

A Multi-Aspect Permutation Test for Goodness-of-Fit Problems

 Rosa Arboretti ^{1,*}, Elena Barzizza ², Nicolò Biasetton ², Riccardo Ceccato ² , Livio Corain ² and Luigi Salmaso ²
¹ Department of Civil, Environmental and Architectural Engineering, University of Padova, 35131 Padova, Italy

² Department of Management and Engineering, University of Padova, 36100 Vicenza, Italy; elena.barzizza@phd.unipd.it (E.B.); nicolo.biasetton@phd.unipd.it (N.B.); riccardo.ceccato.1@unipd.it (R.C.); livio.corain@unipd.it (L.C.); luigi.salmaso@unipd.it (L.S.)

* Correspondence: rosa.arboretti@unipd.it

Abstract: Parametric techniques commonly rely on specific distributional assumptions. It is therefore fundamental to preliminarily identify the eventual violations of such assumptions. Therefore, appropriate testing procedures are required for this purpose to deal with a the goodness-of-fit (GoF) problem. This task can be quite challenging, especially with small sample sizes and multivariate data. Previous studies showed how a GoF problem can be easily represented through a traditional two-sample system of hypotheses. Following this idea, in this paper, we propose a multi-aspect permutation-based test to deal with the multivariate goodness-of-fit, taking advantage of the non-parametric combination (NPC) methodology. A simulation study is then conducted to evaluate the performance of our proposal and to identify the eventual critical scenarios. Finally, a real data application is considered.

Keywords: multi-aspect; NPC; goodness-of-fit



Citation: Arboretti, R.; Barzizza, E.; Biasetton, N.; Ceccato, R.; Corain, L.; Salmaso, L. A Multi-Aspect Permutation Test for Goodness-of-Fit Problems. *Stats* **2022**, *5*, 572–582. <https://doi.org/10.3390/stats5020035>

Academic Editor: Fulvia Mecatti

Received: 28 March 2022

Accepted: 13 June 2022

Published: 17 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this study, we propose a permutation-based methodology, based on multi-aspect testing, to address the goodness-of-fit problems when the sample size is small and numeric multivariate data are available.

Parametric techniques rely on specific assumptions about the distribution of the population from which the parameter of interest is drawn. When such assumptions are violated, inference can be highly unreliable. For this reason, appropriate tests need to be preliminarily conducted to detect the eventual departure from the required distribution.

Considering the multivariate scenario, in the literature, a few solutions have been proposed to evaluate the multivariate normality.

For example, Mardia [1] proposed a pair of solutions based on multivariate versions of skewness and kurtosis measures. Let us suppose that $Skew = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{ij}^3$, $Kurt = \frac{1}{n} \sum_{i=1}^n m_{ii}^2$, $m_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})' S^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$, $S = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$, $df = \frac{V(V+1)(V+2)}{6}$, $\mathbf{x} = \{x_{ij}, i = 1, \dots, n, j = 1, \dots, V\}$ is the sample of interest, n is the sample size, and V is the number of multivariate components. Equation (1) displays the proposed skewness test, while a kurtosis test is reported in Equation (2).

$$\frac{n}{6} Skew \sim X_{df}^2 \quad (1)$$

$$Kurt \sim N\left(V(V+2), \frac{8V(V+2)}{n}\right) \quad (2)$$

Given that for small samples, the power and the type I error could be violated, the author afterwards proposed a corrected version of the Skewness test, using $\frac{nc}{6} Skew$ as test statistic, where $c = \frac{(n+1)(n+3)(V+1)}{n(n+1)(V+1)-6}$.

A small sample size indeed makes multivariate goodness-of-fit problems quite challenging, given that we cannot rely on asymptotic properties. To deal with such circumstances, Arboretti et al. [2] proposed a permutation-based method relying on the nonparametric combination (NPC) methodology [3]. Given a random sample \mathbf{x} from a population with distribution \mathbf{f} and a theoretical distribution \mathbf{f}_0 , the authors followed the approach suggested by Friedman [4], addressing a goodness-of-fit problem as a two-sample equality in distribution problem. They drew an additional sample \mathbf{x}_0 from the theoretical distribution \mathbf{f}_0 , and used \mathbf{x} and \mathbf{x}_0 to test whether $\mathbf{f} = \mathbf{f}_0$. Arboretti et al. [2] also recommended the use of permutation tests, highlighting their non-parametric, distribution-free nature and their power.

Considering a generic equality in distribution problem, there are many aspects that can determine a difference and need to be monitored. We can have completely different distribution functions, but also differences in location, in variability or in shape parameters. For this reason, multi-aspect permutation tests represent a solution worth considering. These tests follow the idea proposed by Fisher in 1947, which affirmed that different tests can be adopted to evaluate different aspects of the same null hypothesis [5,6]. Multi-aspect testing procedures indeed are aimed at simultaneously testing multiple features of the same H_0 .

The NPC methodology allows us to easily extend such procedures to the multivariate scenario and thereafter to multivariate goodness-of-fit problems. In Section 2, we propose a possible extension, providing a detailed description of the underlying algorithm and of a possible competing technique. Then, Section 3 is devoted to the investigation of its performance through a simulation study. In Section 4, a real data application is proposed. Finally, in Section 5, we make conclusions about the conducted study.

2. Multi-Aspect Permutation Solution

Arboretti et al. [2] showed that goodness-of-fit problems can be easily converted into a two-sample equality in the distribution problem. For this reason, the nonparametric combination methodology can provide suitable and quite powerful solutions.

The NPC essentially requires three steps to be undertaken:

1. The decomposition of the global system of hypotheses (see System 3) into multiple sub-problems:

$$\begin{cases} H_0 : \mathbf{X} \stackrel{d}{=} \mathbf{X}_0 \\ H_1 : \mathbf{X} \neq \mathbf{X}_0 \end{cases} ; \quad (3)$$

2. The application of partial test statistics to address each sub-problem and computation of partial p -values.
3. The combination of partial p -values using appropriate combining functions to finally compute a global p -value to test System 3.

In a multivariate scenario, the first step implies the following decomposition:

$$\begin{cases} H_0 : \bigcap_{v=1}^V H_{0v} \\ H_1 : \bigcup_{v=1}^V H_{1v} \end{cases} \quad \text{where} \quad \begin{cases} H_{0v} : \mathbf{x}_v \stackrel{d}{=} \mathbf{x}_{0v} \\ H_{1v} : \mathbf{x}_v \neq \mathbf{x}_{0v} \end{cases}$$

where we create a sub-system of hypotheses for each of the V components \mathbf{x}_v of the multivariate outcome.

On the other hand, NPC-based solutions for multi-aspect testing [7–10] also require an initial decomposition of the system of hypotheses, defining a sub-problem for each aspect to be considered. For the sake of simplicity, in this study, we focus on three aspects and report the related sub-systems:

- location

$$\begin{cases} H_{0\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \\ H_{1\mu} : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \end{cases} ;$$

- variability

$$\begin{cases} H_{0\sigma} : \boldsymbol{\sigma} = \boldsymbol{\sigma}_0 \\ H_{1\sigma} : \boldsymbol{\sigma} \neq \boldsymbol{\sigma}_0 \end{cases} ;$$

- cumulative distribution function

$$\begin{cases} H_{0F} : \mathbf{F} = \mathbf{F}_0 \\ H_{1F} : \mathbf{F} \neq \mathbf{F}_0 \end{cases} .$$

To address the aforementioned multivariate goodness-of-fit problem, we propose a multivariate multi-aspect test, and therefore we need to combine the two different decompositions:

$$\begin{cases} H_{0\mu} : \bigcap_{v=1}^V H_{0\mu v} = \bigcap_{v=1}^V \mu_v = \mu_{0v} \\ H_{1\mu} : \bigcup_{v=1}^V H_{1\mu v} = \bigcup_{v=1}^V \mu_v \neq \mu_{0v} \end{cases} \tag{4}$$

$$\begin{cases} H_{0\sigma} : \bigcap_{v=1}^V H_{0\sigma v} = \bigcap_{v=1}^V \sigma_v = \sigma_{0v} \\ H_{1\sigma} : \bigcup_{v=1}^V H_{1\sigma v} = \bigcup_{v=1}^V \sigma_v \neq \sigma_{0v} \end{cases} \tag{5}$$

$$\begin{cases} H_{0F} : \bigcap_{v=1}^V H_{0Fv} = \bigcap_{v=1}^V F_v = F_{0v} \\ H_{1F} : \bigcup_{v=1}^V H_{1Fv} = \bigcup_{v=1}^V F_v \neq F_{0v} \end{cases} \tag{6}$$

For each individual aspect, we then identify a suitable test statistic. In particular, we detect differences in the following:

- For location, we use the absolute difference between sample means:

$$T_{\mu v} = |\bar{x}_v - \bar{x}_{0v}|$$

- For variability, we adopt the ratio of the two estimated variances s_1^2 and s_0^2 :

$$T_{\sigma v} = \max\left(\frac{s_1^2}{s_0^2}, \frac{s_0^2}{s_1^2}\right)$$

- For the cumulative distribution function, we use the Anderson–Darling test statistic [3]:

$$T_{Fv} = \sum_{i=1}^N [\hat{F}(z_{vi}) - \hat{F}_0(z_{vi})]^2 / \{\bar{F}(z_{vi})[1 - \bar{F}(z_{vi})]\}$$

where $\mathbf{z} = \{\mathbf{x}, \mathbf{x}_0\}$ is the pooled sample, n and m are individual sample sizes, $N = n + m$, $\hat{F}(t) = \sum_{i=1}^n \mathbb{I}(x_{vi} \leq t) / n$, $\hat{F}_0(t) = \sum_{i=1}^m \mathbb{I}(x_{0vi} \leq t) / m$, $\bar{F}(t) = \sum_{i=1}^N \mathbb{I}(z_{vi} \leq t) / N$, $\mathbb{I}(t) = \{1 \text{ if } t \text{ is TRUE}; 0 \text{ otherwise}\}$, and $t \in \mathcal{R}^1$.

The second step of the NPC methodology can thereafter be undertaken. We apply each test statistic to each univariate component of the outcome and compute the related partial p -values via multivariate permutation.

The adopted algorithm is as follows:

- Apply the three test statistics to the original pooled data set $\mathbf{z} = \{\mathbf{x}, \mathbf{x}_0\}$. Observed values $T_{\mu v}^o$, $T_{\sigma v}^o$, and T_{Fv}^o are achieved.
- For $b = 1, \dots, B$:
 - Shuffle rows of \mathbf{z} (i.e., the same permutation scheme is applied to each component), implicitly taking into account the existing correlation among variables.
 - Apply the three test statistics and retrieve $T_{\mu v}^b$, $T_{\sigma v}^b$, and T_{Fv}^b .

- Compute the partial p -values $\lambda'_{\mu v}$, $\lambda'_{\sigma v}$, and λ'_{Fv} , comparing the values of the observed test statistics to those of the permuted test statistics (e.g., $\lambda'_{\mu v} = \frac{\sum_b \mathbb{I}(T_{\mu v}^b \geq T_{\mu v}^o)}{(B+1)}$, $v = 1, \dots, V$), and their permutation distributions $\lambda^{b}_{\mu v}$, $\lambda^{b}_{\sigma v}$, and λ^{b}_{Fv} , $b = 1, \dots, B$ (for further details, see Pesarin and Salmaso [3]).

The last step consists of the combination of the partial p -values and the computation of the global p -value. To do that, the following procedure needs to be followed:

- For each aspect, apply a combining function $\phi(\cdot)$ to the V vectors of partial p -values and their permutation distributions to achieve second-order test statistics $T''_{\mu} = \phi(\lambda'_{\mu 1}, \dots, \lambda'_{\mu V})$, $T''_{\sigma} = \phi(\lambda'_{\sigma 1}, \dots, \lambda'_{\sigma V})$, $T''_F = \phi(\lambda'_{F1}, \dots, \lambda'_{FV})$ and their estimated distributions $T''_{\mu}{}^b = \phi(\lambda^{b}_{\mu 1}, \dots, \lambda^{b}_{\mu V})$, $T''_{\sigma}{}^b = \phi(\lambda^{b}_{\sigma 1}, \dots, \lambda^{b}_{\sigma V})$, $T''_F{}^b = \phi(\lambda^{b}_{F1}, \dots, \lambda^{b}_{FV})$, $b = 1, \dots, B$.
- Compute second-order p -values λ''_{μ} , λ''_{σ} , and λ''_F (and the related distributions $\lambda''_{\mu}{}^b$, $\lambda''_{\sigma}{}^b$, $\lambda''_F{}^b$, $b = 1, \dots, B$) comparing T''_a to the permuted values $T''_a{}^b$, $b = 1, \dots, B$, with $a \in \{\mu, \sigma, F\}$.
- Apply a combining function $\theta(\cdot)$ to the second-order p -values and their permutation distributions to achieve a third-order test statistic $T''' = \theta(\lambda''_{\mu}, \lambda''_{\sigma}, \lambda''_F)$ and its estimated distribution $T'''{}^b = \theta(\lambda''_{\mu}{}^b, \lambda''_{\sigma}{}^b, \lambda''_F{}^b)$, $b = 1, \dots, B$.
- Compute the global p -value λ''' comparing T''' to the permuted values $T'''{}^b$, $b = 1, \dots, B$.

The choice of the combining function can represent a key factor in determining the power of the proposed test. According to Pesarin and Salmaso [3], a combining function should satisfy four fundamental properties, which are as follows:

- It should be a non-increasing and possibly symmetric function;
- It should reach its supremum value even when only a single partial p -value attains 0;
- For each significance level α , the related critical value should be finite and lower than the supremum value;
- The rejection region of the resulting combined test should be convex.

Among the functions satisfying these requirements, we have the following:

- Fisher’s [11]:

$$\psi_{Fisher} : \mathbb{R}^K \longrightarrow \mathbb{R} \quad (\lambda_1, \dots, \lambda_K) \mapsto -2 \cdot \sum_{k=1}^K \log(\lambda_k)$$

- The truncated product method [12], a modification of Fisher’s combining function which generally helps in gaining power with highly dependent data [13]:

$$\psi_{TPM} : \mathbb{R}^K \longrightarrow \mathbb{R} \quad (\lambda_1, \dots, \lambda_K) \mapsto -2 \cdot \sum_{k=1}^K \log(\lambda_k) \cdot \mathbf{1}_{[0, \tau]}(\lambda_k)$$

- Tippett’s [14]:

$$\psi_{Tippett} : \mathbb{R}^K \longrightarrow \mathbb{R} \quad (\lambda_1, \dots, \lambda_K) \mapsto 1 - \min\{\lambda_1, \dots, \lambda_K\}$$

In this study, we decided to set $\theta(\cdot) = \psi_{Tippett}(\cdot)$ and to investigate the impact of choosing the truncated product method over Fisher’s $\psi_{Fisher}(\cdot)$ as the first combining function $\phi(\cdot)$ used to combine V vectors of p -values related to the V components of the multivariate outcome. A simulation study was indeed conducted to evaluate the power of our proposal, implementing two different versions of the test, one using $\phi(\cdot) = \psi_{TPM}(\cdot)$ (indicated as NPC—Fisher) and one using $\phi(\cdot) = \psi_{Fisher}(\cdot)$ (indicated as NPC—Truncated).

A Competing Method

For a better evaluation of the performance of the proposed method, we decided to consider a possible competing method. In particular, we focused on the two-sample energy tests introduced by Székely et al. [15] in order to deal with the equality in distribution in high-dimensional problems. The test statistic they suggested is

$$\epsilon = \frac{nm}{n+m} \left(\sum_{i=1}^n \sum_{i'=1}^m \|x_i - x_{i'}\| + \frac{1}{m^2} \sum_{i=1}^n \sum_{i'=1}^m \|x_{0i} - x_{0i'}\| - \frac{2}{nm} \sum_{i=1}^n \sum_{i'=1}^m \|x_i - x_{0i'}\| \right)$$

where $\|\cdot\|$ is the Euclidean distance. Large values of this statistic lead to rejection and in order to propose an adequate p -value, they rely on a permutation approach.

3. Simulation Study

In this study, we considered several different scenarios to accurately evaluate the performance of the proposed NPC-based approach.

Firstly, we decided to consider three different multivariate distributions:

- Multivariate normal distribution with mean μ and variance-covariance matrix Σ ;
- Multivariate log-normal distribution with the mean vector of the log of the distribution equal to μ and variance-covariance matrix of the log of the distribution equal to Σ ;
- Multivariate Student's t distribution with 3 degrees of freedom, location parameter μ and scale matrix Σ .

Data generation was conducted, taking advantage of the *rmvnorm* [16], *LaplacesDemon* [17] and *compositions* [18] packages implemented in R. The *energy* package [19] was adopted to apply the competing method, while R codes implementing the two versions of the multi-aspect test are available upon request.

Initially, the sizes of the observed and the theoretical samples were both fixed to 20, while we decided to consider two possible numbers of variables V , namely 6 and 10.

Under the null hypothesis, the observed and the theoretical distributions are expected to be the same. Therefore, under H_0 , we set the V -dimensional vector $\mu = [10, \dots, 10]$ and the $(V \times V)$ -dimensional matrix

$$\Sigma = \begin{bmatrix} 1 & \omega & \cdots & \omega \\ \omega & 1 & \cdots & \omega \\ \cdots & \cdots & \cdots & \cdots \\ \omega & \omega & \cdots & 1 \end{bmatrix}$$

for both samples. We considered three possible values of ω , i.e., 0, 0.25, and 0.5, to introduce different degrees of correlation.

Under the alternative hypothesis, we focused on scenarios where the observed and the theoretical distributions were different in terms of both the location and scale parameters. For the theoretical sample X_0 , the aforementioned μ and Σ were adopted. On the other hand, to generate the observed sample X , we used $\mu_1 = [9.5, \dots, 9.5]$ and

$$\Sigma_1 = \begin{bmatrix} 3 & \omega & \cdots & \omega \\ \omega & 3 & \cdots & \omega \\ \cdots & \cdots & \cdots & \cdots \\ \omega & \omega & \cdots & 3 \end{bmatrix}.$$

Again, three possible values of ω , i.e., 0, 0.25, and 0.5, were considered, using each time the same value for both the distributions.

Given the well-known properties of the NPC methodology, increasing the size m of the theoretical sample could allow us to increase the power of our procedure. To better illustrate such a phenomenon, we also tried to vary m in a final scenario, where for the theoretical sample, we use $\mu = [10, \dots, 10]$ and

$$\Sigma = \begin{bmatrix} 1 & \omega & \cdots & \omega \\ \omega & 1 & \cdots & \omega \\ \cdots & \cdots & \cdots & \cdots \\ \omega & \omega & \cdots & 1 \end{bmatrix},$$

and for the observed sample, we have $\mu_1 = [9.5, \dots, 9.5]$ and

$$\Sigma_1 = \begin{bmatrix} 3 & \omega & \cdots & \omega \\ \omega & 3 & \cdots & \omega \\ \cdots & \cdots & \cdots & \cdots \\ \omega & \omega & \cdots & 3 \end{bmatrix},$$

with $\omega = 0.25$. In particular, we considered three possible values for m (i.e., 20, 30, and 40), while keeping n fixed to 20.

It is worth noting that the current choice of test statistics is not ideal for situations where the observed and the theoretical samples differ in terms of correlation structure. This shortcoming is illustrated and further discussed through an additional simulation scenario, where for the theoretical sample we use $V = 6$, $\mu = [10, \dots, 10]$ and

$$\Sigma = \begin{bmatrix} 3 & \omega & \cdots & \omega \\ \omega & 3 & \cdots & \omega \\ \cdots & \cdots & \cdots & \cdots \\ \omega & \omega & \cdots & 3 \end{bmatrix}$$

with $\omega = 0.5$, while for the observed sample we have $\mu_1 = [10, \dots, 10]$ and

$$\Sigma_1 = \begin{bmatrix} 3 & \omega_1 & \cdots & \omega_1 \\ \omega_1 & 3 & \cdots & \omega_1 \\ \cdots & \cdots & \cdots & \cdots \\ \omega_1 & \omega_1 & \cdots & 3 \end{bmatrix}$$

with $\omega_1 = \{1.5, 2.0, 2.5\}$.

The number of simulation runs was set equal to 5000, while the number of permutations was equal to 2000.

Results and Discussion

Under the null hypotheses, both versions of the NPC-based test (using Fisher's combining function and the truncated product method with $\tau = 0.2$) and the energy test all kept the nominal level. Having fixed the significance level α to 1%, the rejection rates (i.e., the proportion of p -values less than or equal to α) are always quite close to 1%, with some slight random fluctuations (see Table 1).

Under the alternative hypothesis, we can appreciate some differences between the considered methods (see Table 2).

First of all, it appears that the choice of the combining function does not affect considerably the performance of the NPC-based multi-aspect test. This is probably due to the fact that two sequential combination steps are undertaken, mitigating the impact of the use of a specific function during the first step. However, when a larger number of variables and a multivariate log-normal distribution are considered, the solution adopting Fisher's combining function appears to be perform slightly better.

Both the Fisher and truncated product methods show reasonably high rejection rates and outperform the energy test under the vast majority of scenarios. However, it should be noted that for the multivariate log-normal distribution, this is not true. When considering this asymmetric distribution, the methods show similar performance when $V = 6$ and the energy test even has the highest rejection rates when a larger number of variables is considered (i.e., when $V = 10$).

Table 1. Rejection rates with significance level $\alpha = 1\%$ under the null hypothesis.

| ω | V | Method | Multivariate Normal | Multivariate Log-Normal | Multivariate Student's t |
|----------|----|---------------|---------------------|-------------------------|--------------------------|
| 0 | 6 | Energy | 0.010 | 0.009 | 0.010 |
| | | NPC—Fisher | 0.012 | 0.010 | 0.011 |
| | | NPC—Truncated | 0.012 | 0.009 | 0.010 |
| | 10 | Energy | 0.007 | 0.011 | 0.010 |
| | | NPC—Fisher | 0.008 | 0.009 | 0.011 |
| | | NPC—Truncated | 0.007 | 0.009 | 0.013 |
| 0.25 | 6 | Energy | 0.010 | 0.010 | 0.009 |
| | | NPC—Fisher | 0.010 | 0.008 | 0.009 |
| | | NPC—Truncated | 0.009 | 0.009 | 0.011 |
| | 10 | Energy | 0.008 | 0.012 | 0.009 |
| | | NPC—Fisher | 0.009 | 0.013 | 0.009 |
| | | NPC—Truncated | 0.008 | 0.014 | 0.009 |
| 0.5 | 6 | Energy | 0.009 | 0.010 | 0.011 |
| | | NPC—Fisher | 0.009 | 0.010 | 0.009 |
| | | NPC—Truncated | 0.009 | 0.009 | 0.009 |
| | 10 | Energy | 0.012 | 0.011 | 0.011 |
| | | NPC—Fisher | 0.009 | 0.012 | 0.011 |
| | | NPC—Truncated | 0.009 | 0.012 | 0.011 |

Table 2. Rejection rates with significance level $\alpha = 1\%$ under the alternative hypothesis varying ω .

| ω | V | Method | Multivariate Normal | Multivariate Log-Normal | Multivariate Student's t |
|----------|----|---------------|---------------------|-------------------------|--------------------------|
| 0 | 6 | Energy | 0.757 | 0.448 | 0.807 |
| | | NPC—Fisher | 0.965 | 0.453 | 0.976 |
| | | NPC—Truncated | 0.958 | 0.450 | 0.972 |
| | 10 | Energy | 0.949 | 0.677 | 0.970 |
| | | NPC—Fisher | 0.999 | 0.582 | 1.000 |
| | | NPC—Truncated | 0.999 | 0.558 | 1.000 |
| 0.25 | 6 | Energy | 0.647 | 0.425 | 0.734 |
| | | NPC—Fisher | 0.931 | 0.435 | 0.955 |
| | | NPC—Truncated | 0.926 | 0.430 | 0.946 |
| | 10 | Energy | 0.854 | 0.595 | 0.891 |
| | | NPC—Fisher | 0.990 | 0.551 | 0.997 |
| | | NPC—Truncated | 0.987 | 0.535 | 0.995 |
| 0.5 | 6 | Energy | 0.563 | 0.378 | 0.645 |
| | | NPC—Fisher | 0.877 | 0.380 | 0.908 |
| | | NPC—Truncated | 0.876 | 0.374 | 0.901 |
| | 10 | Energy | 0.709 | 0.479 | 0.770 |
| | | NPC—Fisher | 0.961 | 0.461 | 0.959 |
| | | NPC—Truncated | 0.953 | 0.455 | 0.955 |

It is then worth noting how a high correlation among variables appears to be detrimental to the power of the methods. All the considered tests show indeed considerably lower rejection rates when $\omega = 0.5$ with respect to the case $\omega = 0$. The performance of

the NPC-based solutions, however, remains reasonably good, even for $\omega = 0.5$ when the symmetrical distributions are considered.

Additionally, a higher number of informative variables appears to lead to an increase in power. This is a well-known property of NPC-based tests, called finite-sample consistency [20]. This means that whatever the sample size, a reasonably good power can be reached if a considerably high number of informative variables is available. This is a pretty useful property that introduces a potential solution to deal with the shortcomings posed by small-sample scenarios.

Table 3 allows us to appreciate the positive effect on power of an increase in the theoretical sample size m . In particular, we can see that the permutation-based solutions with $m = 40$ are able to outperform the competing method even for $V = 10$ and a multivariate log-normal distribution, i.e., the only case where the energy test was performing the best for smaller sample sizes. To further enhance the power of these tests, the user could therefore consider increasing the size of the theoretical sample, given that it can be freely chosen. However, when adopting such an approach, we should be aware that it could lead to a substantial increase in the computational burden.

Table 3. Rejection rates with significance level $\alpha = 1\%$ under the alternative hypothesis varying m .

| m | V | Method | Multivariate Normal | Multivariate Log-Normal | Multivariate Student's t |
|-----|-----|---------------|---------------------|-------------------------|----------------------------|
| 20 | 6 | Energy | 0.667 | 0.421 | 0.731 |
| | | NPC—Fisher | 0.941 | 0.434 | 0.958 |
| | | NPC—Truncated | 0.935 | 0.444 | 0.952 |
| | 10 | Energy | 0.847 | 0.572 | 0.911 |
| | | NPC—Fisher | 0.994 | 0.506 | 0.996 |
| | | NPC—Truncated | 0.990 | 0.492 | 0.995 |
| 30 | 6 | Energy | 0.711 | 0.432 | 0.800 |
| | | NPC—Fisher | 0.985 | 0.467 | 0.986 |
| | | NPC—Truncated | 0.984 | 0.456 | 0.982 |
| | 10 | Energy | 0.902 | 0.578 | 0.934 |
| | | NPC—Fisher | 0.998 | 0.523 | 1.000 |
| | | NPC—Truncated | 0.997 | 0.515 | 1.000 |
| 40 | 6 | Energy | 0.753 | 0.433 | 0.825 |
| | | NPC—Fisher | 0.993 | 0.524 | 0.993 |
| | | NPC—Truncated | 0.992 | 0.518 | 0.992 |
| | 10 | Energy | 0.929 | 0.580 | 0.959 |
| | | NPC—Fisher | 1.000 | 0.595 | 1.000 |
| | | NPC—Truncated | 1.000 | 0.582 | 1.000 |

Investigating a potential shortcoming due to the current choice of the test statistics, we noticed that the methods indeed fail at detecting differences in the correlation structure. This is true for both the NPC-based solutions and the energy test. Looking at Table 4, we can indeed see that the rejection rates are very close to the nominal level expected under H_0 . However, by including an additional test statistics specifically designed to detect differences in correlation, it could be possible to address even such a scenario [3].

Table 4. Rejection rates with significance level $\alpha = 1\%$ with differences in the correlation structure.

| ω_1 | V | Method | Multivariate Normal | Multivariate Log-Normal | Multivariate Student's t |
|------------|---|---------------|---------------------|-------------------------|--------------------------|
| 1.5 | 6 | Energy | 0.013 | 0.012 | 0.014 |
| | | NPC—Fisher | 0.010 | 0.013 | 0.012 |
| | | NPC—Truncated | 0.010 | 0.015 | 0.010 |
| 2.0 | 6 | Energy | 0.019 | 0.014 | 0.018 |
| | | NPC—Fisher | 0.012 | 0.009 | 0.010 |
| | | NPC—Truncated | 0.015 | 0.013 | 0.009 |
| 2.5 | 6 | Energy | 0.021 | 0.022 | 0.028 |
| | | NPC—Fisher | 0.012 | 0.009 | 0.014 |
| | | NPC—Truncated | 0.013 | 0.009 | 0.014 |

4. Real Data Application

We decided to consider a real data application in order to better show the usefulness of our proposed procedure. In particular, we applied our approach to an industrial problem, where an operator is interested in checking the quality of a production process in terms of multiple key performance indicators. Initially, 25 different bottles (i.e., the output of the process) were randomly selected. Their diameters measured on three key positions were expected to be, on average, equal to 2.5 cm (Diameter A), 5 cm (Diameter B) and 7 cm (Diameter C), respectively (i.e., $\mu_s = [2.5, 5.0, 7.5]$). Additionally, after an application of the Six Sigma methodology, we knew in advance that the expected variance–covariance matrix was as follows:

$$\Sigma_s = \begin{bmatrix} 1.02 \times 10^{-4} & 1.04 \times 10^{-5} & 2.01 \times 10^{-5} \\ 1.04 \times 10^{-5} & 1.12 \times 10^{-4} & 2.03 \times 10^{-5} \\ 2.01 \times 10^{-5} & 2.03 \times 10^{-5} & 2.05 \times 10^{-4} \end{bmatrix},$$

with the diameters values following a multivariate normal distribution.

The gathered sample showed a potential shift from the expected mean value in Diameter A (i.e., the diameter of the neck of the bottle) as we can see in Table 5. We therefore applied both the versions of the NPC-based multi-aspect test and the energy test to further investigate this hypothesis.

Table 5. Descriptive statistics.

| Value | Diameter A | Diameter B | Diameter C |
|----------|-----------------------|-----------------------|-----------------------|
| Average | 2.516 | 4.999 | 7.002 |
| Variance | 1.04×10^{-4} | 1.17×10^{-4} | 2.03×10^{-4} |

Table 6 reports the achieved global p -values. We can see that all the considered methods allow us to reject the null hypothesis with a significance level equal to 5%, which means that the gathered sample does not follow a multivariate normal distribution with mean μ_s and a matrix of variance and covariance Σ_s . Looking at adjusted partial p -values (see Table 7) we can also identify which aspects lead to this rejection. In particular, we can see that a significant shift in mean did happen.

Table 6. Global p -values.

| Energy | NPC—Fisher A | NPC—Truncated |
|-----------------------|-----------------------|-----------------------|
| 4.99×10^{-4} | 1.14×10^{-2} | 2.79×10^{-2} |

Table 7. Partial p -values.

| Test Statistics | NPC—Fisher A | NPC—Truncated |
|-----------------|--------------|---------------|
| T_μ | 0.014 | 0.019 |
| T_σ | 0.095 | 0.147 |
| T_F | 0.014 | 0.019 |

5. Conclusions

In this paper, we introduced a multi-aspect permutation test to deal with the multivariate goodness-of-fit (GoF). First of all, we adopted the approach already proposed by Arboretti et al. [2], transforming the GoF problem into a traditional two-sample one. Then, we simply introduced an extension of the nonparametric combination (NPC) methodology [3], which is able to detect differences in location, scale and cumulative distribution function between the observed sample distribution and the theoretical distribution.

To evaluate the performance of this solution, we proposed a simulation study, which allowed us to appreciate the goodness of our proposal, even when compared to a possible competing testing procedure (i.e., the energy test proposed by Székely et al. [15]). Its power appears to be negatively affected by high correlation among variables, but at the same time, it tends to substantially increase when the number of informative variables increases. It also emerged that the choice of the combining function adopted in the first combination step required by the NPC methodology does not appear to significantly affect the performance of the proposed test.

The conducted simulation study showed the benefits of choosing a large size m of the sample drawn from the theoretical distribution, which appears to lead to an increase in power. Future studies could therefore consider providing guidelines about the appropriate ratio between the observed and the theoretical samples sizes.

On the other hand, it emerged also a shortcoming of the current configuration of the proposed approach, which fails at detecting differences in the correlation structure. For this reason, future studies could focus on the introduction of a further test statistic, specifically designed to detect such differences, which could allow us to improve the performance under such scenarios.

A real data application was also proposed, which allowed us to show the usefulness of our approach.

Overall, our proposal demonstrated to be a quite powerful solution to goodness-of-fit problems, which shows high flexibility and leaves room for further improvement and investigation.

Author Contributions: Conceptualization, R.C., N.B. and E.B.; Methodology, R.C., N.B. and E.B.; Software, R.C., N.B. and E.B.; Validation, R.C., N.B. and E.B.; Formal Analysis, R.C., N.B. and E.B.; Investigation, R.C., N.B. and E.B.; Resources, R.C., N.B. and E.B.; Data Curation, L.C., L.S. and R.A.; Writing—Original Draft Preparation, L.C., L.S. and R.A.; Writing—Review and Editing, L.C., L.S. and R.A.; Visualization, L.C., L.S. and R.A.; Supervision, L.C., L.S. and R.A.; Project Administration, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mardia, K.V. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā Indian J. Stat. Ser. B* **1974**, *36*, 115–128.
2. Arboretti, R.; Ceccato, R.; Salmaso, L. Permutation testing for goodness-of-fit and stochastic ordering with multivariate mixed variables. *J. Stat. Comput. Simul.* **2021**, *91*, 876–896. <https://doi.org/10.1080/00949655.2020.1836182>.
3. Pesarin, F.; Salmaso, L. *Permutation Tests for Complex Data: Theory, Applications and Software*; John Wiley & Sons: Hoboken, NJ, USA, 2010.
4. Friedman, J. *On Multivariate Goodness-of-Fit and Two-Sample Testing*; SLAC National Accelerator Lab.: Menlo Park, CA, USA, 2004.
5. Fisher, R.A. *The Design of Experiments*, 4th ed.; Hafner Press: New York, NY, USA, 1947.
6. Lehmann, E.L. The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *J. Am. Stat. Assoc.* **1993**, *88*, 1242–1249.
7. Salmaso, L.; Solari, A. Multiple aspect testing for case-control designs. *Metrika* **2005**, *62*, 331–340.
8. Brombin, C.; Salmaso, L. Multi-aspect permutation tests in shape analysis with small sample size. *Comput. Stat. Data Anal.* **2009**, *53*, 3921–3931.
9. Brombin, C.; Salmaso, L.; Ferronato, G.; Galzignato, P.F. Multi-aspect procedures for paired data with application to biometric morphing. *Commun. Stat. Comput.* **2010**, *40*, 1–12.
10. Corain, L.; Salmaso, L. Improving power of multivariate combination-based permutation tests. *Stat. Comput.* **2015**, *25*, 203–214.
11. Fisher, R. *Statistical Methods for Research Workers*; Oliver and Boyd: Edinburgh, UK, 1932.
12. Zaykin, D.V.; Zhivotovsky, L.A.; Westfall, P.H.; Weir, B.S. Truncated product method for combining P-values. *Genet. Epidemiol. Off. Publ. Int. Genet. Epidemiol. Soc.* **2002**, *22*, 170–185.
13. Arboretti Giancristofaro, R.; Bonnini, S.; Corain, L.; Salmaso, L. Dependency and truncated forms of combinations in multivariate combination-based permutation tests and ordered categorical variables. *J. Stat. Comput. Simul.* **2016**, *86*, 3608–3619.
14. Tippett, L.H.C. *The Methods of Statistics. An Introduction Mainly for Workers in the Biological Sciences*; Williams & Norgate: London, UK, 1931.
15. Székely, G.J.; Rizzo, M.L. Testing for equal distributions in high dimension. *InterStat* **2004**, *5*, 1249–1272.
16. Genz, A.; Bretz, F.; Miwa, T.; Mi, X.; Leisch, F.; Scheipl, F.; Hothorn, T. *Mvtnorm: Multivariate Normal and t Distributions*; R Package Version 1.1-3; R Foundation for Statistical Computing: Vienna, Austria, 2021.
17. Statisticat LLC. *Bayesian Inference*; R Package Version 16.1.6; R Foundation for Statistical Computing: Vienna, Austria, 2021.
18. van den Boogaart, K.G.; Tolosana-Delgado, R.; Bren, M. *Compositions: Compositional Data Analysis*; R Package Version 2.0-4; R Foundation for Statistical Computing: Vienna, Austria, 2022.
19. Rizzo, M.; Székely, G. *Energy: E-Statistics: Multivariate Inference via the Energy of Data*; R Package Version 1.7-8; R Foundation for Statistical Computing: Vienna, Austria, 2021.
20. Pesarin, F.; Salmaso, L. Finite-sample consistency of combination-based permutation tests with application to repeated measures designs. *J. Nonparametric Stat.* **2010**, *22*, 669–684.