Università degl Studi di Milano-Bicocca
PhD ECOSTAT - Curriculum Statistics
Course: Statistical Inference II

Post Selection Inference: fMRI Data analysis

Alice Giampino

29 June 2020

Abstract

The problem of post selection inference arises when the same data are used to select interesting pattern and to perform inference about the selection. Due to this reason, overoptimism problem persist with data-driven selection. To overcome this issue All-Resolutions Inference (ARI) can be applied. This technique is based on Closed Testing principle. In particular, data of fMRI experiments are analysed. The subjects are scanned during two session where they are asked to look at famous or non-famous faces and to press one of two keys [8] ('first-level' analysis). A 'second-level' analysis is presented and it allows to make inferences about the population from which the subjects were drawn. Moreover, the results of selected areas, after the correction for the overoptimism, correspond to brain regions of interest.

1 fMRI Data

The data set contains 12 contrast images from single-subject fMRI analyses or 'first-level' analyses from the repetition priming experiment described in Henson, R.N.A., Shallice, T., Gorno-Tempini, M.-L. and Dolan, R.J. (2002) [8], where the subjects are scanned during two session where they are asked to look at famous or non-famous faces and to press one of two keys. The images are then used in a 'second-level' analysis. Every scan (one for each subjects) presents value related to MRI measurement of oxygen blood flow in brain (brain activity). We obtain the correspondent p-values representing the significance for brain activity at each location (voxel, i.e. 3D pixel). The total number of voxel is 154680. However, only 51560 correspond to the ones contained in the mask (i.e. the mask of the brain that exclude the background to standardize the size of every image).
The goal of the analysis is to find the regions of brain activity. Hence, the cluster size inference allows us to aggregate voxels (micro-scale inference) and perform larger-scale inferences (regions).

2 Methodologies

Presentation of the used methodologies: Cluster Size Inference and ARI, with a presentation of some customized function.

2.1 Cluster Size Inference

Voxel-level inferences make no use of any spatial information in the image, such as the fact that activated voxels might be clustered together in space [12]. For this reason, we perform cluster size inference and declare a cluster as 'significant' if its size is larger than 'chance'. We

have to define the size threshold ($(1-\alpha)$ quantile of the null distribution of the maximum size of clusters) in order to control the familywise error rate at level α. A cluster is a set of contiguous voxel with p-values all under some predefined threshold.
Since discovering a cluster means that 'there exists at least one voxel with an evoked response in the cluster', and not that 'all the voxels in the cluster have an evoked response', it follows that the larger the detected cluster, the less information we have on the location of the activation. Moreover, cluster-based inference gives no information on the extent of the activation within the cluster [14].
Hence, we start the analysis setting different value for the level threshold, i.e. the level at which to threshold the array values, and for the cluster size threshold we choose 50. Deciding the threshold for the array values of the statistics calculated for each voxel has been opitimized using function described in the following sections. As neighbourhood for continuity and dependency, we decide to use a matrix that considers 3 neighbours in 3D for each voxel.

2.2 All-Resolutions Inference

The All-Resolutions Inference (ARI) allows the researcher to apply any data-driven region selection and estimate the proportion of true discoveries (TDP[1]) of any subregion—clusters in our case—all from the same dataset [7]. ARI controls the FWER over all possible subsets of the brain, large or small, contiguous or non-contiguous, using closed testing [11]. ARI exploits the powerful Simes test (Appendix A), which exhibits the necessary correlation structure of the voxel. Since the positive regression dependency on subsets condition (PRDS) is established for brain maps, Simes test is valid under the assumption of the Simes inequality.
A test statistic for activation has been calculated for each voxel, which can be converted into a voxel-wise p-value. ARI allows to construct lower confidence bounds $\bar{\pi}(S)$ for the set-wise proportion of active voxels, simultaneously for all possible sets. The $(1-\alpha)$ lower confidence bound is such that $P(\text{for all } S \in \mathcal{S} : \bar{\pi}(S) \leq \pi(S)) \geq 1-\alpha$, where $\mathcal{S} = 2^B$ is the collection of all $|S| = 2^m$ voxel sets of the brain B.
This procedure allows to obtain a percentage (and a number) of active voxel for each cluster, based on a restrictive threshold. In ARI the singleton sets for which $\bar{\pi}(S) = 1$ correspond to the voxels rejected by the procedure of Hommel [9], a uniform improvement of Bonferroni; the latter is more conservative.
We use the R package *hommel*[6] as the main tool to obtain adjusted p-values and the true discoveries proportion. For sake of completeness, if we want compare the method of Hommel, with the method of Bonferroni and Benjamini-Hochberg (BH), we will find a number of hypothesis with adjusted p-value lower than α (probability of type I error) equals to 5%, lower or higher based on the chosen method. We obtain 248 p-values for Hommel, 212 for Bonferroni and 16258 with Benjamini-Hockberg. Considering the p-values lower than 0.05 without any adjustment for multiple testing, the number of hypothesis are 21870. Bonferroni is more conservative, whereas Benjamini-Hochberg rejects a number of hypothesis higher than 65 times the number of hypothesis of Hommel.

2.3 Optimization

Since we want to conduct an analysis which allows us to find the optimal clusters that correspond to interesting region of the brain, we create some customized function to reach the goal. The algorithms are shortly described in Appendix B.

[1] TDP $= 1 -$ FDP, where FDP is the false discovery proportion.

The idea is to optimize the level threshold of the cluster in a way that exclude voxels which do not increase the percentage of activation of the relative cluster. We base the function on the fact that if a statistic assume a higher value, the correspondent p-value must be small. In this way we obtain sub-clusters. We try three different function: the first one is based on the index of Gini. It finds the concentration of the distribution of the statistics in each cluster based on the concentration of adjusted p-values lower than 0.05 in the different clusters. We want heterogeneous clusters. However, this procedure does not increase significantly the percentage of activation even if it permits to exclude an amount of voxels from the original clusters.
The second and the third function are based only on the optimization of the percentage of activation of each cluster. The former considers a fixed number of clusters equals to 4, and the latter considers the number of clusters in a set of given numbers. These function improve considerably the percentage of activation in each cluster, finding the optimal sub-clusters. The obtained results are described in the next section.

3 Analysis

All the analysis is done using the R software, version 3.6.2[2]. The code is published on a github repository[3]. Cluster size inference initially finds 4 regions of interest (Figure 1).

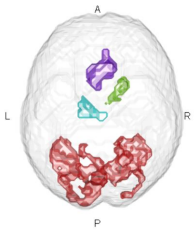

Figure 1: Cluster-size inference.

Two regions correspond to more superficial areas: visual cortex - in particular the infotemporal area (red area) - and part of the area of the motor cortex (violet area).
While others correspond to the basal ganglia, in particular the caudate nucleos (green area) and to the thalamus (lightblue area).

Let S be the selected voxels (clusters), $|S|$ its cardinality, $m_1(S)$ the cardinality of non-null voxels and $\pi_1(S)$ the TDP of each cluster. Table 1a reports the results for this procedure.
Since the main problem of cluster-size inference is that it gives no information on the extent of the activation within the cluster, but it only implies that 'at least one voxel in the cluster is active', we perform ARI. The results can be seen in Table 1b, where a lower bound is calculated for each cluster. Bonferroni is too conservative and finds a tiny percentage of activation (Table 1c). In this case, let p indicates the p-values, $p < 0.00003$.

S	$\lvert S\rvert$	$\mathrm{m}_1(S)$
C_1	1409	≥ 1
C_2	93	≥ 1
C_3	89	≥ 1
C_4	172	≥ 1

(a) Cluster-size inference.

S	$\lvert S\rvert$	$\mathrm{m}_1(S)$	$\pi_1(S)$
C_1	1409	≥ 1382	$\geq 98.1\%$
C_2	93	≥ 66	$\geq 71\%$
C_3	89	≥ 63	$\geq 70.8\%$
C_4	172	≥ 146	$\geq 84.9\%$

(b) ARI results.

S	$\lvert S\rvert$	$\mathrm{m}_1(S)$	$\pi_1(S)$
C_1	1409	≥ 183	$\geq 13\%$
C_2	93	≥ 1	$\geq 1.08\%$
C_3	89	≥ 5	$\geq 5.62\%$
C_4	172	≥ 14	$\geq 8.14\%$

(c) Bonferroni results.

[2]The main packages used for our data are *fmri*[15] and *AnalyzeFMRI*[2] to open image files, *misc3d*[5] and *rgl*[1]: plot 3D images, *hommel*[6] to obtain adjusted p-values and true discoveries proportion, *DescTools*[4] for the gini function, *dplyr*[16] to manipulate data and *tictoc*[10] to calculate how time consuming are the implemented functions.

[3]Github repository: `https://github.com/AliceGiampino/fMRI-Analysis`

ARI legitimates post-selection inference with full-flexibility. The user can both select interesting region data-driven or knowledge-based or a mix of the two. For this reason, we opt to use data-driven selection using the customized function based on a fixed number of cluster. This allows us to obtain sub-clusters from the initial ones selecting a more restrictive threshold (t') to zoom into the cluster (Figure 2). We select a threshold that eliminates some clusters due to the corresponding brain areas. The new threshold is such that $p < 0.000005 < 0.00003$.

S	threshold	$\|S\|$	active	% active
C_1	t	1409	1382	$\geq 98.1\%$
sub_1	t'	282	279	$\geq 98.9\%$
sub_2	t'	142	139	$\geq 97.9\%$
sub_3	t'	90	87	$\geq 96.7\%$
C_2	t	93	66	$\geq 71\%$
sub_1	t'	0	0	$\geq 0\%$
C_3	t	89	63	$\geq 70.8\%$
sub_1	t'	0	0	$\geq 0\%$
C_4	t	172	146	$\geq 84.9\%$
sub_1	t'	57	54	$\geq 94.7\%$

Table 2: Sub-clusters.

3.1 Final results

The selected clusters find a corresponding regions of the Brodmann areas[3]. In fact, looking at Figure 2, we find activation of the area 20-21 and 18 corresponding to the color red, green and light-blue. These regions are related to the visual cortex, in particular they include the fusiform gyrus which is connected to the face recognition. Moreover, we find activation of the inferior temporal gyrus and the middle temporal visual cortex. The violet area corresponds to the supplementary motor area (part of the Brodmann's area 4). These regions are confirmed by a specialist of the field[4]. Considering the structure of the experiment, the correspondence between the highlighted areas and the activities done by the subjects is strongly evident.

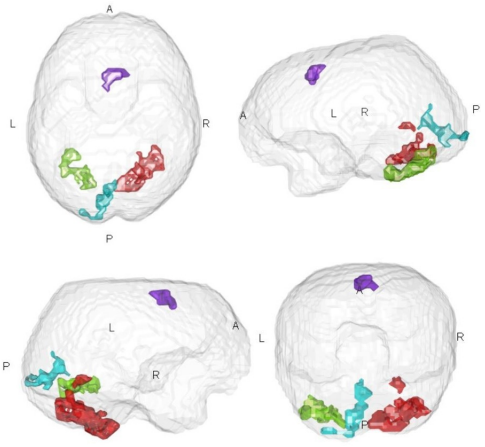

Figure 2: Selected regions

[4]Lidia Fortaner Uyá, Università Vita-Salute San Raffaele Milano, Corso di Laurea Magistrale in Psicologia, Curriculum: Neuroscienze. Using the software *mricron* and AAL (Automated anatomical labelling) Atlas that update the Brodmann areas.

A Simes test

Let $p_1, \ldots, p_m \overset{i.i.d.}{\sim} U(0,1)$ under H_0.
Sort the p-values:

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$$

The null distribution of j-th ordered p-value is

$$p_{(j)} \overset{H_0}{\sim} Beta(j, m-j+1)$$

The Simes test p-value

$$p_s = \min_{j=1,\ldots,m} \left\{ p_{(j)} \frac{m}{j} \right\} \overset{H_0}{\sim} U(0,1)$$

It rejects H_0 if $\exists\ j : p_{(j)} \leq \frac{\alpha j}{m}$.

B Algorithms

Algorithm 1: Gini concentration

Result: gini index, activation percentage, value of the threshold
Input: map, coordinates, clusters, sequence of values for the threshold, p-values;
for *each value of the sequence* **do**
    calculate sub-clusters;
    calculate activation % for each sub-cluster;
    calculate the concentration of active p-values in each clusters;
    calculate the gini concentration;
end

Algorithm 2: Threshold

Result: Average percentage of activation of the clusters
Input: statistics, sequence of values for the threshold, size threshold of clusters, coordinates, number of cluster;
for *each value of the sequence* **do**
    calculate sub-clusters;
    calculate activation % for each sub-cluster;
    compute the mean of the percentage for each sub-cluster;
end

Algorithm 3: Threshold and K

Result: Average percentage of activation of the clusters, number of cluster
Input: statistics, sequence of values for the threshold, size threshold of clusters, coordinates, set of number of cluster;
for *each value of the sequence* **do**
    for *each different number of cluster* **do**
        calculate sub-clusters;
        calculate activation % for each sub-cluster;
        compute the mean of the percentage for each sub-cluster;
    end
end

Additional customized function can be found on the github repository[5].

References

[1] Daniel Adler, Duncan Murdoch, et al. *rgl: 3D Visualization Using OpenGL*, 2020. R package version 0.100.54.

[2] Cécile Bordier, Michel Dojat, and Pierre Lafaye de Micheaux. Temporal and spatial independent component analysis for fmri data sets embedded in the AnalyzeFMRI R package. *Journal of Statistical Software*, 44(9):1–24, 2011.

[3] Korbinian Brodmann. *Brodmann's: Localisation in the cerebral cortex*. Springer Science & Business Media, 2007.

[4] Andri Signorell et mult. al. *DescTools: Tools for Descriptive Statistics*, 2020. R package version 0.99.36.

[5] Dai Feng and Luke Tierney. Computing and displaying isosurfaces in R. *Journal of Statistical Software*, 28(1), 2008.

[6] Jelle Goeman, Rosa Meijer, and Thijmen Krebs. *hommel: Methods for Closed Testing with Simes Inequality, in Particular Hommel's Method*, 2019. R package version 1.5.

[7] Jelle J Goeman, Aldo Solari, et al. Multiple testing for exploratory research. *Statistical Science*, 26(4):584–597, 2011.

[8] RNA Henson, Timothy Shallice, ML Gorno-Tempini, and Raymond J Dolan. Face repetition effects in implicit and explicit memory tests as measured by fmri. *Cerebral Cortex*, 12(2):178–186, 2002.

[9] Bernhard Hommel. Event files: Evidence for automatic integration of stimulus-response episodes. *Visual cognition*, 5(1-2):183–216, 1998.

[10] Sergei Izrailev. *tictoc: Functions for timing R scripts, as well as implementations of Stack and List structures.*, 2014. R package version 1.0.

[11] Ruth Marcus, Peritz Eric, and K Ruben Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.

[12] Russell A Poldrack, Jeanette A Mumford, and Thomas E Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

[13] Jean-Baptiste Poline and Bernard M Mazoyer. Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters. *Journal of Cerebral Blood Flow & Metabolism*, 13(3):425–437, 1993.

[14] Jonathan D Rosenblatt, Livio Finos, Wouter D Weeda, Aldo Solari, and Jelle J Goeman. All-resolutions inference for brain imaging. *Neuroimage*, 181:786–796, 2018.

[15] Karsten Tabelow and Jörg Polzehl. Statistical parametric maps for functional mri experiments in R: The package fmri. *Journal of Statistical Software*, 44(11):1–21, 2011.

[5] Github repository: `https://github.com/AliceGiampino/fMRI-Analysis`

[16] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2020. R package version 1.0.0.