TDP Inference in General Linear models

Samuel Davenport, Bertrand Thirion, Pierre Neuvial

2023 OHBM Educational Course

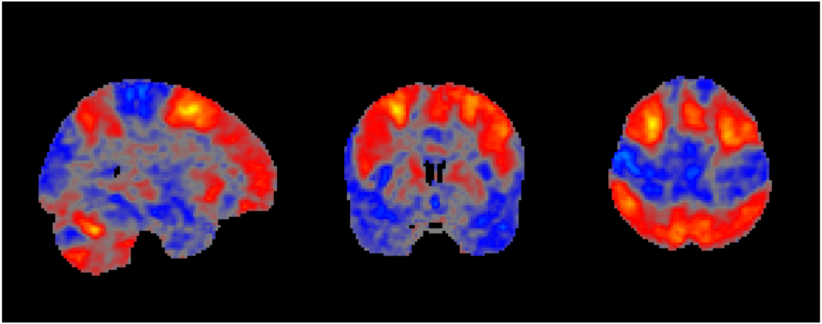
July 7, 2023

Multiple Testing over multiple contrasts

Random Images

Definition

Given $D, L \in \mathbb{N}$ and a set of voxels $\mathcal{V} \subset \mathbb{R}^D$, we define a **random image** on $\mathcal{V}$ to be a random function $f : \mathcal{V} \to \mathbb{R}^L$.



Linear Model

Suppose that we observe random images $y_i : \mathcal{V} \to \mathbb{R}$, for $1 \leq i \leq n$ and some number of subjects n. At each voxel we assume that

$$Y_n(v) = X_n\beta(v) + E_n(v)$$

- $Y_n(v) = [y_1(v), \ldots, y_n(v)]^T$: the response at each $v \in \mathcal{V}$
- $\beta : \mathcal{V} \to \mathbb{R}^p$: vector of parameters
- X_n: design matrix (which is itself random)
- $E_n = [\epsilon_1, \ldots, \epsilon_n]^T$ - the noise where $(\epsilon_m)_{m\in\mathbb{N}}$ are i.i.d. random images.

Testing contrasts

Then given contrasts, $c_1, \dots, c_L \in \mathbb{R}^p$ for some number of contrasts $L \in \mathbb{N}$, we are interested in testing the null hypotheses:

$$H_{0,l}(v) : c_l^T \beta(v) = 0$$

for $1 \leq l \leq L$ and each $v \in \mathcal{V}$.
We can test these using the t-statistic:

$$T_{n,l}(v) = \frac{c_l^T \hat{\beta}_n(v)}{\sqrt{\hat{\sigma}_n(v)^2 c_l^T (X_n^T X_n)^{-1} c_l}}. \tag{1}$$

Obtaining p-values

For $n \in \mathbb{N}$, $1 \leq l \leq L$ and $v \in \mathcal{V}$ we can define two-sided p-values,

$$p_{n,l}(v) = 2(1 - \Phi_{n-r_n}(|T_{n,l}(v)|)) \tag{2}$$

where Φ_{n-r_n} is the CDF of a t-statistic with $n - r_n$ degrees of freedom.

- These are asymptotically valid
- Under an additional assumption of Gaussianity they are valid in the finite sample

Defining the hypothesis space and FWER

- Let $\mathcal{H} = \{(l, v) : 1 \leq l \leq L \text{ and } v \in \mathcal{V}\}$ and $m = |\mathcal{H}|$.
- For $H \subseteq \mathcal{H}$, let $|H|$ denote the number of elements within H.
- let $\mathcal{N} \subset \mathcal{H}$ index the null hypotheses.

Then in order to control for multiple testing we want to control the

$$\text{FWER} = \mathbb{P}(\text{at least one error})$$

To control the FWER over multiple contrasts we can reject at (l, v) if $|T_{n,l}(v)| > u$. So we need to find a threshold u such that

$$\text{FWER} = \mathbb{P}\left(\max_{(l,v) \in \mathcal{N}} |T_{n,l}(v)| > u\right) \leq \alpha.$$

FWER control over contrasts

To do so, for $1 \leq l \leq L$ and $v \in \mathcal{V}$, let

$$S_{n,l}(v) = \frac{c_l^T(\hat{\beta}_n(v) - \beta(v))}{\sqrt{\hat{\sigma}_n(v)^2 c_l^T (X_n^T X_n)^{-1} c_l}}. \tag{3}$$

Then $T_{n,l}(v) = S_{n,l}(v)$ for $(l, v) \in \mathcal{N}$ and so,

$$\begin{aligned}\mathbb{P}\left(\max_{(l,v)\in\mathcal{N}} |T_{n,l}(v)| > u\right) &= \mathbb{P}\left(\max_{(l,v)\in\mathcal{N}} |S_{n,l}(v)| > u\right)\\ &\leq \mathbb{P}\left(\max_{(l,v)\in\mathcal{H}} |S_{n,l}(v)| > u\right).\end{aligned}$$

So we can control the FWER to a level α by ensuring that $\mathbb{P}\big(\max_{(l,v)\in\mathcal{H}} S_{n,l}(v) > u\big) \leq \alpha$.

Resampling in the Linear Model

Resampling in the presence of multiple contrasts

There are several possible ways to resample over multiple contrasts in the linear model.

- Bootstrapping the residuals $Y_n - X_n\hat{\beta}_n$
- Sign-flipping the residuals $Y_n - X_n\hat{\beta}_n$
- Freedman Lane (see (Winkler, Ridgway, Webster, Smith, & Nichols, 2014)), either shuffling or sign-flipping.

Note for Freedman Lane, separate models need to be fit for each contrast of interest. As such it scales as $O(nL)$ instead of $O(n)$.

Bootstrapping

Let

$$\hat{E}_n = Y_n - X_n\hat{\beta}_n = (I_n - X_n(X_n^TX_n)^{-1}X_n^T)E_n.$$

where I_n is the $n \times n$ identity matrix and

$$\hat{\beta}_n = (X_n^TX_n)^{-1}X_n^TY_n = \beta + (X_n^TX_n)^{-1}X_n^TE_n.$$

Given $B \in \mathbb{N}$ for each $1 \leq b \leq B$, we sample from the rows of $\hat{E}_n$ with replacement to get bootstrapped noise E_n^b. Let

$$Y_n^b = X_n\hat{\beta}_n + E_n^b$$

and let

$$\hat{\beta}_n^b = (X_n^TX_n)^{-1}X_n^TY_n^b$$

be the bootstrapped parameter estimates.

Consistency of the bootstrap

For large enough n, the distribution of

$$T_{n,l}^b = \frac{c_l^T(\hat{\beta}_n^b - \hat{\beta}_n)}{\hat{\sigma}_n^b \sqrt{c_l^T(X_n^T X_n)^{-1} c_l}},$$

can be used to approximate the distribution of

$$S_{n,l}(v) = \frac{c_l^T(\hat{\beta}_n(v) - \beta(v))}{\sqrt{\hat{\sigma}_n(v)^2 c_l^T(X_n^T X_n)^{-1} c_l}}. \tag{4}$$

Controlling the FWER

In particular, for each u and bootstrap b,

$$\mathbb{P}\left(\max_{(l,v)\in\mathcal{H}} S_{n,l}(v) > u\right) \approx \mathbb{P}\left(\max_{(l,v)\in\mathcal{H}} T^b_{n,l}(v) > u\right)$$

So we can choose u based on the bootstraps! Ie take u^* to be the upper α quantile of the distribution of

$$\max_{(l,v)\in\mathcal{H}} T^1_{n,l}(v), \dots, \max_{(l,v)\in\mathcal{H}} T^B_{n,l}(v).$$

and reject at (l,v) if $T_{n,l}(v) > u^*$.

FDP Control in the Linear Model

Simultaneous coverage

- Let $\mathcal{H} = \{(l, v) : 1 \leq l \leq L \text{ and } v \in \mathcal{V}\}$ and $m = |\mathcal{H}|$.
- For $H \subseteq \mathcal{H}$, let $|H|$ denote the number of elements within H.
- let $\mathcal{N} \subset \mathcal{H}$ index the null hypotheses.

Given $0 < \alpha < 1$ we want,

$$V : \{H : H \subset \mathcal{H}\} \to \mathbb{N}$$

such that

$$\mathbb{P}(|S \cap \mathcal{N}| \leq V(S), \ \forall S \subset \mathcal{H}) \geq 1 - \alpha. \tag{5}$$

If (5) holds then, with probability $1 - \alpha$, simultaneously over all $S \subset \mathcal{H}$, $V(S)$ provides a upper bound on the number of false positives within S. Importantly $V(S)$ is valid for all S including data-selected subsets.

Joint Error Rate (JER)

Define the **joint error rate (JER)** of the collection $(R_k)_{1\leq k\leq K} \subset \mathcal{H}$

$$\text{JER}((R_k(\lambda))_{1\leq k\leq K}) := \mathbb{P}(|R_k \cap \mathcal{N}| > k-1, \text{ some } 1 \leq k \leq K) \tag{6}$$

(Blanchard, Neuvial, Roquain, et al., 2020) showed that if

$$\text{JER}((R_k)_{1\leq k\leq K}) \leq \alpha$$

then the bound $\overline{V}_\alpha : \{H : H \subset \mathcal{H}\} \to \mathbb{R}$, sending $S \subset \mathcal{H}$ to

$$\overline{V}_\alpha(S) = \min_{1\leq k\leq K} (|S \setminus R_k| + k - 1) \wedge |S|, \tag{7}$$

satisfies (5) and thus provides an α-level bound over the number of false positives within each chosen rejection set.

Template Families

Let $K \in \mathbb{N}$ and suppose we have a set of, strictly increasing and continuous template functions

$$t_k : [0,1] \to \mathbb{R} \tag{8}$$

for each $1 \leq k \leq K$. Given $n \in \mathbb{N}$, define

$$R_k(\lambda) = \{(l,v) \in \mathcal{H} : p_{n,l}(v) \leq t_k(\lambda)\},$$

for each $\lambda \in [0,1]$. We will refer to the collection $(R_k(\lambda))_{1 \leq k \leq K}$ as the canonical reference family. The simplest example is the linear template family i.e. $t_k(\lambda) = \frac{\lambda k}{m}$.

Controlling the JER

Let $p^n_{(k:\mathcal{N})}$ be the kth smallest p-value in the set $\{p_{n,l}(v) : (l, v) \in \mathcal{N}\}$ (and set $p^n_{(k:\mathcal{N})} = 1$ if $k > |\mathcal{N}|$).

Claim

For each $\lambda \in [0, 1]$,

$$JER((R_k(\lambda))_{1\leq k\leq K}) = \mathbb{P}\left(\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \lambda\right).$$

Bootstrapped quantile

Let $f_n : \left\{g : \mathcal{V} \to \mathbb{R}^L\right\} \to \mathbb{R}$ send

$$T \mapsto \min_{1 \leq k \leq K \wedge |\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{H})}(T))$$

For each $n, B \in \mathbb{N}$ and $0 < \alpha < 1$, let $\lambda^*_{\alpha,n,B}(\mathcal{H})$ be α-quantile of the bootstrap distribution of $f_n(T_n)$.

Valid simultaneous inference

In particular, using resampling gives us asymptotic control of the JER, i.e.

$$\text{Then, } \lim_{n\to\infty}\lim_{B\to\infty} \mathrm{JER}\Big((R_k(\lambda^*_{\alpha,n,B}(\mathcal{H})))_{1\le k\le K}\Big)$$
$$= \lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\left(\min_{1\le k\le K\wedge|\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{N})}) \le \lambda^*_{\alpha,n,B}(\mathcal{H})\right) \le \alpha$$

Moreover, letting $\overline{V}_{\alpha,n,B}(H)$ be the corresponding post-hoc bound,

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}\Big(|H\cap\mathcal{N}| \le \overline{V}_{\alpha,n,B}(H),\ \forall H\subset\mathcal{H}\Big) \ge 1-\alpha.$$

So $\overline{V}$ can be used to provide simultaneous inference. As with regular inference this procedure can be iterated to yield a step down.

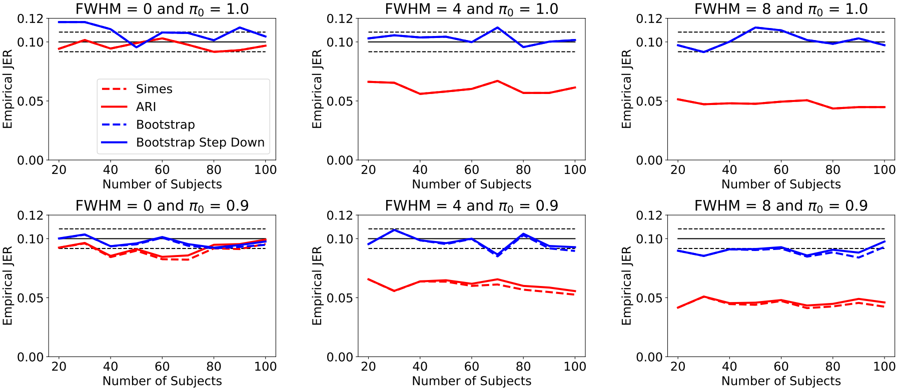
Results

Simulation description

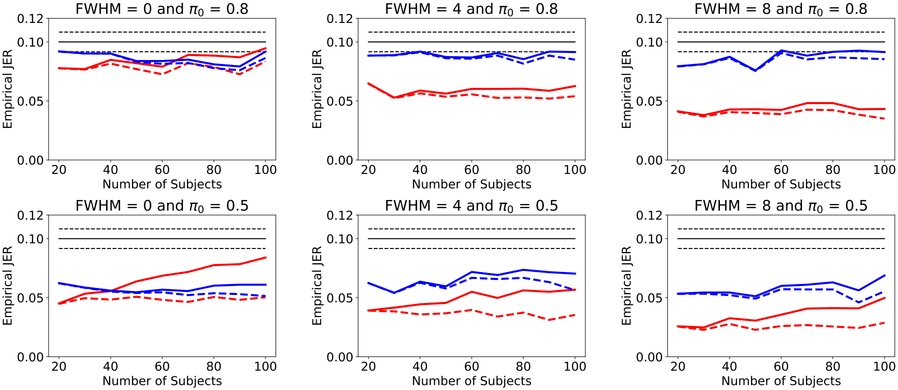
We ran 2D simulations to test the performance of the methods.

- 50×50 GRFs smoothed with FWHM $= 0, 4, 8$
- $N = \{20, 30, \ldots, 100\}$ subjects
- randomly divided the subjects into 3 groups
- tested the difference between the first and the second and between the second and the third group at each pixel
- Randomly assigned a proportion $\pi_0 \in \{0.5, 0.8, 0.9, 1\}$ of the hypotheses to have non-zero mean 1.
- Compared the parametric and bootstrap methods.
- Uses 1000 bootstraps

Empirical JER



Empirical JER - continued



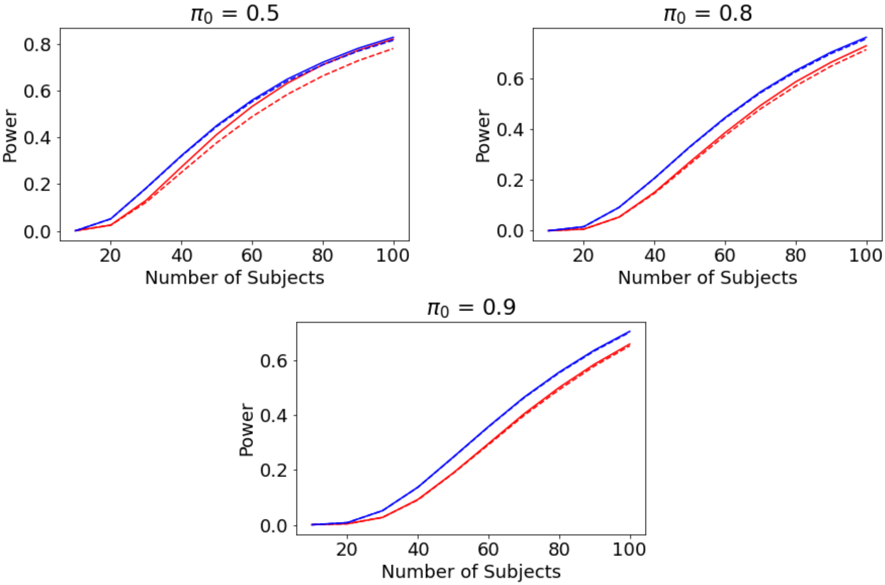
Power - definition

Define the power to be

$$\mathrm{Pow}(R) := \mathbb{E}\left[\frac{|\mathcal{H}| - \overline{V}(\mathcal{N})}{|\mathcal{N}^C|}\middle| |\mathcal{N}^C| > 0\right]$$

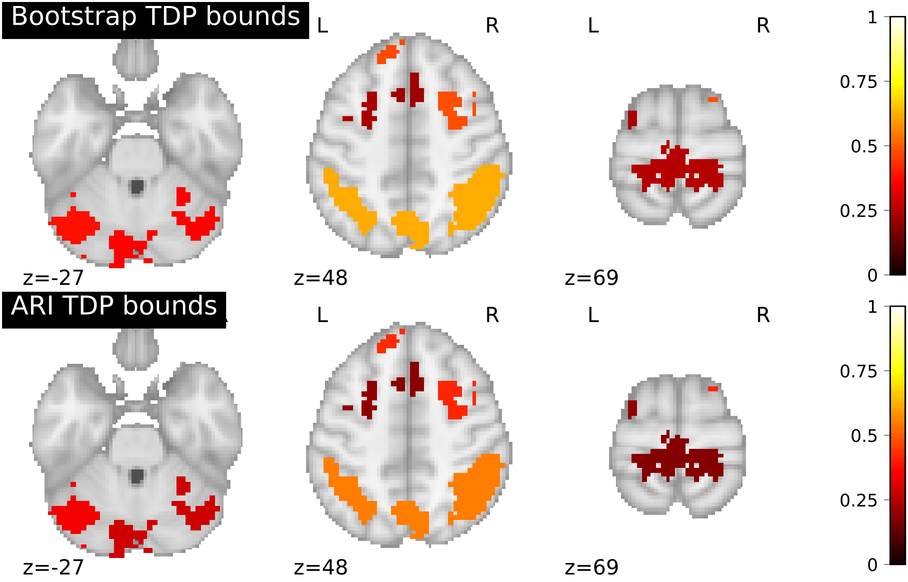
- This is a measure of the bounds on the true discovery proportion and so serves as a measure of power.
- Same notion of power as that of (Blanchard et al., 2020).
- Consider the same simulation setting where the FWHM $= 4$

Power - Results (In the FWHM = 4 setting)

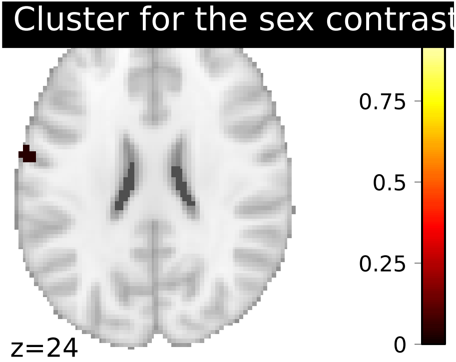


fMRI data model

- fMRI data from 365 unrelated subjects from the HCP
- Subjects take the PMAT the results of which are measured numerically.
- We consider the working memory task
- At each voxel we fit a linear model of the fMRI data against: Age, Sex, Height, Weight, BMI, Blood pressure and the intelligence measure
- Test contrasts for Sex and intelligence
- Used 1000 bootstraps


Bootstrap TDP bounds
L
R
L
R
z=-27
z=48
z=69
1
0.75
0.5
0.25
0
ARI TDP bounds
L
R
L
R
z=-27
z=48
z=69
1
0.75
0.5
0.25
0

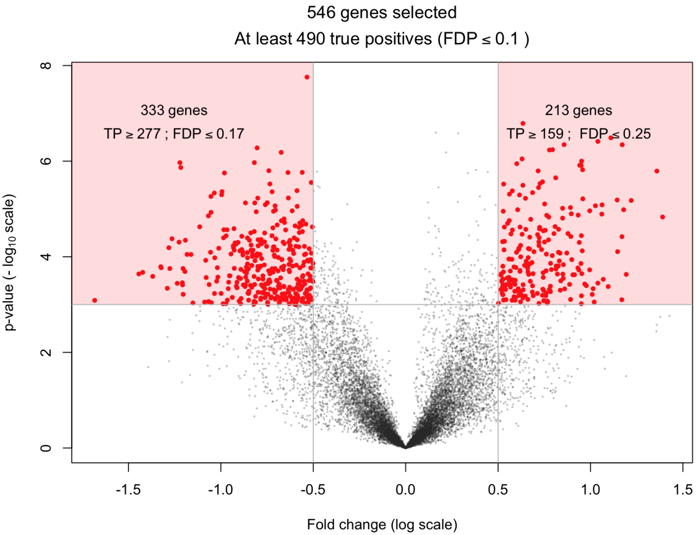
TDP for the HCP - Sex contrast



Transcriptomic data analysis

- Have genetics data from 135 subjects from Bahr et al (2013).
- Subjects had chronic obstructive pulmonary disease (COPD)
- Have a measure of gene expression at 12531 genes.
- Consider a linear model regressing gene expression against age, sex, lung function, BMI, parental history of COPD, and and two smoking variables (smoking status and pack-years).
- We considered the contrast for lung function

Volcano plot



Conclusions

- Using resampling approaches allows for large power gains when doing inference under dependence.
- Non-parametric approaches are typically more powerful than parametric ones.
- ARI assumes positive dependence which may not be valid when there are multiple contrasts
- The method is flexible and extends to other resampling approaches
- Code for implementation is available at github.com/sjdavenport/pyperm, see practical
- Pre-print available on arxiv (and from my website): (Davenport, Thirion, & Neuvial, 2022).

WARNING: Manly based permutation is not valid

We need to be a bit careful when resampling in the linear model and accounting for multiple contrasts because not all methods work.

- Manly permutation permutes Y_n by pre-multiplying by a permutation matrix P and regressing $X_n\beta$ on PY_n.
- This is valid for testing the null hypothesis that $\beta(v) = 0$ but is not valid for testing that e.g. $c^T\beta(v) = 0$ for some contrast c as

$$PY_n = PX_n\beta + PE_n \not\sim PE_n.$$

- Instead we need to target $\max_{(l,v)\in\mathcal{H}} S_{n,l}(v)$.

Step down

Algorithm 1 Step down algorithm

$j \leftarrow 0$
$H_n^{(0)} \leftarrow \mathcal{H}$
repeat
    $j \leftarrow j + 1$
    $\lambda_{n,j} = \lambda^*_{\alpha,n,B}(H_n^{(j-1)})$
    $H_n^{(j)} \leftarrow \{(l,v) : p_{n,l}(v) \geq t_1(\lambda_{n,j})\}$
until $H_n^{(j)} = H_n^{(j-1)}$
$\hat{H}_n \leftarrow H_n^{(j)}$
return $\hat{H}_n$

Using $(R_k(\lambda^*_{\alpha,n,B}(\hat{H}_n)))_{1\leq k\leq K}$ as our reference sets we can derive a valid step down post-hoc bound.

Simes Bound

Under positive dependence, for $0 < \alpha < 1$, the Simes inequality implies that

$$\mathbb{P}\left(\exists k \in \{1, \dots, m\} : p^n_{(k:\mathcal{N})} < \frac{\alpha k}{m}\right) \leq \frac{\alpha|\mathcal{N}|}{m}.$$

Thus defining the linear template family as $t_k(x) = \frac{xk}{m}$, it follows that

$$\text{JER} = \mathbb{P}\left(\min_{1 \leq k \leq K \wedge |\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \alpha\right) \leq \alpha.$$

Thus $\overline{V}_\alpha$ (constructed using the sets $R_k(\alpha)$) is a valid post-hoc bound.

- This works best under independence as then the inequality becomes exact.
- Positive dependence may not hold between contrasts, e.g. when testing the differences of 3 groups.

ARI

(Rosenblatt, Finos, Weeda, Solari, & Goeman, 2018) introduced a version of this that estimates $|\mathcal{N}|$ using the hommel value h. It can be shown that under PRDS,

$$\text{JER} = \mathbb{P}\left(\min_{1 \leq k \leq K \wedge |\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \frac{\alpha m}{h} \right) \leq \alpha.$$

- The $\overline{V}_{\frac{\alpha m}{h}}$ (constructed using the sets $R_k(\frac{\alpha m}{h})$) is thus a valid post-hoc bound.
- Known as All Resolutions Inference or (ARI)
- It's the step down version of the Simes bound

Bibliography

Blanchard, G., Neuvial, P., Roquain, E., et al. (2020). Post hoc confidence bounds on false positives using reference families. *Annals of Statistics*, *48*(3), 1281–1303.

Davenport, S., Thirion, B., & Neuvial, P. (2022). FDP control in multivariate linear models using the bootstrap. *arXiv preprint arXiv:2208.13724*.

Rosenblatt, J. D., Finos, L., Weeda, W. D., Solari, A., & Goeman, J. J. (2018). All-resolutions inference for brain imaging. *Neuroimage*, *181*, 786–796.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. doi: 10.1016/j.neuroimage.2014.01.060