1.5em 0pt

On the Convergence of Transposition Test

Duohan Zhang[1]

[1]University of Wisconsin-Madison, Department of Statistics

May 8, 2023

Abstract

The permutation test is a widely employed method for determining statistical significance without making prior assumptions about the underlying distribution. However, generating all possible permutations becomes infeasible due to computational limitations. Prior research has introduced the transposition test to address this issue. In this report, we investigate the convergence behavior of permutation procedure, and propose two novel transposition tests for the three-sample setting. We then compare the performance of these tests to the standard permutation test using simulated data.

1 Introduction

The permutation test is a powerful non-parametric statistical technique for hypothesis testing. It is particularly useful in situations where traditional parametric tests are not appropriate due to the violation of assumptions, such as normality assumption, or lack of knowledge about the underlying distribution. Permutation tests do not rely on any distributional assumptions, making them applicable to a wide range of real-world data. It is also less sensitive to outliers, making them robust and reliable to noisy data.

In recent years, the permutation test has attracted increasing attention from many fields, including biomedical research, psychology, genomics, neuroscience, and finance [Chu+18; Van+22; Tho+01; RT22]. When the total number of permutation group is very large, scientists use resampling techniques to estimate the exact p-value. However, due to computational constraints, only a small proportion of possible permutation group is generated, which may hurt the usage of the method. In this report, our objective is to introduce the transposition test as an alternative to the conventional permutation test, offering the advantage of generating a larger amount of data with the same computational effort. This report reviews the proposed two-sample transposition test approach [Chu+19] and analyses its asymptotic behavior. We show that after certain amount of sampling, the underlying distribution will get close to the data distribution under the permutation procedure. Then we generalize the approach to the three-sample setting, while we propose a circle-way alternative that can achieve faster convergence. We conduct simulations using both methods under various data size configurations to illustrate the advantages of the latter approach in their mixing proportions. Ultimately, we simulate three sets of data, each following distinct normal distributions, and evaluate their convergence in real-time comparisons.

2 Two-Sample Transposition Test

2.1 Preliminary

The traditional two-sample permutation test [Fis36] is as follows. Considering two ordered-sets

$$x = (x_1, x_2, \cdots, x_m),\ y = (y_1, y_2, \cdots, y_n),$$

we need to measure the distance by a given function $f(x, y)$ such as t-statistic or correlations. Consider the combined ordered set $A = (x_1, \cdots, x_m, y_1, \cdots, y_n)$ and its all possible permutations S_{m+n} with $(m + n)!$ possible permutations. Permutation τ is denoted as

$$\tau = \begin{bmatrix} x_1 & \cdots & x_m & y_1 & \cdots & y_n \\ \tau(x_1) & \cdots & \tau(x_m) & \tau(y_1) & \cdots & \tau(y_n) \end{bmatrix},$$

where the first line denotes the original ordered dataset, while the second line denotes the ordered dataset after one possible permutation. We further split the permutation τ into two groups

$$\tau(x) = (\tau(x_1), \tau(x_2), \cdots, \tau(x_m)),\ \tau(y) = (\tau(y_1), \tau(y_2), \cdots, \tau(y_n)),$$

the exact p-value for testing the null hypothesis is

$$p - value = \frac{1}{(m+n)!} \sum_{\tau \in S_{m+n}} I(f(\tau(x), \tau(y)) > f(x, y)), \tag{1}$$

where I is the indicator function taking value 1 if the argument is true and 0 otherwise.

It is often the case that f is a symmetric function in the sense that $f(x, y) = f(\mu(x), \eta(y))$, for any $\mu \in S_m$ and any $\eta \in S_n$, such as t-statistic and correlation function. In other words, the order in x and y does not matter and what matters is the data x and y contain. Then there are $\binom{m+n}{m}$ possible permutations. If $m = n$ is very large, the number can be approximated by Stirling's formula [Fel91]

$$\binom{2m}{m} \sim \frac{4^m}{\sqrt{\pi m}}.$$

The number of all possible permutations exponentially increases as the sample size increases, and thus it is impractical to compute all possible permutations. For example, $\binom{40}{20} = 1.34 \times 10^{11}$ even when $m = 20$ is not extremely large. In practice, scientists uniformly generate each permutation with possibility $1/\binom{m+n}{m}$ using Monte Carlo simulation, and estimate the p-value.

2.2 Transposition

In the last section, we discussed permutation test and its computation for approximating exact p-value. Chung et al [Chu+19] proposed a novel transposition test that exploits the underlying algebraic structure of the permutation group. With the same amount of real-time computation, the transposition procedure can generate much more permutations than the vanilla permutation procedure.

Transposition is the permutation that exchanges i-th and j-th elements between two sets and keeps all others fixed. For example, if we transpose elements between x and y

$$\pi_{ij}(x) = (x_1, \cdots, x_{i-1}, y_j, x_{i+1}, \cdots, x_m),$$

$$\pi_{ij(y)} = (y_1, \cdots, y_{j-1}, x_i, y_{j+1}, \cdots, y_n).$$

The paper[Chu+19] further gave a theoretical result that if f is an algebraic function, there exists a function g such that

$$f(\pi_{ij}(x), \pi_{ij}(y)) = g(f(x, y), x_i, y_j). \tag{2}$$

Thus instead of uniformly generating each permutation from scratch at each step, we can randomly choose two data in each dataset and exchange them. For example, the two-sample t-statistic is computed based on sample mean $v(x) = \sum_{j=1}^{m} x_j/m$ and sample variance $w(x) = \sum_{j=1}^{m}(x_j - v(x))/(m-1)$. Therefore, instead of computing sample mean and sample variance with the new generated permutation, we can update them over transposition as

$$v(\pi_{ij}(x)) = v(x) + (y_j - x_i)/m, \tag{3}$$

$$w(\pi_{ij}(x)) = w(x) + \frac{y_j^2 - x_i^2}{m-1} + \frac{m}{m-1}(v(x)^2 - v(\pi_{ij}(x))^2). \tag{4}$$

Transposition can rapidly generate a large number of permutations via transposition procedure. However, people may have two questions regarding this faster procedure. Firstly, can we prove the asymptotic behavior of the transposition sample? Secondly, if it would converge, what's the convergence rate?

Now in order to simplify the question in theory, we only consider 1). the mixing proportion-the proportion of original y-data in x, and 2). the equal sample size $m = n$. We easily find that it is a random walk problem. Another reason we only consider mixing proportion is that f is a symmetric function, so the order of the data itself is irrelevant. If we can achieve 50% mixing proportion averagely, intuitively it is mixing well and thus we believe it is somehow uniformly distributed.

To answer the first question, imagine we have two people and some balls of different colors: initially person 1 has n red balls and person 2 has n blue balls. Now at each time step they exchange balls one by one. Let S_i denotes the number of red balls in person 1's pocket in i-th step. Let $\mu(x)$ be the distribution under the permutation, i.e. $\mu(x) = \binom{n}{x}^2/\binom{2n}{n}$, where $0 \le x \le n$.

Theorem 2.1. *$S_i \xrightarrow{d} \mu$, where $\xrightarrow{d}$ denotes convergence in distribution.*

Proof. It is a random walk problem. At the initial step $S_0 = n$. We define $S_{i+1} = S_i + X_{i+1}$. Now at the (i+1)-th step, person 1 has S_i red balls and $n - S_i$ blue balls; person 2 has $n - S_i$ red balls and S_i blue balls. There are three categories that could happen: 1). Both select red balls or blue balls with probability $2\frac{S_i}{n}(1 - \frac{S_i}{n})$, where in such category the number of red balls in person 1's pocket does not change so $X_{i+1} = 0$. 2). Person 1 selects a blue ball and person 2 selects a red ball with probability $(1 - \frac{S_i}{n})^2$, where in such category $X_{i+1} = 1$. 3). Person 1 selects a red ball and person 2 selects a blue ball with probability $(\frac{S_i}{n})^2$, where in such category $X_{i+1} = -1$.

Now we claim that μ is a stationary distribution under the Markov chain: $P(x, x+1) = (1 - \frac{x}{n})^2$, $P(x, x) = 2\frac{x}{n}(1 - \frac{x}{n})$, $P(x, x-1) = (\frac{x}{n})^2$. Here $P(x, y)$ denotes the one-step transitional probability from x to y. We find that $\mu(x)P(x, x+1) = \mu(x+1)P(x+1, x)$ by checking

$$\begin{aligned}\frac{\mu(x)}{\mu(x+1)} &= (\frac{(x+1)!(n-x-1)!}{x!(n-x)!})^2 \\ &= (\frac{x+1}{n-x})^2 \\ &= P(x+1, x)/P(x, x+1).\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_x \mu(x)P(x,y) &= \sum_x \mu(y)P(y,x) \\ &= \mu(y)\sum_x P(y,x) \\ &= \mu(y)\end{aligned}$$

and by definition μ is a stationary measure. It is a finite irreducible and aperiodic Markov chain. Combining that μ is a stationary distribution and the Markov chain theory [Dur19], we conclude that $P(S_i = x) \xrightarrow{i\to\infty} \mu(x)$. $\square$

The Markov chain theory provides the rigorous theoretical guarantee that after infinite steps, the mixing proportion would converge to the stationary distribution, i.e. the uniform distribution under the random permutation. Chung et al[Chu+19] studied the convergence time of mean mixing proportion, and theoretically the transposition method can converge to the uniform distribution after $O(nlog(n))$ time [BSZ11; AD86].

3 Three-Sample Transposition Test

In this section, we generalize the previous transposition test to three-sample setting and propose two different transposition approaches. Consider three ordered sets

$$x = (x_1, x_2, \cdots, x_m),\ y = (y_1, y_2, \cdots, y_n),\ z = (z_1, z_2, \cdots, z_l).$$

and the combined ordered set $A = (x_1, \cdots, x_m, y_1, \cdots, y_n, z_1, \cdots, z_l)$ and its all possible permutations S_{m+n+l}. Permutation τ is denoted as

$$\tau = \begin{bmatrix} x_1 & \cdots & x_m & y_1 & \cdots & y_n & z_1 & \cdots & z_l \\ \tau(x_1) & \cdots & \tau(x_m) & \tau(y_1) & \cdots & \tau(y_n) & \tau(z_1) & \cdots & \tau(z_l). \end{bmatrix}$$

Now we propose the first transposition test [1] as a trivial generalization of the previous section: randomly selecting two datasets and exchanging i-th and j-th elements between two sets and keeps all other data fixed. For example, if we transpose two elements between x and y

$$\pi_{ij.}(x) = (x_1, \cdots, x_{i-1}, y_j, x_{i+1}, \cdots, x_m),$$
$$\pi_{ij.(y)} = (y_1, \cdots, y_{j-1}, x_i, y_{j+1}, \cdots, y_n),$$
$$\pi_{ij.}(z) = z.$$

Similarly, we use $\pi_{i.k}$ denote the transposition that exchanges i-th and k-th elements between x and z.

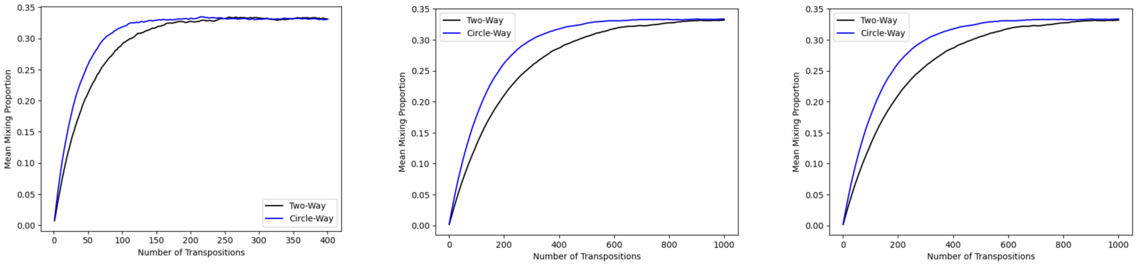
The permutation group mixing time paper [BSZ11] discussed the mixing time for random k-cycles. Informally, a set of n cards need $t_{mix} = (1/k)nlog(n)$ time to shuffle cards uniformly with k-cycle transpositions, which means that a higher k would lead to a lower t_{mix}. Inspired by this result, it is natural to think of a new transposition test procedure [2] with a circle for the three-sample data, which may achieve a faster convergence rate. The only difference within a single transposition is that we randomly select clockwise transposition order or counter-clockwise order (line 8), and that we transpose data in a circle way. This procedure can have faster convergence rate compared to the first algorithm, since three-circle transposition can mix faster than the two-way transposition.

Algorithm 1 Two-Way Transposition For Three-Sample Data.

procedure TRANSPOSE DATA(x_0, y_0, z_0)
    Input: Three datasets x_0, y_0, z_0; number of transpositions, N; test-statistic, f and its corresponding updating function g (2).
    Output: p-value.
    $x \leftarrow x_0, y \leftarrow y_0, z \leftarrow z_0$.
    $t \leftarrow f(x, y, z)$.
    $D \leftarrow \{t\}$.
    for $1 \leq k \leq N$ **do**
        randomly select two sets A,B from x,y,z.
        randomly transpose A_i, B_j between A and B.
        $t \leftarrow g(t, A_i, B_j)$.
        $D \leftarrow D \cup \{t\}$.
    end for
    p-value$=\frac{1}{|D|}\sum_{t\in D} I(t > f(x_0, y_0, z_0))$.
    Return p-value.
end procedure

Algorithm 2 Three-Circle Transposition For Three-Sample Data.

1: **procedure** TRANSPOSE DATA(x_0, y_0, z_0)
2: **Input:** Three datasets x_0, y_0, z_0; number of transpositions, N; test-statistic, f and its corresponding updating function g (2).
3: **Output:** p-value.
4: $x \leftarrow x_0, y \leftarrow y_0, z \leftarrow z_0$.
5: $t \leftarrow f(x, y, z)$.
6: $D \leftarrow \{t\}$.
7: **for** $1 \leq l \leq N$ **do**
8: uniformly select $x \to y \to z \to x$ or $x \to z \to y \to x$, each with probability $\frac{1}{2}$.
9: randomly select a data from x, y, z respectively.
10: Transpose the data following the chosen order in line 8.
11: $t \leftarrow g(t, x_i, y_j, z_k)$.
12: $D \leftarrow D \cup \{t\}$.
13: **end for**
14: p-value$=\frac{1}{|D|}\sum_{t\in D} I(t > f(x_0, y_0, z_0))$.
15: **Return** p-value.
16: **end procedure**



(a) m=n=l=50 (b) m=n=l=100 (c) m=n=l=200

Figure 1: The mixing proportion based on the average of 800 simulations.

3.1 Discussion of Convergence Rate

F-statistic is a example of symmetric functions involving mean and variance of three-sample data. Chung et al [Chu+19]gave the update rule of mean and variance as in (3) and (4). The three-sample F-statistic can be computed as follows:

$$F(x,y,z) = \frac{m+n+l-3}{2}\frac{m(v(x)-a)^2+n(v(y)-a)^2+l(v(z)-a)^2}{(m-1)w(x)+(n-1)w(y)+(l-1)w(z)} \tag{5}$$

where $a = \frac{mv(x)+nv(y)+lv(z)}{m+n+l}$ is the grand mean of the whole data (x, y, z).

Now we compare two proposed approaches in terms of convergence rate - how many transpositions do we need to get uniformly distributed data in Figure 1. We conduct three separate experiments, each utilizing a distinct data size. For each experiment, we perform 800 repetitions to ensure the reliability of our results. The mixing proportions are determined by calculating the sample mean of the 800 proportion mixes. To facilitate a clear and concise presentation, we choose to display only the proportions of the original y-data present in the x-data. From the figures, we see that three-circle transposition needs less iterations to get a good mix, where $\frac{1}{3}$ is the ideal mixing proportion.

In this study, we calculate the distribution of mixing proportion as follows:

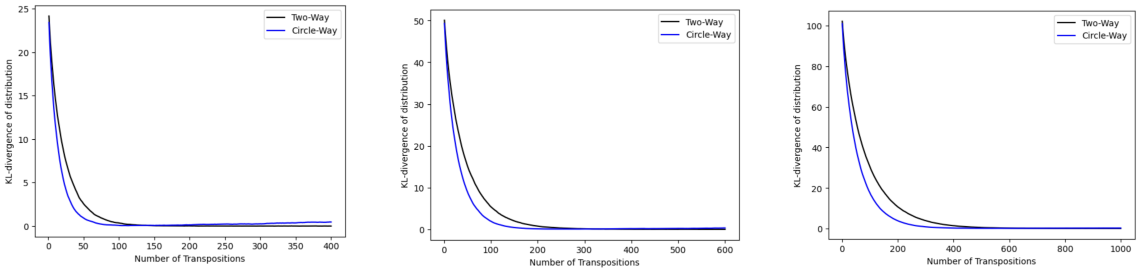
$$\mu(x/n) = \frac{\binom{n}{x}\binom{2n}{n-x}}{\binom{3n}{n}} \tag{6}$$

where $0 \leq x \leq n$. Besides comparing the convergence rate of averaged mixing proportion, we also did simulations to demonstrate the convergence of KL-divergence for mixing proportion's distribution:

$$D_{KL}(P,Q) = \sum_i P(i) log \frac{P(i)}{Q(i)} \tag{7}$$

where P, Q are discrete probability distributions.

Figure 2 further showed the convergence rate of KL-divergence between the distributions on the mixing proportion. The true distribution on the mixing proportion is calculated as (6) and the estimated distribution is calculated as the empirical distribution based on 800 simulations. The circle-way approach behaves better by exhibiting a more rapid convergence towards zero.



(a) m=n=l=50 (b) m=n=l=100 (c) m=n=l=200

Figure 2: The KL-divergence between stationary distribution and simulated distribution.

4 Simulated Study and Discussion

Transposition tests have wide applications in many fields. For example, we can test if there are significant differences of the test scores of students from different schools. One-way ANOVA(Analysis of Variance) is often used to determine if factors have an impact on the students' performance. For the experiments part, we compare two proposed methods against the standard permutation test with the ground truth. We simulate $x_1, ..., x_n \sim N(-0.1, 1)$, $y_1, ..., y_n \sim N(0, 1)$, and $z_1, ..., z_n \sim N(0.1, 1)$. We take $n = 50$. We specify the normal distribution in order to compute the ground truth for the p-value. The simulations are repeated 100 times and we compute the average for ease of demonstration. We write python codes in the Google Colaboratory. Readers can see the codes for more details: https://github.com/duohan0520/768-report. There are the total number of $\binom{150}{50}\binom{100}{50} = 2.03 \times 10^{69}$ permutations. We obtain the F-statistic value and the ground truth of p-value = 0.799. In comparison to the standard permutation test, we discovered that the three-circle transposition method offers reduced computation time for the same amount of iterations. The reduced computation time can be attributed to the fact that the method does not require permuting all data in every single iteration. Instead, it only transposes three data points at each step, which conserves computational resources when updating the F-statistic. Furthermore, when contrasted with the two-way transposition test, the three-circle transposition approach demonstrates superior convergence behavior while maintaining nearly the same level of computational efficiency.

In this study, we introduce two transposition test procedures in the three-sample setting. With respect to the more general k-sample setting, we posit that implementing a k-cycle transposition is a viable approach due to its accelerated convergence rate. Nonetheless, a larger cycle size does not always guarantee less real clock computation time. From one hand, as the cycle size increases, so do the variations in the combinatorial patterns. From the other hand, as k becomes large, there are different approaches other than a cycle way. For example, if we have four groups of data a, b, c, d we may transpose them as $a \to b \to c \to d \to a$ or $a \to b \to a, c \to d \to c$. It is unclear which method way may have better performance.

References

[AD86] David Aldous and Persi Diaconis. "Shuffling cards and stopping times". In: *The American Mathematical Monthly* 93.5 (1986), pp. 333–348.

[BSZ11] Nathanaël Berestycki, Oded Schramm, and Ofer Zeitouni. "Mixing times for random k-cycles and coalescence-fragmentation chains". In: (2011).

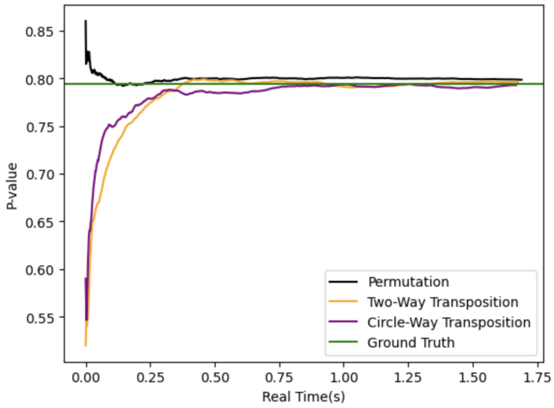


Figure 3: The simulation study for comparing the permutation test and the two proposed transposition test.

[Chu+18] Moo K Chung et al. "Exact combinatorial inference for brain images". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer. 2018, pp. 629–637.

[Chu+19] Moo K Chung et al. "Rapid acceleration of the permutation test via transpositions". In: *Connectomics in NeuroImaging: Third International Workshop, CNI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 3*. Springer. 2019, pp. 42–53.

[Dur19] Rick Durrett. *Probability: theory and examples*. Vol. 49. Cambridge university press, 2019.

[Fel91] William Feller. *An introduction to probability theory and its applications, Volume 2*. Vol. 81. John Wiley & Sons, 1991.

[Fis36] Ronald Aylmer Fisher. "Design of experiments". In: *British Medical Journal* 1.3923 (1936), p. 554.

[RT22] Joseph P Romano and Marius A Tirlea. "Permutation testing for dependence in time series". In: *Journal of Time Series Analysis* 43.5 (2022), pp. 781–807.

[Tho+01] Paul M Thompson et al. "Genetic influences on brain structure". In: *Nature neuroscience* 4.12 (2001), pp. 1253–1258.

[Van+22] Claudia D Van Borkulo et al. "Comparing network structures on three aspects: A permutation test." In: *Psychological Methods* (2022).