True discovery guarantee for brain imaging

4 March 2022 – Psicostat

Anna Vesely

Joint work with Livio Finos and Jelle J. Goeman

University of Padova
anna.vesely@unipd.it

Table of contents

A very important premise

Multiple hypothesis testing

When testing a single null hypothesis H_0:

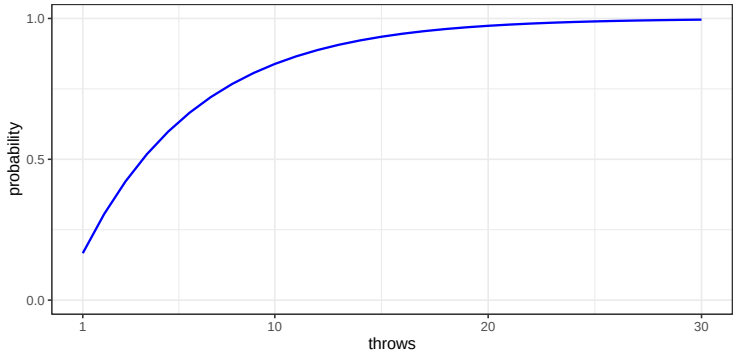
- type I error = we reject H_0 when it is true
- significance level $\alpha = P(\text{type I error})$

When testing multiple hypotheses together, the probability of making at least one type I error can be much greater than α.

Intuition

When throwing dice, the probability of getting a 6 is

- $1/6 \approx 0.167$ for 1 single throw
- $1 - (5/6)^m$ for m throws



Functional Magnetic Resonance Imaging

High-dimensional data

In many fields, the number of features under study is way greater than the sample size:

- neuroimaging - brain activation in voxels
- genomics - expression of genes
- healthcare data - medical records of patients
- ...

Typically, interest is not in single feautures but in detecting, localizing and quantifying signal in sets of features.

Researchers often want to study many sets and select the set of interest post hoc.

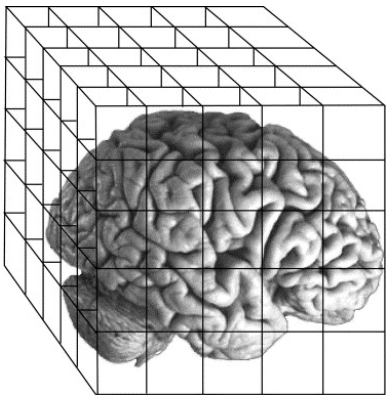
Functional Magnetic Resonance Imaging

fMRI measures brain activation
as changes in blood flow (BOLD)
under a sequence of stimuli.

Activation is measured in voxels,
about 200,000 volume units.

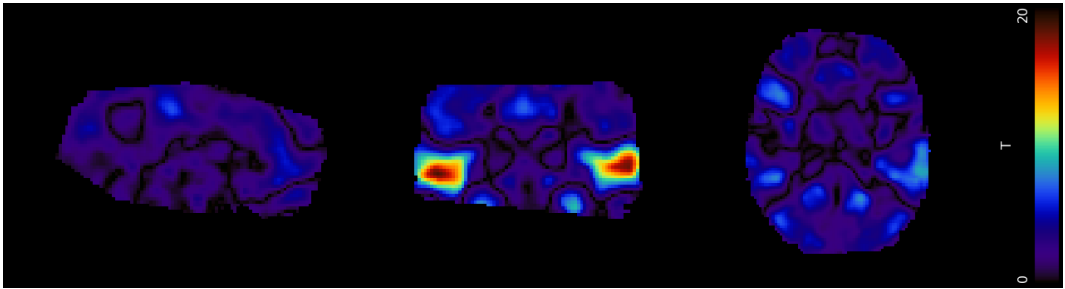
For each voxel j:

- null hypothesis H_j : *no activation in j*
- test statistic T_j
  (from first-level analysis)



Auditory data

140 subjects passively listening to vocal and non-vocal sounds[1]



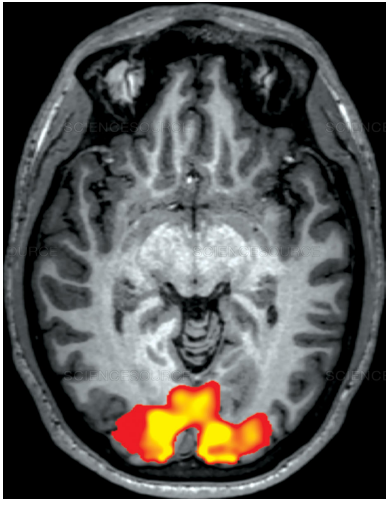
[1]Pernet et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, 2015.

Inference on clusters

Interest generally lies in clusters,
brain regions of (contiguous) voxels.

Supra-threshold clusters: regions of
connected voxels j with $T_j >$ threshold.

$H_S = \bigcap_{j \in S} H_j$: *no activation in cluster* S



Standard cluster inference

Cluster extent thresholding

Standard method for cluster inference that

- adapts to the correlation structure
- strongly controls the FWER at cluster level

H_S is rejected $\Longrightarrow$ S contains at least one active voxel

No information on

- the proportion of active voxels (TDP)
- their spatial location

Cluster extent thresholding

The following statements are not supported

'A large significant cluster...'

- '... contains a substantial number of active voxels.'
  It contains at least one.

- '... is a stronger finding than a small significant cluster.'
  It is a weaker finding (spatial specificity paradox).
  Follow-up inference inside clusters leads to inflated type I error rates.

- ... indicates activity in an anatomical area, if there is substantial overlap.'
  If the cluster is not completely contained, activity may lie outside.

True discovery guarantee

True discovery guarantee

TDP(S) = proportion of truly active voxels in S

Lower $(1-\alpha)$-confidence bounds:

$$P\left(\text{TDP}(S) \geq \text{bound}(S)\right) \geq 1-\alpha$$

Closed testing[1] gives simultaneous bounds:

$$P\left(\text{TDP}(S) \geq \text{bound}(S) \text{ for each cluster } S\right) \geq 1-\alpha$$

[1] Goeman and Solari. Multiple testing for exploratory research. *Stat. Sci.*, 2011.

The computational complexity of closed testing is exponential in the number of voxels.

We[1][2] give a shortcut, valid in many cases, that

- makes inference on the TDP of clusters
- allows for post-hoc selection and follow-up inference
- adapts to the unknown correlation structure of voxels

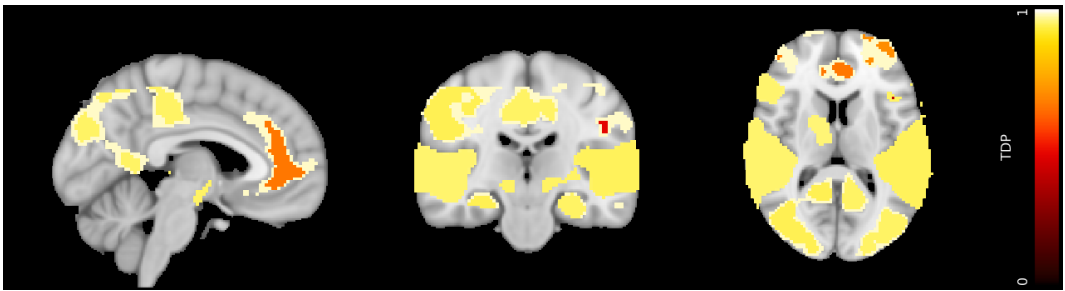
[1] Vesely et al. Permutation-based true discovery guarantee by sum tests *arXiv:2102.11759*, 2021.
[2] R package `sumSome`, `https://CRAN.R-project.org/package=sumSome`

Results

Auditory data

We obtain simultaneous confidence bounds for the TDP of clusters.

Taking clusters with $T_j > 3.2$ and sub-clusters with $T_j > 4$:



Auditory data

Cluster	Threshold	Size	TDP	p-value	Coordinates		
S	thr	s	lower conf. bound	p_{FWER}	x	y	z
FP/CG/SFG/TOF/LO/LG OFG/ITG/SG/AG/NA	3.2	40094	98.21%	< 0.0001	-30	-34	-16
Left LO/TOF	4	8983	94.79%	-	-30	-34	-16
Right LO/LG/ITG	4	7653	93.85%	-	28	-30	-18
Left SFG/FP	4	1523	69.67%	-	-28	34	42
CG	4	1341	65.62%	-	6	40	-2
Right FP	4	1327	66.01%	-	30	56	28
Left SG/AG	4	859	47.85%	-	-50	-56	36
Right STG/PT/MTG HG/PrG/T	3.2	12540	95.41%	< 0.0001	60	-10	0
STG/PT/MTG/HG	4	9533	95.17%	-	60	-10	0
PrG	4	485	25.15%	-	52	0	48
Left STG/PT/MTG/ HG/IFG/T	3.2	10833	94.66%	< 0.0001	-60	-12	2
HG/PT/MTG/STG	4	7894	94.20%	-	-60	-12	2
IFG	4	667	38.98%	-	-40	14	26

Gene expression data

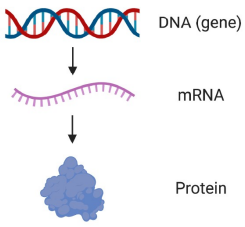
Gene expression is generally measured quantifying levels of the gene product (often a protein).

Expression of genes

H_j : gene j does not influence a response variable Y

Pathway

$H_S = \bigcap_{j \in S} H_i$: *no influence of* S *on* Y



Overview

Making inference on the TDP allows to quantify and localize brain activation in clusters.

`sumSome` is a general closed tesing method to give lower $(1-\alpha)$-confidence bounds for the TDP, simultaneously over all clusters.

This way, results are valid even if

- the cluster of interest is chosen post hoc
- we make follow-up inference inside sub-clusters