mathematics

MDPI

Article

A New Goodness of Fit Test for Multivariate Normality and Comparative Simulation Study

Jurgita Arnastauskaitė [1,2,*], Tomas Ruzgas [2] and Mindaugas Bražėnas [3]

[1] Department of Applied Mathematics, Kaunas University of Technology, 51368 Kaunas, Lithuania
[2] Department of Computer Sciences, Kaunas University of Technology, 51368 Kaunas, Lithuania; tomas.ruzgas@ktu.lt
[3] Department of Mathematical Modelling, Kaunas University of Technology, 51368 Kaunas, Lithuania; mindaugas.brazenas@ktu.lt
* Correspondence: jurgita.arnastauskaite@ktu.lt

Abstract: The testing of multivariate normality remains a significant scientific problem. Although it is being extensively researched, it is still unclear how to choose the best test based on the sample size, variance, covariance matrix and others. In order to contribute to this field, a new goodness of fit test for multivariate normality is introduced. This test is based on the mean absolute deviation of the empirical distribution density from the theoretical distribution density. A new test was compared with the most popular tests in terms of empirical power. The power of the tests was estimated for the selected alternative distributions and examined by the Monte Carlo modeling method for the chosen sample sizes and dimensions. Based on the modeling results, it can be concluded that a new test is one of the most powerful tests for checking multivariate normality, especially for smaller samples. In addition, the assumption of normality of two real data sets was checked.

Keywords: multivariate normality; power of tests; squared radii; skewness; kurtosis

Citation: Arnastauskaitė, J.; Ruzgas, T.; Bražėnas, M. A New Goodness of Fit Test for Multivariate Normality and Comparative Simulation Study. *Mathematics* **2021**, *9*, 3003. https://doi.org/10.3390/math9233003

Academic Editor: María Purificación Galindo Villardón

Received: 12 October 2021
Accepted: 19 November 2021
Published: 23 November 2021

1. Introduction

Much multivariate data is being collected by monitoring natural and social processes. IBM estimates that we all generate 175 zettabytes of data every day. To add, the data were collected at a rapidly increasing rate, i.e., it is estimated that 90% of data has been generated in the last two years. The need to extract useful information from continuously generated data sets drives demand for data specialists and the development of robust analysis methods.

Data analytics is inconceivable without testing the goodness of fit hypothesis. The primary task of a data analyst is to become familiar with the data sets received. This usually starts by identifying the distribution of the data. Then, the assumption that the data follow a normal distribution is usually tested. Since 1990, many tests have been developed to test this assumption, mostly for univariate data.

It is important to use the powerful tests for the goodness of fit hypothesis to test the assumption of normality because an alternative distribution is not known in general. Based on the outcome of normality verification, one can choose suitable analysis methods (parametric or non-parametric) for further investigation. From the end of the 20th century to the present day, multivariate tests for testing the goodness of fit hypothesis have been developed by a number of authors [1–14]. Some of the most popular and commonly used multivariate tests are Chi-Square [8], Cramer von Mises [2], Anderson-Darling [2], and Royston [3].

Checking the assumption of normality of multivariate data is more complex compared to univariate. Additional data processing is required (e.g., standardization). The development of multivariate tests is more complex because they require checking the properties of invariance and contingency. While for the univariate tests, the invariance property is

always satisfied. The properties of invariance, contingency are presented in Section 2 and are discussed in more detail in [2,12,15].

The study aims to perform a power analysis of the multivariate goodness of fit hypothesis tests for the assumption of normality, to find out proposed test performances compared to other well-known tests and to apply the multivariate tests to the real data. The power estimation procedure is discussed in [16].

Scientific novelty. The power analysis of multivariate goodness of fit hypothesis testing for different data sets was performed. The goodness of fit tests were selected as representatives of popular techniques, which had been analyzed by other researchers experimentally. In addition, we proposed a new multivariate test based on the mean absolute deviation of the empirical distribution density from the theoretical distribution density. In this test, the density estimate is derived by using an inversion formula which is presented in Section 3.

The rest of the paper is organized as follows. Section 2 defines the tests for the comparative multivariate test power study. Section 3 presents details of our proposed test. Section 4 presents the data distributions used for experimental test power evaluation. Section 5 presents and discusses the results of simulation modeling. Section 6 discusses the application of multivariate goodness of fit hypothesis tests to real data. Finally, the conclusions and recommendations are given in Section 7.

2. Multivariate Tests for Normality

We denote the p-variate normal distribution as $N_p(\mu, \sum)$, where μ is an expectation vector $\mu = (\mu_1, \ldots, \mu_p)^T$ and $\sum$ is the nonsingular covariance matrix. N_p indicates a set of all possible p-variate normal distributions. Let $X_1, X_2, \ldots, X_n$, where $X_k = \left(X_{k1}, X_{k2}, \ldots, X_{kp}\right)^T$ and $k = 1, 2, \ldots, n$, be a finite sample generated by a random p-variate (column) vector X with distribution function F_X. The mean vector $\overline{X}$ is given by $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, where n is the sample size and the sample covariate matrix is

$$S = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})(X_j - \overline{X})^T.$$

To assess multivariate normality of X (based on the observed sample $X_1, X_2, \ldots, X_n$) a lot of statistical tests have been developed. Before reviewing specific tests, selected for this study, let us consider two essential properties. The set N_p is closed with respect to affine transformations, i.e.,

$$F_{AX+b} \in N_p \Leftrightarrow F_X \in N_p,$$

for any translation vector $b \in \mathbb{R}^p$ and any nonsingular matrix $A \in \mathbb{R}^{p \times p}$. Thus, a reasonable statistic T_n for checking the null hypothesis (H_0) of multivariate normality should have the same value for a sample and its affine transforms, that is

$$T_n(AX_1 + b, \ldots, AX_n + b) = T_n(X_1, \ldots, X_n). \quad (1)$$

An invariant test has a statistic, which satisfies the condition (1). It might seem that a test based on a standardized sample

$$Y_j = S^{-\frac{1}{2}}(X_j - \overline{X}),$$

is invariant, however Henze and Zirkler [2] note that this is not always the case. In practice, for a given sample $X_1, X_2, \ldots, X_n$ the alternative distribution is not know. In such a case it is important to use a test for which the probability of correctly rejecting H_0 tends to one as $n \to \infty$. Such a test is said to be consistent. For more elaborate discussion on these properties we refer the reader to [2]. Other important denotes are given in Appendix A.

2.1. Tests Based on Squared Radii

This section reviews the properties of several measures of squared radii concerning their use for assessing multivariate normality. Squared radii are defined as

$$D_{n,j} = \left(X_j - \overline{X}\right)^T S^{-1} \left(X_j - \overline{X}\right),\ j = 1,\ 2, .., n.$$

$D_{n,j}$ have a distribution which, under normality, is $(n-1)^2/n$ times a $Beta\left(\frac{p}{2}, \frac{n-p-1}{2}\right)$ distribution [9]. Under H_0, the distribution of $D_{n,j}$ is approximately χ^2_p for large n.

2.1.1. Chi-Squared (CHI2)

In 1981, Moore and Stubblebine presented multivariate Chi-Squared goodness of fit test based on order statistics [8]. The statistic of the test is defined as

$$M_{n,k} = \frac{k}{n} \sum_{l=1}^{k} \left(N_{n,l} - \frac{n}{k}\right)^2, \tag{2}$$

where $N_{n,l} = \sum\limits_{j=1}^{n} 1\left\{a_{l-1} < D_{n,j} \leq a_l\right\},\ (l = 1, 2, \ldots, k;\ a_0 = 0,\ a_k = +\infty)$. Since $M_{n,k}$ takes the equivalent form [8]:

$$M_{n,k} = k \sum_{l=1}^{k} (G_n(a_l) - G_n(a_{l-1}))^2,$$

where $G_p(\cdot)$ is the probability distribution function of $\chi^2(p)$. $G_p(a_l) - G_p(a_{l-1}) = k^{-1}$ $\left(l = 1, 2, \ldots, k; G_p(+\infty) = 1\right)$.

2.1.2. Cramer-Von Mises (CVM)

In 1982, Koziol proposed the use of Cramer-von Mises-type multivariate goodness of fit test based on order statistics [2]. This test statistic is defined as

$$CM = \frac{1}{12n} + \sum_{j=1}^{n} \left(G_p\left(D_{(j)}\right) - \frac{2j-1}{n}\right)^2, \tag{3}$$

where $D_{(j)},\ j = 1, 2, \ldots, n$ is order statistics.

2.1.3. Anderson-Darling (AD)

In 1987, Paulson, Roohan and Sullo proposed the Anderson-Darling type multivariate goodness of fit test based on order statistics [2]. The test statistic is defined as

$$AD = -n - \sum_{j=1}^{n} \frac{2j-1}{n} \left(\log G_p\left(D_{(j)}\right) + \log\left(1 - G_p\left(D_{(n+1-j)}\right)\right)\right). \tag{4}$$

2.2. Tests Based on Skewness and Kurtosis

This section reviews the properties of several measures of multivariate skewness and kurtosis regarding their use as statistics for assessing multivariate normality [2]. The skewness and kurtosis are defined as

$$s = \frac{m_3}{\sqrt{m_2}^3},\ k = \frac{m_4}{m_2^2},$$

where $m_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^i \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

2.2.1. Doornik-Hansen (DH)

In 2008, Doornik-Hansen proposed a new multivariate goodness of fit test based on the skewness and kurtosis of multivariate data transformed to ensure independence [6]. The Doornik-Hansen test statistic is defined as the sum of squared transformations of the skewness and kurtosis. Approximately, the test statistic follows a χ^2 distribution

$$DH = Z_1^T Z_1 + Z_2^T Z_2 \sim \chi^2(2p), \tag{5}$$

where $Z_1^T = (z_{11}, \ldots, z_{1p})$ and $Z_2^T = (z_{21}, \ldots, z_{2p})$ are defined as

$$Z_1 = \delta \log\left(y + \sqrt{y^2 - 1}\right) \text{ and } Z_2 = \sqrt{9\alpha}\left(\frac{1}{9\alpha} - 1 + \sqrt[3]{\frac{\chi}{2\alpha}}\right)$$

where

$$\delta = \frac{1}{\sqrt{log(w^2)}},\ w^2 = -1 + \sqrt{2(\beta - 1)},\ \beta = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},\ y = s\sqrt{\frac{(w^2-1)(n+1)(n+3)}{12(n-2)}},$$

$$\alpha = a + c \times s^2,\ a = \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta},\ c = \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta},$$

$$\delta = (n-3)(n+1)\left(n^2+15n-4\right),\ \chi = 2l\left(k - 1 - s^2\right),\ l = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta}.$$

2.2.2. Royston (Roy)

In 1982, Royston proposed a test that uses the *Shapiro-Wilk*/*Shapiro-Francia* statistic to test multivariate normality. If the kurtosis of the sample is greater than 3, then it uses the *Shapiro-Francia* test for leptokurtic distributions. Otherwise it uses the *Shapiro-Wilk* test for platykurtic distributions [3,5]. Let $\mathcal{W}_j$ be the *Shapiro-Wilk*/*Shapiro-Francia* test statistic for the jth variable ($j = 1, 2, \ldots, d$) and Z_j be the values obtained from the normality transformation [3,5].

$$\text{if } 4 \le n \le 11,\ \ x = n \text{ and } \mathcal{W}_j = -\log\left[\gamma - \log\left(1 - \mathcal{W}_j\right)\right],$$

$$\text{if } 12 \le n \le 2000,\ \ x = \log(n) \text{ and } \mathcal{W}_j = \log\left(1 - \mathcal{W}_j\right).$$

Thus, it are observed that x and $\mathcal{W}_j$ change with the sample size. The transformed values of each random variable are obtained by [3,5]

$$Z_j = \frac{\mathcal{W}_j - l}{\sigma},$$

where γ, l and σ are derived from the polynomial approximations. The polynomial coefficients are provided for different [3,5]:

$$\gamma = a_{0\gamma} + a_{1\gamma}x + a_{2\gamma}x^2 + \ldots + a_{d\gamma}x^d,$$

$$l = a_{0l} + a_{1l}x + a_{2l}x^2 + \ldots + a_{dl}x^d,$$

$$\log(\sigma) = a_{0\sigma} + a_{1\sigma}x + a_{2\sigma}x^2 + \ldots + a_{d\sigma}x^d.$$

The Royston's test statistic for multivariate normality is defined as

$$H = \frac{e\sum_{j=1}^{p}\psi_j}{p} \sim \chi_e^2, \tag{6}$$

where e is the equivalent degrees of freedom, $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution such that,

$$e = \frac{p}{[1 + (p-1)\bar{c}]},$$

$$\psi_j = \left\{ \Phi^{-1}\left[\frac{\Phi(-Z_j)}{2} \right] \right\}^2, \; j = 1, \; 2, \ldots, d.$$

Let R be the correlation matrix and r_{ij} is the correlation between ith and jth observations. Then, the extra term $\bar{c}$ is found by

$$\bar{c} = \sum_i \sum_j \frac{c_{ij}}{p(p-1)}, \; \{c_{ij}\}_{i \neq j},$$

where

$$c_{ij} = \begin{cases} g(r_{ij}, n) \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}.$$

When $g(0, n) = 0$ and $g(1, n) = 1$, then $g(\cdot)$ can be defined as

$$g(r, n) = r^{\varrho}\left[1 - \frac{l}{v}(1-r)^l\right],$$

where l, ϱ and v are the unknown parameters, which are estimated by Ross modeling [4]. It was found that $l = 0.715$ and $\varrho = 5$ for sample size $10 \leq n \leq 2000$ and v is a cubic function

$$v(n) = 0.21364 + 0.015124(\log(n))^2 - 0.0018034(\log(n))^3.$$

2.2.3. Mardia (Mar1 and Mar2)

In 1970, K.V. Mardia proposed a new multivariate goodness of fit test based on skewness and kurtosis. The statistic for this test is defined as [17]

$$\begin{aligned} M_S(s) &= \tfrac{n \cdot s}{6} \xrightarrow{p} \chi^2\left(\tfrac{p(p+1)(p+2)}{6}\right), \\ M_k(k) &= \tfrac{n(k - p(p+2))^2}{8p(p+2)} \xrightarrow{d} \chi^2(1). \end{aligned} \tag{7}$$

2.3. Other Tests

This section reviews the properties of several measures of non-negative functional distance, a covariance matrix and Energy distance concerning their use as statistics for assessing multivariate normality. A non-negative functional distance that measures the distance between two functions is defined as

$$D_h(P, Q) = \int \left|\hat{P}(t) - \hat{Q}(t)\right|^2 \varphi_h(t) dt,$$

where $\hat{P}(t)$ is the characteristic function of the multivariate standard normal, $\hat{Q}(t)$ is the empirical characteristic function of the standardised observations, $\varphi_h(t)$ is a kernel (weighting) function

$$\varphi_h(t) = \left(2\pi h^2\right)^{-p/4} e^{\frac{-t^T t}{2h^2}},$$

where $t \in \mathbb{R}^p$ and $h \in \mathbb{R}$ is a smoothing parameter that needs to be selected [10].

2.3.1. Energy (Energy)

In 2013, G. Szekely and M. Rizzo introduced a new multivariate goodness of fit test based on Energy distance between multivariate distributions. The statistic for this test is defined as [18]

$$\mathcal{E}_n = n\left(\frac{2}{n}\sum_{j=1}^{n}\mathbb{E}||\widetilde{Y}_{n,j}-N_1||-\mathbb{E}||N_1-N_2||-\frac{1}{n^2}\sum_{j,k=1}^{n}||\widetilde{Y}_{n,j}-\widetilde{Y}_{n,k}||\right), \tag{8}$$

where $\widetilde{Y}_{n,j} = \sqrt{n/(n-1)}Y_{n,j}$, $Y_{n,j} = S_n^{-\frac{1}{2}}(X_j - \overline{X}_n)$, $j = 1,\ldots,n$ is called scattering residues. N_1 and N_2 are independent randomly distributed vectors according to the normal distribution. $\mathbb{E}||N_1 - N_2|| = 2\Gamma\left(\frac{p+1}{2}\right)/\Gamma\left(\frac{p}{2}\right)$, where $\Gamma(\cdot)$ is a Gamma function. The null hypothesis is rejected when $\mathcal{E}_n$ acquires large values.

2.3.2. Lobato-Velasco (LV)

In 2004, I. Lobato and C. Velasco improved the Jarque and Bera test and applied it to stationary processes. The statistic for this test is defined as [19]

$$\mathcal{G} = \frac{n\hat{\mu}_3^2}{6\hat{F}^{(3)}} + \frac{n(\hat{\mu}_4 - 3\hat{\mu}_2)^2}{24\hat{F}^{(4)}}, \tag{9}$$

where $\hat{F}^{(k)} = \sum_{t=1-n}^{n-1}\hat{\psi}(t)\left[\hat{\psi}(t) + \hat{\psi}(n-|t|)\right]^{k-1}$ is an auto-covariance function.

2.3.3. Henze-Zirkler (HZ)

In 1990, Henze and Zirkler introduced the HZ test [1]. The statistic for this test is defined as

$$HZ = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}e^{-\frac{h^2}{2}\mathcal{D}_{ij}} - 2\left(1+h^2\right)^{-\frac{p}{2}}\sum_{i=1}^{n}e^{-\frac{h^2}{2(1+h^2)}\mathcal{D}_i} + n\left(1+2h^2\right)^{-\frac{p}{2}}, \tag{10}$$

where $\mathcal{D}_{ij} = (X_i - X_j)^T S^{-1}(X_i - X_j)$, $\mathcal{D}_i = (X_i - \overline{X})^T S^{-1}(X_i - \overline{X})$.

$\mathcal{D}_i$ gives the squared Mahalanobis distance of ith observation to the centroid and $\mathcal{D}_{ij}$ gives the Mahalanobis distance between ith and jth observations. If the sample follows a multivariate normal distribution, the test statistic is approximately log-normally distributed with mean [1]

$$1 - \frac{a^{-\frac{p}{2}}\left(1 + ph^{\frac{2}{a}} + \left(p(p+2)h^4\right)\right)}{2a^2},$$

and variance [1]

$$2\left(1+4h^2\right)^{-\frac{p}{2}} + \frac{2a^{-p}\left(1+2ph^4\right)}{a^2} + \frac{3p(p+2)h^8}{4a^4} - 4w_h{}^{-\frac{p}{2}}\left(1 + \frac{3ph^4}{2w_h} + \frac{p(p+2)h^8}{2w_h{}^2}\right),$$

where $a = 1 + 2h^2$ and $w_h = \left(1+h^2\right)\left(1+3h^2\right)$. Henze and Zirkler also proposed an optimal choice of the parameter h in using HZ in the p-variate case as [1]

$$h^* = \frac{1}{\sqrt{2}}\left(\frac{n(2p+1)}{4}\right)^{\frac{1}{p+4}}.$$

A drawback of the Henze-Zirkler test is that, when H_0 is rejected, the possible violation of normality is generally not straightforward. Thus, many biomedical researchers would prefer a more informative and equally or more powerful test than the Henze-Zirkler test [5].

2.3.4. Nikulin-Rao-Robson (NRR) and Dzhaparidze-Nikulin (DN)

In 1981, Moore and Stubblebine suggested a multivariate Nikulin-Rao-Robson (NRR) goodness of fit test [7,8]. This test statistic for a covariance matrix of any dimension is defined as

$$Y_n^2 = \sum V_i^2 + \frac{2pr(\sum V_i p_i)^2}{1 - 2pr\sum p_i^2}, \tag{11}$$

where V_i is a vector of standardized cell frequencies with components

$$V_i = V_{in}(\hat{\theta}_n) = \frac{(N_{in} - n/r)}{\sqrt{n/r}},\ i = 1, \ldots, r,$$

where N_{in} is the number of random vectors $X_1, \ldots, X_n$ falling into $E_{in}(\hat{\theta}_n),\ i = 1, \ldots, r$. Then the limiting covariance matrix of standardized frequencies is $V_n(\hat{\theta}_n) = \sum_l = I - qq^T - BJ^{-1}B^T$, where B is the $r \times m$ matrix with elements

$$B_{ij} = \frac{1}{\sqrt{p_i(\theta)}}\frac{\partial p_i(\theta)}{\partial \theta_j},\ i = 1, \ldots, r,\ j = 1, \ldots, m,$$

where q is a r-vector with its entries as $1/\sqrt{r}$, $\mathbb{mm} = p + p(p+1)/2$ is the number of unknown parameters, $J = J(\theta)$ is the Fisher information matrix of size $\mathbb{mm} \times \mathbb{mm}$ for one observation which evaluated as

$$J(\theta) = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix},$$

where Q is the $p(p+1)/2 \times p(p+1)/2$ covariance matrix of w (a vector of the entries of $\sqrt{nS}$ arranged column-wise by taking the upper triangular elements) [7]:

$$w = \left(s_{11}, s_{12}, s_{22}, s_{13}, s_{23},\ s_{33}, \ldots, s_{pp}\right)^T.$$

The second term of Y_n^2 recovers information lost due to data grouping. Another useful decomposition of Y_n^2 is defined as

$$Y_n^2 = U_n^2 + S_n^2,$$

where U_n^2 is the multivariate statistic defined by Dzhaparidze and Nikulin (1974) [7]. It is defined as

$$U_n^2 = V_n^T(\hat{\theta}_n)\left[I - B_n\left(B_n^T B_n\right)^{-1}B_n^T\right]V_n(\hat{\theta}_n), \tag{12}$$

and in 1985, McCulloch presented a multivariate test statistic [7]:

$$S_n^2 = Y_n^2 - U_n^2 = V_n^T(\hat{\theta}_n)B_n\left[\left(J_n - B_n^T B_n\right)^{-1} + \left(B_n^T B_n\right)^{-1}\right]B_n^T V_n(\hat{\theta}_n).$$

If rank $B = s$, then U_n^2 and S_n^2 are asymptotically independent and distributed in the limit as χ^2_{r-s-1} and χ^2_s, respectively.

3. The New Test

Our test is based on distribution distance and has been derived using an inversion formula. The estimation of a sample distribution density is based on application of the characteristic function and inversion formula. This method is known for its good properties (i.e., low sensitivity) and has been introduced in [20]. Marron and Wand [21] carried out an extensive comparison of density estimation methods (including the adapted kernel method) and concluded that density estimation based on application of characteristic function and inversion is more accurate for non-Gaussian data sets.

The random p-variate vector $X \in \mathbb{R}^p$, which follows a distribution of a mixture model has a density function

$$f(X) = f(X, \theta) = \sum_{k=1}^{q} p_k f_k(X, \theta_k), \tag{13}$$

where q is the number of clusters (i.e., components, classes) of the mixture, and p_k $(k = 1, \ldots, q)$ is the a priori probability which satisfy

$$p_k > 0, \ \sum_{k=1}^{q} p_k = 1. \tag{14}$$

The $f_k(X, \theta_k)$ is a distribution of the kth class and θ is a set of parameters $\theta = \{p_1, \ldots, p_q, \theta_1, \ldots, \theta_q\}$. We denote the p-variate sample of independent and identically distributed random values X.

When examining approximations of parametric methods, it should be emphasized that as the data dimension increases, the number of model parameters increases rapidly, making it more difficult to find accurate parameter estimates. It is much easier to find density of univariate data projections

$$x_\tau = \tau^T x, \tag{15}$$

than multivariate data density f because of mutually unambiguous compliance.

$$f \leftrightarrow \{f_\tau, \ \tau \in \mathbb{R}^p\}. \tag{16}$$

It is quite natural to try to find the multivariate density f using the density estimates $\hat{f}_\tau$ of univariate observational projections [20]. In case of Gaussian mixture model, the projection of the observations (15) is also distributed according to the Gaussian mixture model:

$$f_\tau(x) = f_\tau(x, \theta_\tau) = \sum_{k=1}^{q} p_{k,\tau} \varphi_{k,\tau}(x), \tag{17}$$

where $\varphi_{k,\tau}(x) = \varphi\left(x; \ m_{k,\tau}, \ \sigma^2_{k,\tau}\right)$ is univariate Gaussian density. The parameter set θ of the multivariate mixture and the distribution parameters of the data projections $\theta_\tau = \left(p_{k,\tau}, \ m_{k,\tau}, \ \sigma^2_{k,\tau}\right)$, $k = 1, \ldots, q$ are related by equations:

$$\begin{array}{c} p_{j,\tau} = p_j, \\ m_{j,\tau} = \tau^T M_j, \\ \sigma^2_{j,\tau} = \tau^T \mathcal{R}_j \tau. \end{array} \tag{18}$$

The inversion formula is used

$$f(x) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-it^T x} \psi(t) dt, \tag{19}$$

where

$$\psi(t) = \mathrm{E} e^{it^T x}, \tag{20}$$

where $\psi(t)$ denotes the characteristic function of the random variable X. Given that $u = |t|$, $\tau = t/|t|$ and by changing the variables to a spherical coordinate system we obtain

$$f(x) = \frac{1}{(2\pi)^p} \int_{\tau: \ |\tau|=1} ds \int_0^\infty e^{-iu\tau^T x} \psi(u\tau) u^{p-1} du, \tag{21}$$

where the first integral is the surface integral of the unit sphere. The characteristic function of the projection of the observed random variable is

$$\psi_\tau(u) = \mathrm{E}e^{iu\tau^T X}, \tag{22}$$

and has the property

$$\psi(u\tau) = \psi_\tau(u). \tag{23}$$

By selecting the set T of uniform distributed directions on the sphere and replacing the characteristic function with its estimate, a density estimate is obtained [20,22]:

$$\hat{f}(x) = \frac{A(p)}{\#T} \sum_{\tau \in T} \int_0^\infty e^{-iu\tau^T x} \hat{\psi}_\tau(u) u^{p-1} e^{-hu^2} du, \tag{24}$$

where $\#T$ denotes a size of set T. Using the p-variate ball volume formula

$$V_p(\mathcal{R}) = \frac{\pi^{\frac{p}{2}} \mathcal{R}^p}{\Gamma\left(\frac{p}{2}+1\right)} = \begin{cases} \frac{\pi^{\frac{p}{2}} \mathcal{R}^p}{\left(\frac{p}{2}\right)!}, & \text{when } p \bmod 2 \equiv 0, \\ \frac{2^{\frac{p+1}{2}} \pi^{\frac{p-1}{2}} \mathcal{R}^p}{p!!}, & \text{when } p \bmod 2 \equiv 1, \end{cases} \tag{25}$$

the constant $A(p)$ defined as

$$A(p) = \frac{\left(V_p(1)\right)'_{\mathcal{R}}}{(2\pi)^p} = \frac{p 2^{-p} \pi^{-\frac{p}{2}}}{\Gamma\left(\frac{p}{2}+1\right)}. \tag{26}$$

Computer simulation studies have shown that the density estimates obtained using the inversion formula are not smooth. Therefore, in Formula (24), an additional multiplier e^{-hu^2} is used. This multiplier smoothes the estimate $\hat{f}(x)$ with the Gaussian kernel function. Moreover, this form of the multiplier allows the integral value to be calculated analytically. Monte Carlo studies have shown that its use significantly reduces the error of estimates. Formula (24) can be used to estimate the characteristic function of the projected data. Let us consider two approaches. The first one is based on the density approximation of the Gaussian distribution mixture model. In this case, the parametric estimate of the characteristic function is used:

$$\hat{\psi}_\tau(u) = \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} e^{iu\hat{m}_{k,\tau} - u^2 \hat{\sigma}^2_{k,\tau}/2}. \tag{27}$$

By substituting $\hat{\psi}_\tau(u)$ in (24) by (27), we get

$$\begin{aligned} \hat{f}(x) &= \tfrac{A(p)}{\#T} \textstyle\sum_{\tau \in T} \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} \int_0^\infty e^{iu(\hat{m}_{k,\tau} - \tau^T x) - u^2(h + \hat{\sigma}^2_{k,\tau}/2)} u^{p-1} du \\ &= \tfrac{A(p)}{\#T} \textstyle\sum_{\tau \in T} \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} I_{p-1}\left(\frac{\hat{m}_{k,\tau} - \tau^T x}{\sqrt{\hat{\sigma}^2_{k,\tau} + 2h}}\right) \left(\sqrt{\hat{\sigma}^2_{k,\tau} + 2h}\right)^{-p}, \end{aligned} \tag{28}$$

where

$$I_j(y) = \mathcal{R}e\left[\int_0^\infty e^{iyt - t^2/2} t^j dt\right]. \tag{29}$$

We note, that only the real part of the expression is considered here (the sum of the imaginary parts must be equal to zero) in other words, the density estimate $\hat{f}(x)$ can acquire only the real values. The chosen form of the smoothing multiplier e^{-hu^2} allows relating the smoothing parameter h with the variances of the projection clusters, i.e., in the calculations the variances are simply increased by $2h$. Next, the expression (29) is evaluated.

Let

$$C_j(y) = \int_0^\infty \cos(yt) \cdot e^{-t^2/2} \cdot t^j dt, \tag{30}$$

$$S_j(y) = \int_0^\infty \sin(yt)\cdot e^{-t^2/2}\cdot t^j dt, \tag{31}$$

then (29) can be written as

$$\int_0^\infty e^{-iyt-t^2/2} t^j dt = C_j(y) + iS_j(y). \tag{32}$$

By integrating in parts, we get

$$\begin{gathered} C_j(y) = e^{-\frac{t^2}{2}} t^{j-1}\cos(yt)\Big|_0^\infty + \int_0^\infty e^{-\frac{t^2}{2}}\left((j-1)t^{j-2}\cos(yt) - yt^{j-1}\sin(yt)\right)dt = \\ 1_{\{j=1\}} + (j-1)C_{j-2}(y) - yS_{j-1}(y),\ j \geq 1. \end{gathered} \tag{33}$$

$S_j(y)$ is expressed analogously. With respect to the limitations of the j index, the following recursive equations are obtained:

$$C_j(y) = (j-1)C_{j-2}(y) - yS_{j-1}(y),\ j \geq 2, \tag{34}$$

$$C_1(y) = 1 - yS_0(y), \tag{35}$$

$$S_j(y) = (j-1)S_{j-2}(y) - yC_{j-1}(y),\ j \geq 2, \tag{36}$$

$$S_1(y) = YC_0(y). \tag{37}$$

The initial function $S_0(y)$ is founded by starting with the relation

$$(S_0(y))'_y = \int_0^\infty t\cos(yt)\cdot e^{-t^2/2} dt = C_1(y). \tag{38}$$

From (35) and (38) it follows that S_0 satisfies the differential equation

$$S'_0(y) = 1 - yS_0(y), \quad S_0(0) = 0, \tag{39}$$

which is solved by writing down S_0 as the Taylor series:

$$S'_0(y) = \sum_{l=0}^\infty c_{l+1}(l+1)y^{l+1} = 1 - \sum_{l=2}^\infty c_{l-1}y^l. \tag{40}$$

By equating the coefficients of the same powers, its values are obtained:

$$c_0 = 0,\ c_1 = 1, c_l = -c_{l-2}/l,\ l \geq 2, \tag{41}$$

which gives us

$$S_0(y) = \sum_{l=0}^\infty \frac{(-1)^l y^{2l+1}}{(2l+1)!!} = y - \frac{y^3}{3!!} + \frac{y^5}{5!!} - \frac{y^7}{7!!} + \ldots. \tag{42}$$

C_0 is found from expression (30):

$$\begin{gathered} C_0(y) = \int_0^\infty \cos(yt)\cdot e^{-t^2/2} dt = \frac{1}{2}\int_{-\infty}^\infty \cos(yt)\cdot e^{-t^2/2} dt \\ = \frac{1}{2}\int_{-\infty}^\infty (\cos(yt) - i\sin(yt))\cdot e^{-t^2/2} dt = \sqrt{\frac{\pi}{2}} e^{-y^2/2}. \end{gathered} \tag{43}$$

The value of the integral (24) then is

$$I_j(y) = C_j(y). \tag{44}$$

One of the disadvantages of the inversion formula method (defined by (24)) is that the Gaussian distribution mixture model (13) described by this estimate (for $f_k = \varphi_k$) does not represent density accurately, except around observations. When approximating

the density under study with a mixture of Gaussian distributions, the estimation of the density using the inversion formula often becomes complicated due to a large number of components. Thus, we merge components with small a priori probabilities into one noise cluster.

We have developed and examined a modification of the algorithm which is based on the use of a multivariate Gaussian distribution mixture model. The parametric estimate of the characteristic function of uniform distribution density is defined as

$$\hat{\psi}(u) = \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}, \tag{45}$$

in the inversion Formula (19). In the density estimate calculation Formula (24), the estimation of the characteristic function is constructed as a union of the characteristic functions of a mixture of Gaussian distributions and uniform distribution with corresponding a priori probabilities:

$$\hat{\psi}_\tau(u) = \sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} e^{iu\hat{m}_{k,\tau} - u^2\hat{\sigma}^2_{k,\tau}/2} + \hat{p}_{0,\tau} \frac{2}{(b-a)u} \sin\left(\frac{(b-a)u}{2}\right) \cdot e^{\frac{iu(a+b)}{2}}, \tag{46}$$

where the second member describes uniformly distributed noise cluster, $\hat{p}_0$—noise cluster weight, $a = a(\tau)$, $b = b(\tau)$. Based on the established estimates of the parameters of the uniform distribution and data projections, it is possible to define the range

$$a = \left(\tau^T x\right)_{\min} - \frac{\left(\tau^T x\right)_{\max} - \left(\tau^T x\right)_{\min}}{2(n-1)}, \tag{47}$$

$$b = \left(\tau^T x\right)_{\max} + \frac{\left(\tau^T x\right)_{\max} - \left(\tau^T x\right)_{\min}}{2(n-1)}. \tag{48}$$

By inserting (46) to (24) we obtain

$$\begin{array}{c} \hat{f}(x) = \frac{A(p)}{\#T} \sum_{\tau \in T} \left[\sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} \int_0^\infty e^{iu(\hat{m}_{k,\tau} - \tau^T x) - u^2(h + \hat{\sigma}^2_{k,\tau}/2)} u^{p-1} du \right. \\ \left. + \frac{2\hat{p}_{0,\tau}}{b-a} \int_0^\infty e^{iu(\frac{a+b}{2} - \tau^T x) - u^2 h} \cdot \sin\left(\frac{(b-a)u}{2}\right) \cdot u^{p-2} du \right]. \end{array} \tag{49}$$

Using notations such as (28), we define the density estimate as

$$\begin{array}{c} \hat{f}(x) = \frac{A(p)}{\#T} \sum_{\tau \in T} \left[\sum_{k=1}^{\hat{q}_\tau} \hat{p}_{k,\tau} I_{p-1}\left(\frac{\hat{m}_{k,\tau} - \tau^T x}{\sqrt{\hat{\sigma}^2_{k,\tau} + 2h}}\right) \left(\hat{\sigma}^2_{k,\tau} + 2h\right)^{-\frac{p}{2}} \right. \\ \left. + \frac{2\hat{p}_{0,\tau}}{b-a} J_{p-2}\left(\frac{a+b-2\tau^T x}{2\sqrt{2h}}, \frac{b-a}{2\sqrt{2h}}\right) \cdot (2h)^{-\frac{p-1}{2}} \right], \end{array} \tag{50}$$

where $I_j(y)$ is given in (29) which is evaluated by (44) and

$$J_j(y,t) = \mathcal{R}e\left[\int_0^\infty e^{iyu - u^2/2} \cdot \sin(tu) \cdot u^j du\right]. \tag{51}$$

By integrating, we get

$$\begin{array}{c} \int_0^\infty e^{iyu - \frac{u^2}{2}} \cdot \sin(tu) \cdot u^j du = \int_0^\infty (\cos(yu) + i\sin(yu)) \cdot \sin(tu) \cdot e^{-\frac{u^2}{2}} \cdot u^j du = \\ \int_0^\infty \left(\frac{\sin((y+t)u) + \sin((t-y)u)}{2} + i\frac{\cos((y-t)u) - \cos((y+t)u)}{2}\right) \cdot e^{-\frac{u^2}{2}} \cdot u^j du = \frac{1}{2}S_j(y+t) + \\ \frac{1}{2}S_j(t-y) + i\frac{1}{2}C_j(y-t) - i\frac{1}{2}C_j(y+t), \end{array} \tag{52}$$

where $S_j(y)$ and $C_j(y)$ are defined in (30) and (31). Then the integral (51) evaluates to

$$J_j(y,t) = \frac{1}{2}S_j(y+t) + \frac{1}{2}S_j(t-y). \tag{53}$$

The above procedure is called *a modified inversion formula density estimate*. Our proposed normality test is based on the distance function

$$\mathcal{T} = \int_{\mathbb{R}^p} \left| f(z) - \hat{f}(z) \right| dG(z), \tag{54}$$

where z is a standardized value, $\hat{f}(z)$ is an estimate of density function.

The choice of *G(z)* (54) is influenced by three aspects [23]:

- *G(z)* assigns high weight where $|f(z) - \hat{f}(z)|$ is large, *f(z)* pertaining to the alterative hypothesis. the distribution density is related to the alternative hypothesis.
- *G(z)* gives high weight where the $\hat{f}(z)$ is a relatively precise estimator of *f(z)*.
- *G(z)* is such that the integral (54) has a closed form.

For the *distribution free* method, the first two aspects are fulfilled by adequately selecting the smoothness parameter h, in addition it yields a closed (54) integral form

$$\mathcal{T} = n^{-1} \sum_{t=1}^{n} \left| f(z_t) - \hat{f}(z_t) \right|. \tag{55}$$

$\mathcal{T}$ does not depend on a moderate sample volume ($\geq$32) but depends on the data dimension. It is convenient to use the test statistics $\mathcal{T}^* = -log(\mathcal{T})$ which had the lowest sensitivity based on the exploratory study. Under the null hypothesis statistic $\mathcal{T}^*$ approximately follows the Johnson S_U distribution which is specified by the shape ($\delta > 0$, γ), scale ($\lambda > 0$), location (ξ) parameters and has the density function

$$f(\mathbb{X}) = \frac{\delta}{\lambda\sqrt{2\pi}} g'\left(\frac{\mathbb{X}-\xi}{\lambda}\right) \exp\left(-0.5\left[\gamma + \delta g\left(\frac{\mathbb{X}-\xi}{\lambda}\right)\right]^2\right), \text{ for } \mathbb{X} \in (-\infty, +\infty).$$

where $g(y) = \ln\left[y + \sqrt{y^2+1}\right]$, $g'(y) = \frac{1}{\sqrt{y^2+1}}$.

In the middle of the twentieth century, N. L. Johnson [24] proposed certain systems of curve derived by the method of translation, which, retain most of the advantages and eliminate some of the drawbacks of the systems first based on this method. Johnson introduced log-normal (S_L), bounded (S_B), and unbounded (S_U) systems. The bounded system range of variation covers the area between the bounding line $\beta_2 - \beta_1 - 1 = 0$ and the Pearson Type III distribution; where (β_1, β_2) points are obtained from the distribution moments defined by Wicksell [25]:

$$\mu'_r(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(r(z-\gamma))/\delta} e^{-\frac{1}{2}z^2} dz = e^{\frac{1}{2}r^2\delta^{-2} - r\gamma\delta^{-1}}.$$

It follows that

$$\beta_1 = \left(e^{\delta^{-2}} - 1\right)\left(e^{\delta^{-2}} + 2\right)^2, \ \left(\sqrt{\beta_1} > 0\right),$$

$$\beta_2 = \left(e^{\delta^{-2}}\right)^4 + 2\left(e^{\delta^{-2}}\right)^3 + 3\left(e^{\delta^{-2}}\right)^2 - 3.$$

The S_U system is bounded at one end only (Pearson Type V). The S_L system is lying between S_B and S_U systems. These regions are indicated in Figure 1. The S_U system is presented in detail in [24].

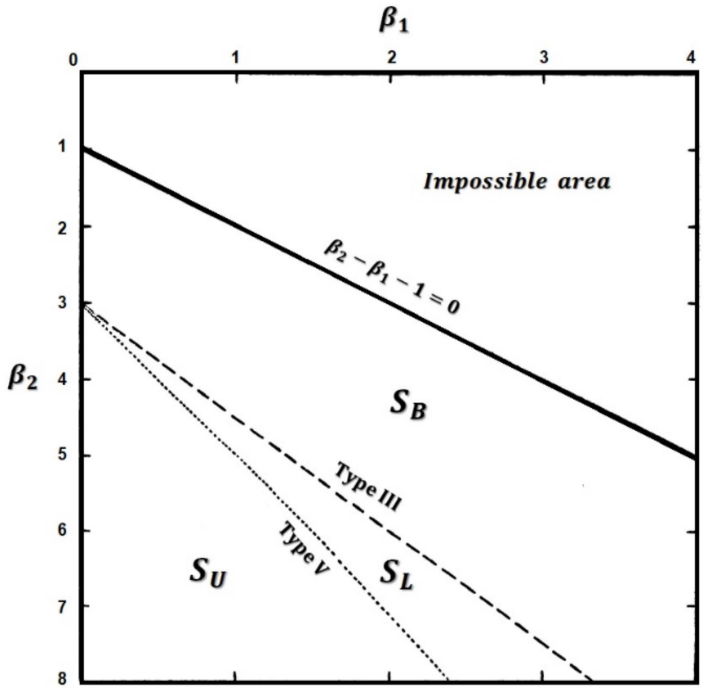

Figure 1. Regions of Johnson's systems.

Estimates of $\mathcal{T}^*$ statistic Johnson S_U distribution parameters for different dimensions are given in Table 1.

Table 1. Statistic $\mathcal{T}^*$ Johnson S_U distribution parameter estimates.

Parameter	Symbol	Estimate
	$p = 2$	
Location	$\hat{\xi}$	4.342807
Scale	$\hat{\lambda}$	0.585038
Shape	$\hat{\delta}$	1.498293
Shape	$\hat{\gamma}$	0.764906
	$p = 5$	
Location	$\hat{\xi}$	7.025845
Scale	$\hat{\lambda}$	0.088023
Shape	$\hat{\delta}$	0.895003
Shape	$\hat{\gamma}$	0.400035
	$p = 10$	
Location	$\hat{\xi}$	5.195174
Scale	$\hat{\lambda}$	1.578613
Shape	$\hat{\delta}$	2.24856
Shape	$\hat{\gamma}$	−1.83037

For statistic $\mathcal{T}^*$, the invariance and contingency properties were checked. The invariance property is confirmed because standardized data was used. The contingency property is confirmed experimentally (see Section 5).

4. Statistical Distributions

The overviewed normality tests are assessed by the simulation study of 11 statistical distributions grouped into four groups: symmetric, asymmetric, mixed and normal mixture distributions [5]. A description of these distribution groups is given in the following subsections.

4.1. A Group of Symmetric Distributions

Symmetric multivariate distributions are taken from the research [5]:

- Three cases of the *Beta(a,b)* distribution − *Beta*(1,1),*Beta*(1,2) and *Beta*(2,2), where *a* and *b* are the shape parameters.
- One case of the *Cauchy(t,s)* distribution − *Cauchy*(0,1), where *t* and *s* are the location and scale parameters.
- One case of the *Laplace(t,s)* distribution − *Laplace*(0,1), where *t* and *s* are the location and scale parameters.
- One case of the *Logistic(t,s)* distribution − *Logistic*(0,1), where *t* and s are the location and scale parameters.
- Two cases of the *t-Student(ν)* distribution − *t*(2) and *t*(5), where ν is the number of degrees of freedom.
- One case of the standard normal *N*(0,1) distribution.

4.2. A Group of Asymmetric Distributions

Asymmetric multivariate distributions are taken from the research [5]:

- Five cases of the *Chi-squared(ν)* distribution − χ^2 (1), χ^2 (2), χ^2 (5), χ^2 (10) and χ^2 (15), where ν is the number of degrees of freedom.
- Two cases of the *Gamma(a,b)* distribution − *Gamma*(0.5,1) and *Gamma*(5,1), where a and b are the shape and scale parameters.
- One case of the *Gumbel(t,s)* distribution − *Gumbel*(1,2), where *t* and *s* are the location and scale parameters.
- Two cases of the *Lognormal(t,s)* distribution − *LN*(0,1) and *LN*(0,0.25) where *t* and *s* are the location and scale parameters.
- Tree cases of the *Weibull(β)* distribution − *Weibull*(0.8), *Weibull*(1) and *Weibull*(1.5), where β is the shape parameter.

4.3. A Group of Mixed Distributions

The generated mixed data distribution

$$X_k = \left(X_{k1}, X_{k2}, \ldots, X_{km}, \ldots, X_{kp}\right)^T,\ k = 1, 2, \ldots, n$$

is such that the first *m* variates (i.e., X_{k1}, X_{k2}, ... , X_{km}) follow the standard normal distribution and distribution of the remaining variates is one of the non-normal distributions (*Laplace*(0,1), $\chi^2(5)$, *t*(5), *Beta*(1,1), *Beta*(1,2), *Beta*(2,2)). The experimental research covers the cases for $m = p - 1$, $m = p/2$ and $m = 1$.

4.4. A Group of Normal Mixture Distributions

Normal mixture distributions are considered in this research [5]: nine cases of the multivariate normal mixture distribution *MVNMIX (a,b,c,d)* − *MVNMIX* (0.5,2,0,0), *MVNMIX* (0.5,4,0,0), *MVNMIX* (0.5,2,0.9,0), *MVNMIX* (0.5,0.5,0.9,0), *MVNMIX* (0.5,0.5,0.9,0.1), *MVNMIX*(0.5,0.5,0.9,0.9), *MVNMIX*(0.7,2,0.9,0.3), *MVNMIX*(0.3,1,0.9,0.1), *MVNMIX*(0.3,1,0.9,0.9). The multivariate normal mixture distribution with density:

$$aN\left(0, \sum_1\right) + (1-a)N\left(b1, \sum_2\right),$$

where 1 is the column vector with all elements being 1,

$$\sum_1 = (1-c)I + c11^T \text{ and } \sum_2 = (1-d)I + d11^T.$$

5. Simulation Study and Discussion

This section provides a modeling study that evaluates the power of selected multivariate normality tests. We used the Monte Carlo method to compare our proposed test with 13 multivariate tests described above for dimensions $p = 2, 5, 10$, with sample sizes $n = 32, 64, 128, 256, 512, 1024$ at significance level $\alpha = 0.05$. Power was estimated by applying the tests on 1 000 000 randomly drawn samples from the alternative distribution (Beta, Cauchy, Laplace, Logistic, Student, Standard normal, Chi-Square, Gamma, Gumbel, Lognormal, Weibull, Mixed, Normal mixture).

The values of the test smoothness parameter (h) were selected experimentally: from 0.1 to 5 with a step of 0.1. The value of the test h parameter was determined for each dimension considered. It was found that the best results are obtained (i.e., maximum statistical value) for $p = 2$ with $h = 1.05$, for $p = 5$ with $h = 0.1$, and for $p = 10$ with $h = 2.4$. These smoothness parameter h values were used to carry out the numerical experiments.

The power of 13 (including our proposed test) multivariate goodness of fit hypothesis tests was estimated calculated for different sample sizes, distributions and mixtures. The mean power values for the groups for distributions (given in Section 4), for each test and sample sizes, have been computed and presented in Tables 2–5. It can be determined that the new test for the groups of symmetric and mixed distributions is the most powerful one. In the group of asymmetric distributions, the new (for $p = 2$) and Roy (for $p = 5$ and 10) tests are the most powerful ones. The new (for $p = 2$ and 5) and Roy (for $p = 10$ with sample sizes $n = 256, 512, 1024$) tests are also the most powerful in the group of normal distribution mixtures. Comparing the Mardia (Mar1 and Mar2) tests, based on asymmetry and excess coefficients, it has been found that Mar1 is the most powerful only for the group of asymmetric distributions. For the group of symmetric distributions the power of this test is the lowest (compared to other tests).

Table 2. An average empirical power for a group of symmetric distributions.

	AD	CHI2	CVM	DH	DN	Energy	HZ	LV	New	Mar1	Mar2	NRR	Roy
$p = 2$													
n = 32	0.651	0.57	0.652	0.677	0.565	0.65	0.644	0.696	0.999	0.532	0.605	0.608	0.703
n = 64	0.778	0.692	0.779	0.809	0.671	0.77	0.765	0.815	0.999	0.617	0.751	0.736	0.819
n = 128	0.867	0.798	0.868	0.892	0.768	0.86	0.853	0.893	0.999	0.681	0.857	0.842	0.891
n = 256	0.92	0.873	0.92	0.932	0.847	0.914	0.906	0.932	0.999	0.721	0.917	0.91	0.929
n = 512	0.939	0.912	0.94	0.945	0.903	0.941	0.936	0.945	0.999	0.743	0.942	0.941	0.944
n = 1024	0.945	0.932	0.945	0.949	0.937	0.948	0.947	0.949	0.999	0.758	0.95	0.949	0.95
$p = 5$													
n = 32	0.644	0.531	0.624	0.735	0.585	0.632	0.622	0.763	0.985	0.523	0.637	0.602	0.784
n = 64	0.791	0.656	0.775	0.864	0.7	0.758	0.755	0.871	0.989	0.621	0.792	0.739	0.875
n = 128	0.883	0.773	0.876	0.924	0.806	0.863	0.856	0.925	0.988	0.696	0.89	0.851	0.924
n = 256	0.929	0.864	0.926	0.941	0.886	0.921	0.91	0.941	0.987	0.735	0.934	0.916	0.941
n = 512	0.942	0.916	0.942	0.946	0.932	0.945	0.94	0.946	0.981	0.752	0.948	0.944	0.947
n = 1024	0.946	0.941	0.946	0.949	0.949	0.949	0.949	0.949	0.985	0.764	0.95	0.95	0.95
$p = 10$													
n = 32	0.557	0.473	0.534	0.754	0.599	0.598	0.604	0.791	0.997	0.458	0.65	0.599	0.834
n = 64	0.754	0.604	0.728	0.884	0.704	0.709	0.71	0.893	0.998	0.592	0.802	0.719	0.905
n = 128	0.878	0.726	0.865	0.934	0.817	0.821	0.831	0.935	0.998	0.676	0.899	0.844	0.934
n = 256	0.928	0.824	0.922	0.941	0.896	0.906	0.901	0.941	0.998	0.733	0.94	0.913	0.943
n = 512	0.942	0.891	0.941	0.945	0.936	0.943	0.937	0.945	0.991	0.747	0.951	0.942	0.946
n = 1024	0.945	0.928	0.945	0.948	0.949	0.948	0.949	0.948	0.991	0.756	0.95	0.949	0.95

Table 3. An average empirical power for a group of asymmetric distributions.

	AD	CHI2	CVM	DH	DN	Energy	HZ	LV	New	Mar1	Mar2	NRR	Roy
$p = 2$													
n = 32	0.634	0.631	0.639	0.852	0.55	0.832	0.811	0.87	0.999	0.813	0.63	0.639	0.877
n = 64	0.744	0.767	0.744	0.956	0.657	0.93	0.906	0.961	0.999	0.941	0.776	0.759	0.962
n = 128	0.827	0.861	0.822	0.995	0.724	0.985	0.968	0.995	0.999	0.992	0.876	0.841	0.995
n = 256	0.897	0.931	0.892	0.999	0.774	0.999	0.995	0.999	0.999	0.999	0.947	0.915	0.999
n = 512	0.954	0.977	0.949	0.999	0.816	0.999	0.999	0.999	0.999	0.999	0.988	0.968	0.999
n = 1024	0.985	0.996	0.982	0.999	0.864	0.999	0.999	0.999	0.999	0.999	0.999	0.993	0.999
$p = 5$													
n = 32	0.614	0.6	0.608	0.915	0.551	0.854	0.798	0.932	0.982	0.803	0.623	0.61	0.945
n = 64	0.763	0.779	0.761	0.99	0.675	0.958	0.907	0.992	0.989	0.954	0.791	0.763	0.993
n = 128	0.869	0.892	0.869	0.999	0.748	0.996	0.974	0.999	0.997	0.997	0.908	0.869	0.999
n = 256	0.946	0.965	0.947	0.999	0.812	0.999	0.998	0.999	0.997	0.999	0.978	0.95	0.999
n = 512	0.984	0.994	0.985	0.999	0.872	0.999	0.999	0.999	0.999	0.999	0.997	0.989	0.999
n = 1024	0.995	0.999	0.995	0.999	0.926	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
$p = 10$													
n = 32	0.483	0.443	0.459	0.944	0.532	0.829	0.744	0.96	0.922	0.693	0.573	0.532	0.98
n = 64	0.707	0.712	0.7	0.998	0.679	0.956	0.861	0.998	0.947	0.931	0.746	0.722	0.999
n = 128	0.863	0.87	0.859	0.999	0.776	0.997	0.954	0.999	0.98	0.997	0.898	0.86	0.999
n = 256	0.955	0.96	0.953	0.999	0.858	0.999	0.994	0.999	0.995	0.999	0.978	0.952	0.999
n = 512	0.99	0.994	0.989	0.999	0.93	0.999	0.999	0.999	0.996	0.999	0.998	0.992	0.999
n = 1024	0.996	0.999	0.996	0.999	0.975	0.999	0.999	0.999	0.996	0.999	0.999	0.999	0.999

Table 4. An average empirical power for a group of mixed distributions.

	AD	CHI2	CVM	DH	DN	Energy	HZ	LV	New	Mar1	Mar2	NRR	Roy
$p = 2$													
n = 32	0.469	0.408	0.463	0.436	0.439	0.582	0.572	0.453	0.999	0.476	0.412	0.444	0.451
n = 64	0.572	0.476	0.567	0.51	0.511	0.703	0.697	0.547	0.999	0.577	0.527	0.533	0.513
n = 128	0.683	0.571	0.679	0.591	0.59	0.809	0.807	0.667	0.999	0.659	0.651	0.641	0.572
n = 256	0.78	0.667	0.778	0.66	0.674	0.872	0.871	0.749	0.999	0.717	0.762	0.741	0.643
n = 512	0.848	0.763	0.847	0.746	0.763	0.895	0.894	0.808	0.999	0.76	0.843	0.827	0.72
n = 1024	0.883	0.842	0.883	0.826	0.835	0.902	0.901	0.857	0.999	0.78	0.884	0.878	0.764
$p = 5$													
n = 32	0.626	0.466	0.585	0.545	0.538	0.703	0.706	0.553	0.982	0.584	0.584	0.551	0.47
n = 64	0.749	0.582	0.726	0.631	0.684	0.788	0.815	0.628	0.989	0.675	0.723	0.694	0.524
n = 128	0.805	0.67	0.791	0.695	0.771	0.845	0.864	0.692	0.995	0.722	0.791	0.769	0.589
n = 256	0.852	0.729	0.841	0.747	0.825	0.88	0.885	0.751	0.998	0.75	0.838	0.822	0.669
n = 512	0.88	0.763	0.875	0.777	0.865	0.894	0.895	0.81	0.999	0.766	0.864	0.863	0.73
n = 1024	0.894	0.789	0.893	0.795	0.891	0.9	0.899	0.883	0.999	0.778	0.889	0.889	0.764
$p = 10$													
n = 32	0.688	0.477	0.642	0.58	0.669	0.719	0.745	0.592	0.916	0.614	0.731	0.679	0.475
n = 64	0.753	0.579	0.744	0.69	0.753	0.744	0.78	0.69	0.942	0.68	0.796	0.753	0.529
n = 128	0.775	0.651	0.771	0.736	0.776	0.777	0.821	0.735	0.94	0.722	0.795	0.774	0.602
n = 256	0.802	0.709	0.795	0.761	0.793	0.823	0.87	0.76	0.968	0.745	0.811	0.79	0.689
n = 512	0.833	0.746	0.821	0.778	0.818	0.875	0.892	0.776	0.995	0.764	0.84	0.814	0.745
n = 1024	0.866	0.763	0.853	0.791	0.842	0.897	0.899	0.79	0.997	0.779	0.861	0.837	0.769

In order to supplement and emphasize the results presented in Tables 2–5, the generalized line diagrams were drawn using the Trellis display [26] multivariate data visualization method. The resulting graph is shown in Figure 2 which shows that the New test is significantly more powerful than the other tests. The power of the Mar1 tests is the lowest compared with the other tests. Figure 2 indicate that the power of the tests increases as the sample size increases. By increasing the dimensions of the power of 8 (AD, CHI2, CVM, Energy, HZ, New, Mar1 and NRR) tests decreases while the power of the other (DH, DN, LV, Mar2 and Roy) tests increases slightly. For small sample sizes, the most powerful tests

are New, Roy and DH. For large sample sizes, the most powerful tests are New, Energy, HZ and LV.

Table 5. An average empirical power for a group of normal mixture distributions.

	AD	CHI2	CVM	DH	DN	Energy	HZ	LV	New	Mar1	Mar2	NRR	Roy
$p = 2$													
n = 32	0.465	0.422	0.468	0.56	0.428	0.537	0.529	0.588	0.999	0.433	0.437	0.442	0.607
n = 64	0.576	0.508	0.581	0.74	0.503	0.682	0.672	0.752	0.999	0.544	0.563	0.544	0.778
n = 128	0.71	0.618	0.715	0.893	0.582	0.836	0.823	0.895	0.999	0.633	0.707	0.664	0.908
n = 256	0.844	0.738	0.848	0.974	0.685	0.938	0.926	0.974	0.999	0.701	0.84	0.805	0.978
n = 512	0.943	0.845	0.945	0.998	0.791	0.986	0.977	0.998	0.999	0.733	0.931	0.924	0.998
n = 1024	0.987	0.917	0.988	0.999	0.882	0.999	0.998	0.999	0.999	0.757	0.977	0.985	0.999
$p = 5$													
n = 32	0.45	0.399	0.441	0.594	0.443	0.503	0.485	0.632	0.98	0.384	0.46	0.442	0.672
n = 64	0.574	0.491	0.563	0.782	0.51	0.64	0.621	0.795	0.994	0.516	0.59	0.539	0.828
n = 128	0.699	0.598	0.689	0.916	0.594	0.781	0.761	0.92	0.997	0.619	0.728	0.655	0.934
n = 256	0.806	0.702	0.798	0.979	0.691	0.894	0.877	0.979	0.999	0.694	0.832	0.766	0.984
n = 512	0.889	0.787	0.883	0.998	0.782	0.963	0.95	0.998	0.999	0.736	0.905	0.857	0.998
n = 1024	0.946	0.857	0.942	0.999	0.859	0.992	0.985	0.999	0.999	0.758	0.954	0.925	0.999
$p = 10$													
n = 32	0.402	0.392	0.396	0.62	0.495	0.476	0.487	0.667	0.989	0.287	0.547	0.487	0.735
n = 64	0.556	0.47	0.537	0.8	0.536	0.599	0.581	0.815	0.984	0.472	0.634	0.55	0.855
n = 128	0.709	0.587	0.692	0.915	0.629	0.723	0.708	0.919	0.977	0.582	0.758	0.674	0.939
n = 256	0.801	0.688	0.79	0.973	0.723	0.834	0.815	0.974	0.971	0.669	0.834	0.771	0.98
n = 512	0.853	0.76	0.846	0.995	0.8	0.912	0.887	0.995	0.971	0.711	0.885	0.833	0.997
n = 1024	0.893	0.818	0.886	0.999	0.85	0.96	0.939	0.999	0.973	0.739	0.924	0.875	0.999

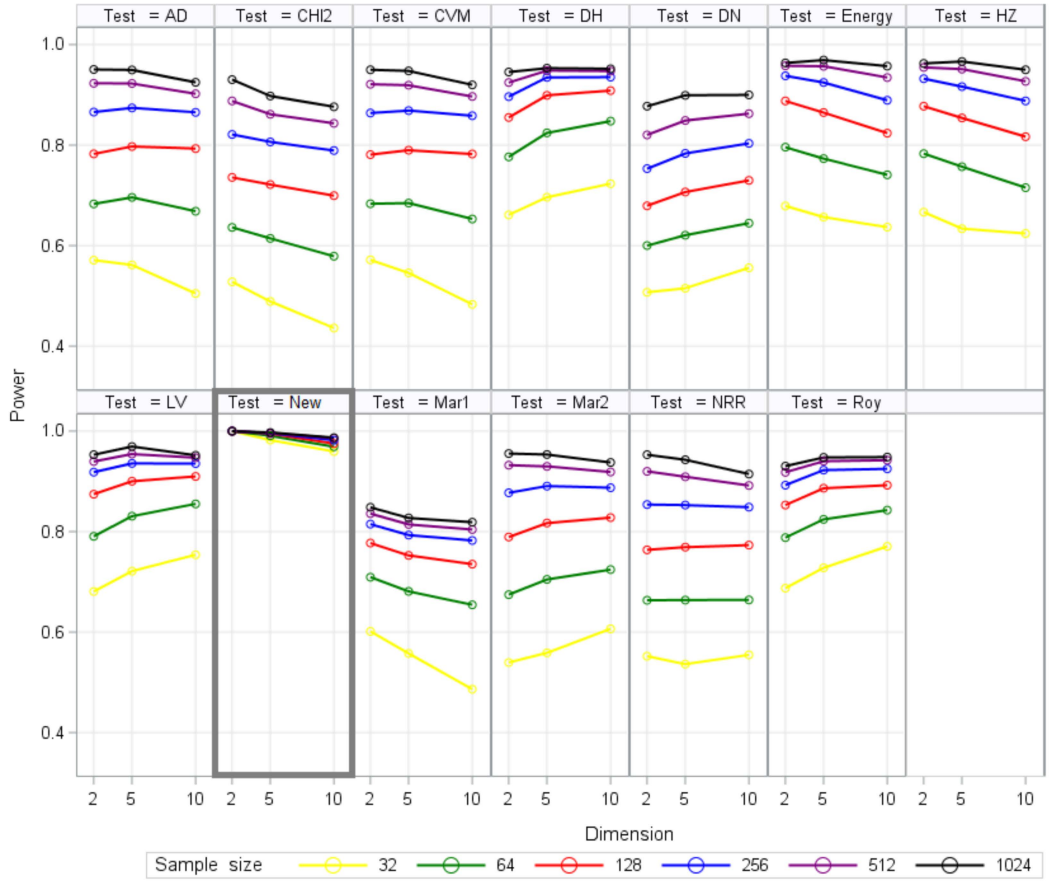

Figure 2. The summary of average empirical power of all examined distribution groups by sample size and dimensionality.

6. Examples

6.1. Survival Data

The data set collected in 2001–2020 by the Head of the Department of Urology Clinic of the Lithuanian University of Health Sciences [27] illustrates the practical application. This dataset consists of study data from 2423 patients and two different continuous attributes (patient age and prostate-specific antigen (PSA)). The assumption of normality was verified by filtering patients' age and PSA by year of death (i.e., deaths during the first 1, 2, 3, 4, 5, 6, 7, 10, and 15 years). The filtered data was standardized. The power and p-value were calculated for the multivariate tests. The significance level of $\alpha = 0.05$ was used for the study. Based on the obtained results, it was found that all the applied multivariate tests rejected the H_0 the hypothesis of normality. The power of tests CHI2, DH, Energy, HZ, LV, New, Mar, NRR and Roy was 0.999 and the p-value was <0.0001. Except for DN test, which power was 0.576 and the p-value was 0.026.

6.2. IQOS Data

In 2017, the data set of pollution research with IQOS and traditional cigarettes [28] was used by Kaunas University of Technology, Faculty of Chemical Technology, and Department of Environmental Technology for practical application. This data set consists of 33 experiments (with different conditions) in which the numerical (Pn10) and mass concentrations (Pm2.5, Pm10) of particles were measured. The assumption of normality was checked by filtering Pn10, Pm2.5, Pm10 according to the number of the experiment in the smoking phase. The filtered data was standardized. The power and p-values of multivariate tests with a significance level of $\alpha = 0.05$ were calculated. Based on the obtained results, it was found that all the applied multivariate tests show that the assumption of normality is rejected. Most of the multivariate tests used (CHI2, DH, Energy, HZ, LV, New, Mar, NRR, and Roy) had a power of 0.999 and p-value of <0.0001. The power of the other tests was also close to 0.99 and the p-value was about 0.0001.

7. Conclusions

In this study, the comprehensive comparison of the power of 13 multivariate goodness of fit tests was performed for groups of symmetric, asymmetric, mixed, and normal mixture distributions. Two-dimensional, five-dimensional, and ten-dimensional data sets were generated to estimate the test power empirically.

A new multivariate goodness of fit test based on inversion formula was proposed. Based on the obtained modeling results, it was determined that the most powerful tests for the groups of symmetric, asymmetric mixed and normal mixture distributions are the proposed test and Roy multivariate test. From two real data examples, it was concluded that our proposed test is stable, even when applied to real data sets.

Author Contributions: Data curation, J.A., T.R.; Formal analysis, J.A., T.R.; Investigation, J.A., T.R.; Methodology, J.A., T.R.; Software, J.A., T.R.; Supervision, T.R.; Writing—original draft, J.A., M.B.; writing—review and editing, J.A., M.B. 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: Generated data sets were used in the study (see in Section 4).

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

Appendix A

$\mathbb{R}^p$ is p-variate set of real numbers,
$X_k = \left(X_{k1}, X_{k2}, \ldots, X_{kp}\right)^T \in \mathbb{R}^p,\ k = 1,\ 2, \ldots, n$ is p-variate vector,
#T denotes a size of set T,
p is dimension,
h is smoothness parameter,
$D_{(j)},\ \ j = 1,\ 2, \ldots, n$ are ordered statistics,
$G_p(\cdot)$ is the probability distribution function of $\chi^2(p)$,
s is skewness,
k is kurtosis,
n is sample size,
$\overline{x}$ is sample mean,
σ^2 is sample variance,
z is a standardized value,
d is number of variables,
e is the equivalent degrees of freedom,
$\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution,
R is the correlation matrix,
r_{ij} is the correlation between ith and jth observations,
V_i is a vector of standardized cell frequencies,
N_{in} is the number of random vectors,
$J = J(\theta)$ is the Fisher information matrix,
Q is the $p(p+1)/2 \times p(p+1)/2$ covariance matrix of w,
$\hat{P}(t)$ is the characteristic function of the multivariate standard normal,
$\hat{Q}(t)$ is the empirical characteristic function of the standardised observations,
$\varphi_\beta(t)$ is a kernel (weighting) function,
$\Gamma(\cdot)$ is a Gamma function,
$\hat{F}^{(k)}$ is an auto-covariance function,
$\mathcal{D}_{ij}$ is Mahalanobis distance between ith and jth observations,
$\mathcal{W}_j$ is the normality transformations,
$f_k(X, \theta_k)$ is a distribution of the kth class,
θ is a set of parameters $\theta = \{p_1, \ldots, p_q, \theta_1, \ldots, \theta_q\}$,
$\psi(t)$ is the characteristic function of the random variable X.
$p_k\ (k = 1, \ldots, q)$ is the a priori probability,
$\mathcal{R}$ is the radius of the ball (bounded sphere),
q_ς is a quantile of standardized normal distribution
δ and γ are shape parameters,
λ is scale parameter,
ξ is location parameter.

References

1. Henze, N.; Zirkler, B. A class of invariant consistent tests for multivariate normality. *Commun. Stat. Theory Methods* **1990**, *19*, 3595–3617. [CrossRef]
2. Henze, N. Invariant tests for multivariate normality: A critical review. *Stat. Pap.* **2002**, *43*, 467–506. [CrossRef]
3. Royston, J.P. An extension of Shapiro and Wilk's W test for normality to large samples. *Appl. Stat.* **1982**, *31*, 115–124. [CrossRef]
4. Ross, G.J.S.; Hawkins, D. *MLP: Maximum Likelihood Program*; Rothamsted Experimental Station: Harpenden, UK, 1980.
5. Korkmaz, S.; Goksuluk, D.; Zararsiz, G. MVN: An R Package for Assessing Multivariate Normality. *R J.* **2014**, *6*, 151–162. [CrossRef]
6. Doornik, J.A.; Hansen, H. An Omnibus Test for Univariate and Multivariate Normality. *Oxf. Bull. Econ. Stat.* **2008**, *70*, 927–939.
7. Voinov, V.; Pya, N.; Makarov, R.; Voinov, Y. New invariant and consistent chi-squared type goodness-of-fit tests for multivariate normality and a related comparative simulation study. *Commun. Stat. Theory Methods* **2016**, *45*, 3249–3263. [CrossRef]
8. Moore, D.S.; Stubblebine, J.B. Chi-square tests for multivariate normality with application to common stock prices. *Commun. Stat. Theory Methods* **1981**, *A10*, 713–738. [CrossRef]
9. Gnanadesikan, R.; Kettenring, J.R. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **1972**, *28*, 81–124.
10. Górecki, T.; Horváth, L.; Kokoszka, P. Tests of Normality of Functional Data. *Int. Stat. Rev.* **2020**, *88*, 677–697. [CrossRef]
11. Pinto, L.P.; Mingoti, S.A. On hypothesis tests for covariance matrices under multivariate normality. *Pesqui. Operacional.* **2015**, *35*, 123–142. [CrossRef]

12. Dörr, P.; Ebner, B.; Henze, N. A new test of multivariate normality by a double estimation in a characterizing PDE. *Metrika* **2021**, *84*, 401–427. [CrossRef]
13. Zhoua, M.; Shao, Y. A Powerful Test for Multivariate Normality. *J. Appl Stat.* **2014**, *41*, 351–363. [CrossRef]
14. Kolkiewicz, A.; Rice, G.; Xie, Y. Projection pursuit based tests of normality with functional data. *J. Stat. Plan. Inference* **2021**, *211*, 326–339. [CrossRef]
15. Ebner, B.; Henze, N. Tests for multivariate normality-a critical review with emphasis on weighted L^2-statistics. *TEST* **2020**, *29*, 845–892. [CrossRef]
16. Arnastauskaitė, J.; Ruzgas, T.; Bražėnas, M. An Exhaustive Power Comparison of Normality Tests. *Mathematics* **2021**, *9*, 788. [CrossRef]
17. Mardia, K. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika* **1970**, *57*, 519–530. [CrossRef]
18. Szekely, J.G.; Rizzo, L.M. Energy statistics: A class of statistics based on distances. *J. Stat. Plan. Inference* **2013**, *143*, 1249–1272. [CrossRef]
19. Lobato, I.; Velasco, C. A simple Test of Normality for Time Series. *Econom. Theory* **2004**, *20*, 671–689. [CrossRef]
20. Ruzgas, T. The Nonparametric Estimation of Multivariate Distribution Density Applying Clustering Procedures. Ph.D. Thesis, Institute of Mathematics and Informatics, Vilnius, Lithuania, 2007; p. 161.
21. Marron, J.S.; Wand, M.P. Exact mMean Integrated Squared Error. *Ann. Stat.* **1992**, *20*, 712–736. [CrossRef]
22. Kavaliauskas, M.; Rudzkis, R.; Ruzgas, T. The Projection-based Multivariate Distribution Density Estimation. *Acta Comment. Univ. Tartu. Math.* **2004**, *8*, 135–141.
23. Epps, T.W.; Pulley, L.B. A test for normality based on the empirical characteristic function. *Biometrika* **1983**, *70*, 723–726. [CrossRef]
24. Johnson, N.L. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* **1949**, *36*, 149–176. [CrossRef] [PubMed]
25. Wicksell, S.D. The construction of the curves of equal frequency in case of type A correlation. *Ark. Mat. Astr. Fys.* **1917**, *12*, 1–19.
26. Theus, M. High Dimensional Data Visualization. In *Handbook of Data Visualization*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 5–7.
27. Milonas, D.; Ruzgas, T.; Venclovas, Z.; Jievaltas, M.; Joniau, S. The significance of prostate specific antigen persistence in prostate cancer risk groups on long-term oncological outcomes. *Cancers* **2021**, *13*, 2453. [CrossRef]
28. Martuzevicius, D.; Prasauskas, T.; Setyan, A.; O'Connell, G.; Cahours, X.; Julien, R.; Colard, S. Characterization of the Spatial and Temporal Dispersion Differences Between Exhaled E-Cigarette Mist and Cigarette Smoke. *Nicotine Tob. Res.* **2019**, *21*, 1371–1377. [CrossRef]