

Testing exchangeability of multivariate distributions

Jan Kalina & Patrik Janáček

To cite this article: Jan Kalina & Patrik Janáček (2023) Testing exchangeability of multivariate distributions, Journal of Applied Statistics, 50:15, 3142-3156, DOI: [10.1080/02664763.2022.2102158](https://doi.org/10.1080/02664763.2022.2102158)

To link to this article: <https://doi.org/10.1080/02664763.2022.2102158>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 Jul 2022.



Submit your article to this journal [↗](#)



Article views: 1116



View related articles [↗](#)



View Crossmark data [↗](#)

Testing exchangeability of multivariate distributions

Jan Kalina ^{a,b} and Patrik Janáček^b

^aInstitute of Information Theory and Automation, The Czech Academy of Sciences, Prague, Czech Republic;

^bInstitute of Computer Science, The Czech Academy of Sciences, Prague, Czech Republic

ABSTRACT

Although there have been a number of available tests of bivariate exchangeability, i.e. bivariate symmetry for bivariate distributions, the literature is void of tests whether a multivariate distribution with more than two dimensions is exchangeable or not. In this paper, multivariate permutation tests of exchangeability of multivariate distributions are proposed, which are based on the non-parametric combination methodology, i.e. on combining non-parametric bivariate exchangeability tests. Numerical experiments on real as well as simulated multivariate data with more than two dimensions are presented here. The multivariate permutation test turns out to be typically more powerful than a bivariate exchangeability test performed only over a single pair of variables, and also more suitable compared to tests exploiting the approaches of Benjamini–Yekutieli or Bonferroni.

ARTICLE HISTORY

Received 15 July 2021

Accepted 26 June 2022

KEYWORDS

Multivariate distribution; exchangeable distribution; multivariate permutation test; multiple testing; non-parametric combination methodology; multiple comparisons

1. Introduction

Let us consider a random vector $X \in \mathbb{R}^p$. Let X_1, \dots, X_n with a fixed n represent p -dimensional random vectors following the probability distribution $\mathcal{L}(X)$. The distribution $\mathcal{L}(X)$ is called exchangeable (interchangeable, permutable) if it holds

$$\mathcal{L}(PX) = \mathcal{L}(X) \quad (1)$$

for all permutation matrices P , i.e. for all square matrices P of size $p \times p$ with values 0 or 1 such that each row and each column contains the value 1 exactly once. In other words, (1) means that indexing of variables of $\mathcal{L}(X)$ is irrelevant and $\mathcal{L}(X)$ is permutation invariant. This work is interested in testing exchangeability of $\mathcal{L}(X)$ based on given data, i.e. testing the null hypothesis H_0 that (1) holds against the general alternative hypothesis H_1 stating that H_0 does not hold.

Attention has been paid to exchangeable distributions mainly for the bivariate case ($p = 2$). For a bivariate random vector $(X_1, X_2)^T$, exchangeability (1) is equivalent to

$$\mathcal{L}(X_1, X_2) = \mathcal{L}(X_2, X_1). \quad (2)$$

Bivariate exchangeability often denoted as bivariate symmetry can be interpreted as symmetry with respect to permutation of axes or symmetry along the axis of the first quadrant

CONTACT J. Kalina  kalina@cs.cas.cz

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

(i.e. line $x_1 = x_2$). We note that (2) in the bivariate case naturally implies

$$\mathcal{L}(X_1) = \mathcal{L}(X_2). \quad (3)$$

Already the first bivariate symmetry tests in the 1970s [18,45] were motivated by a practical question of whether a medical treatment has an effect or not. There have been a number of available bivariate symmetry tests, which are suitable for paired data allowing for a (possibly high) correlation of the two variables within pairs. Still, we are not aware of any test for distributions with more than two dimensions, although these could represent very natural extensions of tests for the bivariate case. Only in the context of copulas, a test of exchangeability of copulas based on empirical processes was proposed in [14] for an arbitrary dimension. To give a motivation example for distributions with $p \geq 3$, the income of five nutrients (calcium, iron, protein, vitamin A, vitamin D) on a sample of women was investigated in [11], where a test of exchangeability of the multivariate distribution of the 5 variables would be useful. Another example may be a study of the effect of a treatment (e.g. of the COVID-19 vaccine) applied in three sequential doses.

Testing exchangeability of multivariate distributions may also be motivated within a broader context of tests of various forms of symmetry of multivariate continuous distributions. It was recently suggested to replace testing the exchangeability of a distribution by testing axial symmetry, e.g. using the test based on (multivariate) directional quantiles of [19]; it namely holds that if a multivariate distribution is exchangeable after a shift, then it is symmetric around the axis of the first orthant. Another connection mentioned in [20] is related to p independent univariate distributions, which are assumed to be the same up to their location; their joint distribution (after a shift) has to be exchangeable. Thus, the test whether their distributions are the same (up to a shift) may be performed as a test of exchangeability of the joint distribution (again up to a shift). Other useful relationships between exchangeability of a distribution and some symmetry concepts (defined in [33,39]) were also described in [19,20].

It is important to stress that this paper is interested in permuting the variables of a multivariate distribution but not in permuting observations. The latter is connected to the concept of an exchangeable sequence of random variables (or arrays; see Section 7 of [25]), which is defined as a sequence invariant to permuting the observations, in other words, as a sequence with an exchangeable distribution [7]. Exchangeability of a sequence represents a weaker property compared to independence of the coordinates and the theory of finite exchangeable random vectors stems from de Finetti's theorem as overviewed in [21]. Exchangeability of a sequence of random variables have found applications within conformal inference, which can be characterized as a method for constructing valid prediction errors for new sequentially added observations; conformal inference has become popular for neural networks or random forests [26].

Section 2 recalls available tests of bivariate symmetry. In the methodological Section 3, multivariate permutation tests of exchangeability of distributions with more than two dimensions are proposed. The performance of the tests is investigated on real data in Section 4 and on simulated data in Section 5. Multivariate permutation tests turn out to outperform tests constructed from multiple comparison procedures of Benjamini–Yekutieli or Bonferroni. Section 6 brings conclusions.

2. Available bivariate symmetry tests

We now consider i.i.d. bivariate random vectors $(X_{11}, X_{21})^T, \dots, (X_{1n}, X_{2n})^T$. Particular available tests of the null hypothesis of bivariate symmetry are overviewed in this section. Testing the exchangeability of distribution is desirable, especially if it can be performed in a non-parametric, distribution-free, coordinate-free way by means of simple and powerful tests not relying on strong assumptions. Bivariate symmetry tests can typically be performed as exact (permutation) tests; advantages of permutation tests in the context of bivariate symmetry were recalled in [3] or [10] and for the context of various other symmetry hypotheses in [35]. We stress that all tests mentioned in this paper are performed on raw data without any normalization.

- Bell and Haller [3] overviewed several equivalent formulations of bivariate symmetry. They derived the likelihood ratio test for normal data in the form

$$LR = \frac{1 - r^2(S, D)}{1 + |\bar{D}|/\sigma_D} \quad \text{with } \bar{D} = \sum_{i=1}^n D_i/n \text{ and } \sigma_D^2 = \sum_{i=1}^n (D_i - \bar{D})^2/n, \quad (4)$$

where $r(S, D)$ denotes the Pearson correlation coefficient between $(S_1, \dots, S_n)^T$ and $(D_1, \dots, D_n)^T$ with $D_i = X_{2i} - X_{1i}$ and $S_i = X_{1i} + X_{2i}$ for $i = 1, \dots, n$. Moreover, all distribution-free tests of bivariate symmetry were proven to be based on permutations in [3], where the Wilcoxon signed-rank test was disqualified from being used for testing bivariate symmetry, as its power was shown to equal only to the specified level α .

- The test of Hollander [18] is based on empirical distribution functions $\hat{F}_n(x_1, x_2)$ and $\hat{F}_n(x_2, x_1)$. The exact size and power of the Hollander test, which represents a rank-based analogue of a Cramér-von Mises type test [11], may be computed for small samples by the algorithm of Hilton and Gee [17], which is implemented in the R package NSM3. Quessy [35] considered a related statistic based on the characteristic function (instead of the empirical distribution function) and used the theory of V-statistics to derive their null asymptotic distribution.
- Yanagimoto and Sibuya [45] proposed and investigated rank tests of bivariate symmetry based on the maximal invariant statistic. Let us now use the notation $(U_1, V_1)^T, \dots, (U_n, V_n)^T$ for the permutation of the random sample $(X_{11}, X_{21})^T, \dots, (X_{1n}, X_{2n})^T$ for that it holds $U_i \leq U_{i+1}$ and $V_i \leq V_{i+1}$ if $U_i = U_{i+1}$. Further, let Z_1, \dots, Z_n be the corresponding permutation of $\text{sgn}(X_1 - Y_1), \dots, \text{sgn}(X_n - Y_n)$. Let R_i be the rank of U_i in the pooled sample $\{X_{11}, X_{21}, \dots, X_{1n}, X_{2n}\}$ and let S_i be the rank of V_i in the pooled sample. The maximal invariant statistic is the vector of triplets $(R_1, S_1, Z_1)^T, \dots, (R_n, S_n, Z_n)^T$ and the paper considered the linear (bivariate) rank test statistic in the form

$$T = \sum_{i=1}^n Z_i(\varphi(R_i) - \varphi(S_i)) \quad (5)$$

with a given score function φ and investigated its unbiasedness.

- Snijders [41] also considered rank tests (5) and derived locally most powerful (LMP) rank tests and their asymptotic normality using a Hájek-type theorem (as in Section 6.1

of [13]). Particularly, he proved (5) with Wilcoxon scores, i.e. the test statistic

$$W = \sum_{i=1}^n Z_i(R_i - S_i), \quad (6)$$

to be the LMP rank test assuming X_1 and X_2 to come from logistic distributions.

- Ernst and Schucany [10] focused on testing the null hypothesis formulated as (3). They proposed a bivariate test statistic so that one component compares the means between X_1 and X_2 and the other compares their variances. The test is based on the Mahalanobis distance of the test statistic from the origin.
- Modarres [32] proposed five tests of bivariate symmetry. The first three are obtained as two-sample tests of equality of distribution functions based on Euclidean interpoint distances; they compare the distribution of the raw data with that of the data reflected along the line $x_1 = x_2$:
 - Runs test (based on evaluating a minimum spanning tree).
 - Nearest neighbor test (application of the test of Henze [15] to bivariate symmetry).
 - Rank test of equality of multivariate distribution functions (application of the test of Maa *et al.* [28] to bivariate symmetry).
 - Sign test (based on dividing the observations to 6 regions, performed as a standard test about the probability of a binomial distribution).
 - Bootstrap test based on $\max\{\hat{F}_n(x_{1i}, x_{2i}) - \hat{F}_n(x_{2i}, x_{1i})\}$, where F_n denotes the empirical distribution function; the test requires bootstrap estimation of the null distribution of the test statistic.

Concerning the rank test, Modarres [32] described it to reject bivariate symmetry for large values of the test statistic M ; however, we can immediately think of a simplistic example, in which the test statistic decreases as we deviate from H_0 . Therefore, we recommend to perform the rank test as a two-sided test, rejecting H_0 for very large or very small test statistics. To justify this, violating bivariate symmetry implies that the test statistic is very small or very large. The null distribution of M is the same as that of the Wilcoxon rank-sum statistic for two samples, where each contains n observations. In analogy, the Wilcoxon rank-sum test (Wilcoxon two-sample test) based on interpoint distances, which was theoretically investigated in [22], is also rejected for very large or very small values of the test statistic.

- Rao and Raghunath [37] developed a non-parametric test for a more general situation of symmetry about a line (possibly different from the line $x_1 = x_2$). The test performs a tedious partition of the sample space to subjectively chosen sets (clusters) and the test statistic is obtained as a deviance measure comparing true and expected counts in the clusters.
- A bivariate symmetry test for competing risks in survival analysis was proposed in [9] and its saddle point approximation was derived in [1].

3. Tests of exchangeability of multivariate distributions

Because there seems to be no direct way for obtaining exchangeability tests for distributions with more than two dimensions, we propose to construct such tests by combining (dependent) permutation tests of bivariate symmetry applied to individual pairs of variables. The

proposed non-parametric combination methodology is presented in Section 3.1. For comparison, tests constructed from multiple comparisons procedures of Benjamini–Yekutieli or Bonferroni are also considered; these are described in Section 3.2. This paper concentrated on data coming from continuous distributions. For categorical data, which can be represented in square contingency tables, exchangeability of the distribution is known as complete symmetry and we refer to Section 10.7 of [2] for its treatment.

Using the notation of Section 1, H_0 (1) is replaced by a different null hypothesis

$$H_0^* : \mathcal{L}(X_j, X_k) = \mathcal{L}(X_k, X_j) \quad \text{for all } j = 1, \dots, p-1, k = j+1, \dots, p, \quad (7)$$

formulated for the total number $D = p(p-1)/2$ of all pairs of variables. We thus consider testing the composite null hypothesis H_0^* , which may be expressed as

$$H_0^* : \bigcap_{j=1}^{D-1} \bigcap_{k=j+1}^D H_{0jk}, \quad \text{where } H_{0jk} : \mathcal{L}(X_j, X_k) = \mathcal{L}(X_k, X_j), \quad (8)$$

against

$$H_1^* : \bigcup_{j=1}^{D-1} \bigcup_{k=j+1}^D H_{1jk}, \quad \text{where } H_{1jk} : H_{0jk} \text{ does not hold.} \quad (9)$$

However, because testing (7) is not equivalent to testing the null hypothesis (1), it may happen that exchangeability (1) fails to be true while pairwise exchangeability is fulfilled for each individual pair.

3.1. Multivariate permutation test

As bivariate symmetry tests are often performed as permutation tests [35], it is the most natural approach to exploit the non-parametric combination methodology using one of the approaches due to Fisher, Lipták or Tippett. In such approach, the dependence of the tests is implicitly taken into account by means of the permutation strategy (Section 1.2 of [6]), irrespective of the dependence relations. The approach, which is free of assumptions about the distribution of the data, yields a global (combined) p -value and may be interpreted as a multivariate permutation test.

The implementation of the non-parametric combination methodology is straightforward following Algorithm 1 formulated for a general p , based on combining the total number of $D = p(p-1)/2$ tests of bivariate symmetry for all pairs of variables. In (11), $\hat{Q}_d(z)$ is used instead of

$$\frac{1}{B} \sum_{b=1}^B \mathbb{1}[|T_d^{*(b)}| \geq z], \quad z > 0, d = 1, \dots, D, \quad (10)$$

where $\mathbb{1}$ denotes indicator function; the reason is a better ability of $\hat{Q}_d(z)$ to keep the probability of type I error in numerical experiments [6].

In the experiments with multivariate data, we use tests obtained by combining these bivariate symmetry tests:

- (1) The test of Hollander [18].
- (2) The likelihood ratio (LR) test of Bell and Haller [3]; this is the only test which is rejected for a small value of the test statistic.
- (3) The rank test of Snijders [41] with Wilcoxon scores.
- (4) The sign test of Modarres [32].
- (5) The nearest neighbor (NN) test of Modarres [32] based on interpoint distances.
- (6) The rank test of Modarres [32] based on interpoint distances (say ID-rank test) with Wilcoxon scores; as explained in Section 2, we (unlike the original test) reject for very large or for very small values of the test statistic.

Algorithm 1 can be described as an adaptation of the general algorithm of [6], using one of these combination functions as a special case:

Algorithm 1 Multivariate permutation test of exchangeability of a distribution with $p \geq 3$.

Input: Data $X = (X_{\bullet 1}, \dots, X_{\bullet p})^T \in \mathbb{R}^{n \times p}$, selected combination function ψ , constant $B \in \mathbb{N}$ (e.g. $B = 1000$)

Input: Selected test of bivariate symmetry based on test statistic (say) T

Output: The combined p -value (say) p^*

- 1: Compute the test statistics T_1, \dots, T_D for $D = p(p - 1)/2$, where each T_d is obtained as $T(X_{\bullet k}, X_{\bullet l})$ (i.e. across $k = 1, \dots, p - 1$ and $l = k + 1, \dots, p$)
- 2: **for** $b = 1$ to B **do**
- 3: Generate (independent) random permutations $(\pi(i, 1), \dots, \pi(i, p))^T$ of values of $(1, \dots, p)^T$ for $i = 1, \dots, n$
- 4: Consider the permuted data (say) $X^{*(b)} = (X_{\bullet 1}^{*(b)}, \dots, X_{\bullet p}^{*(b)})^T$ with elements $X_{ij}^{*(b)} = X_{i, \pi(i, j)} \in \mathbb{R}$ for each $i = 1, \dots, n$ and $j = 1, \dots, p$
- 5: Compute the test statistics $T_1^{*(b)}, \dots, T_D^{*(b)}$, where each $T_d^{*(b)}$ is obtained as $T(X_{\bullet k}^{*(b)}, X_{\bullet l}^{*(b)})$ (i.e. across $k = 1, \dots, p - 1$ and $l = k + 1, \dots, p$)
- 6: **end for**
- 7:

$$\hat{Q}_d(z) := \frac{1}{B + 1} \left(\frac{1}{2} + \sum_{b=1}^B \mathbb{1}[|T_d^{*(b)}| \geq z] \right), \quad z > 0, \quad d = 1, \dots, D \quad (11)$$

8: $q_d := \hat{Q}_d(T_d)$ for each $d = 1, \dots, D$

9: $q_d^{*(b)} := \hat{Q}_d(T_d^{*(b)})$ for each $d = 1, \dots, D$ and $b = 1, \dots, B$

10:

$$p^* := \frac{1}{B} \sum_{b=1}^B \mathbb{1}[|\psi(q_1^{*(b)}, \dots, q_D^{*(b)})| \geq |\psi(q_1, \dots, q_D)|] \quad (12)$$

- Fisher (omnibus) $\psi(q_1, \dots, q_D) = -2 \sum_d \log(q_d)$;
- Lipták $\psi(q_1, \dots, q_D) = \sum_d \Phi^{-1}(1 - q_d)$;
- Tippett $\psi(q_1, \dots, q_D) = \max_d \{1 - q_d\}$.

As discussed in Section 4.2.4 of [34], Tippett's combination is recommendable if only one or a few (but not all) sub-alternatives are true, and that of Lipták [27] performs well when all sub-alternatives are jointly true. Fisher's combination is the most popular in practice [12]; it is considered intermediate between the two others and thus suitable when no prior expectation is available. Finding the optimal combining function for given data appears, however, impossible (Section 4.2.2 of [34]).

Properties of the multivariate permutation tests of this section follow from the non-parametric combination methodology. Particularly, the constructed tests hold the probability of type I error and the distribution of the test statistics under H_0^* does not depend on the underlying distribution of the data. This is true in spite of the individual p -values being dependent and the three choices of ψ used here to fulfil the assumptions of Section 1.2 of [6]. If the individual (pairwise) tests are consistent, the multivariate permutation test is consistent. If the individual tests are unbiased, the multivariate permutation test is unbiased (Section 4.3 of [8,34]).

3.2. Testing exchangeability of a distribution based on multiple comparisons

For the sake of comparisons, we also consider exchangeability tests for distributions with dimensionality $p \geq 3$ based on multiple comparisons. Although the methods of this section are not primarily designed for the task and are intended for post hoc comparisons (after a global test of a composite null hypothesis), we use them here for constructing a global test of H_0^* based on (corrected) p -values of pairwise tests of bivariate symmetry. We use here the notation p_1, \dots, p_D for p -values of the pairwise tests, which will be arranged in ascending order as

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(D)}. \quad (13)$$

Benjamini–Yekutieli. The most popular multiple testing procedure keeping the false discovery rate (FDR), defined as the percentage of false positive tests (incorrectly rejecting the null hypothesis) among all significant tests, below the chosen level α is the approach of Benjamini and Hochberg (B-H) [4] for independent tests. An extension of Benjamini and Yekutieli (B-Y) [5] is suitable for (potentially) dependent statistics even if the structure of the dependence is not known. It is convenient to express the (global) test based on the B-Y procedure by means of the scheme

$$H_0^* \text{ is rejected} \iff \prod_{d=1}^D \mathbb{1} \left[p_{(d)} > \frac{d}{D} \left(\sum_{h=1}^D \frac{1}{h} \right)^{-1} \alpha \right] = 0. \quad (14)$$

Some other procedures for testing high-dimensional data controlling for FDR were presented in the overview [23].

The **Bonferroni** method represents the simplest approach to multiple tests ensuring to keep the family-wise error rate (FWER), defined as the probability of at least one incorrect rejecting the null hypothesis (i.e. making at least one type I error) among all tests, under the specified level α . We now use it to construct the (very conservative) global test according to the scheme

$$H_0^* \text{ is rejected} \iff p_{(1)} \leq \alpha/D. \quad (15)$$

Table 1. *P*-values of tests for the gene expressions (GE) dataset of Section 4 with $p = 3$ and $n = 24$.

Index	Hollander	LR Bell and Haller	Rank test Snijders	Sign test Modarres	NN Modarres	ID-rank test Modarres
Bivariate permutation tests						
1-2	0.768	0.078	0.160	0.779	0.112	0.143
1-3	0.287	0.051	0.286	0.835	0.109	0.116
2-3	0.083	0.001	0.024	0.131	0.060	0.024
Multivariate permutation tests						
Fisher	0.039	0.001	0.008	0.341	0.010	0.128
Lipták	0.293	0.000	0.008	0.561	0.014	0.591
Tippett	0.022	0.001	0.003	0.054	0.115	0.075
Bivariate permutation tests using Bonferroni correction						
*						
Bivariate permutation tests using the approach of Benjamini–Yekutieli						
*						

Notes: Bivariate permutation tests and multivariate permutation tests using the approaches of Fisher (F), Lipták (L), or Tippett (T) are presented here. Significance of tests based on Bonferroni correction and B-Y procedure is denoted by star.

4. Analysis of real datasets

We consider two real datasets to illustrate the performance of the tests of exchangeability of distributions with $p = 3$. All computations of this paper were performed in R software [36] exploiting additional packages (NSM3, spdep, purrr, FNN, and BioConductor).

The first dataset contains **gene expression** (GE) measurements acquired in the study described in [31]. Gene expressions for $p = 38\,950$ gene transcripts were measured on 24 individuals having a cerebrovascular stroke and 24 control persons. First, the Limma methodology (Linear Models for Microarray Data of [40]) was applied to find the most differentially expressed genes. The data are used here only for the $n = 24$ patients and with $p = 3$ most important genes that contribute the most to the separation between the two groups of individuals.

We are interested in testing whether the joint distribution of the three considered genes is exchangeable. The results of permutation tests of bivariate symmetry applied to individual pairs of variables (1-2, 1-3, and 2-3, i.e. with $D = 3$ in the notation of Algorithm 1) for the GE dataset are presented in Table 1. The multivariate permutation tests based on test statistics of bivariate symmetry tests are presented there as well. Tests of bivariate symmetry for the second and third gene have smaller p -values compared to tests evaluated for other pairs of genes. For all the tests, the multivariate permutation tests yield the smallest p -values, if a suitable combination method is chosen; however, the best of the three combination methods in terms of power turns out to be different for different bivariate symmetry tests. Most commonly (although not always), the approach of Tippett yields the best results here.

The second dataset is a subset of the **Australian athletes** (AA) dataset with $p = 3$. This dataset available, e.g. in the R software package DAAG [29] with the red blood cell count (X_1), white blood cell count (X_2), and hemoglobin concentration (X_3) was analyzed, e.g. in [16] or [24]. Graphical visualizations reveal the variables to be very far from being permutable; thus, we further consider 1000 random subsamples from the dataset, where each contains only $n = 20$ measurements. If we consider all tests presented in Table 1 for the GE dataset, the average p -values across the 1000 subsamples of the AA dataset are all highly significant and below 0.001. This also makes the global tests using Bonferroni correction and the B-Y procedure significant in all situations under consideration.

Table 2. Simulation A: Empirical rejection frequencies (in %).

	Hollander	LR Bell and Haller	Rank test Snijders	Sign test Modarres	NN Modarres	ID-rank test Modarres
Contamination proportion 0 (i.e. H_0^* holds)						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.055	0.052	0.046	0.042	0.059	0.054
1-3	0.047	0.055	0.044	0.048	0.055	0.050
2-3	0.053	0.051	0.051	0.049	0.063	0.058
Multivariate procedures						
Fisher	0.052	0.052	0.051	0.046	0.055	0.052
Lipták	0.048	0.050	0.052	0.049	0.054	0.061
Tippett	0.061	0.053	0.044	0.044	0.055	0.054
B-Y	0.011	0.014	0.010	0.004	0.009	0.012
Bonferroni	0.019	0.016	0.014	0.017	0.020	0.015
Contamination proportion 0.1						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.336	0.392	0.367	0.331	0.352	0.413
1-3	0.286	0.335	0.350	0.277	0.318	0.344
2-3	0.081	0.110	0.085	0.074	0.073	0.119
Multivariate procedures						
Fisher	0.412	0.463	0.428	0.403	0.407	0.446
Lipták	0.388	0.439	0.406	0.369	0.394	0.425
Tippet	0.449	0.501	0.460	0.427	0.445	0.508
B-Y	0.024	0.027	0.021	0.016	0.021	0.026
Bonferroni	0.031	0.029	0.023	0.026	0.031	0.028
Contamination proportion 0.2						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.774	0.830	0.742	0.728	0.695	0.846
1-3	0.729	0.806	0.714	0.711	0.647	0.813
2-3	0.122	0.163	0.113	0.091	0.100	0.169
Multivariate procedures						
Fisher	0.894	0.955	0.911	0.874	0.905	0.938
Lipták	0.825	0.921	0.863	0.856	0.857	0.912
Tippett	0.933	0.992	0.947	0.923	0.949	0.990
B-Y	0.998	0.999	0.934	0.432	0.897	0.776
Bonferroni	1.000	1.000	0.985	0.721	0.991	0.952

5. Simulations

The aim of the simulations is to compare the performance of the tests of exchangeability of a distribution proposed in Section 3 for testing the null hypothesis H_0^* against H_1^* . The performance of permutation tests of bivariate symmetry applied to pairs of variables is investigated as well. The simulations are performed for data generated from four different models. In simulations A, B, C, and E, we randomly generate $\kappa = 1000$ samples and for each of them, the permutation tests (bivariate or multivariate) are always performed with 1000 permutations.

Simulation A. Using $n = 20$, we independently generate 3-dimensional data from normal distribution $N_3(0, \mathcal{I}_3)$, where \mathcal{I}_3 denotes the unit matrix of size 3×3 . A selected percentage (ranging from 0 to 20%) of the observations is replaced by values, which are independently generated from $N_3(\mu, \mathcal{I}_3/10)$, where $\mu = (7, -1, 1)^T$. The results in the form of empirical rejection frequencies (rates) are presented in Table 2.¹

Simulation B. Using $n = 30$, we independently generate 3-dimensional data from multivariate t_1 distribution with scale matrix \mathcal{I}_3 . A selected percentage (ranging from 0 to 20%) of the observations is replaced by values, which are independently generated from non-central multivariate t_1 distribution with scale matrix $\mathcal{I}_3/10$ and with non-centrality

Table 3. Simulation B: Empirical rejection frequencies (in %).

	Hollander	LR Bell and Haller	Rank test Snijders	Sign test Modarres	NN Modarres	ID-rank test Modarres
Contamination proportion 0 (i.e. H_0^* holds)						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.053	0.055	0.046	0.057	0.054	0.058
1-3	0.046	0.052	0.051	0.049	0.049	0.050
2-3	0.055	0.044	0.053	0.053	0.052	0.051
Multivariate procedures						
Fisher	0.060	0.052	0.051	0.053	0.050	0.057
Lipták	0.053	0.055	0.054	0.060	0.044	0.052
Tippett	0.051	0.048	0.053	0.052	0.052	0.052
B-Y	0.010	0.011	0.008	0.007	0.010	0.011
Bonferroni	0.015	0.019	0.013	0.014	0.016	0.020
Contamination proportion 0.1						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.270	0.312	0.229	0.226	0.263	0.303
1-3	0.319	0.361	0.268	0.262	0.315	0.364
2-3	0.195	0.234	0.150	0.137	0.169	0.223
Multivariate procedures						
Fisher	0.366	0.431	0.365	0.312	0.362	0.411
Lipták	0.327	0.359	0.313	0.290	0.314	0.376
Tippett	0.419	0.472	0.385	0.375	0.389	0.454
B-Y	0.049	0.051	0.046	0.033	0.047	0.052
Bonferroni	0.044	0.046	0.041	0.038	0.045	0.049
Contamination proportion 0.2						
Permutation tests of bivariate symmetry (for individual pairs of variables)						
1-2	0.708	0.765	0.692	0.678	0.723	0.797
1-3	0.771	0.815	0.734	0.730	0.779	0.832
2-3	0.270	0.310	0.245	0.263	0.276	0.332
Multivariate procedures						
Fisher	0.764	0.818	0.713	0.712	0.769	0.826
Lipták	0.713	0.770	0.697	0.671	0.743	0.769
Tippett	0.814	0.855	0.756	0.730	0.813	0.867
B-Y	0.946	0.972	0.928	0.796	0.904	0.897
Bonferroni	0.962	0.980	0.941	0.848	0.925	0.941

parameter (i.e. mode) equal to $\mu = (12, 5, 0)^T$. The results in the form of empirical rejection frequencies are presented in Table 3.

Simulation C. Using $n = 30$, we independently generate 4-dimensional data from the multivariate logistic distribution [30] with the vector of location parameters $(1, 1, 1, 1)^T$ and the vector of scale parameters $(1, 1, 1, 1)^T$. A selected percentage (ranging from 0 to 25%) of the observations is replaced by values, which are independently generated from multivariate logistic distribution with the vector of location parameters $(6, 6, -1, -1)^T$ and the vector of scale parameters $(0.3, 0.3, 0.3, 0.3)^T$. The results in the form of averaged empirical rejection frequencies are visualized in Figure 1, where the horizontal lines correspond to 5%. The average rejection rates of bivariate symmetry tests applied to variables 1 and 3 are shown in Figure 1(left); testing for the pairs of variables 1-4, 2-3, and 2-4 yields analogous results due to the structure of the data. Average rejection rates of the multivariate permutation test using Tippett’s method is shown in Figure 1(middle) and using Lipták’s method in Figure 1(right).

Simulation D. Simulation D is aimed at comparing the methods for larger values of p . Using $n = 20$, we independently generate p -dimensional data from multivariate normal distribution $N_p(\mu, \mathcal{I}_p)$, where $\mu_i = \mathbb{1}[i \leq m]$ is considered for $i = 1, \dots, p$ with a given m .

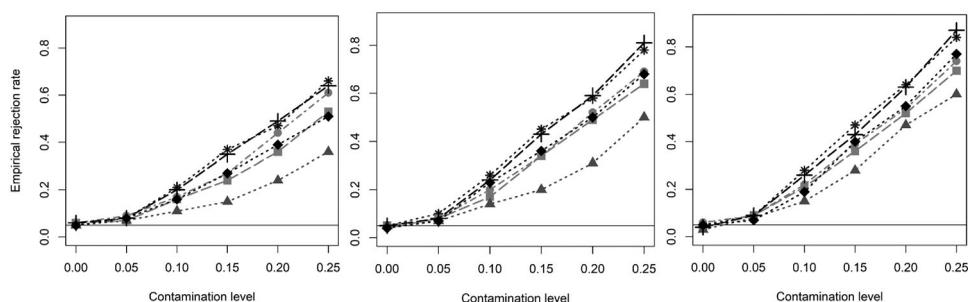


Figure 1. Simulation C: empirical rejection rates for the permutation test of bivariate symmetry for variables 1 and 3 (left), and for the multivariate permutation test using the approach of Fisher (middle) and Tippet (right). The curves correspond to (1) Hollander (stars), (2) LR (circles), (3) rank test of Snijders (squares), (4) sign test (triangles), (5) NN test (diamonds), and (6) ID-rank test (plus signs).

We use always 1000 permutations; $\kappa = 1000$ is used for $p = 10$, but only $\kappa = 100$ is used for $p = 100$. The results are presented in Table 4.

To discuss the results of all the simulations, the multivariate permutation tests hold the probability of type I error at the 5% level. The power of all the tests increases together with increasing contamination of the datasets, i.e. increases as the data become more distant from H_0 . To compare individual tests, the largest powers are obtained with Hollander's test and the ID-rank test (i.e. in the novel version as a two-sided ID-rank test as justified here in Section 2). It seems as the most interesting result that the computations confirm the powers of the multivariate permutation tests to outperform those of bivariate symmetry tests applied to an individual pair of variables. Concerning the non-parametric combination methodology, Tippet's approach yields the largest power and Fisher's is slightly weaker, leaving Lipták's approach behind.

The Benjamini–Yekutieli or Bonferroni approaches, which are primarily designed for post hoc comparisons, without any surprise reduce the type I errors and also attain lower powers. For $p = 3$, the global test based on the B-Y procedure compares $p_{(1)}$ with 0.009, $p_{(2)}$ with 0.018, and $p_{(3)}$ with 0.027, and the test based on the Bonferroni procedure has its probability of type I error equal to precisely 0.017. For a small deviation from H_0^* with $p = 3$, the powers of the B-Y or Bonferroni global tests turn out to be extremely low (much below those of the multivariate permutation tests), and rapidly increase with an increasing deviation from H_0^* . For a larger p , the probability of type I error of B-Y and Bonferroni drops even more, which is very apparent here for $p = 100$ with a corresponding $D = 4950$ pairs to be compared. Tippet's approach outperforms the B-Y or Bonferroni approaches in all considered situations under H_1^* for $p = 10$ as well as for $p = 100$.

Simulation E. In order to remind that the test procedures of Section 3.1 were formulated after replacing H_0 by H_0^* , we present one more simulation for data that violate H_0 and retain close to H_0^* as much as possible. To approximate such situation, we start with the non-contaminated data from Simulation A. These are modified to decrease the density around points $(-m, -m, -m)^T$, $(m, m, -m)^T$, $(-m, m, m)^T$, and $(m, -m, m)^T$ with $m = 1.5$ by finding always 3 closest observations to these points in terms of the Euclidean distance. These are replaced by data around points $(m, -m, -m)^T$, $(-m, m, -m)^T$, $(-m, -m, m)^T$, and (m, m, m) , which are generated as normally distributed with expectation given to these

Table 4. Simulation D: Empirical rejection frequencies (in %).

Method	Hollander	LR Bell and Haller	Rank test Snijders	Sign test Modarres	NN Modarres	ID-rank test Modarres
$p = 10, m = 0$ (i.e. H_0^* holds)						
Fisher	0.03	0.05	0.05	0.03	0.05	0.05
Lipták	0.05	0.04	0.06	0.04	0.05	0.05
Tippett	0.04	0.05	0.05	0.05	0.07	0.06
B-Y	0.00	0.00	0.00	0.00	0.00	0.00
Bonferroni	0.01	0.01	0.00	0.00	0.00	0.01
$p = 10, m = 0.1n$						
Fisher	0.62	0.64	0.59	0.56	0.60	0.62
Lipták	0.58	0.60	0.55	0.53	0.59	0.61
Tippett	0.70	0.67	0.63	0.58	0.66	0.68
B-Y	0.58	0.61	0.53	0.53	0.59	0.61
Bonferroni	0.53	0.55	0.52	0.50	0.54	0.56
$p = 10, m = 0.2n$						
Fisher	0.94	0.95	0.91	0.86	0.92	0.93
Lipták	0.91	0.91	0.87	0.83	0.89	0.91
Tippett	0.96	0.98	0.94	0.88	0.94	0.97
B-Y	0.90	0.91	0.86	0.81	0.86	0.89
Bonferroni	0.87	0.88	0.82	0.77	0.86	0.88
$p = 100, m = 0$ (i.e. H_0^* holds)						
Fisher	0.04	0.04	0.06	0.04	0.05	0.06
Lipták	0.04	0.03	0.06	0.05	0.04	0.05
Tippett	0.05	0.05	0.08	0.05	0.06	0.07
B-Y	0.00	0.00	0.00	0.00	0.00	0.00
Bonferroni	0.00	0.00	0.00	0.00	0.00	0.00
$p = 100, m = 0.1n$						
Fisher	0.65	0.69	0.59	0.59	0.61	0.64
Lipták	0.61	0.64	0.58	0.55	0.61	0.60
Tippett	0.72	0.67	0.62	0.62	0.68	0.69
B-Y	0.62	0.64	0.56	0.52	0.59	0.65
Bonferroni	0.57	0.60	0.59	0.53	0.56	0.59
$p = 100, m = 0.2n$						
Fisher	0.99	0.99	0.93	0.90	0.94	0.97
Lipták	0.95	0.96	0.91	0.87	0.91	0.93
Tippett	0.98	0.99	0.95	0.90	0.97	0.98
B-Y	0.93	0.92	0.89	0.84	0.90	0.93
Bonferroni	0.91	0.92	0.86	0.81	0.85	0.89

Table 5. Simulation E: Empirical rejection frequencies (in %).

	Hollander	LR Bell and Haller	Rank test Snijders	Sign test Modarres	NN Modarres	ID-rank test Modarres
Permutation tests of bivariate symmetry (for a given pair of variables)						
1-2	0.062	0.065	0.060	0.057	0.072	0.067
1-3	0.060	0.061	0.058	0.061	0.065	0.064
2-3	0.059	0.059	0.064	0.061	0.071	0.071
Multivariate procedures						
Fisher	0.125	0.138	0.119	0.106	0.107	0.139
Lipták	0.119	0.132	0.123	0.105	0.102	0.140
Tippett	0.138	0.141	0.131	0.117	0.116	0.144
B-Y	0.023	0.026	0.020	0.009	0.015	0.019
Bonferroni	0.028	0.029	0.025	0.032	0.029	0.024

points and with covariance matrix $\mathcal{I}_3/10$. The results in Table 5 show the pairwise tests to have their powers very slightly above 0.05 being outperformed by multivariate permutation tests. The results of the latter remain nevertheless quite low for such a strong violation of H_0 , so this specific design reveals the limitation of replacing H_0 by H_0^* .

6. Conclusions

While the literature seems void of exchangeability tests for distributions with $p \geq 3$, this paper investigates several possibilities for their construction based on combining bivariate symmetry tests performed over individual pairs of variables. We recommend to perform the multivariate tests as multivariate permutation tests obtained by the non-parametric combination methodology. The computations over the presented real as well as simulated multivariate data reveal the multiple testing procedures to be more powerful than a bivariate exchangeability test performed only over a single pair of variables.

Numerous available studies comparing the Bonferroni correction with B-H and/or B-Y approaches [44] have most often not considered the non-parametric combination methodology. This is because the non-parametric combination methodology is a specific approach tailor-made for combining permutation tests. Multivariate permutations tests keep the probability of type I error under the specified level. This does not, however, hold for tests based on the approaches of B-H, B-Y or Bonferroni, which are all primarily designed for post hoc comparisons; these approaches turn out in the simulations of Section 5 to be unsuitable for testing a composite (global) null hypothesis. All the multivariate tests presented in this paper are computationally demanding for larger values of p , which is a natural property of all tests based on permutations. Let S denote the complexity of computing an individual test statistic and B the number of permutations (as in Algorithm 1). The computational complexity of the multivariate permutation test, which can be expressed as $(D + 1)BS$, is fully comparable to that of the approaches of B-Y or Bonferroni, which both have the complexity of DBS .

Finally, we can say that the multivariate tests of this paper consider p -values as significance probabilities, as it is typical in statistical practice, not reflecting that they are actually random variables. A perspective approach based on expected p -values (EPV) developed in [38] and extended in [42,43] for the context of multiple testing seems, however, not to have been extended to the context of permutation tests (as EPV-based testing needs a specified null distribution of the test statistic and its known or estimated distribution under the alternative) and to the context with a composite alternative hypothesis.

Note

1. A confidence interval for a rejection frequency π based on its estimate $\hat{\pi}$ obtained in simulations may be computed using the standard error $SE(\hat{\pi}) = 1.96 \cdot \sqrt{\hat{\pi}(1 - \hat{\pi})/\kappa}$, if the lower bound is non-negative. When simulating $\kappa = 1000$ random samples, it approximately holds here that $SE(\hat{\pi}) \doteq 0.03$ for $\hat{\pi} \in (0.2, 0.8)$ and $SE(\hat{\pi}) \doteq 0.02$ for $\hat{\pi} \in (0.05, 0.2]$ or for $\hat{\pi} \in [0.8, 0.95)$.

Acknowledgments

The authors would like to thank Miroslav Šiman for the discussion. The authors received valuable input from the reviewers and an associate editor.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work is supported by the grants GA21-05325S (J. Kalina) and GA22-02067S (P. Janáček) of the Czech Science Foundation.

ORCID

Jan Kalina  <http://orcid.org/0000-0002-8491-0364>

References

- [1] E.F. Abd-Elfattah, *Bivariate symmetry tests for complete and competing risks data: A saddle point approach*, J. Stat. Comput. Simul. 87 (2017), pp. 1269–1275.
- [2] A. Agresti, *Categorical Data Analysis*, 2nd ed., Wiley, Hoboken, 2002.
- [3] C.B. Bell and H.S. Haller, *Bivariate symmetry tests: parametric and nonparametric*, Ann. Math. Stat. 40 (1969), pp. 259–269.
- [4] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing*, J. R. Stat. Soc. Ser. B. 57 (1995), pp. 289–300.
- [5] Y. Benjamini and D. Yekutieli, *The control of the false discovery rate in multiple testing under dependency*, Ann. Stat. 29 (2001), pp. 1165–1188.
- [6] S. Bonnini, L. Corain, M. Marozzi, and L. Salmaso, *Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R*, Wiley, New York, 2014.
- [7] D. Commenges, *Transformations which preserve exchangeability and application to permutation tests*, J. Nonparametr. Stat. 15 (2003), pp. 171–185.
- [8] L. Corain and L. Salmaso, *Nonparametric permutation and combination-based multivariate control charts with applications in microelectronics*, Appl. Stoch. Models Bus. Ind. 29 (2013), pp. 334–349.
- [9] J.V. Deshpande, *A test for bivariate symmetry of dependent competing risks*, Biom. J. 32 (2007), pp. 737–746.
- [10] M.D. Ernst and W.R. Schucany, *A class of permutation tests of bivariate interchangeability*, J. Am. Stat. Assoc. 94 (1999), pp. 273–284.
- [11] C. Genest, J. Nešlehová, and J.F. Quessy, *Tests of symmetry for bivariate copulas*, Ann. Inst. Stat. Math. 64 (2012), pp. 811–834.
- [12] R.A. Giancristofaro and C. Brombin, *Overview of nonparametric combination-based permutation tests for multivariate multi-sample problems*, Statistica 74 (2014), pp. 233–246.
- [13] J. Hájek, Z. Šidák, and P.K. Sen, *Theory of Rank Tests*, 2nd ed., Academic Press, San Diego, 1999.
- [14] M. Harder and U. Stadtmüller, *Testing exchangeability of copulas in arbitrary dimension*, J. Nonparametr. Stat. 29 (2017), pp. 40–60.
- [15] N. Henze, *A multivariate two-sample test based on the number of nearest neighbor type coincidences*, Ann. Stat. 16 (1988), pp. 772–783.
- [16] N. Henze, Z. Hlávka, and S.G. Meintanis, *Testing for spherical symmetry via the empirical characteristic function*, Statistics 48 (2014), pp. 1282–1296.
- [17] J.F. Hilton and L. Gee, *The size and power of the exact bivariate symmetry test*, Comput. Stat. Data Anal. 26 (1997), pp. 53–69.
- [18] M. Hollander, *A nonparametric test for bivariate symmetry*, Biometrika 58 (1971), pp. 203–212.
- [19] Š. Hudecová and M. Šiman, *Testing axial symmetry by means of directional regression quantiles*, Electron. J. Stat. 15 (2021), pp. 2690–2715.
- [20] Š. Hudecová and M. Šiman, *Testing symmetry around a subspace*, Stat. Pap. 62 (2021), pp. 2491–2508.
- [21] S. Janson, T. Konstantopoulos, and L. Yuan, *On a representation theorem for finitely exchangeable random vectors*, J. Math. Anal. Appl. 442 (2016), pp. 703–714.
- [22] J. Jurečková and J. Kalina, *Nonparametric multivariate rank tests and their unbiasedness*, Bernoulli 18 (2012), pp. 229–251.

- [23] J. Kalina, *Classification methods for high-dimensional data*, Biocybern. Biomed. Eng. 34 (2014), pp. 10–18.
- [24] J. Kalina, *Common multivariate estimators of location and scatter capture the symmetry of the underlying distribution*, Commun. Stat. Simul. 50 (2021), pp. 2845–2857.
- [25] O. Kallenberg, *Probabilistic Symmetries and Invariance Principles*, Springer, New York, 2005.
- [26] A.K. Kuchibhotla, *Exchangeability, conformal prediction, and rank tests*, preprint (2021). Available at <https://arxiv.org/abs/2005.06095v3>.
- [27] T. Lipták, *On the combination of independent tests*, Magyar Tud. Akad. Mat. Kutató Int. Közl. 3 (1958), pp. 171–197.
- [28] J.F. Maa, D.K. Pearl, and R. Bartoszyński, *Reducing multidimensional two-sample data to one-dimensional interpoint comparisons*, Ann. Stat. 24 (1996), pp. 1069–1074.
- [29] J.H. Maindonald and W.J. Braun, *DAAG: Data analysis and graphics data and functions*, R package version 1.22, 2015. Available at <https://cran.r-project.org/web/packages/DAAG>.
- [30] H.J. Malik and B. Abraham, *Multivariate logistic distributions*, Ann. Stat. 3 (1973), pp. 588–590.
- [31] M. Marozzi, A. Mukherjee, and J. Kalina, *Interpoint distance tests for high-dimensional comparison studies*, J. Appl. Stat. 47 (2020), pp. 653–665.
- [32] R. Modarres, *Tests of bivariate exchangeability*, Int. Stat. Rev. 76 (2008), pp. 203–213.
- [33] H. Oja, *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*, Springer, New York, 2010.
- [34] F. Pesarin and L. Salmaso, *Permutation Tests for Complex Data: Theory, Applications and Software*, Wiley, New York, 2010.
- [35] J.F. Quessy, *On consistent nonparametric statistical tests of symmetry hypotheses*, Symmetry 8 (2016), Article 31.
- [36] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, 2018. Available at <http://www.R-project.org>.
- [37] K.S.M. Rao and M. Raghunath, *A simple nonparametric test for bivariate symmetry about a line*, J. Stat. Plan. Inference 142 (2012), pp. 430–444.
- [38] H. Sackrowitz and E. Samuel-Cahn, *P-values as random variables: Expected p values*, Am. Stat. 53 (1999), pp. 326–3321.
- [39] R. Serfling, *Multivariate symmetry and asymmetry*, in *Encyclopedia of Statistical Sciences*, S. Kotz, N. Balakrishnan, C.B. Read, B. Vidakovic, eds., 2nd ed., Vol. 8, Wiley, New York, 2006, pp. 5338–5345.
- [40] G.K. Smyth, *Limma: Linear models for microarray data*, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber, eds., Springer, New York, 2005, pp. 397–420.
- [41] T. Snijders, *Rank tests for bivariate symmetry*, Ann. Stat. 9 (1981), pp. 1087–1095.
- [42] A. Vexler and J. Yu, *To t-test or not to t-test? A p-values-based point of view in the receiver operating characteristic curve framework*, J. Comput. Biol. 25 (2018), pp. 541–550.
- [43] A. Vexler, J. Yu, Y. Zhao, A.D. Hutson, and G. Gurevich, *Expected p-values in light of an ROC curve analysis applied to optimal multiple testing procedures*, Stat. Methods Med. Res. 27 (2018), pp. 3560–3576.
- [44] T. White, J. van der Ende, and T.E. Nichols, *Beyond Bonferroni revisited: Concerns over inflated false positive research findings in the fields of conservation genetics, biology, and medicine*, Conserv. Genet. 20 (2019), pp. 927–937.
- [45] T. Yanagimoto and M. Sibuya, *Test of symmetry of a bivariate distribution*, Sankhya A 38 (1976), pp. 105–115.