Non-parametric Post-hoc Inference and application to brain imaging

Supervision:

- Bertrand Thirion, Parietal, Inria (bertrand.thirion@inria.fr)
- Pierre Neuvial, Institut de Mathématiques de Toulouse (IMT) (pierre.neuvial@math.cnrs.fr)

Context Large-scale statistical testing is ubiquitous in many scientific fields, where high-dimensional datasets are collected and compared with an outcome of interest. In such high-dimensional contexts, false discovery rate (FDR) control [1] is attractive because it yields reasonable power, while providing an explicit and interpretable control on false detections. Yet the false discovery rate is the expectation of the false discovery proportion(FDP). Controlling the FDR does not mean that the FDP is controlled, a distinction that is most often ignored by practitioners. There is thus interest in methods controlling the FDP.

Such an approach has been developed in the context of neuroimaging, namely the all-resolution inference framework [5] based on the Simes bound that underlies many statistical tests. Yet, the empirical behavior of this methods remains to be assessed. Moreover, it has been clearly established that the procedure is over-conservative in some settings [2].

Methods Indeed, as discussed in [2], the bound of [5] is known to be valid only under certain positive dependence assumptions (PRDS) on the joint p-value distribution. Although the PRDS assumption is widely accepted for fMRI studies (see e.g. [3]), this assumption yields overly conservative post hoc bounds. Indeed, the Simes bound is by construction not adaptive to the specific type of dependence at hand for a particular data set.

To bypass these limitations, [2] have proposed a randomization-based procedure known as λ-calibration, which yields tighter bounds that are adapted to the dependency observed in the data set at hand. I rests on a non-parametric (permutation-based) estimation of the null distribution, leading tight and valid inference under general assumptions. We note that a related approach has been proposed by [4]. In the case of two-sample tests, this calibration can be achieved by permutation of class labels.

In this internship, we propose to fix some of the open issues with the approach described in [2], namely the choice of a template family to calibrate the error distribution in the permutation procedure. We hope to propose a practical choice for this family to avoid putting the burden of choice on practitioners. We propose to assess empirically the benefits of the different estimators available, using many different brain imaging datasets, to provide a clear picture of the merits of each approach. Finally, we want to characterize by simulations and theoretical arguments the behavior of these error control procedures to develop future tests well suited to brain imaging data in particular.

Environment The internship will take place in Inria Saclay, in the Parietal team. This is a large team working focused on mathematical methods for statistical modeling of brain function using neuroimaging data (fMRI, MEG, EEG). Particular topics of interest include machine learning techniques, numerical and parallel optimization, applications to human cognitive neuroscience, and scientific software development.

The internship will be in collaboration with Pierre Neuvial, from Institut de Mathématiques de Toulouse (IMT), that is an expert in FDP control.

Requirements

- Strong mathematical background. Knowledge in statistics and numerical optimization is a plus.

- Good programming skills in Python.
- Willingness to work in a multi-disciplinary environment.
- Prior experience of brain imaging or medical imaging in general is a plus.

References

[1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, 57(1):289–300, 1995.

[2] Gilles Blanchard, Pierre Neuvial, and Etienne Roquain. Post hoc confidence bounds on false positives using reference families. *Ann. Statist.*, 48(3):1281–1303, 06 2020.

[3] C. R. Genovese, N. A. Lazar, and T. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, Apr 2002.

[4] Jesse Hemerik, Aldo Solari, and Jelle Goeman. Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika*, 106:635–649, 07 2019.

[5] J. D. Rosenblatt, L. Finos, W. D. Weeda, A. Solari, and J. J. Goeman. All-Resolutions Inference for brain imaging. *Neuroimage*, 181:786–796, 11 2018.