in practice $B(t)$ is a piecewise constant function and this can be equivalently represented as a finite family). The hypothesis on $B(t)$ implies that this reference family enjoys JER control. Then our bound from (8) is given in this case by

$$\overline{V}(\widetilde{R}_t) = \min_{0\leq\tau\leq 1}(|\widetilde{R}_t \setminus \widetilde{R}_\tau| + B(\tau)) \wedge |\widetilde{R}_t|,$$

and since $|\widetilde{R}_t \setminus \widetilde{R}_\tau| = |\widetilde{R}_t| - |\widetilde{R}_\tau|$ for $\tau \leq t$, we have $\overline{V}(\widetilde{R}_t) \leq \overline{V}^{\mathrm{Mein}}(\widetilde{R}_t)$, with equality under the natural assumptions that $B(\tau)$ is nondecreasing and satisfies $B(\tau) \leq |\widetilde{R}_\tau|$. Hence this is also a particular instance of our general setting.

Finally, note that reference families based on k-FWER controlled sets $(R_k = \{i : p_i \leq t_k\}, \zeta_k = k-1)_{1\leq k\leq m}$ and on regular level sets $(\widetilde{R}_\tau = \{i : p_i \leq \tau\}, \zeta_\tau = \min\{k : t_k \geq \tau\} - 1)_{0\leq\tau\leq 1}$ are equivalent in the sense that they contain the same information and give rise to identical bounds. However, strict equivalence breaks down when considering fixed subranges $\{k \leq K_{\max}\}$, resp. $\{\tau \leq \tau_{\max}\}$ which are more relevant in practice; and from that point of view, we argue that the parametrization by the maximal number of false discoveries $K_{\max}$ is more natural and convenient.

S-1.2. *Local intersection tests setting.* The “inversion procedure” of GW06 as well as the closed testing approach of GS11 are based on the notion of local intersection tests: 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. Let $\mathcal{V}$ denote the collection of individually rejected intersection hypotheses and $\mathcal{U}$ its complement (non rejected intersection hypotheses).

The *inversion* procedure of GW06 returns the bound

$$\overline{V}^{GW}(R) = \max_{J \in \mathcal{U}} |R \cap J| \,.$$

To see why this is a valid bound uniformly over R it suffices to remark that $\Phi_{\mathcal{H}_0}(X) = 0$ holds with probability $1-\alpha$, and on this event $\mathcal{H}_0 \in \mathcal{U}$. (This is indeed a form of the general duality between test and confidence regions, hence the name.)

The *closed testing*-based procedure of GS11 first applies the classical closed testing principle (Marcus et al., 1976) to the family of local intersection tests. In effect, this steps extracts from the set of rejected intersection hypotheses $\mathcal{V}$ a subcollection $\widetilde{\mathcal{V}}$, namely only the rejected intersection hypotheses I such that all $I' \supset I$ are also rejected. After this step, the remaining family $\widetilde{\mathcal{V}}$ of rejected intersection hypotheses, now stable by the superset operation, has family-wise error rate controlled at level α, i.e. denoting $\widetilde{\mathcal{U}}$ the complement of $\widetilde{\mathcal{V}}$ (non rejected intersection hypotheses after extraction step), it holds:

$$\mathbb{P}_{X\sim P}\left(\forall I \subset \mathcal{H}_0(P), I \in \widetilde{\mathcal{U}}\right) \geq 1-\alpha. \tag{S-1}$$

The justification is the same as above: it suffices that the single event $\Phi_{\mathcal{H}_0}(X) = 0$ (of probability $1-\alpha$) is satisfied to ensure that any null intersection hypothesis (which is necessarily a subset of $\mathcal{H}_0$) is also not rejected after the extraction step. The procedure of GS11 then returns the bound

$$\overline{V}^{GS}(R) = \max_{J \in \widetilde{\mathcal{U}}} |R \cap J| = \max\{|J| \,:\, J \in \widetilde{\mathcal{U}}, J \subset R\} \tag{S-2}$$

(the second equality holds because $\widetilde{\mathcal{U}}$ is stable by the subset operation). We observe that the two bounds $\overline{V}^{GW}$ and $\overline{V}^{GS}$ are in fact identical. Since $\mathcal{U} \subset \widetilde{\mathcal{U}}$, the closed testing bound must be larger than or equal to the inversion bound. On the other hand, the closed testing procedure is indeed a closure operation: $\widetilde{\mathcal{U}}$ contains exactly all intersection hypotheses of $\mathcal{U}$ and all their subsets. But adding all subsets does not increase the bound. Hence the two bounds are identical (and the closed testing step is actually not necessary for the GS11 bound).

Finally, note that the NP-hardness result of Proposition 2.2 also applies to the computation of the GS11 bound (S-2) in the following sense: even if the set $\widetilde{\mathcal{U}}$ can be represented in a compact way via the family of its minimal sets $\mathcal{W}$ as $\widetilde{\mathcal{U}} = \text{superset}(\mathcal{W}) = \{I \supset J, J \in \mathcal{W}\}$, the computation of (S-2) is NP-hard with respect to the size of $\mathcal{W}$. This is because the reference families used in the proof of Proposition 2.2 have the specific form $\zeta_k = |R_k| - 1$ and can therefore as well be interpreted as the family $\mathcal{W}$ above.

S-1.3. *Equivalence between JER control and closed testing/inversion bounds.* We discuss a formal equivalence between the JER control approach and post hoc inversion bounds based on local tests of GS11 (S-2) (and therefore equivalently the GW06 bound); in the sense that each approach can be formally represented as an instance of the other.

For this, consider the reference family $\mathfrak{R} = \widetilde{\mathcal{V}} = (R_1, \ldots, R_K)$ returned by the closed testing procedure as in Section S-1.2. The closed testing principle inplies 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

$$\begin{aligned}
\left\{\forall I \subset \mathcal{H}_0(P), I \in \widetilde{\mathcal{U}}\right\}^c &= \left\{\exists I \subset \mathcal{H}_0(P), I \in \mathfrak{R}\right\} \\
&= \left\{\mathcal{H}_0(P) \in \mathfrak{R}\right\} \\
&= \left\{\exists k \in \{1, \ldots, K\} \,:\, |\mathcal{H}_0(P) \cap R_k| = |R_k|\right\} \\
&= \left\{\exists k \in \{1, \ldots, K\} \,:\, |\mathcal{H}_0(P) \cap R_k| > \zeta_k\right\};
\end{aligned}$$

hence (S-1) is indeed the same as the JER control (4)-(5) for this choice of $(R_k, \zeta_k)_{1 \leq k \leq K}$. Concerning the post hoc bound itself, note that

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

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

$$V^*_{\mathfrak{R}}(R) = \max_{A \in \mathcal{A}(\mathfrak{R})} |R \cap A| = \max_{A \in \widetilde{\mathcal{U}}} |R \cap A| = \max_{A \in \widetilde{\mathcal{U}}, A \subset R} |A| = \overline{V}^{GS}(R)\,,$$

since $\widetilde{\mathcal{U}}$ is closed by subset operation.

Conversely, our setting of JER control based on a reference family $\mathfrak{R}$ can be embedded as a particular case of the local intersection test framework. Namely, construct the local tests Φ_I as $\Phi_I(X) = 0$ iff $I \in \mathcal{A}(\mathfrak{R})$. With this representation, it is obvious that the bound obtained via our approach $V^*_{\mathfrak{R}}(R)$ or via the inversion procedure of GW06 coincide (observe namely $\mathcal{U} = \mathcal{A}(\mathfrak{R})$ with this construction). This remark thus generalizes GW06's Theorem 4, which only considered the case of a single reference set with controlled FWER $(R_1, \zeta_1 = 0)$. Similarly, our Proposition 2.5 on the equivalence of V^* and $\overline{V}$ for nested reference families can be interpreted, via the above representation, as a non-trivial extension of GW06's Theorem 5 about the equivalence of augmentation and inversion in specific circumstances.

Is there any interest in the JER approach if it is formally equivalent to the local test inversion bound? Statistically speaking, the equivalence shown above suggests that there is no difference. But we argue that the local test representation can be quite wasteful, and an advantage of the JER control approach is that it allows one to summarize the available information with the reference family alone, while the intersection test setting requires in principle to compute the output of the $(2^m - 1)$ local tests in the first place. Also, even if computation of the optimal bound $V^*(R)$ is possibly hard, the easily computable upper approximation $\overline{V}(R)$ from (8) is always available. Other approximations, e.g. $\widetilde{V}_{\mathfrak{R}}(R)$ (11), can be considered as well. To summarize, the JER point of view allows one to introduce a flexible structure, which in favorable cases (e.g. the reference family is of limited size, and/or has some structural properties such as nestedness, or tree-structure considered by Durand et al., 2018) allows for more transparent representation of the available information and computation of the bounds, features that might be lost when considering the local test equivalent.

In some cases, depending on the specific structure of local tests, so-called *shortcuts* (exact or approximate) are available for the closed testing approach of GS11, allowing to reduce considerably the bulk of computations. In the next section, we analyze a specific shortcut proposed by GS11, and show that it can equivalently be seen as an approximation of the form $\overline{V}(R)$ in our JER-based setting.

S-1.4. *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:

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

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

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

where $S_r = \max\{0 \le s \le r-1 \,:\, p_{(r:R)} \le c^m_{r-s}\}$ (with $\max \emptyset = -\infty$). We argue below in Lemma S-1.2 that this bound is in fact equivalent to the post hoc bound $\overline{V}_{\mathfrak{R}}(R)$ defined in (8), for the family $\mathfrak{R} = (R_k, \zeta_k := k-1)_{1 \le k \le m}$ defined by

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

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

LEMMA S-1.1. *Assume that the tests $(\phi_I)_{I \subset I_m}$ form a family of local tests at level α for the considered model, i.e., for any $P \in H_{0,I}$, it holds $\mathbb{P}_{X \sim P}(\phi_I(X) = 1) \le \alpha$. 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 (4) holds for the reference family $\mathfrak{R} = (R_k, \zeta_k := k-1)_{1 \le k \le m}$.*

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

$$\mathbb{P}_{X \sim P}\left[\exists k \le m_0 : p_{(k:\mathcal{H}_0)} \le c^{m_0}_k\right] \le \alpha\,,$$

implying by the monotonicity assumption $c_i^\ell \le c_i^j$ for $\ell \ge j$:

$$\mathbb{P}_{X \sim P}\left[\exists k \le m_0 : p_{(k:\mathcal{H}_0)} \le c^{m}_k\right] \le \alpha\,.$$

As we argued in Section 2.4, this is equivalent to $\text{JER}(\mathfrak{R}, P) \le \alpha$ for the threshold-based reference family $\mathfrak{R} = (R_k, k-1)_{1 \le k \le m}$ using thresholds $t_k := c^m_k, k \in \mathbb{N}_m$, see (2). □

Now, we establish the equivalence of the two bounds:

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

PROOF. The result comes from

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

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 $\mathfrak{R}_0$ (S-7), and the bound $\overline{V}^{GS}_{\text{Simes}}(R)$ has the simple equivalent form given by (S-8).

S-2. JER control based on classical inequalities. In this section, we present an elementary approach where JER control (4) is derived from probabilistic inequalities that are well-known in multiple testing literature.

PROPOSITION S-2.1 (Simes and Hommel inequalities). *Let $(p_i(X))_{i\in\mathbb{N}_m}$ be a p-value family for the null hypotheses $(H_{0,i})_{i\in\mathbb{N}_m}$, satisfying the characteristic property*

$$\forall P \in \mathcal{P}, \forall i \in \mathcal{H}_0(P), \quad \forall t \in [0,1],\ \mathbb{P}_{X\sim P}(p_i(X) \leq t) \leq t. \tag{S-5}$$

Then it holds that $\forall P \in \mathcal{P}$,

$$\mathbb{P}_{X\sim P}\left(\exists k \in \{1,\dots,m_0\} \,:\, p_{(k:\mathcal{H}_0)} \leq \frac{\alpha k}{m_0 c_m}\right) \leq \alpha\,, \tag{S-6}$$

where:

(i) $c_m = C_m := \sum_{i=1}^m 1/i$ under arbitrary dependency of the p-value family;
(ii) $c_m = 1$ if for all $P \in \mathcal{P}$, the p-value family is positive regression dependent on each element of the subset $\mathcal{H}_0(P)$ (in short, PRDS on $\mathcal{H}_0(P)$).

Moreover, (S-6) is an equality (with $c_m = 1$) when the p_i, $i \in \mathcal{H}_0(P)$, are i.i.d. $U(0,1)$.

The inequalities corresponding to items (i) and (ii) are often referred to as the Hommel inequality (Hommel, 1983) and the Simes inequality (Simes, 1986), respectively. We refer to Benjamini and Yekutieli (2001) for a formal definition of the PRDS property. We recall that in the Gaussian model defined in Section 3.1 (one-sided), the PRDS assumption is valid if $\Sigma_{i,j} \geq 0$ for all $i, j \in \mathbb{N}_m$.

In view of (2), inequality (S-6) implies that the JER control (4) is satisfied for $K = m$ (under the appropriate conditions) by the Simes reference family $\mathfrak{R}^0 = (R_1^0(X), \dots, R_m^0(X))$ given by

$$R_k^0(X) = \left\{ i \in \mathbb{N}_m \,:\, p_i < \frac{\alpha k}{m c_m} \right\}, 1 \leq k \leq m. \tag{S-7}$$

Above, we have upper-bounded m_0 by m because m_0 is generally unknown. The associated post hoc bound

$$\overline{V}_{\mathfrak{R}^0}(R) = \min_{k\in\{1,\dots,m\}} \left\{ \sum_{i\in R} \mathbb{1}\left\{p_i(X) \geq \alpha k/m\right\} + k - 1 \right\}, \quad R \subset \mathbb{N}_m \tag{S-8}$$

corresponds to the bound (3) for the choice $t_k = \alpha k/m$.

The Hommel bound shares some similarity with the procedure introduced in Genovese and Wasserman (2006, Theorem 10), that is designed to work under arbitrary dependence and for which the uniformity in α is solved by an union bound. However, these procedures are generally conservative: for instance, the Hommel bound is achieved in a very unrealistic negative dependent case, see, e.g., Lehmann and Romano (2005).

REMARK S-2.2 (π_0-adaptive version of the Simes bound). A consequence of the Simes bound (S-8) is that $\widehat{m}_0 = \overline{V}_{\mathfrak{R}^0}(\mathbb{N}_m)$ satisfies $\widehat{m}_0 \geq m_0$ on the event described in (S-6). Therefore, the bound

$$\min_{k\in\{1,\dots,\widehat{m}_0\}} \left\{ \sum_{i\in R} \mathbb{1}\left\{p_i(X) \geq \alpha k/\widehat{m}_0\right\} + k - 1 \right\}$$

for $R \subset \mathbb{N}_m$, which is obtained by replacing m by $\widehat{m}_0 = \overline{V}_{\mathfrak{R}^0}(\mathbb{N}_m)$ in the Simes bound (S-8), is a slight but uniform improvement of (S-8). A closely related bound based on Hommel (1988) has recently been proposed by Goeman et al. (2016).

S-3. Properties of linear and balanced procedures.

S-3.1. *Magnitude of $\lambda(\alpha, \mathbb{N}_m)$.* Consider the case of known dependence (therefore, with a λ-calibration given by (19)) 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 Proposition S-2.1, which means that $\mathfrak{R}^L$ reduces to the Simes reference family $\mathfrak{R}^0$ (S-7). 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 (19), as already noted in Section 4.1.

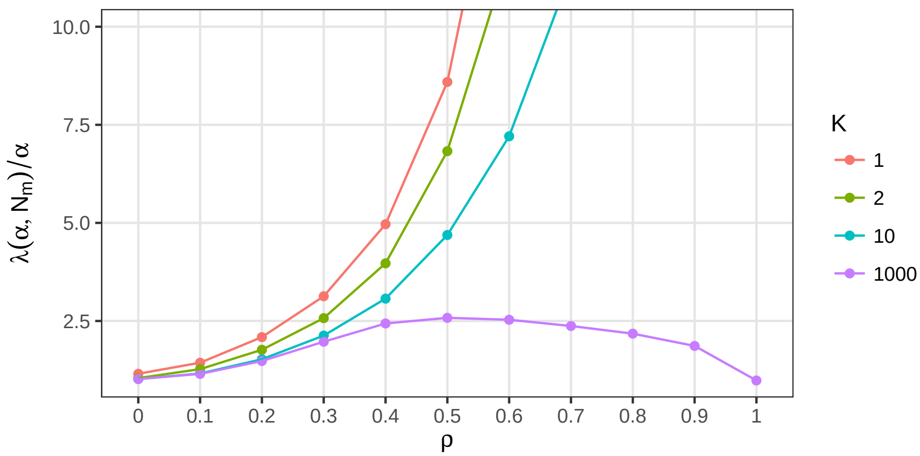

FIG S-1. *Influence of the equi-correlation level* ρ *on the adjustment factor* $\lambda^L(\alpha, \mathbb{N}_m)$ *for linear template. Different values of* K *are used.* $m = 1,000$; $\alpha = 0.2$; $\pi_0 = 1$. $\lambda^L(\alpha, \mathbb{N}_m)$ *was estimated based on* $B = 10^4$ *Monte-Carlo samples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

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

$$\alpha/K \leq \lambda^B(\alpha, \mathbb{N}_m) \leq \alpha. \tag{S-9}$$

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

LEMMA S-3.1. *In the framework of Proposition S-3.3, consider* $\lambda^B(\alpha) = \lambda^B(\alpha, \mathbb{N}_m)$ *for* $K = m$ *(see Section 5.2). Then for* m *large enough, we have*

$$\lambda^B(\alpha, \mathbb{N}_m) \leq 1/(\log m)^{1/4}. \tag{S-10}$$

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-9) 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 $\lambda^B(\alpha, \mathbb{N}_m)$.

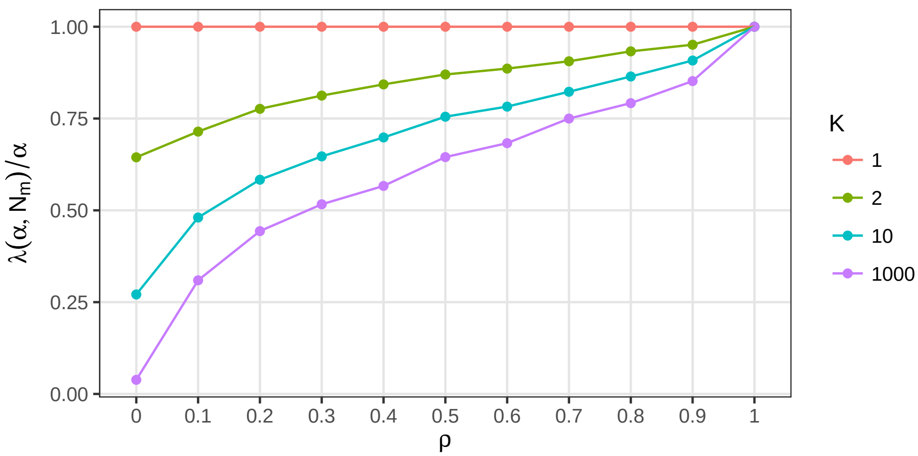

FIG S-2. *Influence of the equi-correlation level ρ on the adjustment factor $\lambda^B(\alpha, \mathbb{N}_m)$ for the balanced template. Different values of K are used. $m = 1,000$; $\alpha = 0.2$; $\pi_0 = 1$. $\lambda^B(\alpha, \mathbb{N}_m)$ was estimated based on $B = 10^4$ Monte-Carlo samples of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

S-3.2. *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-adjustement by replacing $\lambda(\alpha, \mathbb{N}_m)$ by $\lambda(\alpha, \widehat{A})$ for $\widehat{A}$ 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 (S-6), the achieved upper bound (18) on the JER is

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

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-3.2. *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,*

$$\mathbb{P}\Big(\exists k \in \{1,\dots,m_0\} \,:\, q_{(k:m_0)} < t_k\left(\lambda^B(\alpha)\right)\Big) \le \frac{C(\pi_0)}{(\log m)^{1/8}}, \tag{S-11}$$

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-3.3. *Properties of the balanced template under independence.* The following result gathers some of the properties of the balanced template under independence.

PROPOSITION S-3.3. *In the location model (13) under independence, letting $t^B_k(\lambda)$ be the threshold given by (24), we have:*

(i) for all $k \in \{1,\dots,m\}$ and $\lambda \in [0,1]$, $t^B_k(\lambda)$ is the λ-quantile of the distribution $\mathrm{Beta}(k, m+1-k)$.

(ii) for any $\lambda \le 0.5$,

$$t^B_k(\lambda) \le \frac{k}{m+1};$$

$$t^B_k(\lambda) \ge \frac{k}{m+1} - \left\{\frac{k}{m+1}\left(1 - \frac{k}{m+1}\right)\right\}^{1/2}\left(\frac{4\log 1/\lambda}{m}\right)^{1/2}; \tag{S-12}$$

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

$$\mathbb{P}\left(p_{(k:m_0)} \le t^B_k(\lambda)\right) \le \exp\left(-\frac{k}{32}\left(1 - \frac{m_0}{m}\right)^2\right). \tag{S-13}$$

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

$$\mathbb{P}\left(p_{(k:m)} \le \alpha\frac{k}{m}\right) \le \exp\left(-\frac{k}{4}\left(1 - \alpha - \frac{1}{m+1}\right)^2\right). \tag{S-14}$$

S-4. 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 similarities 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:

$$\text{(S-15)} \quad \mathrm{Pow}^\star(\mathfrak{R}, P) = \mathbb{P}(\overline{S}_{\mathfrak{R}}(\mathbb{N}_m) \geq 1) = \mathbb{P}\left(\exists k \in \{1, \dots, K\} \,:\, |R_k| \geq k\right).$$

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-4.1. *Framework.* We consider the location model (13) 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

- $p_i(X) \mid H_i = 0 \sim U(0,1)$;
- $p_i(X) \mid H_i = 1$ 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:

$$\text{(S-16)} \qquad \rho^\star(\beta) = \begin{cases} \beta - 1/2 & \text{if } \beta \in (1/2, 3/4]; \\ (1 - \sqrt{1-\beta})^2 & \text{if } \beta \in (3/4, 1). \end{cases}$$

This is an optimal detection boundary in the following sense: for $r < \rho^\star(\beta)$, any detection procedure will have a risk (type I error rate plus type II error rate) tending to 1, while for $r > \rho^\star(\beta)$, 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-4.2. *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-9) and (S-12), we have the lower bound

(S-17)

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

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-4.1. 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-4.4.

S-4.3. *Optimality results.* By adapting the proof of DJ04, we can show the following result (proved in Section S-7.4):

THEOREM S-4.2. *In the asymptotic setting of Section S-4.1, 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

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

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

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

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-7.4 for completeness.

THEOREM S-4.3. *Consider the asymptotic setting of Section S-4.1. Then the linear reference family $\mathfrak{R}^L$ satisfies the following:*

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

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-4.2 and S-4.3 is not fully reached for such a value of m (while it seems reached for $m = 10^6$ in DJ04).

S-4.4. *Numerical experiments for detection power.* We consider the independent case, and we calibrate the parameter $\overline{\mu}$ and π_0 according to the above-defined regime: $\pi_0 = 1 - m^{-\beta}$ and $\overline{\mu} = \sqrt{2r \log m}$, 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 6. 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

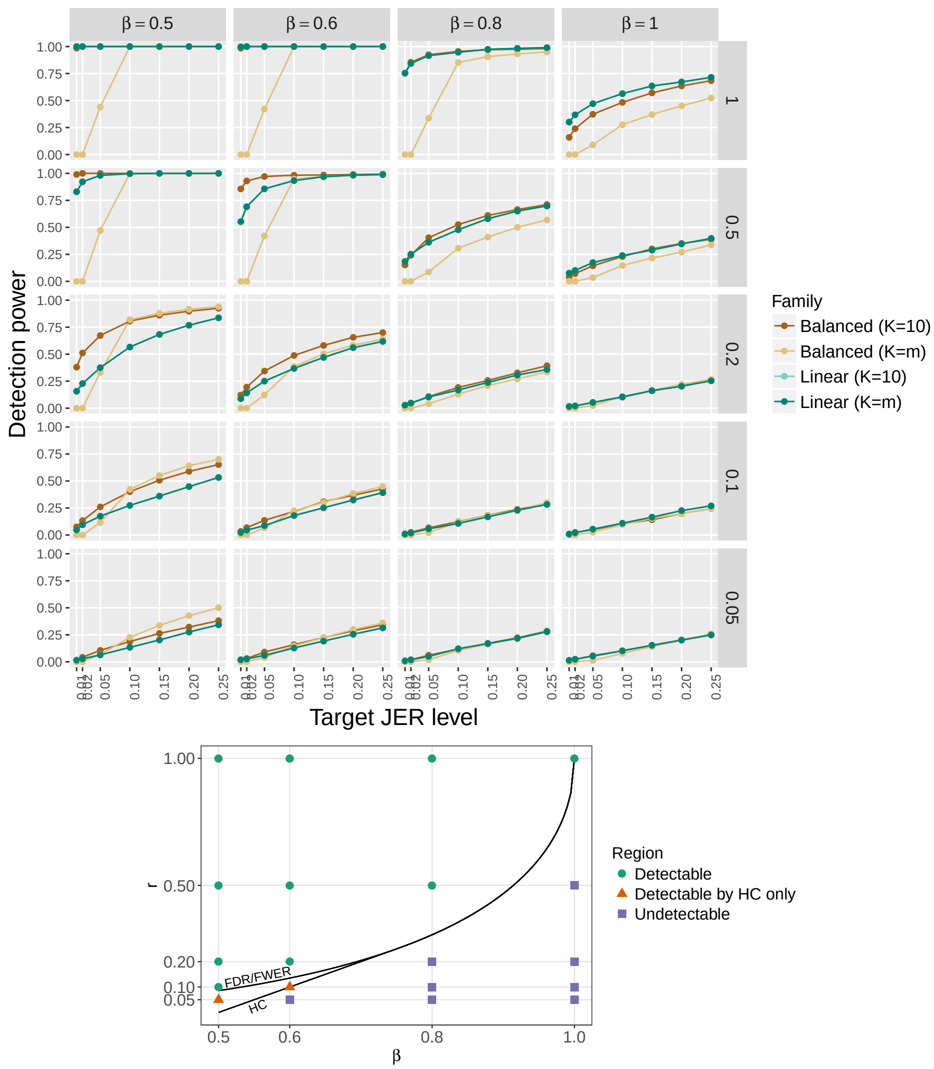

FIG S-3. *Top: detection power of JER controlling procedures for independent test statistics in the sparsity range for* 4×5 *parameter configurations for* (β, r) *in the sparsity range* $[1/2, 1] \times [0, 1]$. *Bottom: these* 4×5 *configurations are positioned with respect to the detection boundaries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

(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 6. 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-4.2 (and of Donoho and Jin, 2004) are quite slow to kick in and not yet reached for $m = 1000$.

S-5. 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

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

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

$$\Big((X_i^{(1)})_{i\in\mathcal{H}_0},\dots,(X_i^{(n)})_{i\in\mathcal{H}_0}\Big) \tag{S-18}$$

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-18) 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-6. Monte-Carlo approximation for balanced reference family. We consider the balanced reference family $\mathfrak{R}^{B,sd}$ given in Section 5.2. 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
$$M_0=\begin{pmatrix} q_1^{(1)} & q_2^{(1)} & \cdots & q_m^{(1)}\\ q_1^{(2)} & q_2^{(2)} & \cdots & q_m^{(2)}\\ \vdots & \vdots & & \vdots\\ q_1^{(B)} & q_2^{(B)} & \cdots & q_m^{(B)}\end{pmatrix};$$
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 $\widetilde{F}_k(x) = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left\{q^{(b)}_{(k:m)} \leq x\right\}$. For each λ, approximate $t^B_k(\lambda)$ by $\widetilde{t}^B_k(\lambda)$, defined as the λ-quantile of the sample
$$\left(q^{(1)}_{(k:m)}, \dots, q^{(B)}_{(k:m)}\right).$$
4. Consider the matrix of "ranks"
$$Z(A) = \begin{pmatrix} Z_{1,1}(A) & Z_{1,2}(A) & \dots & Z_{1,a}(A) \\ Z_{2,1}(A) & Z_{2,2}(A) & \dots & Z_{2,a}(A) \\ \vdots & \vdots & & \vdots \\ Z_{B,1}(A) & Z_{B,2}(A) & \dots & Z_{B,a}(A) \end{pmatrix}$$
where we let $Z_{b,k}(A) = \tilde{F}_k(q^{(b)}_{(k:A)})$, 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 $\widetilde{\lambda}^B(\alpha, A) = U_{(\lceil \alpha B\rceil)}$, i.e., the α empirical quantile of the sample $U = (U_1, \dots, U_B)$.
6. Use Algorithm 1 with $\widetilde{t}^B_k$ and $\widetilde{\lambda}^B$ instead of t^B_k and λ^B, respectively, to obtain $\widetilde{A}$.

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-6.1. Although JER control by λ-calibration for the balanced reference family under unknown dependence is not guaranteed by Proposition 4.8, we have also implemented this λ-calibration. Formally, the construction is identical, using $q^{(b)}_k = p_k(g_b.X)$ for $b = 1 \dots B$ and $k \in \mathbb{N}_m$. In this case, we have $\widetilde{F_k} = F_k$, $\widetilde{t_k} = t_k$, and $\widetilde{\lambda}^B = \lambda^B$.

S-7. Proofs.

S-7.1. *Proofs for Section 2.*

PROOF OF PROPOSITION 2.1. Let A be any subset of $\mathbb{N}_m$. Point 1 of the purported equivalence is $\forall k \in \mathbb{N}_K : |R_k \cap A| \leq \zeta_k$. Obviously this is equivalent to point 2: $A \in \mathcal{A}(\mathfrak{R})$ by the definition (6) of $\mathcal{A}(\mathfrak{R})$. We prove a circular implication of the statements 2 to 4 in the statement of the proposition. The implication from point 2 to point 3 is obvious from the definition (7) of $V^*_{\mathfrak{R}}$, and further the implication to point 4 by specializing to $R := A$. Finally, point 4 and the definition of $V^*_{\mathfrak{R}}$ imply that there must exist $A' \in \mathcal{A}(\mathfrak{R})$ with $A \subset A'$, but since any subset of an element of $\mathcal{A}(\mathfrak{R})$ also belongs to $\mathcal{A}(\mathfrak{R})$, we conclude $A \in \mathcal{A}(\mathfrak{R})$.

Concerning the final statement of the proposition about optimality, let $V : \mathcal{P}(\mathbb{N}_m) \to \mathbb{N}$ be a function such that for any $A \subset \mathbb{N}_m$, $A \in \mathcal{A}(\mathfrak{R})$ implies that $\forall R \subset \mathbb{N}_m$, $|R \cap A| \leq V(R)$. By definition of $V^*_{\mathfrak{R}}$, for any $R \subset \mathbb{N}_m$ there exists $B \in \mathcal{A}(\mathfrak{R})$ such that $V^*_{\mathfrak{R}}(R) = |R \cap B| \leq V(R)$, where the last inequality comes from the assumed implication. □

Proof of Proposition 2.3. Again, we show a circular implication of the three points. First, for $1 \leq k \leq K$, $|R_k \cap A| \leq \zeta_k$, implies that for any $R \subset \mathbb{N}_m$,

$$\begin{aligned} |A \cap R| &= |A \cap R \cap R_k| + |A \cap R \cap R_k^c| \\ &\leq |R \cap R_k| + |A \cap R_k^c| \\ &\leq \zeta_k + |A \setminus R_k|. \end{aligned}$$

which entails $|A \cap R| \leq \overline{V}(R)$ by taking a minimum over all possible values of k. Secondly, specializing the above inequality to $R := A$, we obtain $|A| \leq \overline{V}(A)$. Finally, if the latter is satisfied, it implies that for all k, $|A \setminus R_k| + \zeta_k \geq |A|$, and thus $|A \cap R_k| \leq \zeta_k$. The fact that $V^*(R) \leq \overline{V}(R)$ for all R is a direct consequence of the optimality of V^* from Proposition 2.1. □

Proof of Proposition 2.2. 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. □

PROOF OF PROPOSITION 2.4. Obviously, $\widetilde{\zeta}_k \leq \zeta_k$ and thus $\overline{V}_{\widetilde{\mathfrak{R}}}(R) \leq \overline{V}_{\mathfrak{R}}(R)$. Let us prove the reverse inequality:

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

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

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 $V^*_{\mathfrak{R}}(R) \leq \overline{V}_{\mathfrak{R}}(R)$, 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 $|A| \geq \overline{V}_{\mathfrak{R}}(R)$. For this, let $\tilde{\zeta}_k$ be defined as in (9) 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 (9) 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, $\overline{V}_{\mathfrak{R}}(R) \leq |R \setminus (R_K \cap R)| + \tilde{\zeta}_K$. Therefore, $\overline{V}_{\mathfrak{R}}(R) \leq |R \setminus (R_K \cap R)| + |R_K \cap R \cap A| = |A|$, and the result is proved. □

S-7.2. *Proofs for Section 4.*

PROOF OF PROPOSITION 4.5. Consider the event Ω for which

$$\forall k \in \{1,\dots,K\}, \quad p_{(k:\mathcal{H}_0)}(X) \geq t_k(\lambda(\alpha,\mathcal{H}_0)), \tag{S-19}$$

which occurs with probability at least $1-\alpha$ by (19). 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 $\mathcal{H}_0 \subset \widehat{A}$, which itself entails

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

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

FIRST PROOF OF THEOREM 4.8. We denote in this proof $\lambda(\alpha, X, \mathcal{H}_0)$ instead of $\lambda(\alpha,\mathcal{H}_0)$ to underline the dependence of this functional w.r.t. the data X. By Propositions 4.4 and 4.5, it sufficient to prove that $\lambda(\cdot)$ is a valid λ-calibration, that is, satisfies the requirement of Definition 4.3. Since the monotonic property is clearly satisfied, it remains to establish (17). For this, write

$$\begin{aligned}
\mathbb{P}\Big(\min_{1\leq k\leq K\wedge m_0} \left\{ t_k^{-1}\left(p_{(k:\mathcal{H}_0)}(X)\right)\right\} < \lambda(\alpha,X,\mathcal{H}_0)\Big)& \\
&= \mathbb{P}\left(\Psi(X,\mathcal{H}_0) < \lambda(\alpha,X,\mathcal{H}_0)\right)\\
&\leq \mathbb{P}\Big(B^{-1}\sum_{j=1}^B \mathbb{1}\left\{\Psi(g_j.X,\mathcal{H}_0) \leq \Psi(X,\mathcal{H}_0)\right\} \leq \alpha\Big)\\
&= \mathbb{P}\Big(B^{-1}\sum_{j=1}^B \mathbb{1}\left\{Y_j \leq Y_1\right\} \leq \alpha\Big),
\end{aligned}$$

where we have used in the inequality the definition of $\lambda(\alpha,X,\mathcal{H}_0)$ (see (20)) and we have let $Y_j = \Psi(g_j.X,\mathcal{H}_0)$, $1 \leq j \leq m$. Now, by (Rand), we easily check that $(Y_1,\dots,Y_B)$ is an exchangeable random vector: for any g_0 uniformly distributed on $\mathcal{G}$ (and drawn independently of the other variables),

$$\begin{aligned}
(Y_1,\dots,Y_B) &\sim (\Psi(g_1.g_0.X,\mathcal{H}_0),\dots,\Psi(g_B.g_0.X,\mathcal{H}_0))\\
&\sim \left(\Psi(g_1'.X,\mathcal{H}_0),\dots,\Psi(g_B'.X,\mathcal{H}_0)\right),
\end{aligned}$$

where g_j', $1\leq j\leq B$, are i.i.d. uniform in $\mathcal{G}$ (independent of X). Above, the first equality in distribution holds because it is true conditionally on

$\{g_1,\dots,g_B\}$, and the second one holds because it is true conditionally on X. Since the variables $\Psi(g'_j.X,\mathcal{H}_0)$, $1\leq j\leq m$, are i.i.d. conditionally on X, we deduce that $(Y_1,\dots,Y_B)$ is an exchangeable random vector. Hence, for any independent variable U uniformly distributed on $\{1,\dots,B\}$, we obtain

$$\mathbb{P}\left(B^{-1}\sum_{j=1}^{B}\mathbb{1}\left\{Y_j\leq Y_1\right\}\leq\alpha\right)=\mathbb{P}\left(B^{-1}\sum_{j=1}^{B}\mathbb{1}\left\{Y_j\leq Y_U\right\}\leq\alpha\right).$$

Let σ any permutation (independent of U) such that $Y_{\sigma(1)}\leq\cdots\leq Y_{\sigma(B)}$. Since $\sum_{j=1}^{B}\mathbb{1}\left\{Y_j\leq Y_U\right\}=\sum_{j=1}^{B}\mathbb{1}\left\{Y_{\sigma(j)}\leq Y_U\right\}$ and U and $\sigma(U)$ have the same distribution conditionally on Y, we have

$$\mathbb{P}\left(B^{-1}\sum_{j=1}^{B}\mathbb{1}\left\{Y_j\leq Y_U\right\}\leq\alpha\mid Y\right)=\mathbb{P}\left(B^{-1}\sum_{j=1}^{B}\mathbb{1}\left\{Y_{\sigma(j)}\leq Y_{\sigma(U)}\right\}\leq\alpha\mid Y\right)$$
$$\leq\mathbb{P}\left(B^{-1}\sum_{j=1}^{B}\mathbb{1}\left\{j\leq U\right\}\leq\alpha\mid Y\right)=\mathbb{P}\left(U\leq\alpha B\mid Y\right)=\frac{\lfloor\alpha B\rfloor}{B}\leq\alpha.$$

□

Another argument is possible for this proof using a device recently proposed by Hemerik and Goeman (2017).

Second proof of Theorem 4.8. 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 the previous proof, it is sufficient to prove

$$\mathbb{P}\left(\Psi(X,\mathcal{H}_0)<\lambda(\alpha,X,\mathcal{H}_0,\mathcal{G}')\right)\leq\alpha. \tag{S-20}$$

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 $j\in\{1,\dots,B\}$, $\mathcal{G}'.g_j^{-1}$ has the same distribution as $\mathcal{G}'$ (up to permutation of the elements), which entails that $\mathcal{G}'.g_U^{-1}$ 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, $\lambda(\alpha,X,\mathcal{H}_0,\mathcal{G}'.g_U^{-1})$ has the same distribution as $\lambda(\alpha,X,\mathcal{H}_0,\mathcal{G}')$ (conditionally on X). As a consequence, the LHS of (S-20) is equal to

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

where we used that, by (Rand),

$$\big(p_{\mathcal{H}_0}(X), p_{\mathcal{H}_0}(g_1.g_U^{-1}.X), p_{\mathcal{H}_0}(g_2.g_U^{-1}.X), \dots, p_{\mathcal{H}_0}(g_B.g_U^{-1}.X))\big)$$
$$\sim (p_{\mathcal{H}_0}(g_U.X), p_{\mathcal{H}_0}(g_1.X), p_{\mathcal{H}_0}(g_2.X), \dots, p_{\mathcal{H}_0}(g_B.X))) \,,$$

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

$$\mathbb{P}\bigg(\Psi(g_U.X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}') \,\bigg|\, X, \mathcal{G}' \bigg)$$
$$= B^{-1} \sum_{j=1}^{B} \mathbb{1}\left\{ \Psi(g_j.X, \mathcal{H}_0) < \lambda(\alpha, X, \mathcal{H}_0, \mathcal{G}') \right\} \leq \alpha$$

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

S-7.3. *Proofs for Section S-3.*

Proof of Lemma S-3.1. 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-12) to write for a large m,

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

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-8.2. □

Proof of Lemma S-3.2. Let $N \in \{1, \ldots, m_0 - 1\}$ be some integer to be chosen later. By a union bound argument, we have

$$\mathbb{P}\Big(\exists k \in \{1, \ldots, m_0\} \;:\; q_{(k:m_0)} < t_k\left(\lambda^B(\alpha)\right) \Big) \\ \le \sum_{k=1}^{N} \mathbb{P}\Big(q_{(k:m_0)} < t_k\left(\lambda^B(\alpha)\right) \Big) + \sum_{k=N+1}^{m_0} \mathbb{P}\Big(q_{(k:m_0)} < t_k\left(\lambda^B(\alpha)\right) \Big).$$

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

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

by Lemma S-3.1. For the second term, by (S-13) ($\lambda^B(\alpha)$ begin smaller than 0.5 for large enough m by S-10) and letting $r = e^{-\frac{(1-\pi_0)^2}{32}}$, we have for large enough m:

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

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

Proof of Proposition S-3.3. 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-23) and Lemma S-8.3 in Section S-8. For item (iii), we use item (ii) and $m/(m+1) \ge 1/2$ and $m_0/(m_0+1) \ge 1/2$, to write

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

and we conclude by using (S-23). For (iv), the reasoning is similar, using

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

□

S-7.4. *Proofs for Section S-4.*

PROOF OF THEOREM S-4.2. 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 $\varphi = \mathbb{1}\left\{\exists k : p_{(k)} < t_k\right\}$ 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 $r < \rho^\star(\beta)$. Hence, the asymptotic power of the LRT is less than α.

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

$$\text{Pow}^\star(\mathfrak{R}^B_\alpha, P) \geq \mathbb{P}\left(U_{(k:m)} \leq \pi_{0,m} t_k + \pi_{1,m} F_{1,m}(t_k)\right). \tag{S-21}$$

From (S-17), now write for any $k \in \{1, \dots, m\}$,

$$\text{Pow}^\star(\mathfrak{R}^B_\alpha, P) \geq \mathbb{P}\Bigg(U_{(k:m)} \leq \frac{k}{m+1} - \left\{\frac{k}{m+1}\left(1 - \frac{k}{m+1}\right)\right\}^{1/2} \left(\frac{4\log(m/\alpha)}{m}\right)^{1/2} + \pi_{1,m}\left(F_{1,m}(t_k) - t_k\right)\Bigg)$$

$$\geq \mathbb{P}\left(\max_{1\leq k\leq m}\left\{m^{1/2}\frac{U_{(k:m)} - \frac{k}{m+1}}{\left\{\frac{k}{m+1}\left(1 - \frac{k}{m+1}\right)\right\}^{1/2}}\right\} \leq -\left(4\log(m/\alpha)\right)^{1/2} + \pi_{1,m}\frac{F_{1,m}(t_k) - t_k}{k^{1/2}/(m+1)}\right), \tag{S-22}$$

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 $\overline{\Phi}^{-1}(t_k) = \sqrt{2r_k \log m}$. Then, choosing $k = \lfloor m^{1-q} \log m \rfloor$ for some $q \in (0,1)$, we have $r_k \to q$ as m tends to infinity. To see this, first note that $\lambda^B(\alpha) \to 0$ by Lemma S-3.1, and that $\lambda^B(\alpha) \geq \alpha/m$ by (S-9). Therefore, (S-12) in Proposition S-3.3 entails that $t_k \sim k/m$. Then, recalling that $\overline{\Phi}^{-1}(u) \sim \sqrt{2\log 1/u}$ as $u \to 0$, our choice of r_k yields $r_k \to q$. Furthermore, denoting by ϕ the density of the standard gaussian distribution, we have for $q > r$:

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

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 $r > \rho^\star(\beta)$, by continuity of f there exists $q \in (0,1)$ such that $f(q) > 0$. Now, (ii) comes from (S-22) 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-8.2).

□

PROOF OF THEOREM S-4.3. Let us first prove (i). For any $k \in \{1,\dots,m\}$,

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

Let $r_k > 0$ such that $\overline{\Phi}^{-1}(\alpha k/m) = \sqrt{2r_k\log m}$, so that when $k = \lfloor \log m \rfloor$, $r_k \to 1$ 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

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

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

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

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,

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

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

S-8. Some properties of the Beta distribution. We recall the following result (Shorack and Wellner, 1986, p.454-455):

LEMMA S-8.1. *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*

(S-23)
$$\mathbb{P}\left(n^{1/2}\left(U_{(\ell:n)} - \frac{\ell}{n+1}\right) \leq -\left\{\frac{\ell}{n+1}\left(1-\frac{\ell}{n+1}\right)\right\}^{1/2} x\right) \leq e^{-x^2/4}$$

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

LEMMA S-8.2. *Let $U_1, U_2, \ldots$ i.i.d. $\sim U(0,1)$ and consider*

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

then we have, as n grows to infinity,

$$\mathbb{P}((\log \log n)^{1/2} \le Z_n \le 2(\log \log n)^{1/2}) \to 1 \tag{S-24}$$

LEMMA S-8.3. *Let $U_1, U_2, \ldots$ i.i.d. $\sim U(0,1)$, then, for all $m \ge 2$, for all $k \in \{1, \ldots, m\}$,*

$$\mathbb{P}(U_{(k:m)} \le (k+1)/m) \ge 0.5. \tag{S-25}$$

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

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

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). □

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.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994.

Durand, G., Blanchard, G., Neuvial, P., and Roquain, E. (2018). Post hoc false positive control for spatially structured hypotheses. *arXiv preprint 1807.01470*.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061.

Genovese, C. R. and Wasserman, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.*, 101(476):1408–1417.

Goeman, J., Meijer, R., Krebs, T., and Solari, A. (2016). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *arXiv preprint 1611.06739*.

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)*.

Hommel, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical J.*, 25(5):423–430.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–386.

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.

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

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.

Meinshausen, N. (2006). False discovery control for multiple tests of association under general dependence. *Scand. J. Statist.*, 33(2):227–237.

Meinshausen, N. and Bühlmann, P. (2005). Lower bounds for the number of true null hypotheses for multiple testing of associations under general dependence structures. *Biometrika*, 92:893–907.

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.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

van der Laan, M. J., Dudoit, S., and Pollard, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.*, 3(1):1–25.

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

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

Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, LPSM,
4, Place Jussieu, 75252 Paris cedex 05, France
E-mail: etienne.roquain@upmc.fr