



Department of Data Analysis

## Stability based testing for the analysis of fMRI data

Joke Durnez and Beatrijs Moerkerke

Neurological imaging has become increasingly important in the field of psychological research. The leading technique is functional magnetic resonance imaging (fMRI), in which a correlate of the oxygenlevel in the blood is measured (the BOLD-signal). In an fMRI-experiment, a time series of brain images is taken while participants perform a certain task. By comparing different conditions, the task-related areas in the brain can be localised. An fMRI study leads to enormous amounts of data. To analyse the data adequately, the brain images are devided into a large number of volume units (or voxels). Subsequently, a time series of the measured signal is modelled voxelwise as a linear combination of different signal components, after which an indication of activation can be tested in each voxel. This encompasses an enormous number of simultaneous statistical tests (+/-250 000 voxels). As a result, the multiple testing problem is a serious challenge for the analysis of fMRI data.

In this context, classical multiple testing procedures such as Bonferroni and Benjamini-Hochberg (Benjamini & Hochberg, 1995) have been applied to respectively control the family-wise error rate (FWER) and the false discovery rate (FDR)(Genovese, Lazar, & Nichols, 2002). Random Field Theory (Worsley, Evans, Marrett, & Neelin, 1992) controls the FWER while accounting for the spatial character of the data. Because of the dramatically decrease in power when controlling the FWER, methods to control the topological false discovery rate (FDR) were developed (Chumbley & Friston, 2009; Heller, Stanley, Yekutieli, Rubin, & Benjamini, 2006).

A general shortcoming of current procedures is the focus on detecting non-null activation while a non-null effect is not necessarily biologically relevant. Moreover, failing to reject the hypothesis of no activation is not the same as confidently excluding important effects. Another aspect that remains largely unexplored is the stability of test results which can be defined as selection variability of individual voxels (Qiu, Xiao, Gordon, & Yakovlev, 2006).

Given the need to control both false positives (type I errors) and false negatives (type II errors) in a direct manner (Lieberman & Cunningham, 2009), we approach the multiple testing problem from a different angle. Following the procedure of (Gordon, Chen, Glazko, & Yakovlev, 2009) in the context of gene selection, we present a statistical method to detect brain activation that not only includes information on false positives, but also on power and stability. The method uses bootstrap resampling to extract information on stability and uses this information to detect the most reliable voxels in relation to the experiment. The findings indicate that the method can improve stability of procedures and allows a direct trade-off between type I and type II errors. In this particular setting, it is shown how the proposed method enables researchers to adapt classical procedures while improving their stability. The method is evaluated and illustrated using simulation studies and a real data example.

## References

Benjamini, Y., & Hochberg, Y. (1995). 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.





Department of Data Analysis

- Chumbley, J., & Friston, K. (2009). False discovery rate revisited: Fdr and topological inference using gaussian random fields. *NeuroImage*, 44, 62-70.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*, 870-878.
- Gordon, A., Chen, L., Glazko, G., & Yakovlev, A. (2009). Balancing type one and two errors in multiple testing for differential expression of genes. *Computational Statistics and Data Analysis*, *53*, 1622-1629.
- Heller, R., Stanley, D., Yekutieli, D., Rubin, N., & Benjamini, Y. (2006). *NeuroImage*, *33*, 599-608.
- Lieberman, M. D., & Cunningham, W. A. (2009). Type i and type ii error concerns in fmri research; re-balancing the scale. *Social cognitive and affective neuroscience*, *4*, 423-428.
- Qiu, X., Xiao, Y., Gordon, A., & Yakovlev, A. (2006). Assessing stability of gene selection in microarray data analysis. *bioinformatics*, 7(50).
- Worsley, K., Evans, A., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow Metabolism*, *12*(6), 900-918.