Adaptive Thresholding for fMRI data

Joke Durnez and Beatrijs Moerkerke
Department of Data Analysis, Ghent University

Problem Setting

- In an fMRI data analysis, many brain regions are tested simultaneously for activation.
- Several multiple testing procedures (MTPs) exist that prevent an inflation of type I errors while accounting for the spatial structure of the data.
- Such procedures consider testing peaks or clusters of activation (Chumbley, 2010).
- Not only false positive rates are of interest, but also an estimate of the true positive rate (power).
- We focus on the analysis of maxima (or peaks) of activation.
- The goal is to estimate the proportion of non-active peaks (π_0).
- Given π_0:
  1. The type I error and power following a thresholding technique can be estimated, enabling a direct trade-off between sensitivity and specificity.
  2. Widely used MTPs can be made adaptive and more powerful.

A measure of significance for peaks

- Random Field Theory (RFT) allows to obtain p-values for peaks to test the null hypothesis of no activation (H_0) against the alternative of activation (H_1).
- Consider a t-map with a cluster defining treshold of t_1 with $P(T<t_1)=0.001$. Peaks above t_2 are considered to be activated.

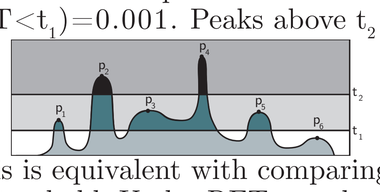

Figure 1

- This is equivalent with comparing p-values of peaks with an α-threshold. Under RFT, p-values for peaks can be obtained.
- Under H_0: RFT p-values are uniformly distributed.
- Results of testing H_0 for peaks can be summarized as follows:

	#selected	**#not selected**	
# true H_0	#FP	#TN	**m_0**
# true H_1	#TP	#FN	**(m-m_0)**
	S	**(m-S)**	**m**

Table 1

- The goal is to estimate $\pi_0=m_0/m$
- Given a specific decision criterion and π_0, the above 2x2 table can be estimated.

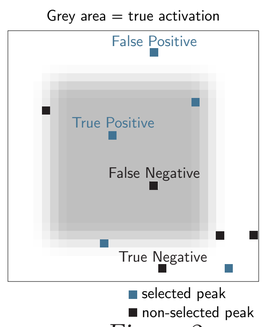

Figure 2

Error Measures and Power

We consider:

- Type I error rate
  - False Positive Rate (FPR): E(#FP)/m_0
  - False Discovery Rate (FDR): E(#FP/S) if S>0 and 0 otherwise
- Power
  - True Positive Rate (TPR):E(#TP)/(m-m_0)

Simulated Data

Data created using `neuRosim` (Welvaert et al., 2011):

- 1000 simulations
- 3D images (32x32x32 voxels) with 3 active clusters (6x6x6 voxels) and 8 time points.
- temporal and spatial noise
- smoothness: HIGH ($\sigma = 3.5$) vs. LOW ($\sigma = 1.5$)
- Signal-to-noise ratio (SNR): HIGH (1) vs. LOW (0.5)
- True and false positives and negatives are defined as in Figure 2, where the grey area equals the activation field *after smoothing*.

Estimation of π_0

Comparison of estimates of π_0, analoguous to the approach of Broberg (2005)

- Smoother (STOREY) (Storey et al., 2003)
- Bootstrap smoother (BOOTSTRAP) (Storey, 2001)
- Beta-uniform model (BUM) (Pounds et al., 2003)
- Spacing loess histogram (SPLOSH) (Hsueh et al., 2003)
- Lowest SLope estimator (LSL) (Scheid et al., 2003)
- Successive Elimination Procedure (SEP) (Scheid et al., 2003)

	smoothness = LOW		smoothness = HIGH	
	SNR = LOW	SNR = HIGH	SNR = LOW	SNR = HIGH
true π_0 (sd)	0.581 (0.044)	0.558 (0.045)	0.594 (0.049)	0.58 (0.05)
bias STOREY (sd)	0.171 (0.178)	0.144 (0.176)	0.1 (0.183)	0.069 (0.285)
bias BOOTSTRAP (sd)	0.057 (0.141)	-0.027 (0.098)	-0.017 (0.133)	-0.084 (0.102)
bias BUM (sd)	0.077 (0.056)	-0.127 (0.037)	-0.027 (0.054)	-0.175 (0.045)
bias SPLOSH (sd)	-0.007 (0.137)	-0.068 (0.108)	-0.058 (0.13)	-0.114 (0.104)
bias LSL (sd)	0.289 (0.081)	0.045 (0.029)	0.179 (0.075)	0.001 (0.036)
bias SEP (sd)	0.133 (0.099)	0.042 (0.055)	0.059 (0.095)	-0.017 (0.07)

Table 2

- All methods rely on the fact that the p-values are uniformly distributed under H_0.
- They differ in the way they estimate the proportion of null p-values.
- Bias = estimated π_0 - true π_0.
- BOOTSTRAP produces a small bias and is the most stable under different conditions.

bootstrap procedure:

1. Compute the number of true null hypotheses
$$\hat{\pi}_0(\lambda) = \frac{m - N(\lambda) + 1}{(1-\lambda)m}$$
where $N(\lambda)$ is the number of p-values less than or equal to λ.
2. For each λ, bootstrap from the sample of p-values to form B bootstrap versions $\pi_0^b(\lambda)$.
3. Choose the λ that yields the minimal variance of $\hat{\pi}_0(\lambda)$.

Application 1: ROC-estimation

	selected	**not selected**	
H_0 true	$a*m_0$	$m_0-(a*m_0)$	m_0
H_1 true	S-($a*m_0$)	(m-S)-(m_0-($a*m_0$))	(m-m_0)
	S	(m-S)	m

Table 3

- Consider a range of α-thresholds for the p-values.
- For a single peak: reject H_0 when its p-value is lower than α.
- For each α, the FPR and TPR can be estimated, according to table 3.
- This allows to produce an estimate of the ROC curve.
- We estimate ROC curves for simulated data and compare them with the true ROC curve.
- Evaluation: AUC (Area Under Curve) and maximal distance:
  - $AUC_{dif} = AUC_{est} - AUC_{true}$
  - $AUC_{restricted} = AUC_{est} - AUC_{true}$ for FPR *in* [0,0.10]
  - maximal distance: the maximal distance between ROC_{est} and ROC_{true}

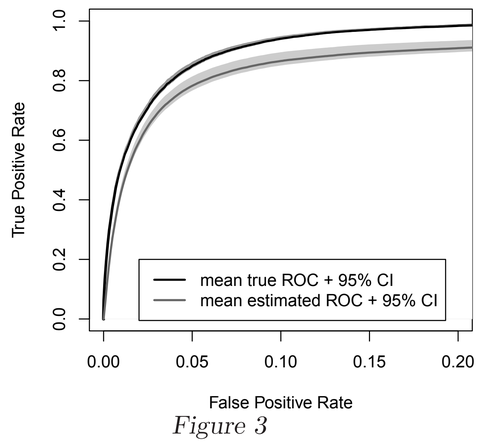

Figure 3

		AUC_{dif} (sd)	$AUC_{restricted}$ (sd)	maximal distance (sd)
Smoothness LOW	SNR LOW	-0.043 (0.095)	0.016 (0.016)	0.042 (0.232)
	SNR HIGH	-0.029 (0.072)	0.008 (0.022)	-0.037 (0.206)
Smoothness HIGH	SNR LOW	-0.045 (0.089)	0.008 (0.022)	-0.039 (0.206)
	SNR HIGH	-0.026 (0.071)	-0.041 (0.044)	-0.112 (0.139)

Table 4

Application 2: Adaptive FDR control

- False discovery rate on peaks: topological FDR (Chumbley, 2010).
- FDR kan be made less conservative when taking into account π_0 (Benjamini et al, 1995).

$$\widehat{FDR}(\alpha) = \frac{\alpha * \hat{\pi}_0 * m}{\sum_{i=1}^{m}(p_i \leq \alpha)} = \frac{\alpha * \hat{m}_0}{S}$$

- However: FDR control on peaks is not well-defined since several evaluations of true activation are possible.
- In our simulation, we find that the FDR is controlled conservatively using our evaluation criteria as in figure 2.
- However, adaptive FDR control leads to less conservative results.
- Results with SNR=0.75 and smoothness 2.5, FDR control of 0.05:

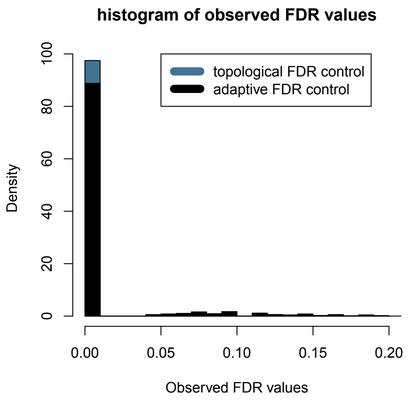

Figure 4

	mean observed FDR (sd)
topological FDR	0.003 (0.040)
adaptive FDR	0.012 (0.019)

Table 5

Discussion and Conclusion

- Estimating the proportion of null voxels to make voxel-based procedures adaptive is not useful since this proportion is close to 1 in fMRI studies (Chen et al., 2009).
- However, when performing inference on other topological features, such as peaks, more active features can be expected, reducing the proportion of null voxels.
- We present a technique to estimate the proportion of activated peaks.
- This allows to estimate error measures and power in a single fMRI-study.
- Overall, we find that π_0 and operating characteristics of selection procedures are estimated adequately.
- Given an estimate of specificity and sensitivity, a direct trade-off between both measures can guide thresholding peaks of brain activation in fMRI studies. This allows researchers to reconsider the balance between true positive and false negative rates in function of study goals.

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 B, 57, 289-300.
- Broberg P (2005). A comparative review of estimates of the proportion unchanged genes and the false discovery rate. BMC Bioinformatics, 6:199.
- Chumbley J, Worsley K, Flandin G, Friston K (2010). Topological FDR for neuroimaging. NeuroImage, 49,3057-3064.
- Chen S, Wang C, Eberly L, Caffo B, Schwartz B (2009). Adaptive control of the false discovery rate in voxel-based morphometry. Human Brain Mapping, 30, 2304-2311.
- Hsueh H, Chen JJ, Kodell RL (2003). Comparison of methods for estimating the number of true null hypotheses in multiplicity testing. Journal of biopharmaceutical statistics, 13, 675-689.
- Pounds S, Morris, SW (2003). Estimating the occurence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. Bioinformatics, 19, 1236-1242.
- Scheid S, Spang R (2004). A stochastic downhill search algorithm for estimating the local false discovery rate. IEEE Transactions on computational biology and bioinformatics, 1, 98-108.
- Storey JD (2001). A direct approach to false discovery rates. Journal of the Royal Statistical Society B, 64, 479-498.
- Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. PNAS, 100,9440-9445.
- Welvaert M, Durnez J, Moerkerke B, Verdoolaege G, Rosseel Y (2011). neuRosim: An R package for generating fMRI data. Journal of statistical software, 44(10), 1-18.

Corresponding author

Joke Durnez, Joke.Durnez@UGent.be
Ghent University, Department of Data Analysis, H. Dunantlaan 1, 9000 Ghent