Cluster Analysis Revisited - Again:
Implementing Nonstationary Cluster Size Inference

FMRIB Technical Report TR08TN1

Thomas E. Nichols[1,2]

[1]GlaxoSmithKline Clinical Imaging Centre, UK
[2]FMRIB (Oxford University Centre for Functional MRI of the Brain), Dept. Clinical Neurology, University of Oxford

Abstract

We review cluster inference results for nonstationary random fields, extending Flitney & Jenkinsons's "Cluster Analysis Revisited" [TF00DF1] to account for spatially-varying smoothness.

Keywords

Clustering, random field theory, inference.

1 Introduction

The purpose of this work is express Random Field Theory (RFT) nonstationary cluster size inference results (Hayasaka *et al.*, 2004) in a notation consistent with (Flitney & Jenkinson, 2000), so as to support their implementation in FSL's randomise program. As our only intention is to derive cluster size measures that are adjusted for local smoothness, and to use permutation to obtain nonparametric P-values, we do not review the non-stationary RFT results to find P-values. For such details the interested reader is referred to (Hayasaka *et al.*, 2004).

2 Theory

While we are not interested in parametric P-values, the non-stationary cluster size statistic is difficult to understand without some basic parametric results, and so we begin with review of the random field theory for stationary cluster size (mainly excerpted from (Flitney & Jenkinson, 2000)).

2.1 Cluster Size Inference Under Stationarity

For an image with V voxels and a cluster-defining threshold u, denote the (random) number of observed clusters m. For a randomly selected cluster, let the cluster size in voxels be n.

The mean of the cluster size is

$$E\{n\} = E\{N\}/E\{m\}. \tag{1}$$

$E\{N\}$ is trivially calculated as number of voxels times the null tail probability of the cluster defining threshold:

$$E\{N\} = V(1 - \Phi(u)) \tag{2}$$

where $\Phi(u)$ is the cumulative distribution function of a standard normal.

$E\{m\}$ is approximated with the expected Euler characteristic, which (for a $D = 3$ dimensional image) is

$$E\{m\} \approx R_3 (4 \ln 2)^{3/2} (2\pi)^{-2} e^{u^2/2} (u^2 - 1) \tag{3}$$

where R_3 is the Resel count[1],

$$R_3 = V \ |\Lambda|^{1/2} (4 \ln 2)^{-3/2}, \tag{4}$$

and Λ is the covariance matrix of the partial derivatives of the random field; Λ can be thought of as a roughness matrix, as larger values along the diagonal correspond to a rougher random field.

Since it is almost impossible to have any intuition about the quantity $|\Lambda|^{1/2}$, you can re-express the Resel count R_3 in terms of σ or FWHM,

$$\begin{aligned} R_3 &= V \frac{1}{\sigma_x \sigma_y \sigma_z} (8 \ln 2)^{3/2} \\ &= V \frac{1}{\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z} \end{aligned} \tag{5}$$

where σ or FWHM describe the size of a Gaussian kernel needed to smooth white noise into a field with roughness matrix Λ.

The null distribution of $n^{2/D}$ is approximated with an exponential distribution (with mean $1/\beta$, (Flitney & Jenkinson, 2000) eq. (6); see also Appendix 1 of (Hayasaka & Nichols, 2003) for more details) and a correction for multiple comparisons is also applied ((Flitney & Jenkinson, 2000) eq. (9), and (Hayasaka & Nichols, 2003) eq. (9)).

2.2 Cluster Size Inference Under Non-Stationarity

While the standard theory assumes that the smoothness is constant over the entire image, the nonstationary methods allow smoothness to vary spatially. We first motivate the method with vague notation, and then give precise results for implementation.

Various schemes for accounting for variable smoothness could be considered, e.g. by dividing n by some local measure of smoothness averaged over the cluster, as in

$$\frac{n}{\text{Avg}\{\text{SMOOTHNESS}\}} \tag{6}$$

where Avg{SMOOTHNESS} is *some* measure of smoothness averaged over the n voxels in the cluster. The specific method proposed by (Worsley *et al.*, 1999) instead adds up the contribution of each voxel to the cluster, with each voxel adjusted for local smoothness, as in

$$\sum_{i \in C} \frac{1}{\text{SMOOTHNESS}_i} \tag{7}$$

where C is the list of n voxels in the cluster.

Specifically, Worsley defines 1/SMOOTHNESS as 'Resels per voxel' or RPV. First note that the total Resel count is $V/(\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)$, and hence, under stationarity, the contribution from each of the V voxels to the total is $1/(\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)$; i.e. in this case the "resels-per-voxel" is just $1/(\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)$. In terms of σ (c.f. eq. (5)) for the stationary case

$$\text{RPV} = (\sigma_x \sigma_y \sigma_z)^{-1} (8 \ln 2)^{-3/2}. \tag{8}$$

[1]Beware that FSL's `smoothest` program reports `RESELS`, which is actually the the size of one Resel, i.e. volume of space with dimensions $\text{FWHM}_x, \text{FWHM}_y, \text{FWHM}_z$, i.e. from `smoothest`, `RESELS` $= \text{FWHM}_x \text{FWHM}_y \text{FWHM}_z$. In SPM, "RESEL" generically refers to the Resel count, R_3. And, for completeness, note that `smoothest`'s `DHL` $= |\Lambda|^{1/2}$, "Determinant Half-Power Lambda."

Under nonstationarity, then, we seek a voxel-wise estimate of RPV

$$\text{RPV}_i = (\sigma_{ix}\sigma_{iy}\sigma_{iz})^{-1}(8\ln 2)^{-3/2}. \tag{9}$$
$$\tag{10}$$

FSL's estimate of σ is given by [(Flitney & Jenkinson, 2000), eq. (21)],

$$\hat{\sigma} = \sqrt{\frac{-1}{4\ln\left(\frac{SS_-}{S^2}\right)}} \tag{11}$$
$$= \left(4\ln\left(\frac{S^2}{SS_-}\right)\right)^{-1/2}. \tag{12}$$

where SS_- is the correlation of two neighbouring voxels and S^2 is the sample variance of a voxel. For a specific direction, say x, at voxel i, define the smoothness estimate

$$\hat{\sigma}_{ix} = \left(4\ln\left(\frac{S_i^2}{SS_{ix-}}\right)\right)^{-1/2} \tag{13}$$

where S_i^2 is the sample variance at voxel i, and SS_{ix-} is the sample correlation between voxel i and its neighbor in the x direction.

The voxel-wise RPV can be calculated as the product of RPV's, one for each direction

$$\widehat{\text{RPV}}_i = \widehat{\text{RPV}}_{ix}\widehat{\text{RPV}}_{iy}\widehat{\text{RPV}}_{iz} \tag{14}$$

where

$$\widehat{\text{RPV}}_{xi} = \left(4\ln\left(\frac{S_i^2}{SS_{ix-}}\right)\right)^{1/2}(8\ln 2)^{-1/2}. \tag{15}$$

Thus armed with a voxel-wise map of RPV, the smoothness adjusted cluster size (i.e. cluster size measured in Resels) can be obtained with

$$r = \sum_{i\in C}\widehat{\text{RPV}}_i. \tag{16}$$

2.3 P-values

Once r is computed for a cluster, it can be treated as a usual cluster size measure *in a resampling* framework. The only possible sublty is that the RPV image *must* be recomputed for each permutation, as there is considerable uncertainty in $\widehat{\text{RPV}}$ which must be accounted for over permutations.

The random field theory P-values for r are quite involved as they must account for the uncertainty that RPV introduces into the cluster statistic. For full details see (Hayasaka *et al.*, 2004).

3 Implementation Details

To summarize the nonstationary cluster size inference:

1. Cluster "size" is measured not in voxels, but in Resels. And instead of computing a single Resel measure, RPV is computed for each permutation, at every voxel, using equations (14) and (15).

   Note that the current code which does smoothness estimation is `cluster`'s `smoothest.cc` function, and that the variable names there directly correspond to S^2 and SS_- (but currently are pooled over all in-mask, non-edge voxels).

2. At edges of the mask, $\widehat{\text{RPV}}_i$ may not be able to be computed since the differences are not available. However, to reduce the number of voxels with no RPV data, the following scheme can be used when at least one of the x, y or z directions are available:

   If only 1 direction is available (say x), estimate $\widehat{\text{RPV}}_i$ as

$$\widehat{\text{RPV}}_i = (\widehat{\text{RPV}}_{xi})^3 \tag{17}$$

   If 2 directions are available (say x & y), estimate $\widehat{\text{RPV}}_i$ as

$$\widehat{\text{RPV}}_i = (\widehat{\text{RPV}}_{xi}\widehat{\text{RPV}}_{yi})^{3/2} \tag{18}$$

3. When summing the RPV within a cluster, some values may be missing, as just mentioned. The appropriate action is to just impute the missing values with the mean of the other RPV values. Computationally, this can be done by scaling the sum: If n is the number of voxels in the cluster, C_r is the set of n_r voxels with non-missing RPV, the cluster resel size is

$$r = \frac{n}{n_r} \sum_{i \in C_r} \widehat{\text{RPV}}_i. \tag{19}$$

4. If a cluster is on the edge of the mask, it may occur that there are *no* non-missing RPV values available. In such cases, a fall-back strategy is to simply assume the global average RPV, and compute the cluster resel size as

$$r = n\frac{1}{V} \sum_{i} \text{RPV}_i. \tag{20}$$

References

Flitney, D.E., & Jenkinson, M. 2000. *Cluster Analysis Revisited*. Tech. rept. Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK. TR00DF1.

Hayasaka, S., & Nichols, T. E. 2003. Validating Cluster Size Inference: Random Field and Permutation Methods. *NeuroImage*, **20**(4), 2343–2356.

Hayasaka, Satoru, Phan, K. Luan, Liberzon, Israel, Worsley, Keith J., & Nichols, Thomas E. 2004. Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage*, **22**(2), 676–687.

Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., & Evans, A.C. 1999. Detecting changes in nonisotropic images. *Human Brain Mapping*, **8**, 98–101.