1 Supplement A: Proof of theorem 2.3:

Proof (Thm. 2.3):

Assume w.l.o.g. that the first m_0 hypothesis are null, and the next $m - m_0$ are the alternative. As is hinted from the bound, the key idea used in the proof is to consider a modified (and more liberal) estimator $\hat{m}_0^{(\not 1)}$, which does not take into account one of the null p-values, which we assume w.l.o.g. is p_1 (in practice the researcher does not know which of the p-values are null and which are not, but assuming that at least one p-value is null, we can still consider this hypothetical estimator as a liberal estimator for m_0. In the case where non of the p-values is null, the theorem follows immediately). We also define the following modified q_k: $q_k^{(1)} \equiv \frac{qk}{\hat{m}_0^{(\not 1)}}$. From the monotonicity of $\hat{m}_0$ we have:

$$\hat{m}_0^{(\not 1)} \leq \hat{m}_0, \quad q_k^{(1)} \geq q_k \tag{1}$$

In addition, we define the following events:

$$C_k^{(1)} \equiv \big\{ \max\{j : p_{(j-1)}^{(1)} \leq q_j\} = k \big\}$$

$$D_k^{(1)} \equiv \bigcup_{j \leq k} C_j^{(1)} = \{p_{(j-1)}^{(1)} > q_j, \, \forall j = k+1, .., m\} \tag{2}$$

where $p_{(1)}^{(1)} \leq .. \leq p_{(m-1)}^{(1)}$ are the ordered $m-1$ p-values excluding p_1.

Since $C_k^{(1)}$ and $\hat{m}_0^{(\not 1)}$ depend only on $p_2, .., p_m$, the following conditional independence relation holds:

$$C_k^{(1)} \perp\!\!\!\perp p_1 | \hat{m}_0^{(\not 1)} \tag{3}$$

Before giving our proof, we need the following two lemmas:

Lemma 1 *If p_1 is independent of $p_{2..m}$ then:*

$$Pr(D_k^{(1)} | \{\hat{m}_0^{(\not 1)}, p_1 \leq q_j^{(1)}\}) \leq Pr(D_k^{(1)} | \{\hat{m}_0^{(\not 1)}, p_1 \leq q_k^{(1)}\}), \quad \forall j \leq k \tag{4}$$

Proof (lemma 1):

The lemma's statement involves conditioning on $p_1 \leq p$. We first prove a point-wise auxiliary claim, conditioning on $p_1 = p$. Let $g(p_2, .., p_m | \hat{m}_0^{(\not 1)})$ be the conditional density function of $p_2, .., p_m$ given $\hat{m}_0^{(\not 1)}$. Denote $f_{p_{2..m}}(p) = Pr(D_k^{(1)} | \{\hat{m}_0^{(\not 1)}, p_1 = p\})$. We prove below that f is monotonically non-decreasing. Let $p \leq p'$. Then:

$$f_{p_{2..m}}(p) = Pr(D_k^{(1)} | \{\hat{m}_0^{(\not 1)}, p_1 = p\}) = Pr(\bigcap_{j=k+1}^{m} \{p_{(j-1)}^{(1)} > q_j\} | \{\hat{m}_0^{(\not 1)}, p_1 = p\}) =$$

$$Pr(\bigcap_{j=k+1}^{m} \{p_{(j-1)}^{(1)} > \frac{qj}{\hat{m}_0}\} | \{\hat{m}_0^{(\not 1)}, p_1 = p\}) =$$

$$\int_{p_{2..m}} Pr(\bigcap_{j=k+1}^{m} \{p_{(j-1)}^{(1)} > \frac{qj}{\hat{m}_0(p, p_{2..m})}\} | \{p_1 = p, p_{2..m}\}) g(p_2, .., p_m | \hat{m}_0^{(\not 1)}) dp_2..dp_m \leq$$

$$\int_{p_{2..m}} Pr(\bigcap_{j=k+1}^{m} \{p_{(j-1)}^{(1)} > \frac{qj}{\hat{m}_0(p', p_{2..m})}\} | \{p_1 = p', p_{2..m}\}) g(p_2, .., p_m | \hat{m}_0^{(\not 1)}) dp_2..dp_m =$$

$$Pr(\bigcap_{j=k+1}^{m}\{p_{(j-1)}^{(1)} > \frac{qj}{\hat{m}_0}\}|\{\hat{m}_0^{(\not 1)}, p_1 = p'\}) = Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 = p'\}) = f_{p_{2..m}}(p') \tag{5}$$

Thus f is a monotonically non-decreasing function. Therefore, the average of f over $[0,p]$ is also non-decreasing in p:

$$\frac{1}{p}\int_0^p f_{p_{2..m}}(x)dx = [\frac{1}{p'} + (\frac{1}{p} - \frac{1}{p'})]\int_0^p f_{p_{2..m}}(x)dx \leq$$
$$\frac{1}{p'}\int_0^p f_{p_{2..m}}(x)dx + (1 - \frac{p}{p'})f_{p_{2..m}}(p) \leq$$
$$\frac{1}{p'}\int_0^p f_{p_{2..m}}(x)dx + \frac{1}{p'}\int_p^{p'} f_{p_{2..m}}(x)dx = \frac{1}{p'}\int_0^{p'} f_{p_{2..m}}(x)dx \tag{6}$$

And this in fact shows:

$$Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq p\}) \leq Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq p'\}), \quad \forall p \leq p' \tag{7}$$

From here the lemma's claim follows by simple integration, noting that any $q_j^{(1)}$ can be treated as a constant given $p_2, .., p_m$:

$$Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_j^{(1)}\}) = \int_{p_{2..m}} g(p_2, .., p_m|\hat{m}_0^{(\not 1)})\frac{1}{q_j^{(1)}}\int_{p_1=0}^{q_j^{(1)}} f_{p_{2..m}}(x)dp_1dp_2..dp_m \leq$$

$$\int_{p_{2..m}} g(p_2, .., p_m|\hat{m}_0^{(\not 1)})\frac{1}{q_k^{(1)}}\int_{p_1=0}^{q_k^{(1)}} f_{p_{2..m}}(x)dp_1dp_2..dp_m = Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_k^{(1)}\}) \tag{8}$$

■

The next lemma follows the spirit of [1].

Lemma 2

$$\sum_{j=1}^{k} Pr(C_j^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_j\}) \leq Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_k\}) \quad \forall k = 1, .., m \tag{9}$$

Proof (lemma 2):

The proof is done by induction. For $k = 1$, eq. (9) is reduced to:

$$Pr(C_1^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_1\}) \leq Pr(D_1^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_1\}) \tag{10}$$

And the two quantities are equal, since by definition $C_1^{(1)} = D_1^{(1)}$. Assuming the correctness of eq. (9) for k, we prove it for $k+1$ using lemma 1:

$$\sum_{j=1}^{k+1} Pr(C_j^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_j\}) \leq Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_k\}) +$$

$$Pr(C_{k+1}^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_{k+1}\}) \leq Pr(D_k^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_{k+1}\}) +$$

$$Pr(C_{k+1}^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_{k+1}\}) = Pr(D_{k+1}^{(1)}|\{\hat{m}_0^{(\not 1)}, p_1 \leq q_{k+1}\}) \tag{11}$$

■

By using eqs. (1, 3) and lemma 2 we are able to express the FDR as:

$$E\left[\frac{V}{R^+}\right] = \sum_{k=1}^{m} Pr(R=k)E[\frac{V}{R}|R=k] = \sum_{k=1}^{m}\frac{1}{k}Pr(R=k)\sum_{j=1}^{m_0} Pr(p_i \le q_k|R=k) =$$

$$m_0\sum_{k=1}^{m}\frac{1}{k}Pr(R=k, p_1 \le q_k) \le m_0\sum_{k=1}^{m}\frac{1}{k}Pr(R=k, p_1 \le q_k^{(1)}) =$$

$$m_0\sum_{k=1}^{m}\frac{1}{k}Pr(C_k^{(1)}, p_1 \le q_k^{(1)}) = m_0\sum_{k=1}^{m}\frac{1}{k}\int_{\hat{m}_0^{(\mathcal{Y})}} Pr(C_k^{(1)}, p_1 \le q_k^{(1)}|\hat{m}_0^{(\mathcal{Y})}) f_{\hat{m}_0^{(\mathcal{Y})}} d\hat{m}_0^{(\mathcal{Y})} =$$

$$m_0\int_{\hat{m}_0^{(\mathcal{Y})}} \frac{q}{\hat{m}_0^{(\mathcal{Y})}}\sum_{k=1}^{m} Pr(C_k^{(1)}|\{p_1 \le q_k^{(1)}, \hat{m}_0^{(\mathcal{Y})}\}) f_{\hat{m}_0^{(\mathcal{Y})}} d\hat{m}_0^{(\mathcal{Y})} \le$$

$$m_0\int_{\hat{m}_0^{(\mathcal{Y})}} \frac{q}{\hat{m}_0^{(\mathcal{Y})}} Pr(D_m^{(1)}|\{p_1 \le q_m^{(1)}, \hat{m}_0^{(\mathcal{Y})}\}) f_{\hat{m}_0^{(\mathcal{Y})}} d\hat{m}_0^{(\mathcal{Y})} = m_0\int_{\hat{m}_0^{(\mathcal{Y})}} \frac{q}{\hat{m}_0^{(\mathcal{Y})}} f_{\hat{m}_0^{(\mathcal{Y})}} d\hat{m}_0^{(\mathcal{Y})} =$$

$$m_0 q E\left[\frac{1}{\hat{m}_0^{(\mathcal{Y})}}\right] \quad (12)$$

■

2 Supplement B: Designing the IBHsum estimator

There are many ways to design the estimator $\hat{m}_0$ such that it will satisfy Thm. 2.3. Here we choose to define $\hat{m}_0$ using $\hat{m}_0'$ as:

$$\hat{m}_0 = C(m) \cdot min\left[m, (max(s(m), 2\sum_{j=1}^{m} p_j)\right] \quad (13)$$

Our goal is to calculate the optimal correction factors $C(m)$ and $s(m)$ such that $\hat{m}_0$ will still satisfy eq. (3.1). Setting $C = 1$ and $s = 0$ gives the (uncorrected) unbiased estimator $\hat{m}_0'$. We can bound $E[\frac{1}{\hat{m}_0^{(1)}}]$ by neglecting the alternative p-values:

$$E\left[\frac{1}{\hat{m}_0^{(\mathcal{Y})}}\right] = \frac{1}{C(m)}E\left[\frac{1}{min\left[m, max(s(m), 2\sum_{j=2}^{m} p_j)\right]}\right] \le \frac{1}{C(m)}E\left[\frac{1}{min\left[m, max(s(m), 2\sum_{j=2}^{m_0} p_j)\right]}\right] \quad (14)$$

Define the r.v. $z_{m_0} = 2\sum_{j=2}^{m_0} p_j$ and denote its density by $h^{(m_0)}(z_{m_0})$. Then:

$$E\left[\frac{1}{\hat{m}_0^{(\mathcal{Y})}}\right] \le \frac{1}{C(m)}\left[\frac{1}{s}\int_0^s h_z^{(m_0)}(t)dt + \int_s^m \frac{h_z^{(m_0)}(t)}{t}dt + \frac{1}{m}\int_m^{2m} h_z^{(m_0)}(t)dt\right] \quad (15)$$

We want to find an optimal pair (C, s) satisfying the above inequality. First, assume that we know the value of s and find the optimal (smallest possible) C for this s. Had we known m_0, and since we want $E[1/\hat{m}_0^{(\mathcal{Y})}] \le 1/m_0$ we would have chosen C to be:

$$C(m, m_0, s) = m_0\left[\frac{1}{s}\int_0^s h_z^{(m_0)}(t)dt + \int_s^m \frac{h_z^{(m_0)}(t)}{t}dt + \frac{1}{m}\int_m^{2m} h_z^{(m_0)}(t)dt\right] \quad (16)$$

Since in the above equation m_0 is unknown we must maximize over all of its possible values $C(m, s) \equiv \max_{m_0} C(m, m_0, s)$. We are now left with the choice of s. As we increase s from zero, the maximal C decreases but at some point it remains constant, since when $m_0 = m$ our bound for C is independent of s - we set this

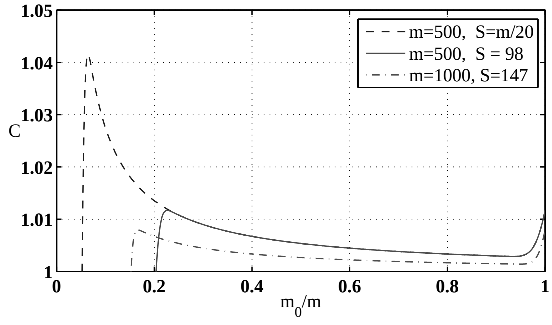

Figure 1: An example of the dependency of $C(m, m_0, s)$ on m_0 for different values of s. The value of s affects both the location and height of the left maximum, whereas the right maximum (at $m_0 = m$) is independent of s. We choose s such that the maximum of C is the smallest possible, and take the minimal s which achieves this. This gives $s = 98$ for $m = 500$ and $s = 147$ for $m = 1000$.

point as the optimal s, $s(m) = min\{s : s = \underset{s' \in [0,m]}{argmin}[C(m, s')]\}$. Fig. 1 presents an example for the dependency of C on the m_0/m and s.

Using numerical integration we calculate $C(m_0)$ for fixed m, its behavior for different values of m and s is presented in Fig. 1. The value of s controls the location of the left maximum and we choose s to be such that the maximal C is minimal, this happen when the left maximum is the same as the value at $m_0/m = 1$. The resulting $s(m), C(m)$ are presented in Fig. 2a and b, and several values of interest are listed in table 2. We also provide a MATLAB function for computing these values in the companion code.

In the above formula for $C(m, m_0)$ the density $h_z^{(m_0)}(z)$ is the density of the uniform sum distribution. Since calculation of the above integrals with the exact uniform sum distribution cause numerical difficulties we approximated it by a Gaussian distribution. For large values of m, the approximation converges to the exact distribution according to the central limit theorem. For small values of m ($m \leq 40$), we were able to compare the $C(m)$ values calculated by the exact uniform-sum distribution with the $C(m)$ which were calculated by the Gaussian approximation. This comparison shows that $C_{approximate}(m) > C_{exact}(m)$, and that the rate of convergence is faster that $1/m^{1.1}$, (see Fig. 2). Thus calculating $C(m)$ using the normal approximation is conservative (gives higher C) and as expected converges to the values of C calculated by the exact distribution.

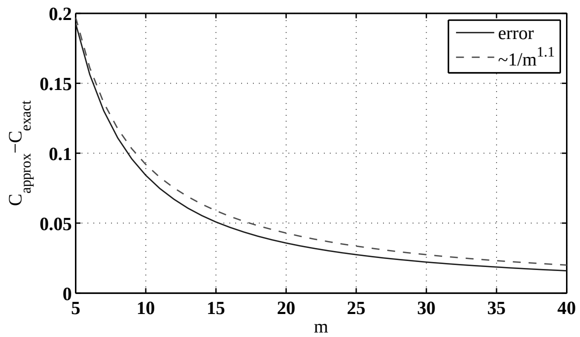

Figure 2: Difference between $C(m)$ calculated using the normal approximation and the $C(m)$ calculated by the exact uniform sum distribution. The difference is always positive (thus the approximation is conservative) and the rate of convergence is faster than $1/m$.

m	C	s
10	1.096981	5
100	1.030604	35
200	1.019915	55
300	1.015671	72
400	1.013267	86
500	1.011709	98
600	1.010554	109
700	1.009688	119
800	1.009	129
900	1.008441	138
1000	1.007968	147
2000	1.00549	217
3000	1.004426	272
4000	1.003808	318
5000	1.003386	359
6000	1.003085	396
7000	1.002844	430
8000	1.002654	462
9000	1.002502	491
10000	1.002366	521
15000	1.001922	645
20000	1.001662	750
25000	1.001482	843
30000	1.001349	928
40000	1.001168	1077
50000	1.001041	1211
60000	1.000949	1332
70000	1.000879	1439
80000	1.000821	1543
90000	1.000774	1641
100000	1.000734	1731

Table 1: Values of correction factors C, s for selected values of m

3 Supplement C: Proof of claim 3.1

Proof ***(claim 1):***

The proof is accomplished by bounding $E[1/\tilde{m}_0^{(\mathcal{Y})}]$ and using Thm. 2.3.

For $c \geq 0$, the function $\phi(x) = 1/(x+c)$ is convex. Therefore, we can use Jensen's inequality with this function and get:

$$E[1/\tilde{m}_0^{(\mathcal{Y})}] = E\Big[\frac{1}{2 - \sum_{i=2}^{m} \log(1-p_i)}\Big] = \int_{p_{2..m}} f_{p_{2..m}}(p_{2..m}) dp_{2..m} \frac{1}{2 - \sum_{i=2}^{m} \log(1-p_i)} =$$

$$\int_{p_{2..m}} f_{p_{2..m}}(p_{2..m}) dp_{2..m} \frac{1}{E[-\log(1-p_0) - \log(1-p_1)] - \sum_{i=2}^{m} \log(1-p_i)} \leq$$

$$\int_{p_{2..m}} f_{p_{2..m}}(p_{2..m}) dp_{2..m} E\Big[\frac{-1}{\sum_{i=0}^{m} \log(1-p_i)}\Big] = E\Big[\frac{-1}{\sum_{i=0}^{m} \log(1-p_i)}\Big] \leq$$

$$E\Big[\frac{-1}{\sum_{i=0}^{m_0} \log(1-p_i)}\Big] = E\Big[\frac{-1}{\sum_{i=0}^{m_0} \log p_i}\Big] \tag{17}$$

where $p_0 \sim U[0,1]$ is an auxiliary random variable defined to be independent of $p_{1..m}$.

Define $Y_{m_0} = \prod_{i=0}^{m_0} p_i$. Since $p_{0..m_0}$ are i.i.d. $U[0,1]$, Y_{m_0} has the following density function:

$$h_{Y_{m_0}}(t) = \frac{(-1)^{m_0}}{m_0!} (\log t)^{m_0} \tag{18}$$

Define also $X_{m_0} = -2 \log Y_{m_0}$, then $H_{X_{m_0}}(t) = 1 - H_{Y_{m_0}}(e^{-t/2})$ and

$$h_{X_{m_0}}(t) = \frac{e^{-t/2}}{2} h_{Y_{m_0}}(e^{-t/2}) = \frac{e^{-t/2} t^{m_0}}{2^{m_0+1} m_0!} \tag{19}$$

Therefore X_{m_0} is a chi-square r.v., $X_{m_0} \sim \chi^2(2m_0+2)$. Using this fact, we get

$$E[1/\tilde{m}_0^{(\mathcal{Y})}] \leq E[2/X_{m_0}] = \int_0^\infty \frac{e^{-t/2} t^{m_0}}{2^{m_0} m_0!} \cdot \frac{1}{t} dt = \frac{1}{m_0} \int_0^\infty \frac{e^{-t/2} t^{m_0-1}}{2^{m_0}(m_0-1)!} dt = \frac{1}{m_0} \tag{20}$$

Therefore, according to Thm. 2.3, we immediately get:

$$FDR \leq m_0 q \frac{1}{m_0} = q \tag{21}$$

■

4 Supplement D: Proof of monotonicity theorem

Proof (Thm. 4.1):

Assume w.l.o.g. that the first m_0 hypothesis are null, and the next $m - m_0$ are the alternative. For each procedure, R is some function of $\vec{p} = (p_1, .., p_m)$ which depends only on the order statistics $\vec{p}_{()} = (p_{(1)}, .., p_{(m)})$. We therefore need to prove:

$$\int_{\vec{p}} f_{p_{1...m}}(\vec{p}) d\vec{p} \frac{V_1(\vec{p})}{R_1^+(\vec{p})} \leq \int_{\vec{p}} f_{p_{1...m}}(\vec{p}) d\vec{p} \frac{V_2(\vec{p})}{R_2^+(\vec{p})} \tag{22}$$

The V_i's depend on the exact realization $\vec{p}$ while the R_i's depend only on the order statistics $\vec{p}_{()}$, and thus we can write equivalently:

$$\int_{\vec{p}_{()}} g(\vec{p}_{()}) d\vec{p}_{()} \frac{E[V_1(\vec{p})|\vec{p}_{()}]}{R_1^+(\vec{p}_{()})} \leq \int_{\vec{p}_{()}} g(\vec{p}_{()}) d\vec{p}_{()} \frac{E[V_2(\vec{p})|\vec{p}_{()}]}{R_2^+(\vec{p}_{()})} \tag{23}$$

Where g is the joint density of the order statistics $\vec{p}_{()}$, given by:

$$g(\vec{p}_{()}) = \sum_{\sigma \in S_m} f_{p_{1\ldots m}}(\sigma^{-1}(\vec{p}_{()})) \tag{24}$$

That is, g is obtained by summing over all $m!$ permutations σ on m elements in the symmetric group S_m, each permutation σ transferring different configuration of the p_i's into the same order statistics vector $\vec{p}_{()} = \sigma(\vec{p}) = (p_{\sigma_1}, .., p_{\sigma_m})$, and thus $\vec{p}$ is given by applying the inverse permutation σ^{-1} to $\vec{p}_{()}$. Under the assumption that the p-value are independent and the null p-values are $U[0,1]$, g can be written as:

$$g(\vec{p}_{()}) = \sum_{\sigma \in S_m} f_{p_{1\ldots m}}(\sigma^{-1}(\vec{p}_{()})) = \sum_{\sigma \in S(m)} \Big[\prod_{i=m_0+1}^{m} f(p_{(\sigma_i^{-1})}) \Big] \tag{25}$$

In order to show that the inequality (23) holds for the integral, it is enough to show it for each realization of the order statistics $\vec{p}_{()}$. Thus, we want to show:

$$\frac{E[V_1(\vec{p})|\vec{p}_{()}]}{R_1^+(\vec{p}_{()})} \leq \frac{E[V_2(\vec{p})|\vec{p}_{()}]}{R_2^+(\vec{p}_{()})}, \ \forall \vec{p}_{()} \tag{26}$$

Or:

$$E[V_1|\vec{p}_{()}] \leq \frac{R_1^+(\vec{p}_{()})}{R_2^+(\vec{p}_{()})} E[V_2|\vec{p}_{()}], \ \forall \vec{p}_{()} \tag{27}$$

It is enough to show that eq. (27) holds for the case $R_1(\vec{p}) = k, R_2(\vec{p}) = k+1$ for some $0 \leq k \leq m-1$, and then it will follow by induction for the case $R_2(\vec{p}) - R_1(\vec{p}) > 1$. Define the r.v. $x_j(\vec{p}) = 1_{\{p_{(j)} null\}}$, i.e. the indicator for the event that the j-th order statistic is null.

$$x_j(\vec{p}) = \begin{cases} 1 & \sigma_j \leq m_0 \\ 0 & \sigma_j > m_0 \end{cases} \tag{28}$$

where σ here is the permutation transferring $\vec{p}$ to $\vec{p}_{()}$. We thus need to prove:

$$E[V|k;\vec{p}_{()}] = \sum_{j=1}^{k} E[x_j(\vec{p})|\vec{p}_{()}] \leq \frac{k}{k+1} \sum_{j=1}^{k+1} E[x_j(\vec{p})|\vec{p}_{()}] = \frac{k}{k+1} E[V|k+1;\vec{p}_{()}] \tag{29}$$

It is therefore enough to prove

$$E[x_j(\vec{p})|\vec{p}_{()}] \leq E[x_{k+1}(\vec{p})|\vec{p}_{()}], \ \forall j < k+1 \tag{30}$$

or in other words, that $E[x_j|\vec{p}_{()}]$ is monotonically non-decreasing in j. We will show that $E[x_k|\vec{p}_{()}] \leq E[x_{k+1}|\vec{p}_{()}]$ and then the claim follows again by induction.

$$E[x_k(\vec{p})|\vec{p}_{()}] = \sum_{\sigma \in S_m} Pr(\vec{p} = \sigma^{-1}(\vec{p}_{()})|\vec{p}_{()}) x_k\big(\sigma^{-1}(\vec{p}_{()})\big) = \frac{1}{Z(\vec{p}_{()})} \sum_{\sigma \in S_m} f\big(\sigma^{-1}(\vec{p}_{()})\big) x_k\big(\sigma^{-1}(\vec{p}_{()})\big) =$$

$$= \frac{1}{Z(\vec{p}_{()})} \sum_{\sigma \in S_m} \Big[\prod_{i=m_0+1}^{m} f(p_{(\sigma_i^{-1})}) \Big] 1_{\{\sigma_k \leq m_0\}} \tag{31}$$

Where $Z(\vec{p}_{()})$ is a normalization constant depending on the order statistics, and we have used the independence of all p-values. For each permutation on m elements σ, we define σ' to be the permutation identical to σ, except that σ_k and σ_{k+1} are swapped, i.e. $\sigma'_k = \sigma_{k+1}, \sigma'_{k+1} = \sigma_k$. Then we can write:

$$E[x_k|\vec{p}_{()}] - E[x_{k+1}|\vec{p}_{()}] =$$

$$\frac{1}{Z(\vec{p}_{()})}\sum_{\sigma\in S_m}\left\{\Big[\prod_{i=m_0+1}^{m} f(p_{(\sigma_i^{-1})})\Big]1_{\{\sigma_k\le m_0\}}-\Big[\prod_{i=m_0+1}^{m} f(p_{(\sigma_i'^{-1})})\Big]1_{\{\sigma'_{k+1}\le m_0\}}\right\} \tag{32}$$

The usage of the swapped permutation in the above sum makes the value of the two indicators identical, and thus we sum only over permutations σ such that $\sigma_k \le m_0$, i.e. when $p_{(k)}$ is null. In the case where $p_{(k+1)}$ is also null (i.e. $\sigma_{k+1} \le m_0$) the difference is zero and we can omit this case from the sum, while in the case where $p_{(k+1)}$ is alternative ($\sigma_{k+1} > m_0$) one element in the product is different and we get:

$$E[x_k|\vec{p}_{()}] - E[x_{k+1}|\vec{p}_{()}] = \frac{1}{Z(\vec{p}_{()})}\sum_{\sigma\in S_m} 1_{\{\sigma_k\le m_0<\sigma_{k+1}\}}\Big[\prod_{i=m_0+1}^{m} f(p_{(\sigma_i^{-1})})\Big]\Big[1-\frac{f(p_{(k)})}{f(p_{(k+1)})}\Big] \le 0 \tag{33}$$

where the last inequality follows from the monotonicity assumption on $f(p)$.

■

5 Supplement E: Simulations study details

A simulation study was done in order to determine the performance of the proposed procedure and compare it to existing procedures. We generated multivariate Gaussian random variables (and corresponding p-values) in similar to [2] and previous works. First randomize a vector of i.i.d. r.v.s. $Y_1,,,Y_{m+1} \sim N(0,1)$; then, given the parameters m, m_0, μ_1 and ρ, build the vector $X_1,,,X_m$ (which is the test statistics vector) as follows: the first m_0 elements are $X_i = \sqrt{\rho}Y_{m+1} + \sqrt{1-\rho}Y_i$, and the remaining $m - m_0$ elements are $X_i = \sqrt{\rho}Y_{m+1} + \sqrt{1-\rho}Y_i + \mu_1$. Here m_0/m is the fraction of true hypotheses, ρ is a dependency factor (the correlation coefficient between X_i and X_j for $i \neq j$), and μ_1 is the mean of the false hypotheses test statistics (the signal intensity). The resulting vector X is such that its first m_0 variables come from the $N(0,1)$ distribution, and the remaining $m - m_0$ variables come from $N(\mu_1, 1)$ distribution, where for any X_i and X_j (either both, one or none of them are null) their correlation coefficient is ρ. The p-values were calculated using 2 tailed z-test ($p = 2\Phi(-|x|)$). The number of simulations for each case was 50000, which provided highly accurate and reproducible results. Since the simulation results depend on several parameters, $m_0/m, \mu_1, \rho, m$, we have chosen to vary two parameters at a time, and present the results using isolines of the actual FDR (or any other quantity). These isoline plotted in Fig 3 describe the performance of the IBHsum and IBHlog procedures, respectively, on simulated data in the $(m_0/m, \mu_1)$ plane.

References

[1] Benjamini, Y. and Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1168, 2001.

[2] Gavrilov, Y., Benjamini, Y. and Sarkar, S.K. An adaptive step-down procedure with proven fdr control under indepemdence. *The Annals of Statistics*, in press, 2008.

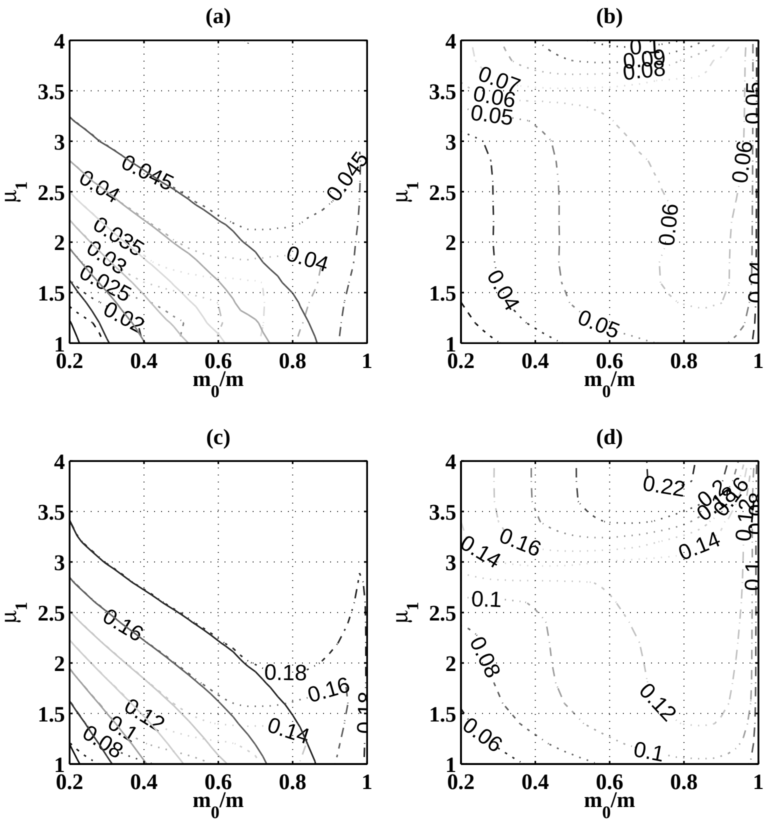

Figure 3: Isolines of $E(V/R^{+})$, measured for the IBHlog procedure by simulations, presented in the $(\mu_1, m_0/m)$ plane. The solid lines are for the step-up procedure and the dashed lines for the step-down procedure. (a) and (c) are for the independent case ($\rho = 0$). (b) and (d) are for the positive dependency case $\rho = 0.8$. The FDR levels are $q = 0.05$ in (a),(b) and $q = 0.2$ in (c),(d). In (b) we find $E(V/R^{+}) > 0.05$ for large μ_1, in violation of the bound $q = 0.05$. In similar to the behavior for IBHsum, the step-up and step-down procedures tend to coincide under dependency, while for independent p-values the step-down procedure is more conservative, especially for weak signal (small μ_1).

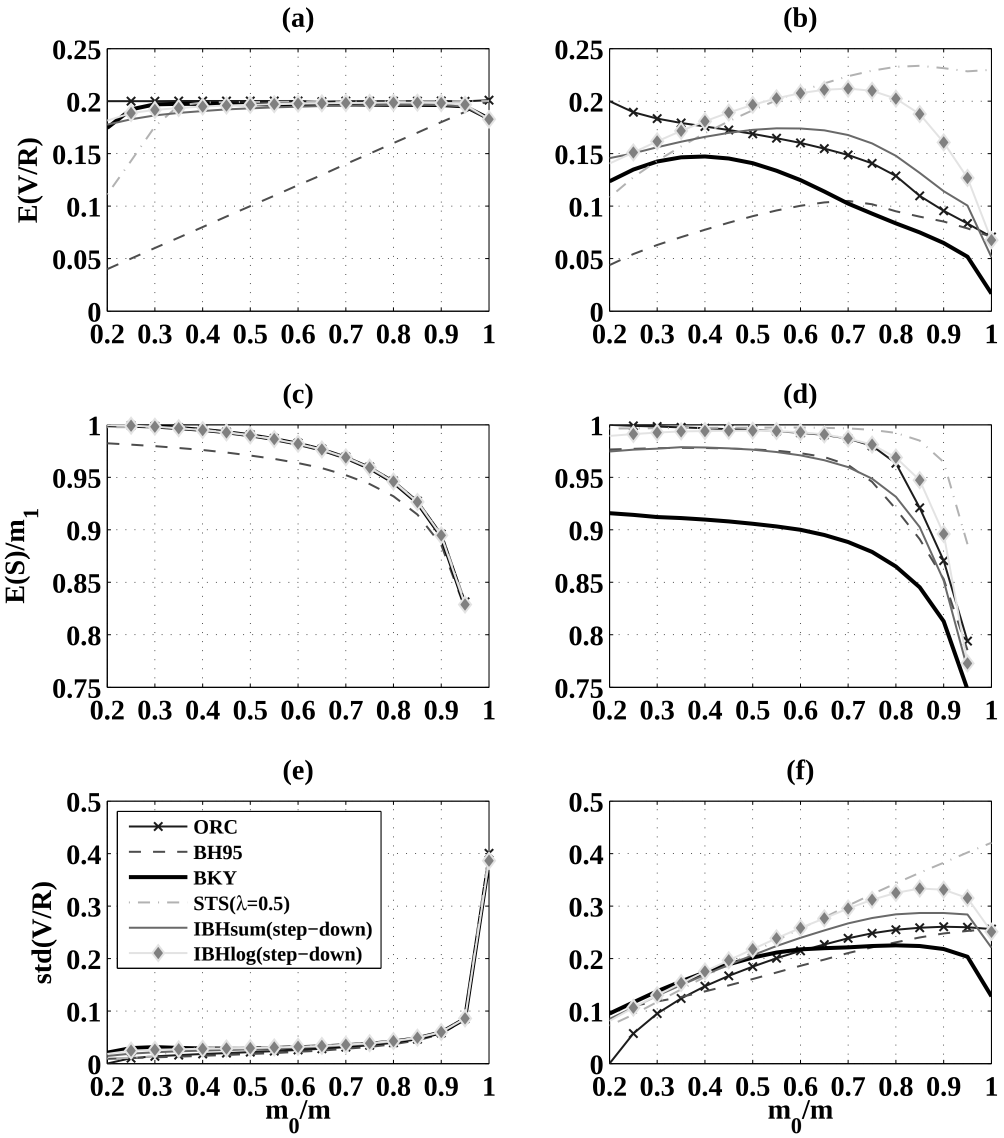

Figure 4: Results obtained for synthetic data with $m = 500$ hypotheses; m_0 was varied, the FDR was set at $q = 0.2$, the mean of the distributions P_1 was $\mu_1 = 3.5$ and the data were drawn either with covariance $\rho = 0$ [(a), (c) and (e)] or $\rho = 0.8$ [(b), (d) and (f)]. Six methods were compared: oracle (ORC), BH95, BKY, STS and our two IBH procedures (in a step down manner), showing $E(V/R^{+})$ in (a) and (b), the power $E(S)/m_1$ in (c) and (d), and the standard deviation (st.d.) of V/R^{+} in (e) and (f), for the independent case and positively dependent cases, respectively.

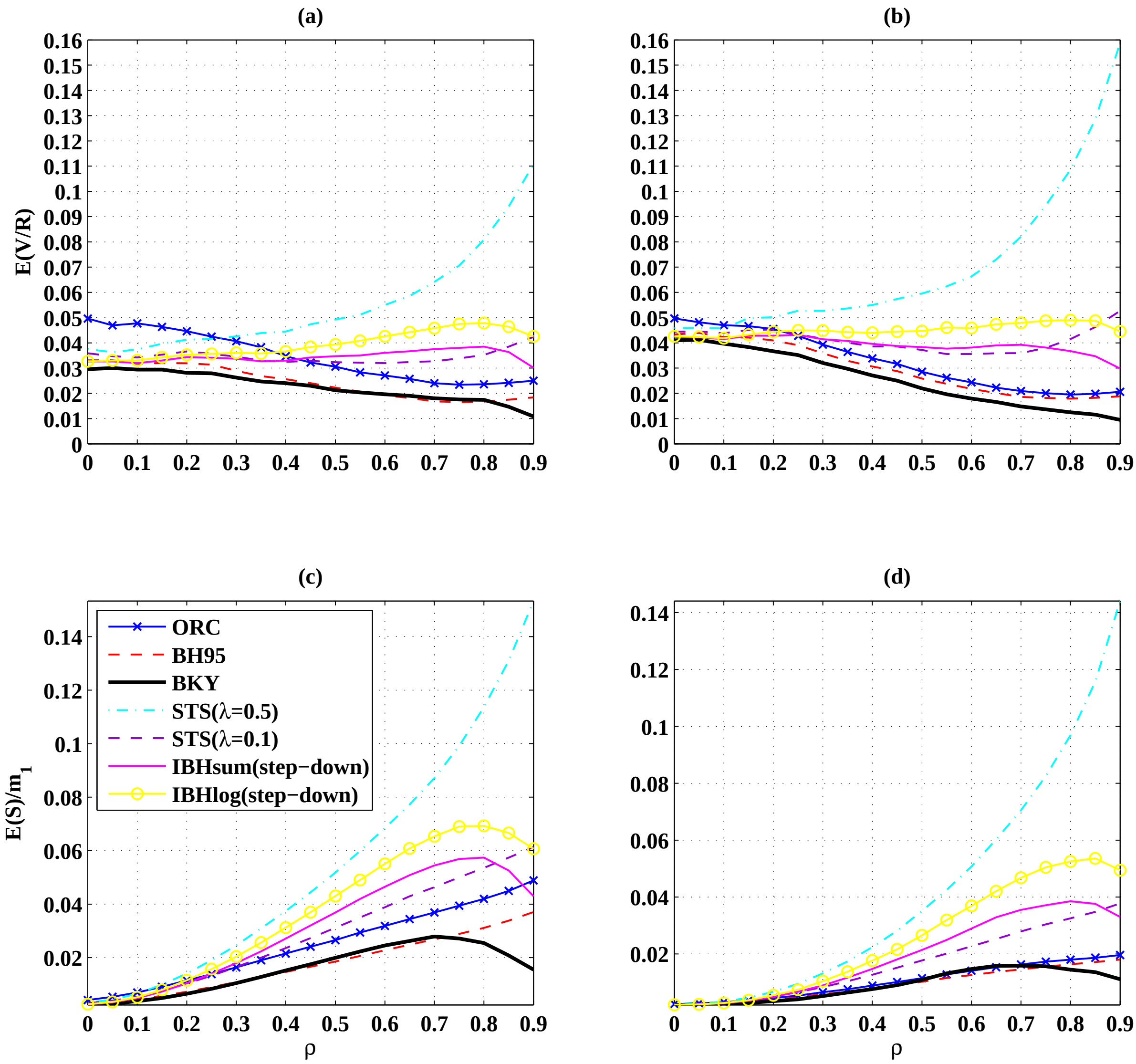

Figure 5: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^{+})$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(S)/m_1$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.05$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 1$ (weak signal). Seven methods were compared: Oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).

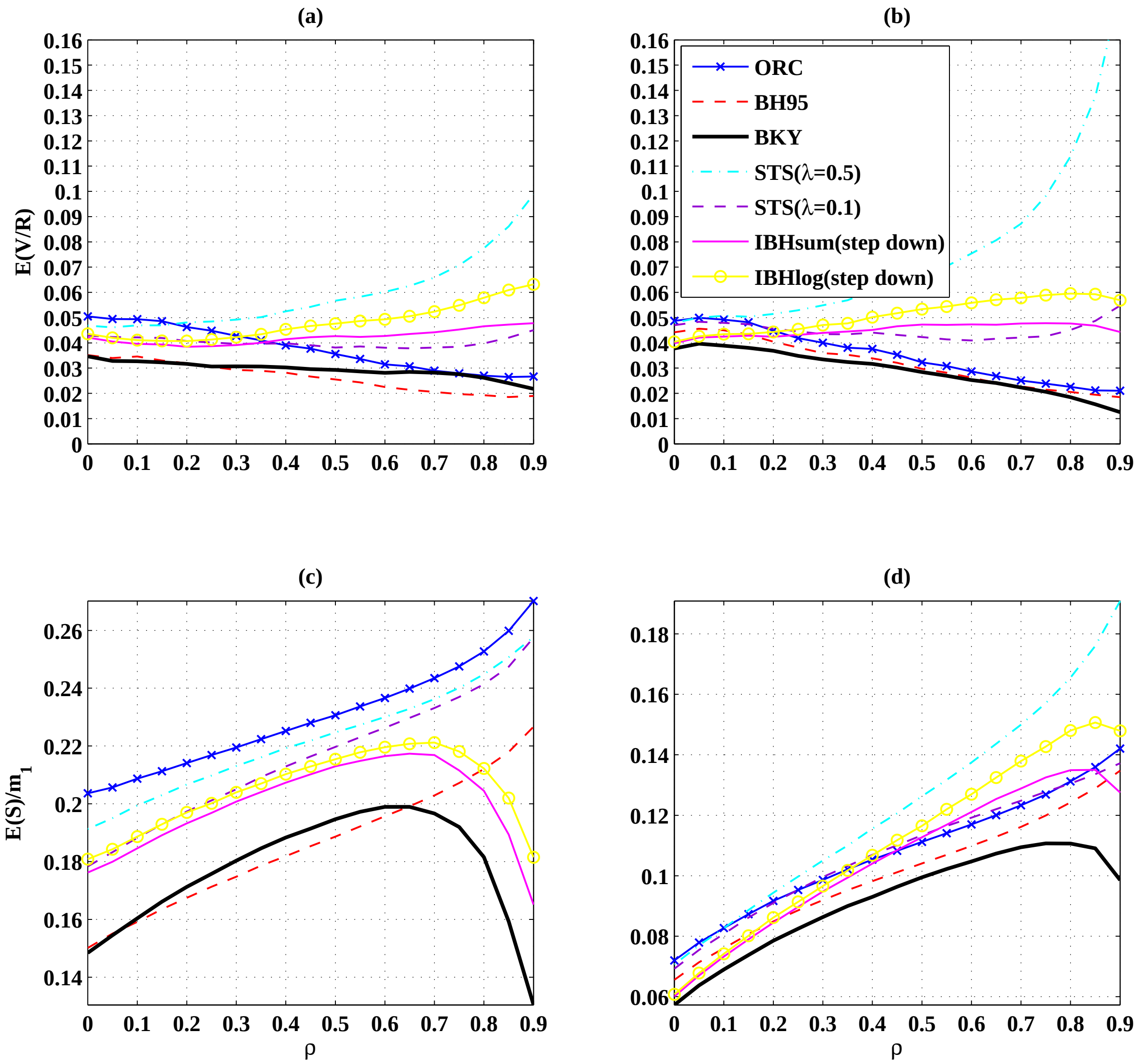

Figure 6: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^+)$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(S)/m_1$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.05$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 2$ (intermediate signal). Seven methods were compared: Oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).

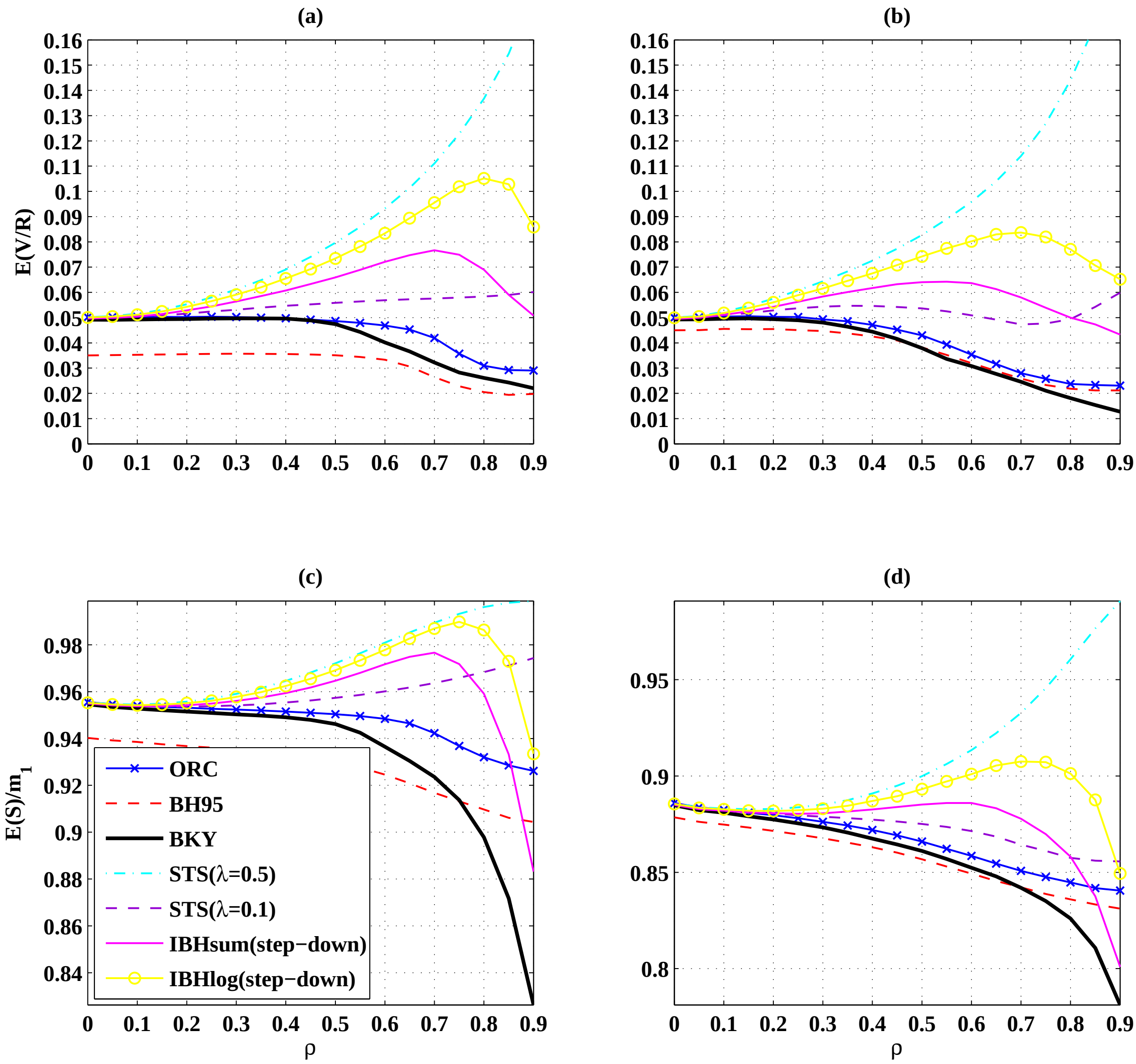

Figure 7: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^+)$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(S)/m_1$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.05$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 4$ (strong signal). Seven methods were compared: Oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).

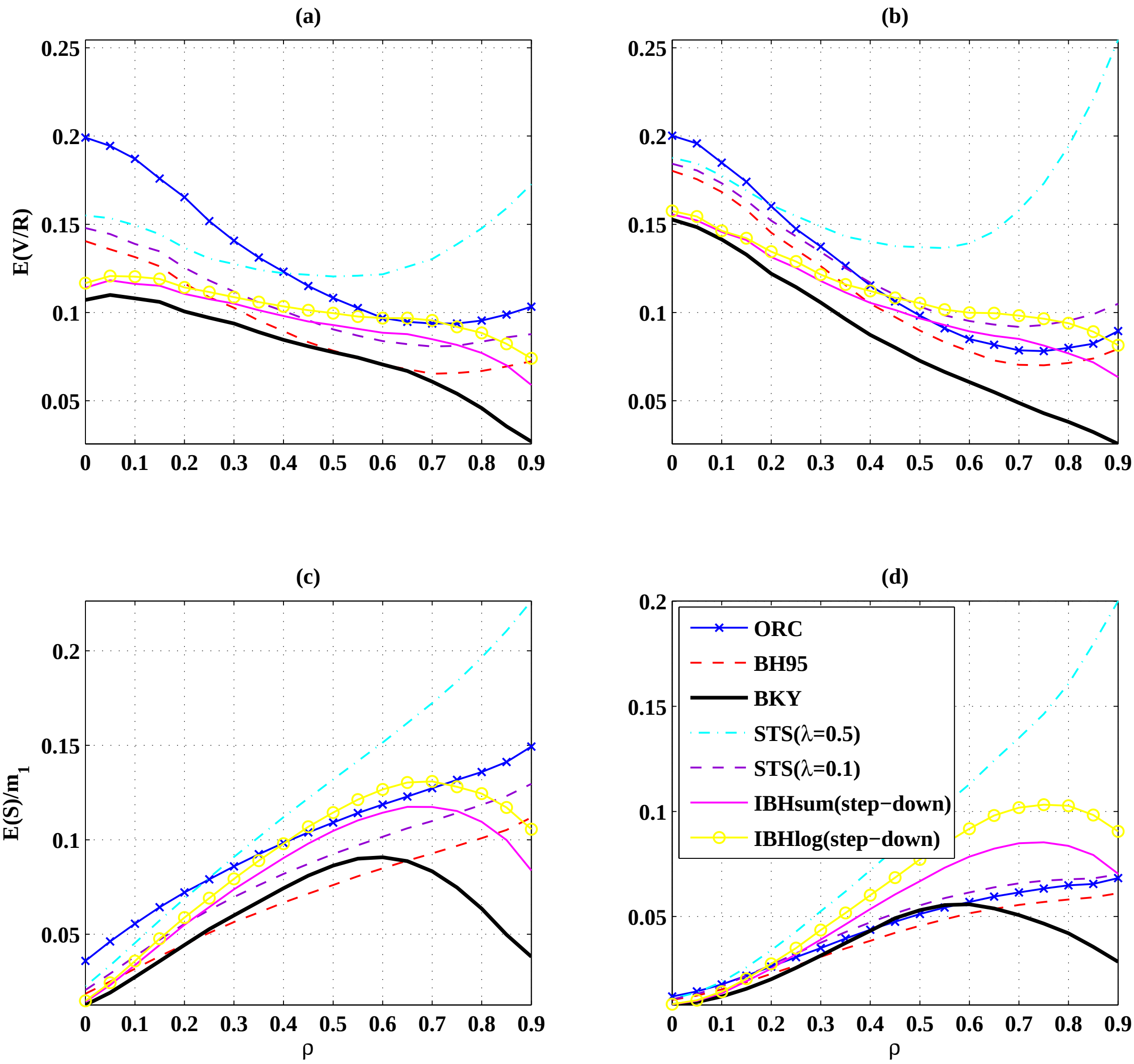

Figure 8: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^+)$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(S)/m_1$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.2$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 1$, (weak signal). Seven methods were compared: Oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).

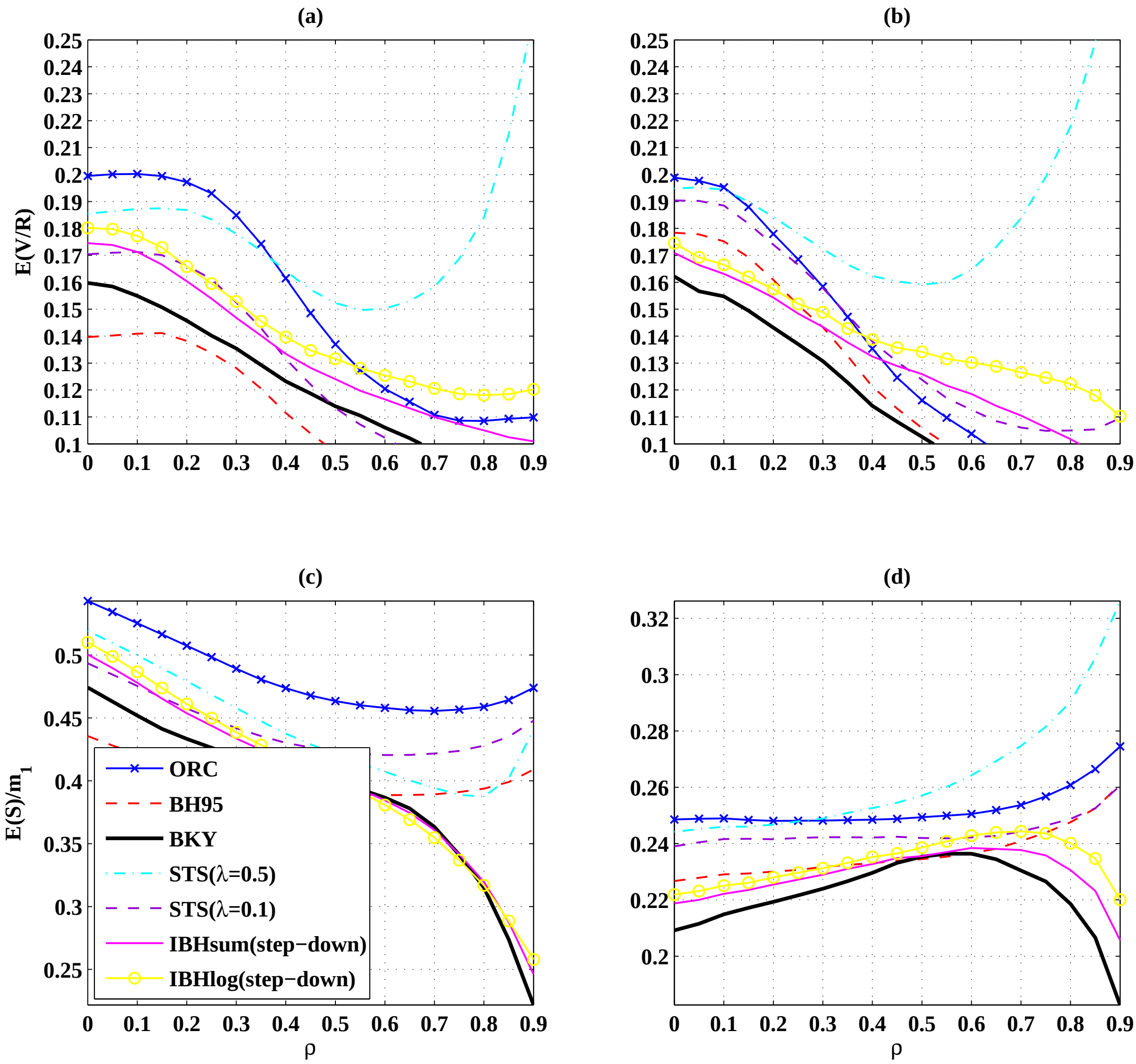

Figure 9: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^+)$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(S)/m_1$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.2$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 2$, (intermediate signal). Seven methods were compared: Oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).

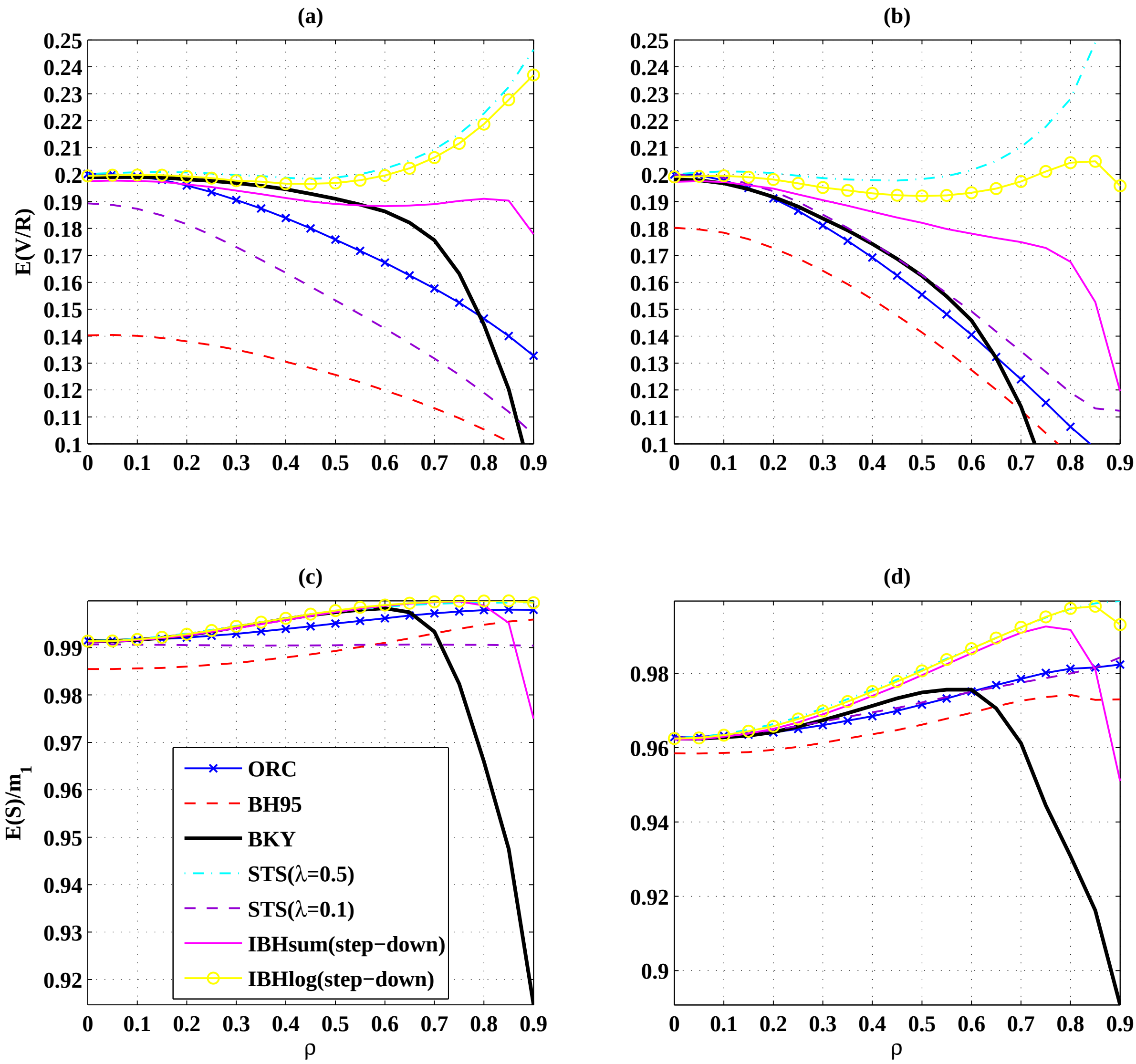

Figure 10: Results obtained for synthetic data with $m = 500$ hypotheses, (a) and (b) showing the actual FDR levels ($E(V/R^{+})$), versus the correlation between test statistics (ρ); (c) and (d) showing the actual power ($E(V/R^{+})$), versus the correlation between test statistics (ρ). The FDR was set to $q = 0.2$, the fraction of true hypotheses set to $m_0/m = 0.7$ in (a)-(c) or $m_0/m = 0.9$ in (b)-(d), the mean of the distributions P_1 was $\mu_1 = 4$, (strong signal). Seven methods were compared: oracle (ORC), BH95, BKY, STS ($\lambda = 0.5$), STS ($\lambda = 0.1$) and our two IBH procedures (in a step down manner).