A semi-parametric approach to estimate *p*-values for activation in fMRI correcting for multiple testing and low frequency processes

R. R. Nandy[1], D. Cordes[1]

[1]Radiology, University of Washington, Seattle, WA, United States

Introduction

An important consideration in constructing the activation map for an fMRI dataset is to choose the right threshold. In a parametric approach, typically a pre-assigned *p*-value is chosen as the threshold. However, in fMRI, it is necessary to make a few corrections to obtain an adjusted *p*-value. First, the voxel timecourses usually have high temporal autocorrelation and it is necessary to estimate the temporal autocorrelation structure to obtain the *p*-value. SPM2 uses an ReML approach to address this problem [1]. The second problem is known as the multiple comparison problem, which arises from the fact that we are testing the same hypothesis at tens of thousands of voxels and the probability of at least one voxel being detected to be active incorrectly is much larger than the single-voxel *p*-value. The simplest way to address this problem is to use the Bonferroni correction, but this approach is too conservative considering the spatial dependence in the fMRI data. A more popular method which utilizes the spatial smoothness of the data is based on the theory of Random Fields (RFT) and is implemented in SPM2.

However, both of these corrections have certain weaknesses. It has been established that even resting-state data has some low frequency behaviors [2] and there is a reasonable chance that there is some overlap between the inherent low frequencies in the brain and the paradigm frequency. Unfortunately, even a standard high-pass filter cannot eliminate the low frequencies matching the paradigm since the cutoff is half the frequency of the actual paradigm. As a result, it is not unusual to have decent "activation maps" with resting-state data when SPM2 is implemented. Regarding the RFT approach to the multiple comparison problem, the problem is that it is imperative to use spatial smoothing to implement the RFT. Furthermore, even when spatial filter is used, the estimates using RFT is often no better than Bonferroni.

We extend a previously proposed semi-parametric method using order statistics on resting-state data to overcome the problems mentioned above [3]. The proposed method will have two advantages over conventional approaches; (i) since resting-state data is used as null (which already has the appropriate temporal autocorrelation), the correction for temporal autocorrelation becomes redundant and (ii) it indeed offers a threshold better than Bonferroni even without any spatial filter.

Theory and methods

<u>*Adjusting for the low frequency effects*</u> – Since resting-state data is used as null, the low frequency effects are already there. However, there may be a phase mismatch and hence the null distribution will depend on the phase of the paradigm. To overcome this problem, we will estimate the null distribution for all possible phases and take an average, since all phases are equally likely when the activation paradigm begins.

<u>*Adjusting for multiple comparison*</u> – To adjust for the multiple comparison, it is necessary to estimate the distribution of the maximum value of the statistic used in the analysis over all voxels [3]. Define $d_i = i(X^{(i)} - X^{(i+1)}), i = 1,..., k$ as the normalized sample spacings for the $k+1$ largest order statistics $X^{(i)}$ from the resting-state data. For each X_i , let Z_i be the corresponding *p*-value using the parametric distribution. If the parametric distribution is accurate, the function $-\log(Z_i)$ is exponentially distributed since Z_i is uniformly distributed. The transformation to an exponential random variable is necessary to establish the validity of resampling, but a detailed theoretical discussion is beyond the scope of this abstract. In the following, without loss of generality, we can assume that $X_1,..., X_N$ are already transformed using the log transform. It can be shown that it is valid to resample the normalized spacings for exponential random variables using bootstrap or permutation and for each resample of normalized spacings, we can reconstruct a resample for the maximum statistic. Unlike commonly used resampling techniques in fMRI, our method is extremely fast and easy to generate a large number of resamples and obtain the distribution of the maximum statistic.

Results

We first demonstrate the inadequacy of temporal autocorrelation correction to obtain the true null distribution. We acquired a whole brain resting-state data (145 time frames, TR=2 sec) and implemented SPM2 (corrected for AR) to obtain the t-map for the whole brain using a canonical HRF convolved boxcar function (period of length 30 timeframes). In Figure 1 (left), we plotted the distribution of the t-map over the whole brain (solid line) and the actual parametric t-distribution (dotted line). It is obvious that for this null dataset, even the corrected t-map has a much heavier tail at the right and even at an FWE adjusted *p*-value threshold of 0.05, SPM2 detected more than 50 voxels to be active. In Figure 1 (middle), we have plotted the tails of the empirical distributions for the same resting-state data, but with varying phases for the convolved boxcar function which demonstrates the importance of averaging over phases to obtain the correct null distribution. Finally, in Figure 1 (right), we have plotted estimated adjusted correlation threshold at *p*-value 0.05 using the proposed semi-parametric approach using the resting-state dataset. Clearly, the estimate depends on the number of normalized spacings used in the estimate, but the dependence is not strong. We have also plotted the corresponding Bonferroni estimates by fitting the tail of null resting-state data with a smooth function and a parametric threshold using exponential properties of the normalized spacings (not discussed here).

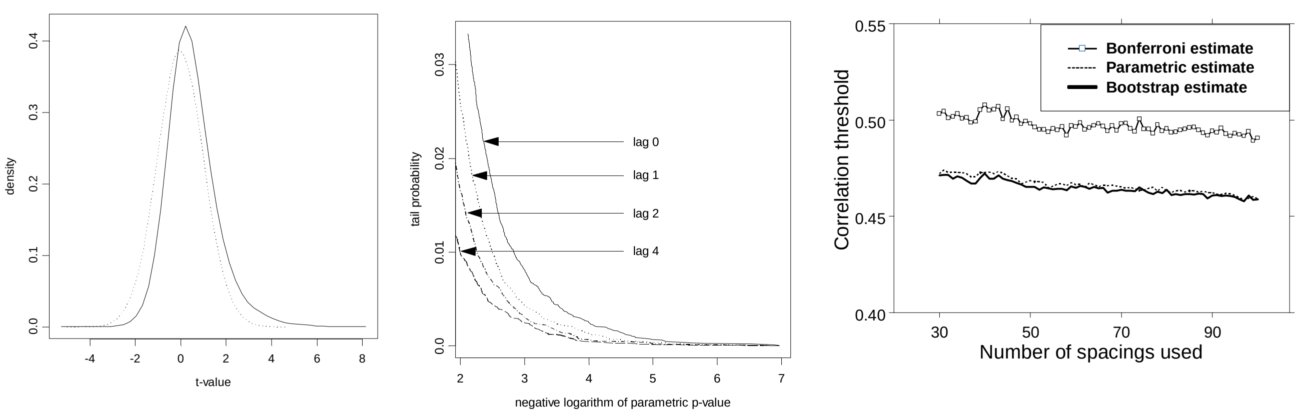

Figure 1.

References

1. Friston, K. *Neuroimage*. 16:484-512 (2002).
2. Cordes, D et. al. *Magn. Reson. Imaging.* 20:305-317 (2002).
3. Nandy, R and Cordes, D. *ISMRM Proceedings.* 1089 (2004).