Flexible Inference for fMRI Data

Aldo Solari

CLADAG 2017
Big Data - Big Knowledge

Milan, September 14, 2017

Joint work with

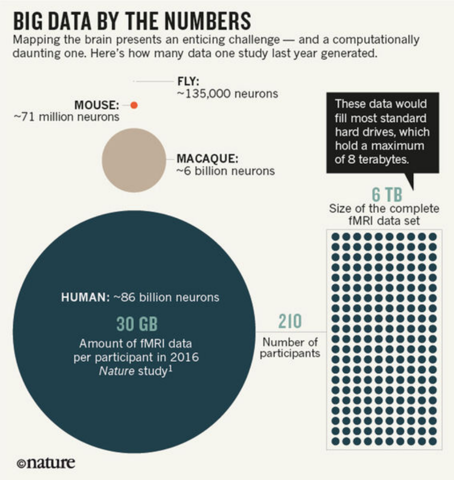
Co-author	*University*	*Country*
Jonathan D. Rosenblatt	Ben-Gurion	Israel
Livio Finos	Padua	Italy
Wouter D. Weeda	Leiden	The Netherlands
Jelle J. Goeman	LUMC	The Netherlands

Outline

1. Introduction
2. Cluster-based Inference
3. All-Resolutions Inference
4. Discussion

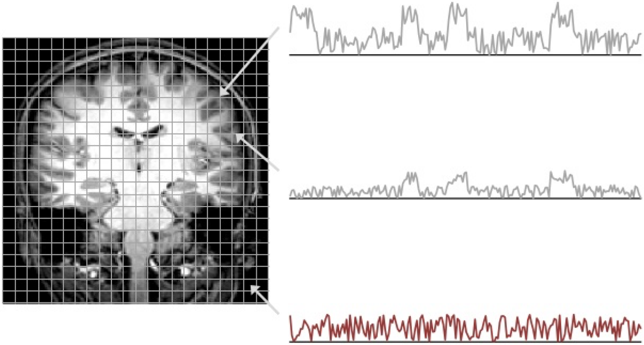
NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Big brain



Source: Landhuis (2017)

Brain activity



- Brain activity is measured on a 3D grid of voxels
- Voxel = 3mm × 3mm × 3mm cube

Go/No-go data

Task

- Go: press a button when you see an happy face ☺
- No-go: hold when you see a neutral face 😐

High-dimensional data

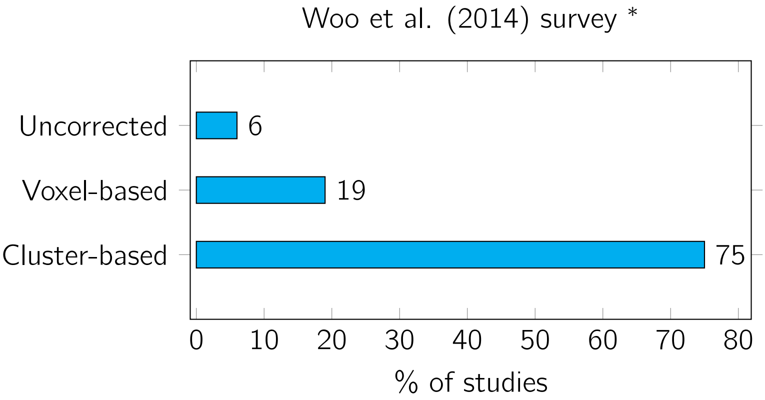
- $n = 34$ subjects
- $m = 225212$ voxels

Test statistics

- Null hypothesis: Go $=$ No-go
- A test statistic Z is computed for each voxel

NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Standard approaches



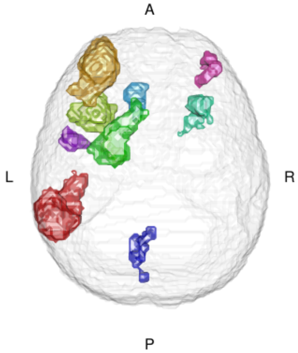
* 814 fMRI studies published in 2010 and 2011 from Cerebral Cortex, Nature, Nature Neuroscience, NeuroImage, Neuron, PNAS, and Science

Outline

Cluster-based inference

- Cluster-based inference is data-driven: clusters are both defined and tested with the same data
- Issues with circular analysis solved by random field theory
- Two steps:
    1. Cluster definition: Z-threshold
       e.g. cluster = contiguous voxels with $Z > 3.2$
    2. Cluster significance: α-threshold
       e.g. a cluster is significant if RFT p-value $< 5\%$

Significant clusters



9 significant clusters with $Z > 3.2$ and $\alpha = 5\%$

Low spatial resolution

- Discovering a cluster means that

  "there exists at least one active voxel in the cluster"

  and not that

  "all the voxels in the cluster are active"

- Spatial specificity paradox: the larger the detected cluster, the less information we have on the location of the activation

- No information on the % of activation of each cluster

NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Low spatial resolution

cluster	*size*	*# active*	*% active*
A	2191	?	?
B	1835	?	?
C	1400	?	?
D	698	?	?
E	421	?	?
F	304	?	?
G	245	?	?
H	232	?	?
I	187	?	?

Notation

- $B = \{1, 2, \ldots, m\}$: brain, collection of m voxels
- $S \subseteq B$: voxel set. A cluster C is a particular case
- $\mathcal{S} = \{S : S \subseteq B\}$ with $|\mathcal{S}| = 2^m$: collection of all voxel sets
- $A \subseteq B$: (unknown) set of truly active voxels

Parameter of interest

- $a(S) = |A \cap S|$: # of truly active voxels in S
- $\pi(S) = a(S)/|S|$: % of truly active voxels in S

Cluster null hypothesis

- Given a pre-specified cluster forming Z-threshold (e.g. $Z > 3.2$), we obtain a collection of candidate clusters $\mathcal{C}$
- Both the number of clusters $|\mathcal{C}|$ and each cluster $C \in \mathcal{C}$ are random quantities, because are determined by the data
- Cluster null hypothesis

$$H_C : \pi(C) = 0$$

  Rejecting H_C implies $\pi(C) > 0$, at least one active voxel in C
- However, the null hypotheses are random

Outline

NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Proposed approach: ARI

- Rather than testing random hypotheses, we propose a more classical approach
- Construct simultaneous lower bounds $\underline{\pi}(S)$ for the parameter of interest $\pi(S)$ satisfying

$$\mathrm{P}(\text{for all } S \in \mathcal{S} : \underline{\pi}(S) \leq \pi(S)) \geq 1 - \alpha$$

- The bound is valid for all S, and therefore for one or more selected S, regardless of how they were selected

Confidence bounds

Based on Goeman and Solari (2011), Meijer et al. (2016) and Rosenblatt et al. (2017) showed that

$$\underline{\pi}(S) = \min\left\{0 \leq k \leq |S| : \min_{1 \leq i \leq |S|-k} \frac{h}{i} p_{(i+k:S)} > \alpha\right\} / |S|$$

where $p_{(i:S)}$ is the ith smallest p-value in S and

$$h = \max\left\{i \in \{0, \dots, m\} : i p_{(m-i+j:B)} > j\alpha, \ \text{for } j = 1, \dots, i\right\}$$

Assumption: Simes inequality

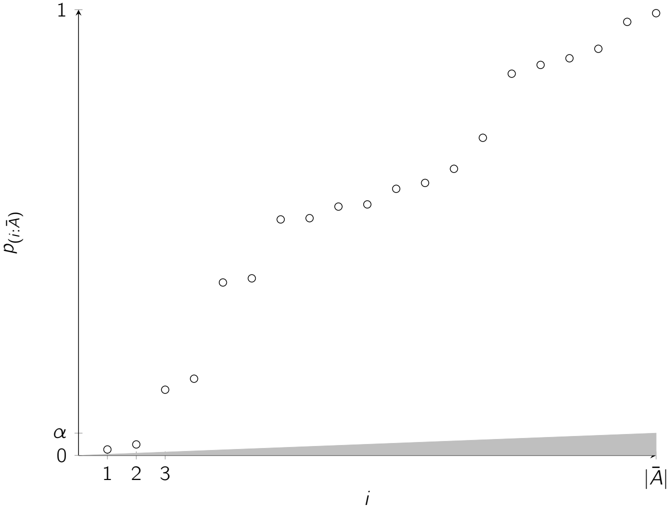
For the set $\bar{A} = B \setminus A$ of all non-active voxels, we assume the Simes inequality

$$\mathrm{P}\left(\bigcap_{i=1}^{|\bar{A}|}\left\{p_{(i:\bar{A})} > \frac{i\alpha}{|\bar{A}|}\right\}\right) \geq 1-\alpha$$

- It has been shown to hold for independent p-values, and under various conditions implying non-negative correlations between p-values, one of which is the PRDS condition
- For instance, Simes inequality is necessary for the Benjamini and Hochberg (1995) procedure
- Nichols and Haysaka (2003) suggested that PRDS, and therefore the Simes inequality, is plausible for brain maps

NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Simes inequality



Computation time

- Sorting the p-values: linearithmic $O(m \log m)$
- Computing h (only once): linearithmic $O(m \log m)$
- Computing $\underline{\pi}(S)$: linear in S
- With $m = 225212$ takes seconds

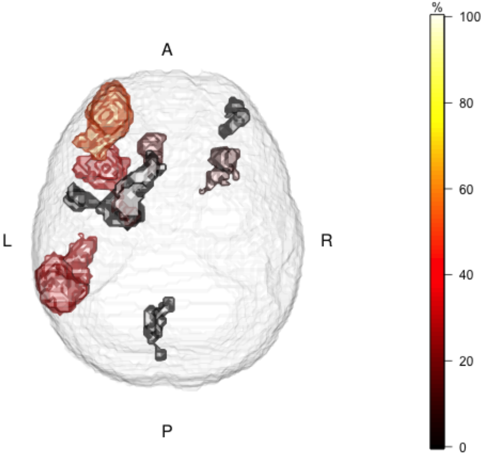
ARI: quantify and localize

1. quantify the activation of each cluster
2. localize the source of the activation within the cluster: a "drill-down" from discovered clusters to sub-clusters

Quantify

cluster	*size*	*# active* $\underline{a}(C)$	*% active* $\underline{\pi}(C)$
A	2191	624	29 %
B	1835	847	46 %
C	1400	454	32 %
D	698	0	0 %
E	421	25	6 %
F	304	33	11 %
G	245	0	0 %
H	232	0	0 %
I	187	0	0 %

Quantify



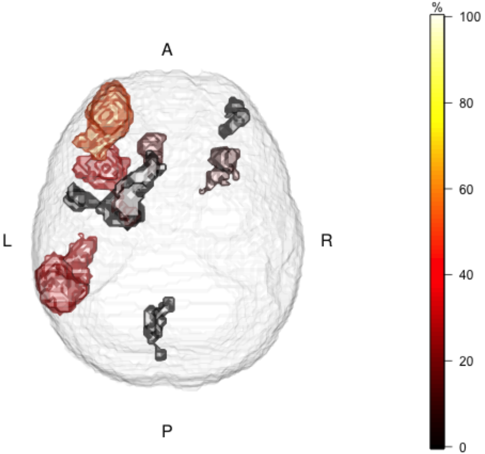
NeuroMI
MILAN CENTER FOR NEUROSCIENCE

Localize

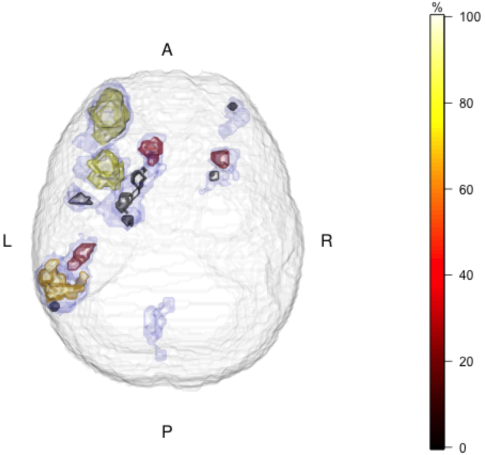
cluster	*threshold*	*size*	*# active*	*% active*
A	$Z > 3.2$	2191	624	29 %
1	$Z > 4$	405	267	66 %
2	$Z > 4$	133	31	23 %
3	$Z > 4$	6	0	0 %
B	$Z > 3.2$	1835	847	46 %
1	$Z > 4$	963	826	86 %
C	$Z > 3.2$	1400	454	32%
1	$Z > 4$	583	449	77 %
2	$Z > 4$	4	0	0 %
3	$Z > 4$	1	0	0 %

$\vdots$

Localize



Localize



Outline

Discussion

- ARI brings unprecedented flexibility to the analysis of fMRI data
- With ARI, users can iterate the process of
  1. choosing regions in any way, also after seeing the data
  2. quantify the % of activation
  3. refine the regions

  without compromising the validity of the inference
- ARI is implemented in `R` package `hommel`

References

Goeman JJ and Solari (2011)
Multiple Testing for Exploratory Research
Statistical Science, 26:584?597

Meijer RS, Krebs TJP, Solari A and Goeman JJ (2016)
Simultaneous Control of All Discovery Proportions by an Extension of Hommel's Method
Arxiv

Rosenblatt JD, Finos L, Weeda WW, Solari A and Goeman JJ (2017)
All-Resolutions Inference for Brain Imaging
Submitted