Biometrika (2021), **xx**, x, *pp*. 1–15
Printed in Great Britain

Supplementary Material for Smoothed Nested Testing on Directed Acyclic Graphs

1. A GENERAL THEORY OF POSITIVE REGRESSION DEPENDENCE

We establish a general theory of positive regression dependence that extends the results of Benjamini & Yekutieli (2001) to a much broader class of statistics. Section 1.1 presents a generic result that connects the concepts of stochastic ordering with positive regression dependence. In Section 1.2 and 1.3, we prove results which imply Theorem 1 and Theorem 2, respectively. Finally, the technical lemmas are presented in Section 1.4.

1.1. A Generic Result

We start by two definitions related to stochastic orderings.

DEFINITION 1. *Given two vectors $x, y \in \mathbb{R}^n$, $x \preceq y$ iff $x_i \leq y_i$ for all $i \in \{1, \ldots, n\}$. A function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is non-decreasing iff $f(x) \leq f(y)$ for any $x \preceq y$. Further, given two probability distributions P_1, P_2 on $\mathbb{R}^d$, $P_1 \preceq P_2$ iff for any nondecreasing function f,*

$$\int_{\mathbb{R}^d} f dP_1 \leq \int_{\mathbb{R}^d} f dP_2.$$

DEFINITION 2. *A random vector X is said to be stochastically increasing in the random variable Y, denoted by $X \uparrow^{\mathrm{st}} Y$, if the regular conditional probability $\mathbb{P}(X \in \cdot \mid Y = y)$ exists (e.g. the underlying measurable space is a Polish space), and $\mathbb{E}[g(X) \mid Y = y]$ is bounded nondecreasing in y for every nondecreasing function g.*

Indeed, $(X_1, \ldots, X_n)$ satisfy positive regression dependence on a subset $T \subset \{1, \ldots, n\}$ iff

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} X_i, \quad \forall i \in T.$$

The "if" part is straightforward because the function $x \mapsto I(x \in \mathcal{C})$ for any non-decreasing set is non-decreasing. The "only if" part can be proved by approximating each non-decreasing function by sums of indicator functions on non-decreasing set; See Theorem 1 of Kamae et al. (1977) (Proposition 3).

The following theorem provides a broad class of multivariate statistics that satisfy positive regression dependence. The intuition is quite simple: under the condition $(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} Y_j$, that Y_j "increases" would imply that $(X_1, \ldots, X_n)$ "increases" and thus all other Y_i's because f_i's are non-decreasing.

THEOREM 1.1. *Let $X_1, \ldots, X_n$ be random variables on $\mathbb{R}$ and $\{f_j : j \in [M]\}$ be a collection of (entrywise) non-decreasing functions on $\mathbb{R}^n$. For each $j \in [M]$, define*

$$Y_j = f_j(X_1, \ldots, X_n).$$

Then $(Y_1, \ldots, Y_M)$ is positive regression dependence on the subset $\{j\}$ if

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} Y_j.$$

Proof. It is left to prove that, for any $y_1 < y_2$ and a bounded non-decreasing function h on $\mathbb{R}^M$,

$$\mathbb{E}[h(Y_1, \ldots, Y_M) \mid Y_j = y_1] \leq \mathbb{E}[h(Y_1, \ldots, Y_M) \mid Y_j = y_2]. \tag{1.1}$$

Because $\mathbb{R}^n$ is a Polish space, the regular conditional probability $\mathbb{P}((X_1,\ldots,X_n)\in\cdot \mid f_j(X_1,\ldots,X_n)=y)$ exists. Let P_1 and P_2 denotes two measures with

$$P_k(\cdot)=\mathbb{P}((X_1,\ldots,X_n)\in\cdot \mid f_j(X_1,\ldots,X_n)=y_k),\quad k=1,2.$$

Our condition ensures that $P_1\preceq P_2$. By Proposition 3, there exists $X_1^{(1)},\ldots,X_n^{(1)}$ and $X_1^{(2)},\ldots,X_n^{(2)}$ such that $(X_1^{(1)},\ldots,X_n^{(1)})\sim P_1$, $(X_1^{(2)},\ldots,X_n^{(2)})\sim P_2$, and

$$(X_1^{(1)},\ldots,X_n^{(1)})\preceq(X_1^{(2)},\ldots,X_n^{(2)}),\quad a.s.,$$

and for any Borel set $A\subset\mathbb{R}^n$,

$$\mathbb{P}\left((X_1^{(k)},\ldots,X_n^{(k)})\in A\right)=P_k(A)=\mathbb{P}((X_1,\ldots,X_n)\in A\mid Y_j=y_k).$$

Let

$$Y_i^{(k)}=f_i(X_1^{(k)},\ldots,X_n^{(k)}),\quad \forall i\in[M],k=1,2.$$

Since each f_i is non-decreasing, we have

$$Y_i^{(1)}\le Y_i^{(2)},\quad \forall i\in[M],\ a.s., \tag{1.2}$$

and for any Borel set $B\subset\mathbb{R}^M$ and $k=1,2$,

$$\mathbb{P}\left((Y_1^{(k)},\ldots,Y_M^{(k)})\in B\right)=\mathbb{P}((Y_1,\ldots,Y_M)\in B\mid Y_j=y_k). \tag{1.3}$$

Finally, since h is non-decreasing, (1.2) implies that

$$h(Y_1^{(1)},\ldots,Y_M^{(1)})\le h(Y_1^{(2)},\ldots,Y_M^{(2)}),\quad a.s.,$$

which further implies that

$$\mathbb{E}h(Y_1^{(1)},\ldots,Y_M^{(1)})\le\mathbb{E}h(Y_1^{(2)},\ldots,Y_M^{(2)}).$$

The proof of (1.1) is then completed by (1.3). □

1.2. Weighted averages of independent log-concave random variables

We first present a powerful result proved by Efron (1965).

PROPOSITION 1. *[Theorem 1 of Efron (1965); see also Block et al. (1985)] Suppose $X_1,\ldots,X_n$ are independent random variables on $\mathbb{R}$ or $\mathbb{Z}$ with (potentially distinct) log-concave densities. Then*

$$(X_1,\ldots,X_n)\uparrow^{\mathrm{st}}\sum_{i=1}^n X_i.$$

Remark 1. Efron (1965) assumes each X_i has a PF2 density, which is equivalent to a log-concave density; see e.g. Balabdaoui & Wellner (2014).

THEOREM 1.2. *Let $X_1,\ldots,X_n$ be random variables on $\mathbb{R}$ or $\mathbb{Z}$, and $T\subset\{1,\ldots,n\}$ such that*

(i) X_i has a log-concave density for each $i\in T$;
(ii) $\{X_i:i\in T\}$ are mutually independent;
(iii) $\{X_i:i\in T\}$ are independent of $\{X_i:i\notin T\}$.

Let

$$Y_1 = g\left(\sum_{i\in T} a_i X_i\right), \tag{1.4}$$

where $a_i \geq 0$ for all $i \in T$ and g is non-decreasing. For any $j = 2, \ldots, M$, let

$$Y_j = f_j(X_1, \ldots, X_n)$$

for some (entrywise) non-decreasing function f_j. Then $(Y_1, \ldots, Y_M)$ are positive regression dependence with respect to $\{1\}$.

Proof. By Theorem 1.1, it is left to show that

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} Y_1.$$

Since g is non-decreasing, this is equivalent to

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} \sum_{i\in T} a_i X_i$$

By condition (iii) and Lemma 1.1, it remains to prove

$$(X_i)_{i\in T} \uparrow^{\mathrm{st}} \sum_{i\in T} a_i X_i. \tag{1.5}$$

Let

$$\tilde{X}_i = b_i X_i, \quad \text{where } b_i = \begin{cases} a_i \ (a_i \neq 0) \\ 1 \ (a_i = 0) \end{cases},$$

and define

$$\tilde{f}_j(x_1, \ldots, x_n) = f_j\left(\frac{x_1}{b_1}, \ldots, \frac{x_n}{b_n}\right)$$

Since all a_i's are non-negative and b_i's are positive, each $\tilde{f}_j$ is non-decreasing. For each $i \in T$, by condition (i), $\tilde{X}_i$ has a log-concave density since $X_i \mapsto \tilde{X}_i$ is a linear mapping. Let $T_+ = \{i \in T : a_i \neq 0\}$, then

$$\sum_{i\in T} a_i X_i = \sum_{i\in T_+} \tilde{X}_i.$$

By Proposition 1,

$$(\tilde{X}_i)_{i\in T_+} \uparrow^{\mathrm{st}} Y_1.$$

By condition (ii), $\tilde{X}_i$'s are independent. Then, by Lemma 1.1,

$$(\tilde{X}_i)_{i\in T_+} \uparrow^{\mathrm{st}} Y_1 \Longrightarrow (\tilde{X}_i)_{i\in T} \uparrow^{\mathrm{st}} Y_1.$$

The equation (1.5) is then proved. □

Now we prove Theorem 1 as a special case of Theorem 1.2.

*Proof (**Theorem 1**).* Fix any null node v. For each node $c \in \mathcal{C}_v$, let $X_c = H_{v,c}^{-1}(p_c)$. Since $H_{v,c}$ is monotone increasing and $p_c \sim \mathrm{Uniform}([0,1])$, for any $x \in \mathbb{R}$,

$$\mathbb{P}(X_c \leq x) = \mathbb{P}(p_c \leq H_{v,c}(x)) = H_{v,c}(x).$$

The density function of p_c is then $H'_{v,c}$, which is log-concave. Since G_v and $H_{v,c}^{-1}$ are monotone increasing for any (v, c), each smoothed p-value and original p-value is a monotone increasing

transformation of $(X_1, \ldots, X_n)$. Setting $T = |\mathcal{C}_b|$, $g = G_v$ and $a_c = 1$ for all $c \in T$ in Theorem 1.2, we conclude that the smoothed p-values satisfy positive regression dependence on the subset $\{v\}$. Since this holds for any null node v, th p-values satisfy positive regression dependence on the subset of nulls. □

1.3. Local order statistics

We first present a powerful result proved by Block et al. (1987).

PROPOSITION 2 (COROLLARY IN SECTION 3 OF BLOCK ET AL. (1987)). *Let $X_1, \ldots, X_n$ be independent random variables on $\mathbb{R}$ with a common continuous distribution. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the ordered statistics. For any given subsets $\{k_1, \ldots, k_r\}$ of $\{1, \ldots, n\}$,*

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} (X_{(k_1)}, \ldots, X_{(k_r)}).$$

Theorem 2 is a direct consequence of the following theorem.

THEOREM 1.3. *Let $X_1, \ldots, X_n$ be random variables on $\mathbb{R}$ or $\mathbb{Z}$, and $T \subset \{1, \ldots, n\}$ such that*

(i) $\{X_i : i \in T\}$ are i.i.d.;
(ii) $\{X_i : i \in T\}$ are independent of $\{X_i : i \notin T\}$.

Let

$$Y_1 = g\left(X_{(k);T}\right), \tag{1.6}$$

where g is an non-decreasing function, T is a subset of $[n]$ and $X_{(k);T}$ denotes the k-th order statistics of $(X_i)_{i \in T}$. For any $j = 2, \ldots, M$, let

$$Y_j = f_j(X_1, \ldots, X_n)$$

for some non-decreasing function f_j. Then $(Y_1, \ldots, Y_M)$ are PRDS with respect to $\{1\}$.

Proof. Similar to (1.5) in proof of Theorem 1.2, it is left to prove that

$$(X_1, \ldots, X_n) \uparrow^{\mathrm{st}} X_{(k);T}.$$

By condition (ii) and Lemma 1.1, it remains to prove that

$$(X_i)_{i \in T} \uparrow^{\mathrm{st}} X_{(k);T}.$$

This is proved by condition (i) and Proposition 2. □

1.4. Technical lemmas

We first present a coupling property of stochastic orderings.

PROPOSITION 3 (THEOREM 1 OF KAMAE ET AL. (1977)). *Let E be a partially ordered Polish space (i.e. complete separable metrizable topological space). For any two distributions P_1 and P_2 on E, $P_1 \preceq P_2$ iff there exists a distribution P on $E \times E$ equipped with the product topology, which is supported on the set $\{(x, y) \in E \times E\}$ with first marginal P_1 and second marginal P_2. Equivalently, there exists a random vector (X, Y) on $E \times E$, such that $X \preceq Y$ almost surely.*

Next we prove a useful property of stochastic monotonicity.

LEMMA 1.1. *Assume that the sample space of the random element (X, Y, Z) is partially ordered. If $X \uparrow^{\mathrm{st}} Y$ and $Z \perp (X, Y)$, then*

$$(X, Z) \uparrow^{\mathrm{st}} Y, \quad (X, Z) \uparrow^{\mathrm{st}} (Y, Z)$$

Proof. Let h be any bounded non-decreasing function on the domain of (X, Z). Since $Z \perp (X, Y)$,

$$\mathbb{E}[h(X,Z) \mid Y] = \mathbb{E}[\mathbb{E}_Z[h(X,Z)] \mid Y] = \mathbb{E}\left[\int h(X,z) dP_Z(z) \mid Y\right].$$

Let $\tilde{h}(x) = \int h(x,z) dP_Z(z)$. Since $h(x,z)$ is non-decreasing in x for any z, $\tilde{h}(x)$ is also bounded non-decreasing in x. Since $X \uparrow^{\text{st}} Y$, $\mathbb{E}[\tilde{h}(X) \mid Y = y]$ is non-decreasing in y. This proves the first result.

For the second result, given any y, z,

$$\mathbb{E}[h(X,Z) \mid Y = y, Z = z] = \mathbb{E}[h(X,z) \mid Y = y, Z = z] = \mathbb{E}[h(X,z) \mid Y = y].$$

For any $y \preceq y'$ and $z \preceq z'$,

$$\mathbb{E}[h(X,z) \mid Y = y] \leq \mathbb{E}[h(X,z') \mid Y = y] \leq \mathbb{E}[h(X,z') \mid Y = y']$$

where the first inequality uses the fact that $h(X,z)$ is non-decreasing in z, and the second inequality uses the fact that $h(x,z')$ is non-decreasing in x and $X \uparrow^{\text{st}} Y$. Therefore,

$$\mathbb{E}[h(X,Z) \mid Y = y, Z = z] \leq \mathbb{E}[h(X,Z) \mid Y = y', Z = z'].$$

2. Other Technical Details

We here recall some notation for the benefit of the reader. For each $v \in \{1, \cdots, n\} = \mathcal{V}$ we can observe a random variable p_v. Let $\bar{\mathcal{S}} \subset \mathcal{V}$; we assume $\{p_v\}_{v \in \bar{\mathcal{S}}} \overset{\text{i.i.d}}{\sim} \text{Uniform}[0,1]$. $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ whose edges encode logical constraints on the hypotheses: $v \to w \implies (v \in \bar{\mathcal{S}} \implies w \in \bar{\mathcal{S}})$. For each v we have that $\mathcal{C}_v$ denotes the union of v with all of its descendants in the graph $\mathcal{G}$ and $f_v(x_{\mathcal{C}_v}, x_{\mathcal{V} \setminus \mathcal{C}_v})$ denotes any function. The smoothed $\tilde{p}$-value for node v is then,

$$F_v(c; x_{\mathcal{V} \setminus \mathcal{C}_v}) \triangleq \text{pr}(f_v(u_{\mathcal{C}_v}, x_{\mathcal{V} \setminus \mathcal{C}_v}) \leq c), \quad \tilde{p}_v \triangleq F_v(f_v(p_{\mathcal{C}_v}, p_{\mathcal{V} \setminus \mathcal{C}_v}); p_{\mathcal{V} \setminus \mathcal{C}_v}),$$

where $\{u_w\}_{w \in \mathcal{C}_v}$ are independent and identically distributed as $\text{Uniform}[0,1]$.

2.1. Smoothed test statistics

The starting point for this work is Lemma 1, which we prove here.

Proof. Fix any $v \in \bar{\mathcal{S}}$. The meaning of the graph structure indicates that $\mathcal{C}_v \subset \bar{\mathcal{S}}$

(a) Our assumption that the null p-values are independent and identically distributed as $\text{Uniform}[0,1]$ and independent of non-null p-values imply that

$$p_{\mathcal{C}_v} \overset{d}{=} u_{\mathcal{C}_v} \mid p_{\mathcal{V} \setminus \mathcal{C}_v}.$$

As a result,

$$\tilde{p}_v \overset{d}{=} F_v(f_v(u_{\mathcal{C}_v}; p_{\mathcal{V} \setminus \mathcal{C}_v})) \mid p_{\mathcal{V} \setminus \mathcal{C}_v}.$$

By definition, $F_v(f_v(u_{\mathcal{C}_v}; p_{\mathcal{V} \setminus \mathcal{C}_v}))$ is super-uniform.

(b) Since null p-values are independent and super-uniform ,

$$p_{\mathcal{C}_v} \preceq u_{\mathcal{C}_v} \mid p_{\mathcal{V} \setminus \mathcal{C}_v},$$

where $\preceq$ denotes entrywise stochastic dominance. Since f_v is nondecreasing in $p_{\mathcal{C}_v}$,

$$f_v(p_{\mathcal{C}_v}, p_{\mathcal{V} \setminus \mathcal{C}_v}) \succeq f_v(u_{\mathcal{C}_v}, p_{\mathcal{V} \setminus \mathcal{C}_v}) \mid p_{\mathcal{V} \setminus \mathcal{C}_v}.$$

Further, since F_v is nondecreasing,

$$\tilde{p}_v \succeq F_v(f_v(u_{\mathcal{C}_v}, p_{\mathcal{V}\setminus\mathcal{C}_v})) \mid p_{\mathcal{V}\setminus\mathcal{C}_v}.$$

As in case (a), we conclude that $\tilde{p}_v$ is super-uniform. □

2.2. *Familywise error rate control*

We apply the method of Meijer & Goeman (2015) for familywise error rate control based on the Sequential Rejection Principle developed in Goeman & Solari (2010). For the benefit of the reader, we sketch the main ideas below.

The starting-place for this analysis is an application of the Sequential Rejection Principle of Goeman and Solari.

DEFINITION 3. *Fix $n \in \mathbb{N}$, let $\mathcal{P}$ denote the power set $\{1 \cdots n\}$, let $\pi : \{1 \cdots n\} \times \mathcal{P} \to \mathbb{R}$. Let $\hat{\mathcal{S}}_0 \triangleq \emptyset$ and recursively define*

$$\hat{\mathcal{S}}_i(p, \pi) \triangleq \hat{\mathcal{S}}_{i-1}(p, \pi) \cup \{v : p_v \leq \pi(v, \hat{\mathcal{S}}_{i-1}(p, \pi))\} \tag{2.1}$$

The limit of this process, $\hat{\mathcal{S}}(p, \pi) \triangleq \cup_i^n \hat{\mathcal{S}}_i(p, \pi)$, is said to be the final result of sequential rejection on p via π.

Intuitively, π is an object which uses the current set of rejections to produce a weight for each hypothesis. These weights are then used to decide whether further rejections may be made. Goeman and Solari provide conditions on π such that this algorithm conserves familywise error.

THEOREM 2.1 (GOEMAN AND SOLARI). *Let $\mathbb{M} \subset \mathcal{P}$. For each $\mathcal{S} \in \mathbb{M}$ let $P_{\mathcal{S}}$ denote a probability distribution for the variables $p_1, p_2, \cdots, p_n \in \mathbb{R}$. Let $\pi : \{1 \cdots n\} \times \mathcal{P} \to \mathbb{R}$ satisfy $P_{\mathcal{S}}(\{v : p_v \leq \pi(v, \mathcal{S})\} \subset \mathcal{S}) \geq 1 - \alpha$ for every $\mathcal{S} \in \mathbb{M}$ (this will be referred to as the "single-step condition"). Let us further assume that $\pi(v, \hat{\mathcal{S}}_i(p, \pi)) \leq \pi(v, \mathcal{S})$ for any i, p and any $\mathcal{S} \in \mathbb{M}$ such that $\mathcal{S} \supseteq \hat{\mathcal{S}}_i(p, \pi)$ and any $v \notin \mathcal{S}$ (this will be referred to as the "monotonicity condition"). It follows that*

$$P_M(\hat{\mathcal{S}}(p, \pi) \subset \mathcal{S}) \geq 1 - \alpha$$

for any $\mathcal{S} \in \mathbb{M}$.

Proof. Consider the event that $\{v : p_v \leq \pi(v, \mathcal{S})\} \subset \mathcal{S}$, i.e.

$$p_v > \pi(v, \mathcal{S}) \qquad \forall v \in \bar{\mathcal{S}}.$$

Here we adopt the notation $\bar{\mathcal{S}} = \{1 \cdots n\}\setminus\mathcal{S}$. The single-step condition guarantees that this event occurs with probability at least $1 - \alpha$. Thus to prove our point it suffices to show that $\hat{\mathcal{S}} \subset \mathcal{S}$ whenever this event occurs.

We argue by induction. Clearly $\hat{\mathcal{S}}_0 \subset M$. Now suppose that $\hat{\mathcal{S}}_{i-1} \subset \mathcal{S}$; the monotonicity assumption then yields that $\pi(v, \hat{\mathcal{S}}_{i-1}) \leq \pi(v, \mathcal{S})$. On the other hand, we have already assumed that $p_v > \pi(v, \mathcal{S})$ for every $v \in \bar{\mathcal{S}}$. Together, these facts imply that $p_v \geq \pi(v, \hat{\mathcal{S}}_{i-1})$ for every $v \in \bar{\mathcal{S}}$. It follows that

$$\hat{\mathcal{S}}_i = \hat{\mathcal{S}}_{i-1} \cup \{v : p_v \leq \pi(v, \hat{\mathcal{S}}_{i-1})\} \subset \mathcal{S}$$

By induction, it follows that $\hat{\mathcal{S}} \subset \mathcal{S}$. □

Meijer & Goeman (2015) provide several weight functions that satisfy the single-step and monotonicity conditions of the Theorem above. For simplicity, in this work we focus on their "all-parents" weight function. This function produces weights by using a particular "water-filling" procedure, described below.

Algorithm 1. All-parents water-filling procedure

input : a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a subset $\hat{\mathcal{S}} \subseteq \mathcal{V}$
$\mathcal{Z}(\hat{\mathcal{S}}) \leftarrow |\{v \notin \hat{\mathcal{S}} : \ v \text{ is a leaf}\}|$;
foreach $v \in \mathcal{V}$ **do**
  if *v is a leaf and $v \notin \hat{\mathcal{S}}$* **then**
    $g_v \leftarrow 1/|\mathcal{Z}(\hat{\mathcal{S}})|$;
  else
    $g_v \leftarrow 0$;
while *there exists a node v such that $g_v \geq 0$ and* $\mathrm{parents}(v) \backslash \hat{\mathcal{S}} \neq \emptyset$ **do**
  foreach $w \in \mathrm{parents}(v) \backslash \hat{\mathcal{S}}$ **do**
    $g_w \leftarrow g_w + \frac{g_v}{|\mathrm{parents}(v) \backslash \hat{\mathcal{S}}|}$;
  $g_v \leftarrow 0$;
return g

This algorithm can be equivalently defined via the following recursive relations:

$$\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}) = \begin{cases} 0 & \text{if } v \in \hat{\mathcal{S}} \\ 1/|\{v \notin \hat{\mathcal{S}} : \ v \text{ is a leaf of } \mathcal{G}\}| & \text{if } v \notin \hat{\mathcal{S}} \text{ is a leaf of } \mathcal{G} \\ \sum_{w \in \mathrm{children}(w)} \frac{\tilde{g}_w(\mathcal{G}, \hat{\mathcal{S}})}{|\mathrm{parents}(w) \backslash \hat{\mathcal{S}}|} & \text{otherwise} \end{cases}$$

$$g_v(\mathcal{G}, \hat{\mathcal{S}}) = \begin{cases} \tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}) & \text{if } \mathrm{parents}(v) \subset \hat{\mathcal{S}} \\ 0 & \text{otherwise} \end{cases}$$

Both points of view are helpful in the proofs below.

For any directed acyclic graph $\mathcal{G}$ and any given target error level, α, this weight-filling function can be used to produce a satisfactory weight function by taking $\pi(v, \hat{\mathcal{S}}) = \alpha g_v(\mathcal{G}, \hat{\mathcal{S}})$. This weight function is guaranteed to satisfy the conditions of Goeman and Solari.

THEOREM 2.2 (MEIJER AND GOEMAN). *Fix a directed graph $\mathcal{G}$ and let $\mathbb{M} \subset \mathcal{P}$ denote the set of all collections of hypotheses consistent with that graph (i.e. if $\mathcal{S} \in \mathbb{M}$ and $v \in \mathcal{S}$ then all ancestors of v are also in $\mathcal{S}$). Fix any $\mathcal{S} \subset \mathbb{M}$ and assume $P_{\mathcal{S}}(p_v \leq c) \leq c, \forall v \notin \mathcal{S}$. Then the weight function $\pi(v, \hat{\mathcal{S}}) = \alpha g_v(\mathcal{G}, \hat{\mathcal{S}})$ satisfies the single-step and monotonicity conditions of Theorem 2.1.*

Proof. Let us begin with the monotonicity condition. This monotonicity condition follows from the corresponding monotonicity property on the water-filling function g, defined above. Let us fix any i, p and assume $\hat{\mathcal{S}}_i(p, \pi) \subset \mathcal{S}$.

- If $v \notin \mathcal{S}$ is a leaf, then $\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i), \tilde{g}_v(\mathcal{G}, \mathcal{S})$ denote the reciprocal of the number of unrejected leaves in each of their respective cases. Since $S \subset S'$ it follows that $\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i) \leq \tilde{g}_v(\mathcal{G}, \mathcal{S})$ for all $v \notin \mathcal{S}$.
- Now consider the case that $w \notin \mathcal{S}$ isn't a leaf. Let us apply the inductive hypothesis that $\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i) \leq \tilde{g}_v(\mathcal{G}, \mathcal{S})$ for every child $v \notin \mathcal{S}$. Now note that $\mathcal{S} \in \mathbb{M}$ and $w \notin \mathcal{S}$; it follows that all of the children of w also lie outside of $\mathcal{S}$. Thus, in fact, since all children of w lie outside $\mathcal{S}$, we have that $\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i) \leq \tilde{g}_v(\mathcal{G}, \mathcal{S})$ for every child v. We obtain that

$$\tilde{g}_w(\mathcal{G}, \hat{\mathcal{S}}_i) = \sum_{v \in \mathrm{children}(w)} \frac{\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i)}{\left|\mathrm{parents}(v) \backslash \hat{\mathcal{S}}_i\right|} \leq \sum_{v \in \mathrm{children}(w)} \frac{\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i)}{|\mathrm{parents}(v) \backslash \mathcal{S}|} \leq \tilde{g}_w(\mathcal{G}, \mathcal{S})$$

Putting these two facts together and applying an induction argument on the nodes in topological order from the bottom of the tree, it follows that $\tilde{g}_v(\mathcal{G}, \hat{\mathcal{S}}_i) \leq \tilde{g}_v(\mathcal{G}, \mathcal{S})$ for every $v \notin \mathcal{S}$. Finally, observe that

$$\text{parents}(v) \subset \hat{\mathcal{S}}_i \implies \text{parents}(v) \subset \mathcal{S}$$

In combination with the inequality on $\tilde{g}$, this yields that $g_v(\mathcal{G}, \hat{\mathcal{S}}_i) \leq g_v(\mathcal{G}, \mathcal{S})$ for all $v \notin \mathcal{S}$, as desired.

Now let us turn to the single-step condition. First observe that the water-filling function g always satisfies $\sum_v g_v = 1$. Indeed, observe that Algorithm 1 initializes the values of g_v to satisfy this property, and the total value of $\sum_v g_v$ is unchanged at each iteration of the while loop. The single-step condition then follows from a union bound:

$$\begin{aligned}
\text{pr}\left(\{v : p_v \leq \pi(v, \mathcal{S})\} \subseteq \mathcal{S}\right) &= \text{pr}(p_v > \alpha g_v(\mathcal{G}, \mathcal{S}) \; \forall v \in \bar{\mathcal{S}}) \\
&\geq 1 - \sum_{v \in \bar{\mathcal{S}}} \text{pr}(p_v \leq \alpha g_v(\mathcal{G}, \mathcal{S})) \\
&\geq 1 - \sum_{v \in \bar{\mathcal{S}}} \alpha g_v(\mathcal{G}, \mathcal{S}) \geq 1 - \alpha.
\end{aligned}$$

2.3. *False exceedance control*

We combine the method of Meijer & Goeman (2015) and the generic procedure of Genovese & Wasserman (2006), which turns any familywise error rate controlling method into a false exceedance controlling method. The procedure is outlined in Alg. 2. Here `topological_sort` refers to a topological ordering of the vertices of the graph such that any node appears later than its parents (Cormen et al., 2009, Section 22.4).

Algorithm 2. Hybrid method to control false exceedance rate

input : Smoothed p-values $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_n)$, directed acyclic graph $\mathcal{G}$, exceedance level γ, α

$\hat{\mathcal{S}}_0 \leftarrow$ Algorithm 1 with inputs $\tilde{p}, \mathcal{G}, \alpha$;

$\hat{\mathcal{S}}' \leftarrow$ first $\lfloor |\hat{\mathcal{S}}| \gamma / (1 - \gamma) \rfloor$ elements of $\text{topological\_sort}(\mathcal{G} \setminus \hat{\mathcal{S}}_0)$;

return $\hat{\mathcal{S}}_0 \cup \hat{\mathcal{S}}'$

2.4. *False discovery rate control*

For false discovery rate control, we use the unreshaped version of the Ramdas et al. (2019) method on our $\tilde{p}$-values. It is a recursive step-up procedure with graph-specific thresholds. In particular, for a directed acyclic graph, the depth of each node d_v is defined as the length of the longest path from v to any root, the node without parents. Let $\mathcal{V}_d = \{v : d_v = d\}$. For each node $v \in \mathcal{V}_d$, let $\{\alpha_{d,v}(r) : r = 1, 2, \ldots\}$ be a sequence of thresholds which will be specified later. The procedure starts by applying a generalized step-up procedure on $\mathcal{V}_1$ to obtain an initial rejection set:

$$\hat{\mathcal{S}}_1 = \{v \in \mathcal{V}_1 : \tilde{p}_v \leq \alpha_{1,v}(R_1)\},$$

where

$$R_1 = \max \left\{ 1 \leq r \leq |\mathcal{V}_1| : \sum_{v \in \mathcal{V}_1} I(\tilde{p}_v \leq \alpha_{1,v}(r)) \geq r \right\}.$$

Once $\hat{\mathcal{S}}_1, \ldots, \hat{\mathcal{S}}_{d-1}$ are decided, let $\mathcal{V}'_d$ be the set of nodes with depth d and all parents rejected, i.e.

$$\mathcal{V}'_d = \left\{ v \in \mathcal{V}_d : \texttt{parents}(v) \subset \bigcap_{j=1}^{d-1} \hat{\mathcal{S}}_j \right\}.$$

Then the procedure computes $\hat{\mathcal{S}}_d$ as

$$\hat{\mathcal{S}}_d = \left\{ v \in \mathcal{V}'_d : \tilde{p}_v \leq \alpha_{d,v}(R_d) \right\},$$

where

$$R_d = \max \left\{ 1 \leq r \leq |\mathcal{V}'_d| : \sum_{v \in \mathcal{V}'_d} I(\tilde{p}_v \leq \alpha_{d,v}(r)) \geq r \right\}.$$

The final rejection set of the Ramdas et al. (2019) method is then

$$\hat{\mathcal{S}} = \bigcup_{d=1}^{D} \hat{\mathcal{S}}_d,$$

where D is the maximal depth. By definition, every node in $\hat{\mathcal{S}}_d$ has all their parents rejected in $\hat{\mathcal{S}}_1 \cup \ldots \cup \hat{\mathcal{S}}_{d-1}$. Therefore, $\hat{\mathcal{S}}$ obeys the logical constraint.

To define the threshold $\alpha_{d,v}(r)$, Ramdas et al. (2019) define the effective number of leaves ℓ_v and the effective number of nodes m_v for each node v. Similar to the weight in Meijer-Goeman algorithm, ℓ_v and m_v are computed via a "water-filling" algorithm in a bottom-up fashion. Specifically, they set $\ell_v = m_v = 1$ for each leaf node and then proceed up the tree, from leaves to roots, recursively calculating

$$\ell_v = \sum_{w \in \texttt{children}(v)} \frac{\ell_w}{|\texttt{parents}(w)|}, \quad m_v = 1 + \sum_{w \in \texttt{children}(v)} \frac{m_w}{|\texttt{parents}(w)|}.$$

Let L is the total number of leaves. Then

$$\alpha_{d,v}(r) = \alpha \frac{\ell_v}{L} \frac{m_v + r + \sum_{j=1}^{d-1} R_j - 1}{m_v}.$$

2.5. Dependency

Let Φ^{-1} denote the quantile function of the standard normal. We here consider the case that $Z_v = \Phi^{-1}(p_v)$ satisfies $Z_{\bar{\mathcal{S}}} \sim \mathcal{N}(0, R)$ for some correlation matrix R. In this case some of the methods presented in the main text may fail due to the correlation between the p-values. However, we can account for this dependency, even if we do not know R. Throughout this section we consider this case, and apply what we call conservative Stouffer smoothing.

$$\tilde{p}_v \leftarrow \begin{cases} 1 & \text{if } \sum_{w \in \mathcal{C}_v} \pi_{vw} Z_w \geq 0 \\ \Phi\left(\sum_{w \in \mathcal{C}_v} \pi_{vw} Z_w\right) & \text{otherwise,} \end{cases}$$

where Φ denotes the standard normal CDF and π satisfies $\pi_{vw} \geq 0, \sum_{w \in \mathcal{C}_v} \pi_{vw} = 1$. As in the independent case, the existing techniques of Meijer & Goeman (2015) and Ramdas et al. (2019) can be now be applied to these smoothed values. The target error rate will still be controlled, even though the null p-values are no longer independent.

Proof of Lemma 2.

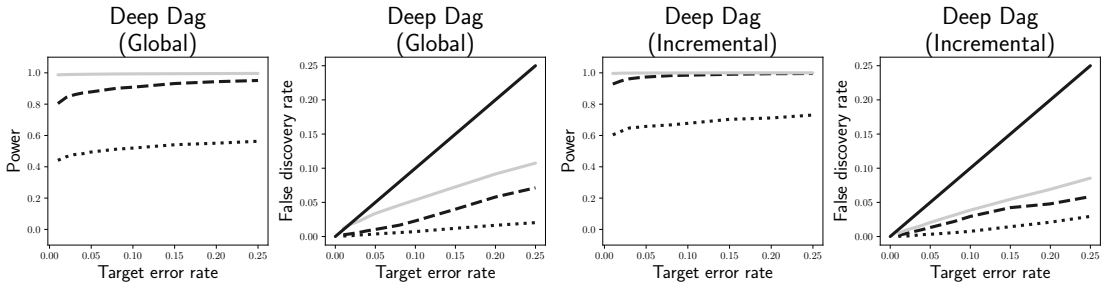

Fig. 1: Power and empirical error rate for three different multiple hypothesis testing methods on two different simulations. Gray lines indicate a Fisher smoothing version of DAGGER, dashed lines indicate a conservative Stouffer smoothing version of DAGGER, dotted lines indicate DAGGER, and the solid black line indicates the maximum allowable error.

Proof. Fix any $v \in \bar{\mathcal{S}}$. Let $Y_v = \sum_{w \in \mathcal{C}_v} \pi_{vw} Z_w$. The assumption of Gaussian copula tells us that $Y_v \sim \mathcal{N}(0, \sigma^2)$. We can get an upper bound for σ^2 using the observation that $R_{w,w'} \in [-1, 1]$ and π lies on the simplex:

$$\operatorname{var}(Y_v) = \sum_w \pi_w \left(\sum_{w'} R_{w,w'} \pi_{w'} \right) \leq \sum_w \pi_w(1) = 1.$$

Thus $\sigma \leq 1$. We can now prove the Lemma, considering three different cases. First, let $\alpha \leq 1/2$. Then

$$\begin{aligned} \operatorname{pr}(\tilde{p}_v \leq \alpha) &= \operatorname{pr}(Y_v \leq \Phi^{-1}(\alpha)) = \Phi(\Psi^{-1}(\alpha)/\sigma) \\ &\leq \Phi(\Phi^{-1}(\alpha)) = \alpha. \end{aligned}$$

This only works because we can assume $\Phi^{-1}(\alpha) \leq 0$; otherwise the inequality goes the other way. Second, let $1/2 < \alpha < 1$. Then

$$\begin{aligned} \operatorname{pr}(\tilde{p}_v \leq \alpha) &= \operatorname{pr}(\tilde{p}_v \leq 1/2) + \operatorname{pr}(\tilde{p}_v \in (1/2, \alpha)) \\ &= \operatorname{pr}(\tilde{p}_v \leq 1/2) + 0 \leq 1/2 \leq \alpha. \end{aligned}$$

Finally, let $\alpha = 1$. Then $\operatorname{pr}(\tilde{p}_v \leq 1) = 1 = \alpha$. □

Proof of Lemma 3.

Proof. As in the previous proof, let $Y_v = \sum_{w \in \mathcal{C}_v} \pi_{vw} Z_w$. Note that $Y_{\bar{\mathcal{S}}}$ is jointly Gaussian. Since R has no negative entries and π is also non-negative, the covariance of $Y_{\bar{\mathcal{S}}}$ is positive. It follows that Y are positive regression dependent (Benjamini & Yekutieli, 2001). Moreover, $\tilde{p}$ is an elementwise monotone nondecreasing transformation of Y; each $\tilde{p}_v$ can be expressed as a monotone nondecreasing function of Y_v; it follows that $\tilde{p}$ are also positive regression dependent.

3. ADDITIONAL EXPERIMENTAL RESULTS

3.1. Dependent null statistics

Here we perform numerical experiments to investigate the performance of various methods in the presence of dependent null statistics with Gaussian copulas. Several extant methods can provably control error in this scenario:

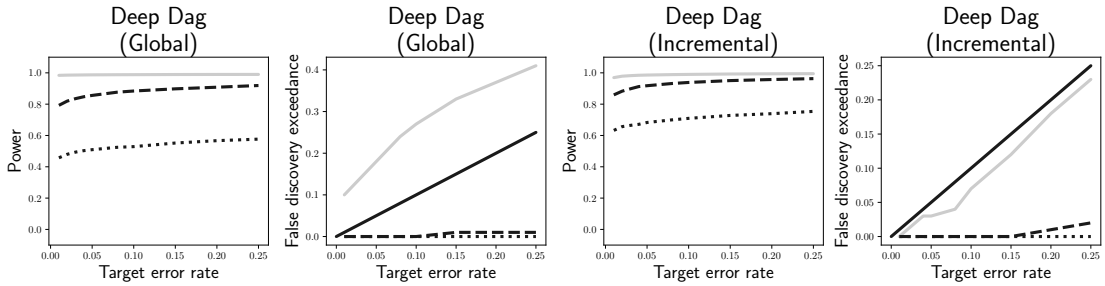

Fig. 2: Power and empirical error rate for three different multiple hypothesis testing methods on two different simulations. Gray lines indicate a Fisher smoothing version of the FDX extension of Meijer and Goeman, dashed lines indicate a conservative Stouffer smoothing version of the FDX extension of Meijer and Goeman, dotted lines indicate the FDX extension of Meijer and Goeman, and the solid black line indicates the maximum allowable error.

- The estimators of Meijer and Goeman are still guaranteed to control familywise error.
- It follows that the FDX extension of Meijer & Goeman (2015) (proposed in Algorithm 3) still guarantees false discovery exceedance.
- The DAGGER estimator is still guaranteed to control false discovery rate – as long as the null statistics have nonnegative correlations.

Algorithm 3. Meijer-Goeman procedure

input : smoothed p-values $\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_n)$, directed acyclic graph $\mathcal{G}$, target level α

$\hat{\mathcal{S}} \leftarrow \emptyset$;

repeat

  $(\pi_v)_{v \notin \hat{\mathcal{S}}} \leftarrow$ Algorithm 1;

  $\hat{\mathcal{S}} \leftarrow \hat{\mathcal{S}} \cup \{v \notin \hat{\mathcal{S}} : \texttt{parents}(v) \subset \hat{\mathcal{S}} \texttt{ and } p_v \leq \alpha \pi_v\}$

until $\hat{\mathcal{S}}$ *does not change*;

return $\hat{\mathcal{S}}$

Smoothing techniques can be used with each of these methods, yielding improved power in some cases. We will consider two smoothing methods.

1. The conservative Stouffer smoothing method:

$$\tilde{p}_v \leftarrow \begin{cases} 1 & \text{if } \sum_{w \in \mathcal{C}_v} \pi_{vw} \Phi^{-1}(p_w) \geq 0 \\ \Phi\left(\sum_{w \in \mathcal{C}_v} \pi_{vw} \Phi^{-1}(p_w)\right) & \text{otherwise.} \end{cases}$$

   where Φ indicates the cumulative distribution function for a standard normal distribution. In the experiments here, we focus on a specific choice for π, namely one which averages each node with its direct children:

$$\pi_{vw} \propto \begin{cases} 1 & \text{if } v = w \\ 1 & \text{if } v \to w \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_w \pi_{vw} = 1$$

2. Fisher smoothing:

$$\tilde{p}_v \leftarrow 1 - F_{\chi^2_{|2\mathcal{C}_v|}}\left(-2\sum_{w\in\mathcal{C}_v}\log p_w\right)$$

where $F_{\chi^2_d}$ indicates the cumulative distribution function for a χ^2 distribution with d degrees of freedom.

Hypothesis testing can be performed by applying existing methods directly to the smoothed values. As described in the main text and proved in 1.1, the conservative Stouffer smoothing versions of several methods provably control type I errors:

- The estimators of Meijer and Goeman applied to conservative Stouffer smoothed statistics are still guaranteed to control familywise error.
- The FDX extension of Meijer of Goeman applied to conservative Stouffer smoothed statistics still guarantees false discovery exceedance.
- The DAGGER estimator applied to conservative Stouffer smoothed statistics is still guaranteed to control false discovery rate – as long as the null statistics have nonnegative correlations.

To quantify the performance of these smoothing-based algorithms, we created simulated datasets in which the ground truth was known. The simulated data was created in three stages:

1. We first designed a five-layer directed acyclic graph structure with 50 nodes at each layer. Each node has three randomly-selected parents from the layer above it.
2. We then selected which nodes would be considered nonnull and sampled a value for each of the nonnull nodes. We performed this selection and sampling using two different methods:
   - Global alternative. The value at each nonnull node is $\mathrm{Beta}(\exp(-4), 0.5)$. directed acyclic graphs are populated starting at the leaves and null nodes are flipped to nonnull with probability 0.2.
   - Incremental alternative. The value at each nonnull node is $\mathrm{Beta}(\exp(-4 - 0.3 \times (D - d)), 0.5)$, where d is the depth of the node and D is the maximum depth of the directed acyclic graph. The graph is populated starting at the leaves with nonnull probability 0.2 and internal nodes are intersection hypotheses that are null if and only if all their child nodes are null.
3. Finally, we sampled values for the null nodes. We constructed a Gaussian process on the graph structure:

$$Z_v | Z_{\mathcal{A}_v} \sim \mathcal{N}\left(\frac{1}{|\mathcal{P}_v \cap \bar{\mathcal{S}}|}\sum_{w\in\mathcal{P}_v\cap\bar{\mathcal{S}}} Z_w, 1,\right)$$

Here $\mathcal{P}_v$ denotes the direct parents of the node v and $\mathcal{A}_v$ denotes all ancestors of v in the graph. We then computed marginally uniform values for each node; letting $F_v(c) = \mathrm{pr}(Z_v \le c)$, we took $p_v = F_v(Z_v)$.

For each type of alternative (global and incremental) we constructed 100 trials (yielding 200 trials in all). We applied six different hypothesis testing methods to each trial and calculated the power and empirical error rates of each method. The six methods are listed below.

- DAGGER (which provably controls the false discovery rate in this setting).

- A Fisher smoothing version of DAGGER (which is not guaranteed to control the false discovery rate).
- A conservative Stouffer version of DAGGER (which provably controls the false discovery rate).
- The FDX extension of Meijer and Goeman (which provably controls the false discovery exceedance), tuned to limit the false discovery proportion below 10%.
- A Fisher smoothing version of the FDX extension of Meijer and Goeman (which is not guaranteed to control false discovery exceedance), tuned to limit the false discovery proportion below 10%.
- A conservative Stouffer version of the FDX extension of Meijer and Goeman (which provably controls the false discovery exceedance).

The results are shown in Figure 1 and Figure 2. In each simulation we find that that the smoothed versions of algorithms have higher power than the original unsmoothed versions. However, as expected, the Fisher smoothing versions occasionally violate their target error rates. For example, the Fisher smoothing version of the FDX extension of Meijer and Goeman has an empirical false discovery exceedance which is roughly double its target error rate when applied to simulations based on the global alternative scheme.

3.2. Intuitions for achieving the greatest power with smoothing techniques

There are some cases in which some smoothed versions of extant algorithms have less power than the corresponding unsmoothed versions. Trouble will arise if the smoothed value for a nonnull hypothesis is heavily influenced by null p-values. However, there are many kinds of smoothing, and in most cases we found that there is *some* form of smoothing which yields higher power. Ideally the user may use a-priori knowledge to choose a smoothing method wisely. For example, if user suspects many of the null hypotheses to be correct, it may be unwise to include smoothing functions which are heavily influenced by the p-values for those hypotheses.

We designed a numerical experiment to demonstrate the power of different smoothing methods in different contexts. We hope this may help guide users in their thinking about what smoothing techniques may be appropriate for different kinds of data. We created four simulated datasets. Each was created in three stages:

1. We first designed a five-layer directed acyclic graph structure with 50 nodes at each layer. Each node has three randomly-selected parents from the layer above it.
2. We then selected which nodes would be considered nonnull: for the first simulation only the nodes in the first layer were considered nonnull, for the the second simulation only the nodes in the first two layers were considered nonnull, and so-on.
3. We then sampled a value for each node. Values at nonnull nodes were sampled according to $\mathrm{Beta}(0.1, 0.5)$ and values at null nodes were sampled uniformly between 0 and 1.

We then ran several different hypothesis selection algorithms on each of the four simulations. We investigated three types of smoothing:

1. No smoothing, i.e. with smoothing function $f_v(p) = p_v$
2. Fisher smoothing using only direct children, i.e. with smoothing function

$$f_v(p) = -2 \log p_v - \sum_{w:\ v \in \mathcal{P}_w} 2 \log p_w,$$

where $\mathcal{P}_w$ indicates the parents of node w.

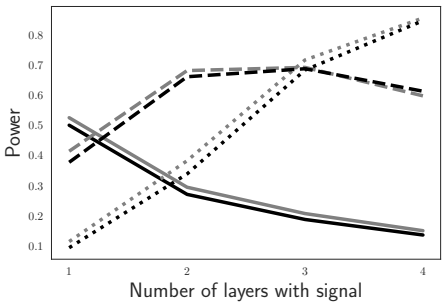

Fig. 3: To get the highest possible power, the hypothesis selection algorithm must be chosen differently for different simulation types. Gray lines indicate algorithms based on Dagger and black lines indicate algorithms based on Meijer and Goeman. Solid lines indicate the original algorithms, dashed lines indicate smoothed versions of those algorithms which use information from each node along with its direct descendants, and dotted lines indicate smoothed versions which use information from all descendants of each node. In simulations where more layers include nonnull hypothesis, algorithms with more smoothing perform better.

3. Fisher smoothing using all children, i.e. with smoothing function

$$f_v(p) = \sum_{c \in \mathcal{C}_v} 2 \log p_c.$$

We investigated two classes of hypothesis selection algorithms: the DAGGER method and the FDX extension of Meijer and Goeman. In total, this yielded six hypothesis selection algorithms:

1. the DAGGER method,
2. the FDX extension of Meijer and Goeman,
3. a Fisher smoothing version of the DAGGER method which only uses the direct children of each node,
4. a Fisher smoothing version of the FDX extension of Meijer and Goeman which only uses the direct children of each node,
5. a Fisher smoothing version of the DAGGER method which uses all descendants of each node, and
6. a Fisher smoothing version of the FDX extension of Meijer and Goeman which uses all descendants of each node.

The DAGGER-based algorithms were tuned to limit the false discovery rate to at most 5%. The Meijer and Goeman algorithms were tuned to limit the false discovery proportion below 10% with probability at least 95%.

The results are shown in Figure 3. To get the highest possible power, the hypothesis selection algorithm must be chosen differently for different simulation types. When the hypotheses in layers 2-5 are all correct, the best choices would be methods 1 and 2 from the list above (i.e. algorithms without any smoothing). When only the hypotheses in layers 3-5 are correct, methods 3 and 4 are best (i.e. algorithms which involve only a local smoothing over the direct children of each node). When only the hypotheses in layers 4-5 are correct, methods 5 and 6 are best (i.e. algorithms which smooth over all descendants of each node).

In summary, smoothing methods can help most when the smoothed values for nonnull hypotheses are most heavily influenced by p-values from other nonnull hypothesis. It follows that the user should select smoothing functions such that this property holds as often as possible. Unfortunately, this may be difficult to do because the user does not know a-priori which hypotheses

are correct and which are not. We caution that some users may feel tempted to choose smoothing techniques based on the data itself; we emphasize such an approach is completely unacceptable unless the user has sufficient data to create a clean split between data used to choose the hypothesis selection algorithm and data used to conduct the final hypothesis tests. A deeper investigation of how users should make these decisions is merited, but beyond the scope of the present work.

REFERENCES

BALABDAOUI, F. & WELLNER, J. A. (2014). Chernoff's density is log-concave. *Bernoulli* **20**, 231.

BENJAMINI, Y. & YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188.

BLOCK, H. W., BUENO, V., SAVITS, T. H. & SHAKED, M. (1987). Probability inequalities via negative dependence for random variables conditioned on order statistics. *Naval Research Logistics* **34**, 547–554.

BLOCK, H. W., SAVITS, T. H. & SHAKED, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics & Probability Letters* **3**, 81–86.

CORMEN, T., LEISERSON, C., RIVEST, R. & STEIN, C. (2009). *Introduction To Algorithms*.

EFRON, B. (1965). Increasing properties of pólya frequency function. *The Annals of Mathematical Statistics* , 272–279.

GENOVESE, C. R. & WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *Journal of the American Statistical Association* **101**, 1408–1417.

GOEMAN, J. J. & SOLARI, A. (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics* , 3782–3810.

KAMAE, T., KRENGEL, U. & O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *The Annals of Probability* , 899–912.

MEIJER, R. J. & GOEMAN, J. J. (2015). A multiple testing method for hypotheses structured in a directed acyclic graph. *Biometrical Journal* **57**, 123–143.

RAMDAS, A., CHEN, J., WAINWRIGHT, M. J. & JORDAN, M. I. (2019). A sequential algorithm for false discovery rate control on directed acyclic graphs. *Biometrika* **106**, 69–86.