CORRELATION-ADJUSTED *t*-SCORES IN APPLICATION TO FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA

Verena Zuber and Korbinian Strimmer

Institute for Medical Informatics, Statistics and Epidemiology (IMISE),
University of Leipzig, Härtelstr. 16–18, D-04107 Leipzig, Germany

ABSTRACT

The correlation-adjusted t-score (cat score) is a modification of the Student t-statistic to account for dependencies among variables. Recently, we have shown (1) that the cat score improves ranking of genes to detect differential expression in the presence of correlation. Noting the similarity of structure between high-dimensional gene expression and image analysis data, we apply the cat score using shrinkage estimation to functional magnetic resonance imaging data. We show that the cat score is a simple, yet effective, means to accommodate correlation among voxels and to improve standard t-type tests of neural activation.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a noninvasive, indirect imaging technique that allows to monitor brain-activity. Using the difference in magnetic susceptibility of oxygenated and deoxygenated blood MRI measures the so-called "blood oxygen dependent" (BOLD) signal, which is closely connected to neuronal activity. Thus, the neuronal signals of interest are only observed indirectly – see (2) for further information. Functional magnetic resonance imaging (fMRI) provides not only a snapshot of neural activity, but a whole time series to describe dynamics of neural activation. With modern techniques the BOLD-Signal can be measured accurately with high spatial resolution in 3d-cubes of 3mm^3 volume and low temporal resolution (sampling rate: about 0.5 Hz); this 3d-extension of the pixel is called voxel.

The two main concepts of brain architecture are *functional specialization* and *functional integration* (3). Functional integration posits that there are neurons, possible over larger distances, working together as a neural network to perform certain tasks. Thus, functional integrationist approaches focus on voxel to voxel interaction. In contrast, the functional specialization paradigm is based on the assumption that these networks are spatially fixed and that it is possible to map brain-areas responsible for certain tasks, e.g., emotions, vision, or hearing. Here, the focus is on constructing study-designs in which contrasts of neural activation under conditions of stimulus and rest are investigated.

There are two main challenges in analyzing fMRI data. First, the signal to noise ratio is very low (4). Second, the dimensionality of the data is in general much higher than the available sample size. For example, the number of voxels may easily exceed $p = 100,000$, but there is only a limited set of measurements (around $t = 100$).

This points directly towards analogies between the data structure of fMRI and gene expression studies. Data from high-throughput microarray studies are also characterized by large dimensions and small sample size. Specifically, there are many thousands of genes and due to cost and tissue availability the number of gene chips is restricted. Hence, the so-called "large p, small n" set-up is common to both microarray and fMRI data. Moreover, like gene expression data, fMRI data also exhibit a rich correlation structure among both variables as well as among measurements. In contrast to gene expression data, fMRI data contain correlation due to the spatial structure. In addition to these short range dependencies, fMRI data include signal-specific dependencies on long distances, too.

In this note, our aim is to exploit this connection to derive a new tool for fMRI image comparisons, based on a similar method that we have originally developed for gene expression studies, where its favorable performance has already been demonstrated (1). In the following we first give a short introduction to the basic analysis of fMRI data with focus on the parametric approach for two-group comparisons. Subsequently, we present an extension of the t-score under spatial correlation, the so-called correlation-adjusted t-score, defined in section 3. In the final section 4 we apply the cat score to real imaging data, and discuss its merits.

2. ANALYSIS OF FMRI DATA

There are two contrary approaches to analyze fMRI data depending on the task at hand. Integrationist approaches focus on the analysis of correlation and dependency structures. A simple tool is correlation analysis, where a seed voxel is selected and then the (temporal) correlation to every of the other voxel is computed. This tool depends strongly on the a priori choice of the seed voxel, but can provide useful and easy to interpret information.

A related, but more refined approach is independent component analysis, where the data are decomposed into independent rather than only uncorrelated components. The key assumption here is that the observed data are a mixture of the intrinsic signals of neural networks and different noise signals, mechanical and physical. Using the independence between the different neural networks, as well as the independence to and between the sources of noise, it is possible to recover the hidden signals from the observed mixture. For further information, see (4).

In contrast, functional specialization focuses primarily on parametric methods based on regression approaches such as generalized linear models, where the time-series $\boldsymbol{Y}_i$ of voxel $i \in \{1, ..., p\}$ is modeled by

$$\boldsymbol{Y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{X}$ is the design matrix given by the experiment and $\boldsymbol{\epsilon}_i$ is a voxel-specific, well behaved error. The most popular approach is statistical parametric mapping (SPM), where first each voxel is analyzed separately, resulting in univariate statistics, such as the t-score or z-score, that are presented in 3d

maps. Spatial correlation is only considered in the inference of the given maps and in the choice of interesting or activated voxels.

The comparison of images belonging to different classes is a common task in fMRI studies. For one, there are experiments, where the proband undergoes a given experimental set-up of alternating stimulus and rest sequences. This so-called block design is a simple way of localizing areas of neural activation triggered by a given task. Furthermore, proper studies usually do not rely on only one individual, but include a set of probands possibly in differing groups of interest. Therefore, after the analysis of individual data, a second-level of analysis might be needed, either to compare groups or to test hypotheses.

In addition to case-control studies, imaging data may also be used for classification and prediction. For instance, images with a given class-label are analyzed to extract regions that contain valuable information to discriminate among the groups of interest. In an early example by DeCarli et al. (5) discriminant analysis was used to determine the presence of dementia of the Alzheimer type. More recently, LaConte et al. (6) proposed support vector machines for temporal classification of block design fMRI data.

3. THE CORRELATION-ADJUSTED t-SCORE (CAT SCORE)

Since fMRI data are characterized by a rich spatial correlation structure, it is worthwhile to investigate methods that include knowledge of the correlation among voxels. In (1) we use the intrinsic link between discriminant analysis in the two group case and two-group comparisons via t-scores to derive a new statistic to measure group differences, the so-called correlation-adjusted t-score (abbreviated: cat score). Due to the multivariate nature of our approach the cat score is defined as a vector $\boldsymbol{t}_{\text{cat}}$ of length p

$$\boldsymbol{t}_{\text{cat}} = \boldsymbol{P}^{-1/2}\,\boldsymbol{t}_{\text{stud}}$$

with the ordinary Student t-score vector

$$\boldsymbol{t}_{\text{stud}} = \left((\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{V}\right)^{-1/2}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2),$$

the vector of variances

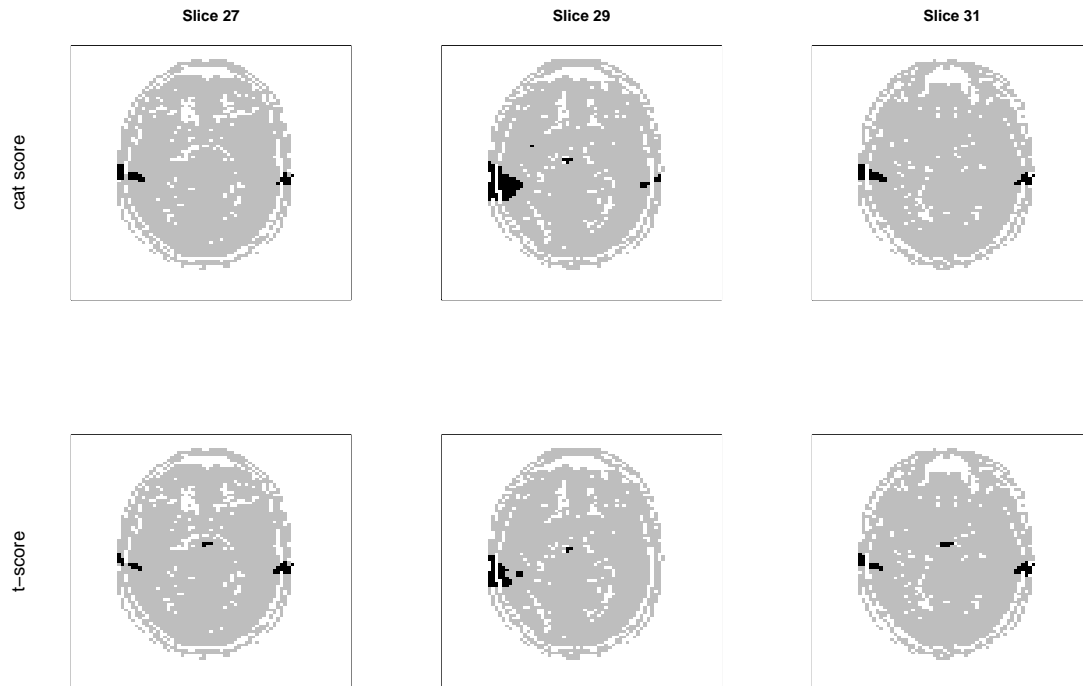$$\boldsymbol{V} = \text{diag}(\hat{\sigma}_1^2, ..., \hat{\sigma}_p^2),$$

Figure 1. Comparison of the cat score (top) and the t-score (bottom) on three exemplary slices on the auditory paradigm data. Both scores were estimated via shrinkage procedures using the package "st" in the R language. Thresholds for activation were chosen using the false discovery rate criterion.

and the (spatial) correlation matrix P of dimension $p \times p$. For one particular voxel $i \in \{1, ..., p\}$ the cat score $t_{\text{cat}}(i)$ is given as the ith element of the cat score vector.

In contrast to the univariate t-score for voxel i the cat score $t_{\text{cat}}(i)$ is a multivariate measure, given as weighted mean over the t-scores of all p voxels. The weights are assigned according to the square root of the inverse correlation matrix, which in turn is linked to partial correlation. Thus, for a given voxel i tightly connected voxels contribute more to $t_{\text{cat}}(i)$ than uncorrelated voxel. If there is no correlation present, the cat score reduces to the ordinary t-score. Another perspective on the cat score is that it performs prewhitening in the spatial domain of fMRI data. Hence, the cat score is the standardized decorrelated mean difference and thus estimates the contribution of each voxel in discriminating between the two conditions, after removing the effect of all other voxels.

In practice, to compute the cat score V and P need to be estimated from data. Due to the "large p, small n" data structure we propose to use regularized estimation procedures, like shrinkage. For further information on the cat score and a specific shrinkage estimation procedure we refer to (1).

4. ANALYSIS OF FMRI DATA FROM AN AUDITORY PARADIGM

As an illustrating example we investigate a standard block design to detect differences in neural activation between rest and auditory stimulation. In particular, our aim is to evaluate the performance of the cat score in comparison to the t-score.

Auditory stimulation was presented in bisyllabic words to one proband simultaneously on both ears at a rate of 60 words per minute. The whole brain fMRI images, each including $64 \times 64 \times 64$ voxels of 27mm^3 volume, for one proband were collected by Geriant Rees under the direction of

Karl Friston and the FIL methods group at a modified 2T Siemens MAGNETOM Vision system with a scan repetition time of 7 seconds.

After preprocessing, including realignment, spatial normalization and spatial smoothing, one image consisted of $79 \times 95 \times 68 = 510,340$ voxels, each of volume 8mm^3. Subsequently, using structure information we discerned the relevant voxels inside the brain from voxels outside the brain. Altogether the single-subject data consist of each 7 blocks of rest and stimulus with 6 measurements in each block. Since we focus on the comparison between stimulus and rest-sequences we reduced the data set to only one measurement for each block (note this thinning also is simple mean to reduce temporal autocorrelation). Considering haemodynamic properties of neural activation we selected the second measurement in each block. Hence, for further analysis we arrived at $n = 7 + 7 = 14$ measurements of 68 slices, each containing 7,505 voxels.

We performed the analysis using the R language for statistical analysis, in which we implemented shrinkage estimators of the cat and the t-scores. Computation was performed for each slice separately excluding voxel that describe space outside the brain. Significant voxels are chosen using a false discovery rate cutoff, as described in (7). Figure 1 shows the results for the exemplary slices 27, 29 and 31. Voxels that show significant differences are indicated in black, the rest of the brain is given in gray. Thus, the black voxels represent those parts of the brain where neural activation is triggered by the auditory stimulus. Most of the significant voxels are found by the cat score and the t-score alike in symmetric regions, near the ears where the auditory cortex is located (Brodman area 22, 41, 42). Nonetheless, there are small, but notable differences between the voxels declared significant. It is the cat score that gives centers of activation with a better structure: there are larger coherent clusters of activation and less single significant voxels.

In a previous study (1) we have shown that for simulated gene expression data, where the "ground truth" was known, the genes declared significant by the cat score were on average more correct, with higher power for fixed false discovery rate, and with lower false discovery rates for fixed power.

Therefore we conclude for the present data that the single voxel activations found by the t-score, but not by the cat score are likely to be false positives. In turn, this implies that taking into account of correlation in the fashion of cat scores may be advantageous not only for gene expression, but also for fMRI data.

APPENDIX

The "shrinkage cat" estimator is contained in the package "st" which is freely available under the terms of the GNU General Public License from CRAN (`http://cran.r-project.org/`).

References

[1] V. Zuber and K. Strimmer, "Gene ranking and biomarker discovery under correlation," *arXiv*, vol. stat.AP, pp. 0902.0751, 2009.

[2] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, pp. 869–878, 2008.

[3] J. D. Cohen and F. Tong, "Neuroscience. the face of controversy," *Science*, vol. 293, pp. 2405–2407, 2001.

[4] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[5] C. DeCarli, D. G. M. Murphy, A. R. McIntosh, D. Teichberg, M. B. Schapiro, and B. Horwitz, "Discriminant analysis of MRI measures as a method to determine the presence of dementia of the alzheimer type," *Psychiatry Research*, vol. 57, pp. 119–130, 1995.

[6] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu, "Support vector machines for temporal classification of block design fMRI data," *Neuroimage*, vol. 26, pp. 317–329, 2005.

[7] K. Strimmer, "A unified approach to false discovery rate estimation," *BMC Bioinformatics*, vol. 9, pp. 303, 2008.