# New Blockwise Permutation Tests Preserving Exchangeability in Functional Neuroimaging

Chunxiao Zhou and Yongmei Michelle Wang

*Abstract*— In this paper, we present a new blockwise permutation test approach based on the moments of the test statistic. The method is of importance to functional neuroimaging studies. In order to preserve the exchangeability condition required in permutation, we divide the time series into certain exchangeability blocks. In addition, efficient moments-based permutation tests are performed by approximating the permutation distribution of the test statistic with the Pearson distribution series. This involves the calculation of the first four moments of the permutation distribution within each block and then over the whole time series. The accuracy and efficiency of the proposed method are demonstrated using both simulated time series and fMRI data.

## I. INTRODUCTION

HYPOTHESIS Testing has been widely used in functional neuroimaging data analysis, such as brain activation detection and inference, functional integration and connectivity [1] - [5]. Traditionally, researchers perform the statistical analysis by using parametric hypothesis testing, including commonly used $F$ test, $T$ test, $Z$ test and Hotelling's $T^2$ test [1], [2]. In general, a parametric method models the distribution of a test statistic with a parametric form which is mathematically tractable. The parametric methods thus work well when data is independent and normally distributed. However, in neuroimaging studies, the data distribution is usually unknown. It is also very expensive to collect large amount of data to satisfy the normal distribution assumption for large sample size independent data. Furthermore, sometimes, a desirable test statistic could be mathematically intractable. Nonparametric hypothesis testing methods are preferable in these cases.

In order to deal with small sample size neuroimaging data with unknown distribution, Holmes et al. [3] introduced nonparametric permutation tests. Permutation tests construct the distribution of a test statistic by resampling data without replacement. They are flexible and distribution-free. The key and only assumption for permutation tests is data exchangeability. In the two-sample hypothesis testing case, data exchangeability means the distributions of two group data are identical under the null hypothesis [4], [6]. We can then randomly permute $n_a$ data to one group and the rest $n_b$

Chunxiao Zhou is with the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, IL 61820, USA. (e-mail: czhou4@illinois.edu).

Yongmei Michelle Wang is with the Statistics Department, Psychology Department, and Bioengineering Department, University of Illinois at Urbana-Champaign, IL 61820, USA. (e-mail: ymw@illinois.edu).

data to the other group. Here, $n_a$ and $n_b$ are the sample sizes of the two groups. As a result, the empirical distribution of a test statistic is constructed using test statistic values for all possible permutations. The original observation can be considered as one of all possible permutation setups. To measure how strongly the observed data support the null hypothesis, we calculate the $p$-value by dividing the frequency of permutations having more extreme test statistic value by the number of all permutations. The statistical decision is made based on whether the $p$-value is less than a pre-chosen significance level. We reject the null hypothesis if the $p$-value is smaller than the pre-chosen significance level since it is unlikely to occur under the null hypothesis.

In real applications, the data exchangeability condition is not always valid. Although permutation tests still work when the exchangeability assumption is slightly violated, it is important to preserve data exchangeability to a reasonable level [4], [6]. In functional neuroimaging data analysis, the main effect (i.e. the effect of interest or the effect to be tested) is often confounded with the undesirable temporal artifacts such as those caused by scanner drift, head motion, and cardio-respiratory effects. Such artifacts can be reduced in the data preprocessing but are unlikely to be taken away completely. Thus, the global data exchangeability does not usually hold. Nichols and Holmes [4] proposed a restrictive permutation scheme by segmenting the entire set of a time series into certain blocks. Since the time duration is relatively short within each block, the data exchangeability approximately holds within a block. The blockwise permutation tests only allow permutations within each block to preserve data exchangeability and prohibit any permutation across blocks.

Another critical issue involved in permutation tests is the computational complexity. There are three common approaches to construct the permutation distribution [6], [7], [8]: (1) exact permutation enumerating all possible arrangements; (2) approximate permutation based on random sampling from all possible permutations; (3) approximate permutation using the analytical moments of the exact permutation distribution under the null hypothesis. The main disadvantage of the exact permutation is the computational cost, due to the factorial increase in the number of permutations with the increasing number of subjects. The second technique often gives inflated type I errors caused by random sampling. When a large number of repeated tests are needed, the random permutation strategy is also

computationally expensive to achieve satisfactory accuracy. Regarding the third approach, the exact permutation distribution may not have moments or moments with tractability. In most applications, it is not the existence but the derivation of moments that limits the third approach. Recently, we have proposed an efficient recursive algorithm to calculate the moments of the permutation distribution by a simple sumproduct of data partition sums and index partition sums [9]. The data partition sums and index partition sums are computed recursively, from the simplest sum to the most complex sum. For the first four moments, the computation can be done in the first order or third order polynomial time for univariate and multivariate test statistics, respectively. Given the first four moments, the permutation distribution can be well fitted by the Pearson distribution series [10].

In this paper, we develop a new moments-based blockwise permutation test method. We first divide the whole scans into certain blocks. For each block, we apply our recursive algorithm to obtain the first four moments. The first four moments of the entire set of scans are computed by combining the first four moments from all blocks through an efficient representation. With this efficient moments-based blockwise permutation tests scheme, we maintain the flexibility of permutation tests, preserve the exchangeability condition, and reduce the computation cost dramatically.

## II. METHODOLOGY

### A. Blockwise Permutation Tests

Let us consider a single subject activation experiment based on functional magnetic imaging (fMRI). The approach is applicable to other functional neuroimaging such as positron emission tomography (PET). Let $x = [x(1), x(2), \ldots, x(n)]$ denote a voxel time series of size $n$ (i.e. $n$ scans). Each scan is associated with a condition, for example, "activate" or "rest". To test the main effect (i.e. the effect of interest or the effect to be tested), we may choose a test statistic to measure the difference between scans of "activate" and scans of "rest". One choice could be the mean difference test statistic, which calculates the difference between the mean of "activate" scans and that of "rest" scans. In this case, we formulate the test statistic as: $T(x) = \sum_i c(i) x(i)$, where

$c(i) = 1/n_a$ if the status of the $i$-th scan is "activate", $c(i) = -1/n_r$ elsewhere. Here, $n_a$ and $n_r$ are respectively the numbers of "activate" and "rest" scans.

Since the main effect is usually confounded with the undesirable temporal effects in the time series, the exchangeability condition may not hold for the entire set of scans. To tackle this, we divide the scans into certain blocks, and assume that the exchangeability is preserved within each block, which can be defined as an *exchangeability block* (EB) [4], i.e., $x = [x(1), \ldots, x(n_1); x(n_1+1), \ldots, x(n_1+ n_2); \ldots ; x(n-n_g+1), \ldots , x(n)] = [x_1; x_2; \ldots ; x_g]$, where $x_1 = [ x(1), \ldots,$

$x(n_1)], \ldots, x_g = [x(n-n_g+1), \ldots , x(n)]$ are $g$ EBs. Next, we perform all possible permutations within each block and conduct blockwise permutation tests. To preserve the exchangeability, no cross-block permutation is allowed. Note that the number of total possible blockwise permutations is equal to the product of the numbers of permutations within each block, i.e., $\#(\pi) = \#(\pi_1)\#(\pi_2)\cdots\#(\pi_g)$, where $\pi$ is the blockwise permutation, and $\pi_i$ is the permutation within the $i$-th block. Although this number is smaller than the number of general non-blockwise permutations, it is still large enough to lead to high computation cost for a typical time series.

### B. Moments-based Blockwise Permutation Tests

To estimate the $p$-value, the permutation distribution needs to be constructed using test statistic values corresponding to all possible permutations. However, it is computationally expensive to enumerate all possible blockwise permutations. To reduce the computation cost, we fit the permutation distribution with the Pearson distribution series [10] without performing any permutation. The Pearson distribution series is a widely used four-parameter system. The four parameters required are the mean, variance, skewness, and kurtosis, which can all be calculated from the first four moments. We describe next about how to calculate the moments of our blockwise permutation distribution.

We have developed an efficient recursive algorithm [9] to calculate the moments of regular (non-blockwise) permutation distribution. To obtain the moments of blockwise permutation distribution for the entire set of time series, the key idea is to formulate it as a combination of the moments of regular permutation distribution from all EBs (see Eq. 1). Here, we assume the test statistic is summable. That is, $T(x, \pi) = T(x_1, \pi_1) + T(x_2, \pi_2) + \cdots + T(x_g, \pi_g)$. This is a reasonable assumption and works for most popular test statistics or their equivalent test statistics [11]. For permutation tests, two test statistics are equivalent if they have the same $p$-value for any observation.

The $r$-th moment of the blockwise permutation tests is:

$$E(T(x,\pi)^r) = \sum_\pi (T(x_1, \pi_1) + \cdots + T(x_g, \pi_g))^r / \#(\pi)$$

$$= \sum_{\pi_1} \cdots \sum_{\pi_g} (T(x_1, \pi_1) + \cdots + T(x_g, \pi_g))^r / \#(\pi)$$

$$= \sum_{\pi_1} \cdots \sum_{\pi_g} \sum_{k_1, \cdots, k_g} \binom{r}{k_1, \cdots, k_g} T(x_1, \pi_1)^{k_1} \cdots T(x_g, \pi_g)^{k_g} / \#(\pi)$$

$$= \sum_{k_1, \cdots, k_g} \binom{r}{k_1, \cdots, k_g} \sum_{\pi_1} T(x_1, \pi_1)^{k_1} \cdots \sum_{\pi_g} T(x_g, \pi_g)^{k_g} / \#(\pi)$$

$$= \sum_{k_1, \cdots, k_g} \binom{r}{k_1, \cdots, k_g} E(\sum_{\pi_1} T(x_1, \pi_1)^{k_1}) \cdots E(\sum_{\pi_g} T(x_g, \pi_g)^{k_g}). \quad (1)$$

In order to further reduce the computation cost, we represent the first four moments by several symmetric functions:

$$m_j(i) = \sum_{\pi_i} (T(x_i, \pi_i)^j) / \#(\pi_i) = E(T(x_i, \pi_i)^j),$$

$$m_j = \sum_i m_j(i),$$

$$E(T(x,\pi)) = m_1,$$

$$E(T(x,\pi)^2) = m_2 + m_1^2 - \sum_i m_1(i)^2,$$

$$E(T(x,\pi)^3) = m_3 + 3m_1m_2 + m_1^3 - 3\sum_i m_1(i)m_2(i) - 3(\sum_i m_1(i)^2)m_1 + 2(\sum_i m_1(i)^3),$$

$$E(T(x,\pi)^4) = m_4 + 4m_1m_3 + 6m_1^2m_2 + 3m_2^2 + m_1^4 - 4m_1\sum_i m_1(i)^3 - 6m_1^2\sum_i m_1(i)^2 - 6(\sum_i m_1(i)^2)m_2 + 8\sum_i m_1(i)m_3(i) + 3(\sum_i m_1(i)^2)^2 - 3\sum_i m_2(i)^2 - 6\sum_i m_1(i)^4,$$

where $m_j(i)$ is the $j$-th moment for the $i$-th block. All within-block moments $m_j(i)$ can be obtained through our newly developed recursive algorithm [9]. Here, the within-block moments are the moments for an EB data using the same test statistic function as that for the complete data. For example, when $n = 8$, and the condition is [activate, rest, activate, rest, activate, rest, activate, rest], we choose the mean difference test statistic for the complete time series, i.e., $T = (x(1)+ x(3)+ x(5)+ x(7))/4 - (x(2)+ x(4)+ x(6)+ x(8))/4$. If we divide the time series into two blocks, $x_1 = [ x(1), x(2), x(3), x(4)]$ and $x_2 = [ x(5), x(6), x(7), x(8)]$ . For each block, for example $x_1$ , $T(x_1)$ should be ( $x(1)+ x(3))/4 - ( x(2)+ x(4))/4$, not the mean difference for the block: ( $x(1)+ x(3))/ 2 - ( x(2)+ x(4))/2$.

In summary, we convert the moments calculation for blockwise permutation tests to a simple combination of a moments calculation for regular permutation tests without restriction. The computation cost due to this simple combination can be ignored, compared with the cost of a moments calculation for regular permutation tests.

### III. RESULTS

#### A. Simulated experiments

To illustrate and validate our moments-based blockwise permutation tests method, we generate a simple time series confounded with two different temporal artifacts. The main effect (Fig. 1(b)) is generated by adding a Gaussian noise to a two-state boxcar function (Fig. 1(a)). Two temporal artifacts, a nonlinear cosine trend (Fig. 1(c)) and an edge trend (Fig. 1(e)) were added to the main effect, leading to two confounded time series in Fig. 1(d) and Fig. 1(f). Here the null hypothesis is that no difference between the two states exists. The exchangeability condition does not hold for the entire set of the time series with 180 samples because of the strong temporal artifact under the null hypothesis. Therefore, we divide the time series into 15 blocks. Each block consists of 6 samples with boxcar function value 1 and another 6 samples with value 0. Within each block, the temporal artifact can be ignored. The number of all possible blockwise permutations is $(_{12}C_6)^{15} \approx 3\times10^{44}$. It is extremely expensive to enumerate all possible blockwise permutations.

We choose the mean difference test statistic and apply the proposed efficient moments-based blockwise permutation method. We obtained the $p$-value in 0.001 seconds on our computer (Intel Core 2 CPU, 2.4GHZ, 2G RAM).
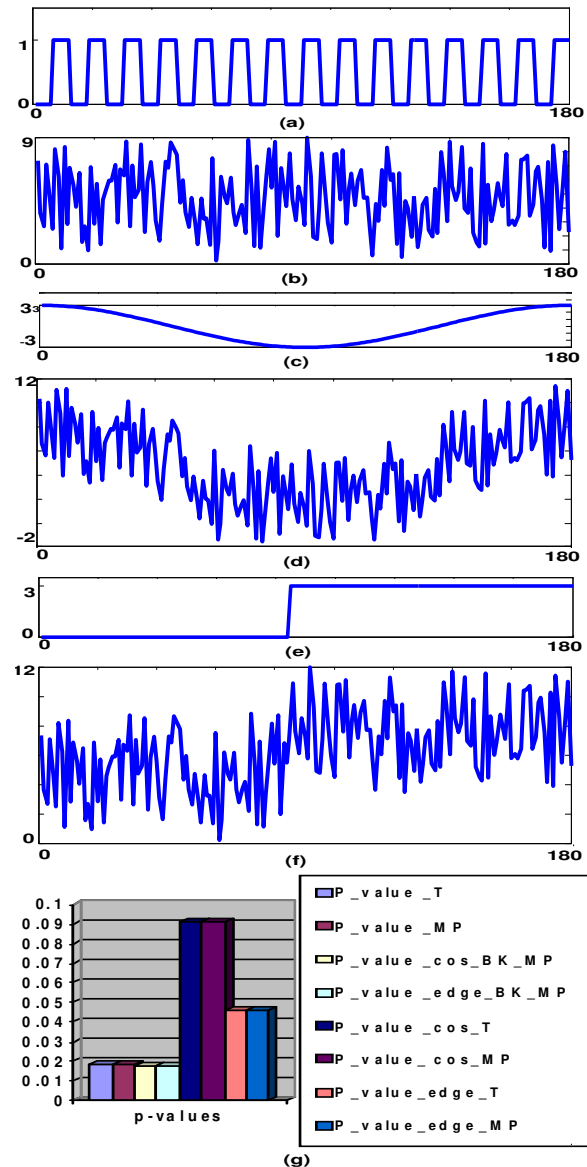


Fig. 1. Comparison of blockwise and regular permutation tests. (a): boxcar function. (b): the main effect time series with Gaussian noise. (c): the nonlinear cosine trend. (d): the mixed time series with cosine temporal effect. (e): the edge trend. (f): the mixed time series with edge temporal effect. (g): estimated $p$-values from $t$-test (_T), regular and blockwise moments-based permutation tests (_MP and_BK_MP).

Since Gaussian noise is added here, $t$-test is a valid test for the main effect and its result can be used as the golden standard. Fig. 1(g) shows the comparison of $p$-values obtained by different methods. The first four $p$-values are respectively calculated by the $t$-test and the regular moments-based permutation tests for the main effect time series, and the moments-based blockwise permutation tests for the time series confounded with a cosine trend and an edge trend. They are all very similar. This demonstrates that our moments-based blockwise permutation tests are robust to

different temporal artifacts. The fifth and sixth *p*-values are results of the *t*-test and the regular moments-based permutation tests for the time series with the cosine trend; the seventh and eighth *p*-values are results of the corresponding two tests for the time series with the edge trend. For the time series confounded with temporal artifacts, the *t*-test and regular moments-based permutation tests are not valid and lead to poor estimation of *p*-values.

### B. Auditory fMRI data

The auditory fMRI data (block) was from the SPM data site, and its description is available at *http://www.fil.ion.ucl.ac.uk/spm/data/auditory.html*. The paradigm consists of a rest period of 42s, followed by auditory stimulus with 42s bi-syllabic words at 1 Hz. 96 acquisitions were made. Each acquisition has 64x64x64 (3x3x3 mm$^3$) voxels. Due to T1 effects, we discarded the first 12 scans. The rest 84 acquisitions were made (RT=7s), in blocks of 6, giving 14 42s segments. The condition for successive segments alternated between rest and auditory stimulation, starting with rest.

In order to have some form of ground truth, we take the pre-processed data (smoothed, normalized and realigned) instead of the raw fMRI data, but add cosine and edge trends to the time series separately to generate two sets of pseudo-fMRI data. For both cases, we divide the time series into 7 blocks. Each block consists of a 42s rest segment and a 42s auditory stimulation segment. The mean difference test statistic is used to detect activation. Due to unknown data distribution and temporal artifacts, the *t*-test is not a valid technique here. We compare the results of moments-based regular and blockwise permutation tests. The blockwise permutation tests have succeeded in detecting the auditory cortex regions. The regular ones are sensitive to the confounded temporal artifacts and have difficulty in identifying auditory cortex regions.

## IV. CONCLUSION AND DISCUSSION

We have developed a new moments-based blockwise permutation test approach based on the moments of the test statistic, and applied it to activation detection in functional neuroimaging. To preserve the exchangeability condition, the time series are first divided into several exchangeability blocks. Next, efficient moments-based permutation tests are performed by approximating the permutation distribution of the test statistic with the Pearson distribution series. This involves the computation of the first four moments of the permutation distribution within each block and then over the entire set of a time series. Experimental results demonstrated the advantages of the proposed method.

Although we focus on fMRI data analysis in this paper, the proposed method is general and applicable to many other situations and biomedical image modalities. It can deal with different artifacts, including the spatial and temporal effects, etc. For example, if the data are collected from different instruments or sites, it would be reasonable to block the data according to the instrument or site in order to preserve exchangeability in the permutation hypothesis tests. Therefore, our blockwise permutation method has potential wide applications in both structural and functional neuroimage analyses involving hypothesis tests and group comparisons. One of issues that needs to be investigated in the future is the dependence of the size and the number of blocks on the performance of the method.
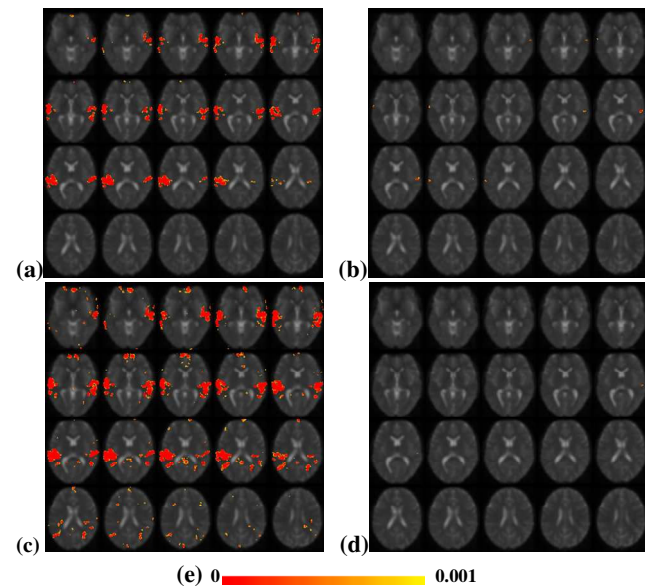


Fig. 2. Auditory fMRI activation detection. Top panel: pseudo-fMRI data with cosine trend. (a) and (b): results of moments-based blockwise and regular permutation tests, respectively. Bottom panel: pseudo-fMRI data with edge trend: (c) and (d): results of moments-based blockwise and regular permutation tests, respectively. (e) *p*-value color bar.

## REFERENCES

[1] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, R.S.J. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Hum. Brain Mapp.*, vol. 2, 1995, pp.189-210.

[2] K.J. Worsley, A.C .Evans, S. Marrett, and P. Neelin, "A three-dimensional statistical analysis for CBF activation studies in human brain," *J Cereb Blood Flow Metab*, vol. 12, 1992, pp.900–918.

[3] A.P. Holmes, R.C. Blair, J.D.G. Watson, and I. Ford, "Nonparametric analysis of statistic images from functional mapping experiments," *J Cereb Blood Flow Metab*. Vol. 16, 1996, pp. 7–22.

[4] T.E. Nichols, and A.P. Holmes, "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Human Brain Mapping*, vol.15, 2001, pp. 1–25.

[5] Y. M. Wang, and J. Xia, "Unified framework for robust estimation of brain networks from fMRI using temporal and spatial correlation analyses," *IEEE Transactions on Medical Imaging*, in press, 2009.

[6] P. Good, *Permutation, Parametric and Bootstrap Tests of Hypotheses.* 3rd ed., New York: Springer, 2005.

[7] L Hubert, *Assignment methods in combinatorial data analysis.* New York: Marcel Dekker, Inc., 1987.

[8] P.W. Mielke, and K.J. Berry, *Permutation Methods: A Distance Function Approach*, New York: Springer, 2001.

[9] C. Zhou, and Y. M. Wang, "Efficient Moments-based Permutation Tests," *Technical report*, UIUC, 2009.

[10] G.J .Hahn, and S.S. Shapiro, *Statistical models in engineering.* John Wiley and Sons, Inc., 1967.

[11] E. Edgington, *Randomization Tests*. CRC, 3rd ed., 1995.