FDP control in multivariate linear models using the bootstrap

Samuel Davenport, Bertrand Thirion, Pierre Neuvial

University of California, San Diego

March 2, 2022

Notation and general framework

Random Fields on a lattice

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and take $\mathbb{N}$ to be the set of positive integers.

Definition

Given $D, m \in \mathbb{N}$ and a finite set $\mathcal{V} \subset \mathbb{R}^D$, we define a **random field** on $\mathcal{V}$ to be a measurable mapping $f : \Omega \to \{g : \mathcal{V} \to \mathbb{R}^m\}$. We will say that f has **dimension** m.

For $\omega \in \Omega$ and $v \in \mathcal{V}$, we will write $f(\omega, v) = f(\omega)(v) = f(v)$ and $f : \mathcal{V} \to \mathbb{R}^m$.

Definition

For $u, v \in \mathcal{V}$ define

$$\mu_f(v) = \mathbb{E}[f(v)]$$

and

$$\mathfrak{c}_f(u, v) = \text{cov}(f(u), f(v))$$

Gaussian Random Fields on a Lattice

Definition

For $1 \leq j \leq m$ define the random fields $f_j : \Omega \to \{g : \mathcal{V} \to \mathbb{R}\}$ which send $\omega \in \Omega$ to $f_j(\omega)(\cdot) = f(\omega)(\cdot)_j = f(\cdot)_j$. We will call $f_1, \dots, f_m$ the **components** of f.

Let $vec(f)$ be the vector $(f_l(v) : l \in \{1, \dots, m\}$ and $v \in \mathcal{V})$.

Definition

Given functions $\mu : \mathcal{V} \to \mathbb{R}^m$ and $\mathfrak{c} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ we write $f \sim \mathcal{G}(\mu, \mathfrak{c})$ if f is a random field on $\mathcal{V}$ with mean μ and covariance $\mathfrak{c}$ and such that $\text{vec}(f)$ has a multivariate Gaussian distribution.

Linear Model

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

$$\mathbf{Y}_n(v) = X_n \beta(v) + \boldsymbol{\epsilon}_n(v)$$

- $\mathbf{Y}_n(v) = [Y_1(v), \ldots, Y_n(v)]^T$: 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)
- $\boldsymbol{\epsilon}_n = [\epsilon_1, \ldots, \epsilon_n]^T$ - the noise - is an n-dimensional random field. We will assume that $(\epsilon_n)_{n \in \mathbb{N}}$ is an i.i.d sequence.

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:

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

Convergence in the linear model - assumptions

Assumption

- (a) For $n \in \mathbb{N}$, $X_n = [x_1, \dots, x_n]^T$ for a sequence of i.i.d vectors $(x_n)_{n\in\mathbb{N}}$ whose multivariate density is bounded above and that $\mathrm{var}(x_1) < \infty$.
- (b) Assume that $\mathrm{var}(\epsilon_1(v)) < \infty$ for all $v \in \mathcal{V}$ and that $(x_n)_{n\in\mathbb{N}}$ and $(\epsilon_n)_{n\in\mathbb{N}}$ are independent.

Lemma

Suppose that $(X_n)_{n\in\mathbb{N}}$ satisfies Assumption (a) and let $\Sigma_X = \mathbb{E}\left[x_1 x_1^T\right]$, then Σ_X is invertible and

$$\left(\frac{X_n^T X_n}{n}\right)^{-1} \xrightarrow{a.s.} \Sigma_X^{-1}.$$

CLT for $\hat{\beta}$

Lemma

Suppose that $(X_n)_{n\in\mathbb{N}}$ *and* $(\epsilon_n)_{n\in\mathbb{N}}$ *satisfy Assumption 1. Then*

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}_\epsilon \Sigma_X^{-1}).$$

Proof.

$$\sqrt{n}(\hat{\beta}_n - \beta) = \sqrt{n}(X_n^T X_n)^{-1} X_n^T \epsilon_n = \left(\frac{X_n^T X_n}{n}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \epsilon_i.$$

□

Convergence of the t-statistics

Let $\mathfrak{c}'(u,v) = \rho_\epsilon(u,v) AC\Sigma_X^{-1}C^T A^T$, where $A \in \mathbb{R}^{L\times L}$ is a diagonal matrix with $A_{ll} = (c_l^T \Sigma_X^{-1} c_l)^{-1/2}$ for $1 \leq l \leq L$.

Theorem

For $n \in \mathbb{N}$, let S_n be the L-dimensional random field on $\mathcal{V}$ defined by

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

for $1 \leq l \leq L$. Then, under the Assumption, as $n \to \infty$,

$$S_n \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')$$

and it follows that

$$T_n|_{\mathcal{N}} \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')|_{\mathcal{N}}.$$

Obtaining p-values

For $n \in \mathbb{N}$, $1 \leq l \leq L$ and $v \in \mathcal{V}$ we can define 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 by the previous theorem
- Under an additional assumption of Gaussianity they are valid in the finite sample

Simultaneous coverage

- Let $\mathcal{H} = \{(l, v) : 1 \leq l \leq L \text{ and } v \in \mathcal{V}\}$.
- 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,

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

such that

$$\mathbb{P}(|S \cap \mathcal{N}| \leq V(S),\ \forall S \subset \mathcal{H}) \geq 1 - \alpha. \tag{3}$$

If (3) 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.

Joint Error Rate (JER)

Suppose that for some $K \in \mathbb{N}$ we have sets $R_1, \ldots, R_K \subset \mathcal{H}$ and constants $\zeta_1, \ldots, \zeta_K \in \mathbb{N}$ and define

$$\text{JER}((R_k, \zeta_k)_{1 \leq k \leq K}) := \mathbb{P}(|R_k \cap \mathcal{N}| > \zeta_k, \text{ some } 1 \leq k \leq K) \tag{4}$$

to be the **joint error rate** of the collection $(R_k, \zeta_k)_{1 \leq k \leq K}$. (Blanchard, Neuvial, Roquain, et al., 2020) showed that if

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

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

$$\overline{V}_\alpha(S) = \min_{1 \leq k \leq K} (|S \setminus R_k| + \zeta_k) \wedge |S|, \tag{5}$$

satisfies (3) and thus provides an α-level bound over the number of false positives within each chosen rejection set.

Template Families

Given a set of template functions $t_k : [0,1] \to \mathbb{R}$ and $\lambda \in [0,1]$, for each $1 \leq k \leq K$ and $n \in \mathbb{N}$, we will take

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

set $\zeta_k = k - 1$, and 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 $k > |\mathcal{N}|$). We will refer to the collection $(R_k(\lambda), k-1)_{1 \leq k \leq K}$ as the canonical reference family.

Claim

For each $\lambda \in [0,1]$, *let* $\zeta_k = k - 1$, *then*

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

Bootstrapping in the Linear Model

Bootstrapping

Let

$$\hat{\boldsymbol{\epsilon}}_n = \mathbf{Y}_n - X_n\hat{\beta}_n = (I_n - X_n(X_n^TX_n)^{-1}X_n^T)\boldsymbol{\epsilon}_n.$$

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

$$\hat{\beta}_n = (X_n^TX_n)^{-1}X_n^T\mathbf{Y}_n = \beta + (X_n^TX_n)^{-1}X_n^T\boldsymbol{\epsilon}_n.$$

Given $B \in \mathbb{N}$ for each $1 \leq b \leq B$, conditional on the data, $\hat{\epsilon}_1^b, \dots, \hat{\epsilon}_n^b$ are chosen independently with replacement from $\{\hat{\boldsymbol{\epsilon}}_{n,1}, \dots, \hat{\boldsymbol{\epsilon}}_{n,n}\}$ resulting in a combined random field $\boldsymbol{\epsilon}_n^b = [\hat{\epsilon}_1^b, \dots, \hat{\epsilon}_n^b]^T$. Let

$$\mathbf{Y}_n^b = X_n\hat{\beta}_n + \boldsymbol{\epsilon}_n^b$$

and let

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

be the bootstrapped parameter estimates.

CLT for $\hat{\beta}_n^b$

Theorem

(Bootstrap convergence.) Suppose that $(X_n)_{n\in\mathbb{N}}$ and $(\epsilon_n)_{n\in\mathbb{N}}$ satisfy Assumption 1. Then conditional on $(Y_n)_{n\in\mathbb{N}}$, for almost every sequence $(Y_n)_{n\in\mathbb{N}}$, for each $1 \leq b \leq B$,

$$\sqrt{n}(\hat{\beta}_n^b - \hat{\beta}_n) \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}_\epsilon \Sigma_X^{-1}).$$

- (Freedman, 1981) proved a version of this in 1D based on convergence in the Mallows metric using ideas from (Bickel & Freedman, 1981).
- (Eck, 2018) extended this proof to the multivariate case.
- We have a (substantially simpler) proof based on the Lindeberg CLT which has not to our knowledge been written down before.

Convergence of the bootstrapped t-statistics

Theorem

(Bootstrap test-statistic convergence.) Suppose that $(X_n)_{n\in\mathbb{N}}$ *and* $(\epsilon_n)_{n\in\mathbb{N}}$ *satisfy Assumption 1 and, for each* $1 \leq b \leq B$*, let* $T_n^b : \mathcal{V} \to \mathbb{R}$ *be the* L*-dimensional random field on* $\mathcal{V}$ *such that, for* $1 \leq l \leq L$*,*

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

Then conditional on $(Y_n)_{n\in\mathbb{N}}$*, for almost every sequence* $(Y_n)_{n\in\mathbb{N}}$*, for each* $1 \leq b \leq B$*,*

$$T_n^b \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')$$

as $n \to \infty$*. In particular it follows that*

$$T_n^b|_{\mathcal{N}} \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')|_{\mathcal{N}}.$$

JER Control in the Linear Model

Convergence of the bootstrapped quantile

Theorem

Let $(f_n)_{n\in\mathbb{N}}, f$ *be functions from* $\left\{g : \mathcal{V} \rightarrow \mathbb{R}^L\right\}$ *to* $\mathbb{R}$ *such that for each* $b \in \mathbb{N}$, *and some random field* $\mathcal{G}$,

$$f_n(T_n^b)|Y \xrightarrow{d} f(\mathcal{G}).$$

and for each $n, B \in \mathbb{N}$ *and* $0 < \alpha < 1$, *let*

$$\lambda^*_{\alpha,n,B} = \inf\left\{\lambda : \frac{1}{B}\sum_{b=1}^{B} 1\left[f_n(T_n^b) \leq \lambda\right] \geq \alpha\right\}.$$

Take F *to be the CDF of* $f(\mathcal{G})$, *i.e. for* $\lambda \in [0,1]$, $F(\lambda) = \mathbb{P}(f(\mathcal{G}) \leq \lambda)$ *and let* $\lambda_\alpha = F^{-1}(\alpha)$. *Then*

$$\lim_{n\rightarrow\infty}\lim_{B\rightarrow\infty}\lambda^*_{\alpha,n,B} = \lambda_\alpha.$$

Main Result

Theorem

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

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

and for $n \in \mathbb{N}$ *let* $\lambda^*_{\alpha,n,B}$ *be the* α*-quantile of the bootstrap distribution (based on* $B \in \mathbb{N}$ *bootstraps) of* $f_n(T_n)$ *conditional on the observed data. Then,*

$$\lim_{n \to \infty} \lim_{B \to \infty} \mathbb{P}\Big(f_n(T_n) \leq \lambda^*_{\alpha,n,B}\Big) = \alpha$$

I.e. the joint error rate is asymptotically bounded.

Idea: conditional on Y, $f_n(T_n) \xrightarrow{d} f(\mathcal{G}(0, \mathfrak{c}'))$ so we can apply the previous theorem to show that the quantiles converge.

Main Result - Generalisation to subsets of $\mathcal{N}$

For any $H \subset \mathcal{H}$, let $p^n_{(k:H)}$ be the kth smallest p-value in the set $\{p_{n,l}(v) : (l, v) \in \mathcal{N}\}$

Theorem

For $H \subset \mathcal{H}$, let $f_{n,H} : \left\{g : \mathcal{V} \rightarrow \mathbb{R}^L\right\} \rightarrow \mathbb{R}$ send

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

and for $n \in \mathbb{N}$ let $\lambda^*_{\alpha,n,B}(H)$ be the α-quantile of the bootstrap distribution (based on $B \in \mathbb{N}$ bootstraps) of $f_{n,H}(T_n)$ conditional on the observed data. Then if $\mathcal{N} \subset H$,

$$\lim_{n\rightarrow\infty} \lim_{B\rightarrow\infty} \mathbb{P}\big(f_{n,\mathcal{N}}(T_n) \leq \lambda^*_{\alpha,n,B}(H)\big) \leq \alpha$$

where the limit holds with equality if $H = \mathcal{N}$. I.e. the joint error rate is asymptotically bounded.

Step down

Algorithm 1 Step down algorithm

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

Step down theorem

Using the ideas similar to those in (Blanchard et al., 2020).

Theorem

Let $\hat{H}_n$ be the set generated by applying Algorithm 1. Then

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}\Big(f_{n,\mathcal{N}}(T_n) < \lambda^*_{\alpha,n,B}(\hat{H}_n)\Big) \leq \alpha$$

Simes Bound

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

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

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

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

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

- This works best under independence as then the inequality becomes exact.
- PRDS may not hold (especially in the contrast cases);

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,

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

- 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

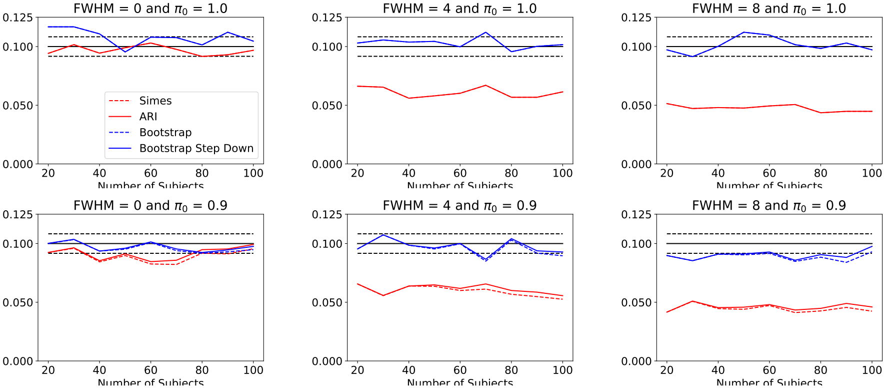
Results

Simulation description

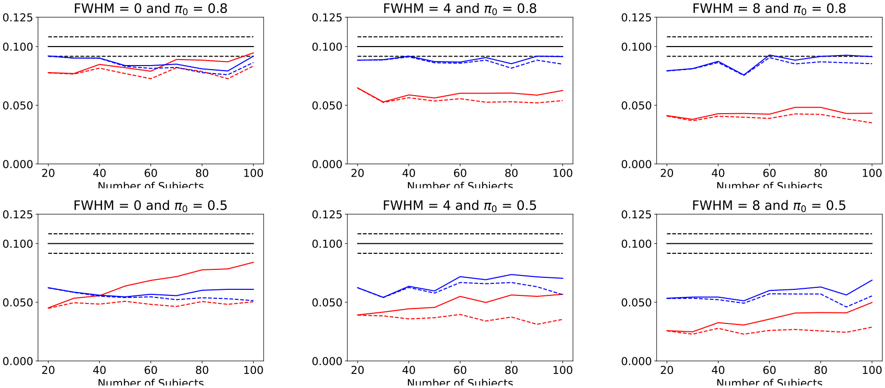
We ran 2D simulations to test the performance of the methods.

- 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 contrasts to have non-zero mean 1.
- Compared the parametric and bootstrap methods.
- Bootstrap uses 100 bootstraps

Empirical JER



Empirical JER - continued



Power - definition

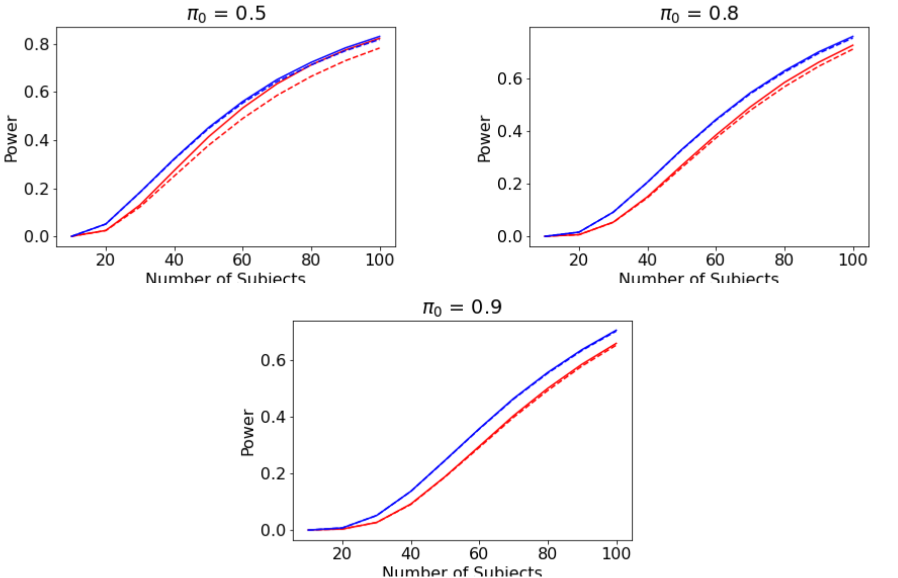
Given a set $R \subset \mathcal{H}$, define

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

we take $R = \mathcal{H}$ (in this talk)

- Same notion of power as that of (Blanchard et al., 2020).
- Consider the same simulation setting where the FWHM = 4

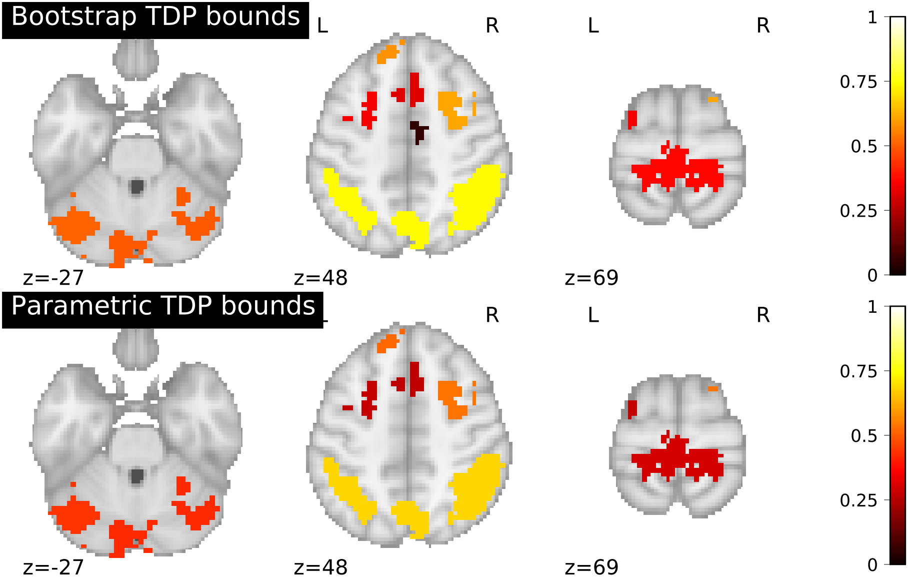
Power - Results



fMRI data model

- fMRI data from 365 unrelated subjects from the HCP
- Subjects take a test the results of which are measured numerically.
- They also perform a 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

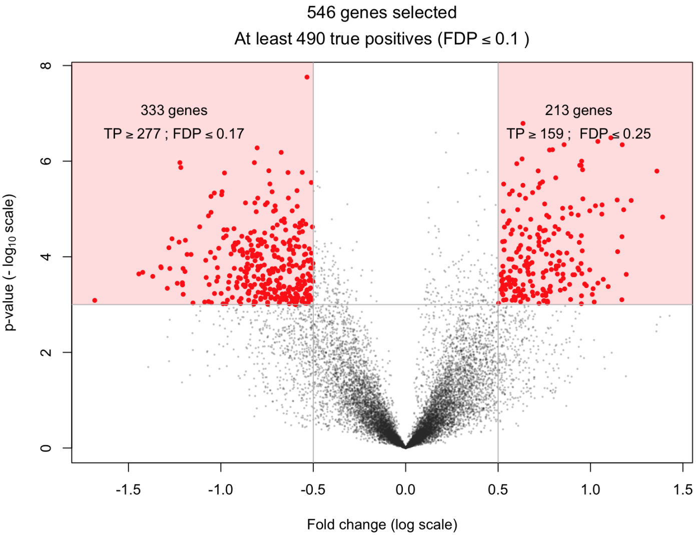
fMRI data analysis



Transcriptomic data analysis

- Have genetics data from 135 subjects
- 12531 genes
- run a regression against some controlled covariates and lung function and considered a single contrast for lung function.

Volcano plot



Conclusions

- Using resampling approaches allows for large power gains when doing inference under dependence.
- Recommend using it over ARI in most cases
- The method is flexible and extends to other settings. I.e. other bootstrap settings.
- Code for implementation is available at github.com/sjdavenport/pyrft
- Hopefully will have a pre-print out soon

Bibliography

Bickel, P. J., & Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *Annals of Statistics*, *9*(6), 1196–1217. doi: 10.1214/aos/1176342871

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.

Eck, D. J. (2018). Bootstrapping for multivariate linear regression models. *Statistics & Probability Letters*, *134*, 141–149.

Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, *9*(6), 1218–1228.

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