Extended Methods

Daniel Kessler, Michael Angstadt, Chandra Sripada

August 29, 2016

1   Data & Code

This document, along with source code and data for all analyses, is available at `http://github.com/mangstad/FDR_permutations`. Anders Eklund graciously shared with us the per-participant "first-level" maps as well as group-level results for each contrast analyzed as part of their study [5]. These contrasts were in turn derived from two studies [4, 8], which were available at openfMRI [7].

2   Overview

We consider each contrast from each study separately and perform the following steps. In order to generate a realization of the null, we randomly flip the sign of each participant's contrast map, conduct a voxelwise, one-sample t-test across participants, apply a cluster defining threshold (CDT) of {.001,.01}, and then record the voxel extents of all clusters that emerge (identified as connected components using the FSL-standard 26-degree connectivity criteria that considers voxels connected if they share any face, edge, or vertex). We repeat this procedure 5,000 times to obtain an empirical mass function of cluster sizes under the null. To obtain a per-cluster uncorrected p-value, we compare the extent for each cluster that arises from the unflipped design to the contrast's null distribution. Finally, we submit the vector of uncorrected p-values to an FDR procedure [1] with $\alpha = .05$.

3   Cluster Extent Test Statistic and Procedure

We desire to assess the statistical significance of each of our clusters. To accomplish this, we select any one of these clusters (at random), and consider the distribution of its corresponding test statistic, Z:

$$Z = \begin{cases} \text{Extent of randomly selected cluster,} & nCluster \geq 1 \\ 0, & \text{else} \end{cases} \quad (1)$$

Let N be the number of participants under consideration and I, J, K represent the number of voxels in the three dimensions. In order to obtain Z,

we require the following: X, a collection of N three-dimensional statistical parameter maps, one per participant; Y, a vector of signs/labelings/covariates corresponding to each participant (because all studies considered here use a one-sample testing framework, all elements of Y are 1); and CDT, a cluster defining threshold (a p-value).

We can repeat the below procedure multiple times for the observed labelings until Z corresponds to each of our clusters.

Algorithm 1 Return Extent of Arbitrary Cluster

Require: $X \in \mathbb{R}^{I \times J \times K \times N}$
Require: $Y \in \mathbb{R}^N$; In current setting, $Y \in \{-1, +1\}^N$
Require: $0 < CDT < 1$
 1: **for** Each voxel **do**
 2:    Fit a GLM, $X_{i,j,k} \sim Y$, return p-value
 3:    **if** $p < CDT$ **then**
 4:       **return** p
 5:    **else**
 6:       **return** 0
 7:    **end if**
 8: **end for**
 9: Clusterize, with 26-degree connectivity, thresholded map
10: **if** $nClusters \geq 1$ **then**
11:    **return** Extent of randomly selected cluster
12: **else**
13:    **return** 0
14: **end if**

---

4   Permutation Strategy

Under the null hypothesis, the values of Y are independent of Z for our data $X = x$. We can thus realize the null distribution by repeatedly altering Y and observing a new Z. Ultimately, we arrive at p-values by comparing the z obtained from the true labelings y with the distribution of Z obtained from altered versions of Y. In the one-sample t-test case (the one studied here), all Y are 1, so we alter Y by flipping the sign of each element at random. An exact test would visit all 2^N possible flippings of the signs of the elements of Y, but for computational reasons, we approximate this by randomly visiting a single possible flipping, and repeating this process 5,000 times.

5   Permutation Aggregation & PMF

We are ultimately interested in $P(Z > z | X = x)$. Importantly, $Var\big[Z | X = x, Y = y\big] > 0$ when $nCluster > 1$ and there are 2 or more unique cluster extents because Z is a randomly selected from among $nCluster$ options. While we could

use repeated realizations of Z for a given x, y to approximate $P(Z|X = x, Y = y)$ we can, instead, directly obtain this distribution as the probability mass function (i.e., the normalized histogram of cluster extents) for a given Y. This enables us to visit each Y only once. Because visiting any Y is equiprobable under our sampling scheme, $P(Z|X = x)$ is simply the average of $P(Z|X = x, Y = y)$ for all y visited.

6 Clusterwise, Uncorrected p-Values

We begin with the extents of all clusters observed under the true labels (not only those previously reported which were restricted to those significant at $\alpha_{\text{RFT-FWE}} < .05$). For each cluster, we assign the observed extent to z and calculate $P(Z > z|X = x)$ (using the null distribution arrived at above) and retain this value as the uncorrected p-value for that cluster.

7 Multiple Testing Correction: FDR

While the uncorrected p-values from above were obtained using a nonparametric rather than parametric procedure, we follow the approach from Chumbley and Friston [3]: We submit the uncorrected p-values as a vector to the FDR controlling procedure established in Bejamini and Hochberg [1], with $\alpha_{FDR} = .05$, retain the q_{FDR}-value, and record whether the null hypothesis ought to be rejected (significant value) or not. As indicated above, we are certainly not the first to apply FDR to cluster properties; cf. [3, 6, 2].

8 Comparing $p_{\text{RFT-FWE}}$ with q_{FDR}

In Figure 1, we plot $p_{\text{RFT-FWE}}$ vs q_{FDR} on a $-log_{10}$ scale for all clusters significant at $\alpha_{\text{RFT-FWE}} = .05$. We emphasize (with shading) those clusters with corresponding $q_{\text{FDR}} \leq .05$, i.e., the clusters that remain significant under $\alpha_{\text{FDR}} = .05$.

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, January 1995.

[2] J. Chumbley, K. Worsley, G. Flandin, and K. Friston. Topological FDR for neuroimaging. *NeuroImage*, 49(4):3057–3064, February 2010.

[3] Justin R. Chumbley and Karl J. Friston. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage*, 44(1):62–70, January 2009.

[4] Keith J. Duncan, Chotiga Pattamadilok, Iris Knierim, and Joseph T. Devlin. Consistency and variability in functional localisers. *NeuroImage*, 46(4):1018–1026, July 2009.

[5] Anders Eklund, Thomas E. Nichols, and Hans Knutsson. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, page 201602413, June 2016.

[6] Ruth Heller, Damian Stanley, Daniel Yekutieli, Nava Rubin, and Yoav Benjamini. Cluster-based analysis of FMRI data. *NeuroImage*, 33(2):599–608, November 2006.

[7] Russell A. Poldrack, Deanna M. Barch, Jason Mitchell, Tor Wager, Anthony D. Wagner, Joseph T. Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham. Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in Neuroinformatics*, 7:12, 2013.

[8] Sabrina M. Tom, Craig R. Fox, Christopher Trepel, and Russell A. Poldrack. The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, 315(5811):515–518, January 2007.