Permutation-based true discovery guarantee by sum tests

28 Sept. 2022 – XXX Conference of the Italian Association of Psychology

Anna Vesely

Joint work with Livio Finos and Jelle J. Goeman

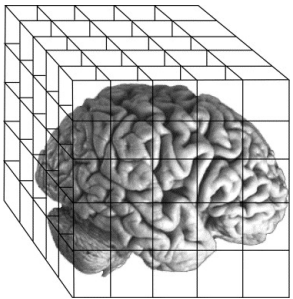
University of Padova
anna.vesely@unipd.it

Overview

Functional Magnetic Resonance Imaging

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

Activation is measured in voxels
$\approx$ 200,000 highly correlated volume units



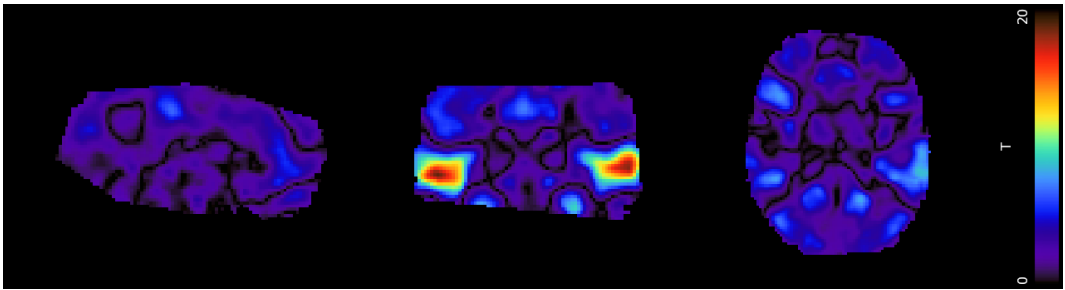
Voxel i:

- null hypothesis H_i : *no activation in voxel i*
- test statistic T_i from first-level analysis

Auditory data

140 subjects passively listening to vocal and non-vocal sounds (168,211 voxels)[1]

H_i : no activation in voxel i $\qquad$ T_i = one-sample t-statistic



[1] Pernet et al. The human voice areas: spatial organisation and inter-individual variability in temporal and extra-temporal cortices. *OpenNeuro dataset*, 2019.

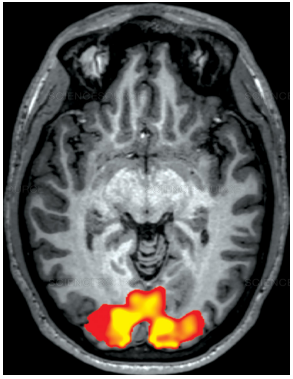
Functional Magnetic Resonance Imaging

Generally interest lies in clusters,
brain regions of (contiguous) voxels

Supra-threshold clusters:
regions of connected voxels i with
$T_i >$ threshold

Clusters S:

- null hypothesis
  $H_S = \bigcap_{i \in S} H_i$: *no activation in cluster* S
- global test statistic T_S



Multiple hypothesis testing

Multiple hypothesis testing

Single null hypothesis H_i:

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

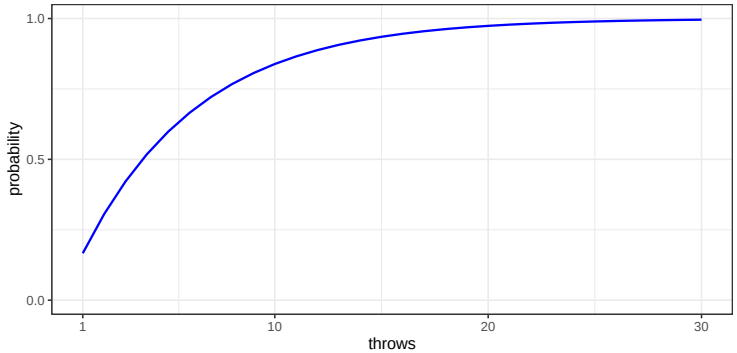
Multiple hypotheses together:

the probability of making at least one type I error can be much greater than α

Intuition: m throws of a dice

The probability of getting at least one 6 is

- $1/6 \approx 0.167$ for $m = 1$
- $1 - (5/6)^m$ in general



Standard cluster inference

Cluster extent thresholding

Standard method for cluster inference that

- adapts to the unknow 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.'
  Spatial specificity paradox.
  Follow-up inference 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 cluster* S

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

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

Closed testing[1][2] gives simultaneous bounds:

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

[1] Genovese and Wasserman. Exceedance control of the false discovery proportion. *JASA*, 2006.
[2] Goeman and Solari. Multiple testing for exploratory research. *Stat. Sci.*, 2011.

Computational complexity of closed testing: exponential in the number of voxels
$\rightarrow$ infeasible for high-dimensional data

We provide a shortcut[1], 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.

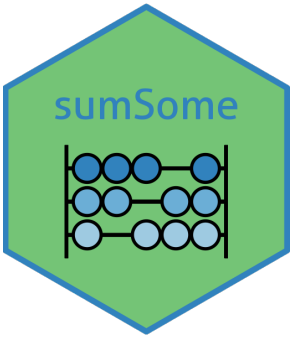
Results

sumSome package

The method is implemented in the R package sumSome[1], with underlying code in C++.

```
permT <- brainScores(copes, alpha=0.05, ...)

brainAnalysis(permT, clusters)
```

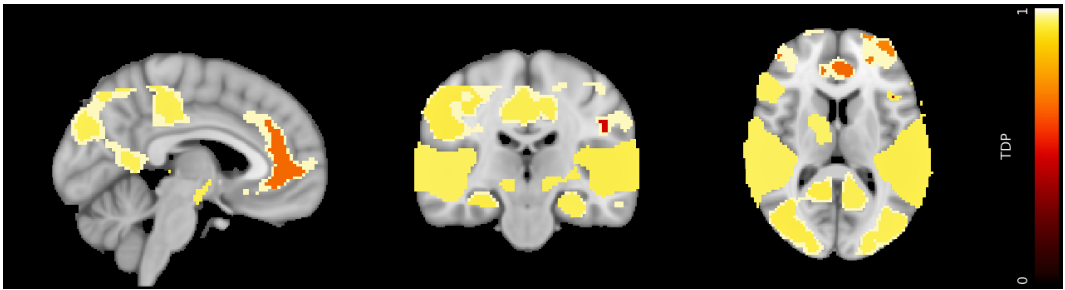


[1] https://CRAN.R-project.org/package=sumSome

Auditory data

We obtain simultaneous confidence bounds for the TDP of clusters:

- clusters with $T_i > 3.2$
- sub-clusters with $T_i > 4$



Auditory data

cluster	threshold	size	TDP (%)	coordinates		
S	*thr*	*s*	lower conf. bound	*x*	*y*	*z*
FP/CG/SFG/TOF/LO/LG OFG/ITG/SG/AG/NA	3.2	40,094	98.21	-30	-34	-16
Left LO/TOF	4	8,983	94.79	-30	-34	-16
Right LO/LG/ITG	4	7,653	93.85	28	-30	-18
Left SFG/FP	4	1,523	69.67	-28	34	42
CG	4	1,341	65.62	6	40	-2
Right FP	4	1,327	66.01	30	56	28
Left SG/AG	4	859	47.85	-50	-56	36
Right STG/PT/MTG HG/PrG/T	3.2	12,540	95.41	60	-10	0
STG/PT/MTG/HG	4	9,533	95.17	60	-10	0
PrG	4	485	25.15	52	0	48
Left STG/PT/MTG/ HG/IFG/T	3.2	10,833	94.66	-60	-12	2
HG/PT/MTG/STG	4	7,894	94.20	-60	-12	2
IFG	4	667	38.98	-40	14	26

Conclusion

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

Questions

https://github.com/annavesely/AIP2022