Article

Modifications to the Jarque–Bera Test

Vladimir Glinskiy [1,2,*], **Yulia Ismayilova** [1], **Sergey Khrushchev** [3], **Artem Logachov** [1,4], **Olga Logachova** [5], **Lyudmila Serga** [1,2], **Anatoly Yambartsev** [6] **and Kirill Zaykov** [1]

1. Department of Business Analytics, Accounting and Statistics and Research Laboratory of Sustainable Development of Socio-Economic Systems, Siberian Institute of Management—Branch of the Russian Presidential Academy of National Economy and Public Administration, 630102 Novosibirsk, Russia; ismayylova-yn@ranepa.ru (Y.I.); logachev-av@ranepa.ru (A.L.); serga-lk@ranepa.ru (L.S.); zaykov-ka@ranepa.ru (K.Z.)
2. Department of Statistics, Novosibirsk State University of Economics and Management, 630099 Novosibirsk, Russia
3. Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Science, 630090 Novosibirsk, Russia; s.e.khrushchev@math.nsc.ru
4. Department of Computer Science in Economics, Novosibirsk State Technical University (NSTU), 630087 Novosibirsk, Russia
5. Department of Higher Mathematics, Siberian State University of Geosystems and Technologies (SSUGT), 630108 Novosibirsk, Russia; o.m.logacheva@sgugit.ru
6. Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo (USP), São Paulo CEP 05508-220, Brazil; yambar@usp.br
* Correspondence: glinskiy-vv@ranepa.ru

Abstract: The Jarque–Bera test is commonly used in statistics and econometrics to test the hypothesis that sample elements adhere to a normal distribution with an unknown mean and variance. This paper proposes several modifications to this test, allowing for testing hypotheses that the considered sample comes from: a normal distribution with a known mean (variance unknown); a normal distribution with a known variance (mean unknown); a normal distribution with a known mean and variance. For given significance levels, $\alpha = 0.05$ and $\alpha = 0.01$, we compare the power of our normality test with the most well-known and popular tests using the Monte Carlo method: Kolmogorov–Smirnov (KS), Anderson–Darling (AD), Cramér–von Mises (CVM), Lilliefors (LF), and Shapiro–Wilk (SW) tests. Under the specific distributions, 1000 datasets were generated with the sample sizes $n = 25, 50, 75, 100, 150, 200, 250, 500,$ and 1000. The simulation study showed that the suggested tests often have the best power properties. Our study also has a methodological nature, providing detailed proofs accessible to undergraduate students in statistics and probability, unlike the works of Jarque and Bera.

Keywords: normality test; Jarque–Bera test; skewness; kurtosis; Monte Carlo simulation

MSC: 60F05; 62F03; 62F05

1. Introduction

C. Jarque and A.K. Bera proposed the following goodness-of-fit test (see [1–3]) to determine whether the empirical skewness and kurtosis match those of a normal distribution. The hypothesis to be tested is as follows:

H0. *If the population the sample presents is normally distributed against the alternative hypothesis.*

H1. *If the population the sample presents follows a distribution from the Pearson family that is not normally distributed.*

More precisely, the null hypothesis is formulated as follows: the sample comes from a population with a finite eighth moment, the odd central moments (up to the seventh) are equal to zero, and the kurtosis is equal to three, $K = 3$. Note that only the normal distribution has these properties within any reasonable family of distributions. For sure, it is true for the Pearson family. In practice, the Pearson family is not typically mentioned in the hypothesis.

The test statistic is a combination of squares of normalized skewness, S, and kurtosis, K:

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right).$$

If the null hypothesis, H_0, is true, then as $n \to \infty$, the distribution of the random variable JB converges to $\chi^2(2)$. Therefore, for a sufficiently large sample size, the following testing rule can be applied: given a significance level α, if $JB < \chi^2_{1-\alpha}(2)$ (where $\chi^2_{1-\alpha}(2)$ is the $1 - \alpha$ quantile of the $\chi^2(2)$ distribution), then the null hypothesis H_0 is accepted; otherwise, it is rejected.

Note that the Pearson family of distributions is quite rich, including exponential, gamma, beta, Student's t, and normal distributions. Suppose it is known that a random variable has a distribution from the Pearson family and has the first four moments. In that case, its specific form is uniquely determined by the skewness, S, and kurtosis, K, see [4]. For an illustration, we present this classification in Figure 1. Due to this property, the Jarque–Bera test is a goodness-of-fit test, i.e., if the alternative hypothesis H_1 holds for the sample elements, the statistic JB converges in probability to ∞ as $n \to \infty$.
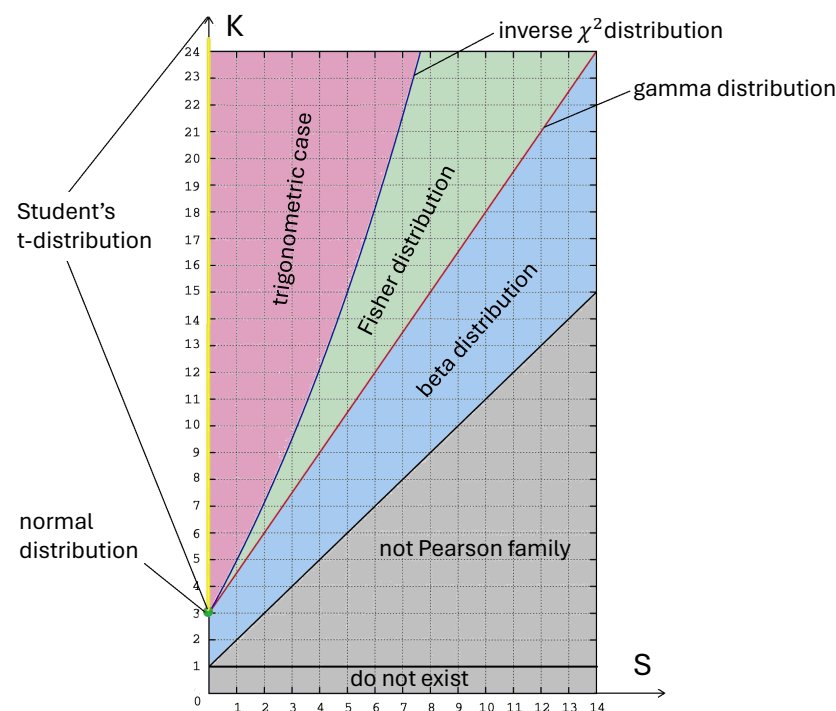
Figure 1. Some distributions from the Pearson family in the $S - K$ plot. Figure created by A. Logachov and A. Yambartsev.

This article emerged as a result of addressing the following question: How does the JB statistic change when the researcher knows the following?

1. The mean of the population distribution (the variance is unknown);
2. The variance of the population distribution (the mean is unknown);
3. The mean and variance of the population distribution.

In the last case, the known mean and variance lead us to the known coefficient of variation. For a discussion on inference when the coefficient of variation is known, we refer the reader to [5,6], and the references therein.

In this paper, we adapt the *JB* statistic for cases where one or both normal distribution parameters are known. As simulations show, the proposed tests also demonstrate good power for many samples not belonging to the Pearson family of distributions.

To conclude this section, we note the following: In practical research, knowing the parameters of the normal distribution is crucial, as it allows us to estimate the probabilities of desired events. Testing the hypothesis of normality with specific parameters is significant because any deviation—whether in the form of outliers (deviations from normality) or change points in stochastic processes (a sudden change in the parameter)—can indicate the presence of unusual or catastrophic events. For example, small but significant parameter changes can signal a disturbance in the production process. Strong and rare deviations are of particular interest when studying stochastic processes with catastrophes. We believe a deeper connection exists between these seemingly distinct fields, which still awaits thorough investigation.

The rest of this paper is organized as follows. The following section, Section 2, presents the main results (the limit theorem and criteria for testing the corresponding statistical hypotheses). In Section 3, we present a Monte Carlo simulation to compare the suggested tests with some existing procedures. We prove Theorem 1 in Section 4. Finally, the last section contains tables of test power resulting from the Monte Carlo numerical simulations.

2. Definitions and Results

Let $X_1, X_2, \ldots, X_n, n \in \mathbb{N}$ be i.i.d. random variables on the same probability space $(\Omega, \mathfrak{F}, \mathbf{P})$. We use $\mathbf{E}$ and $\mathbf{D}$ to denote expectation and variance with respect to the probability measure $\mathbf{P}$. The convergence in distribution we denote as $\xrightarrow[n \to \infty]{d}$.

Recall the definition of empirical skewness, S, and kurtosis, K:

$$S = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X})^3}{\hat{\sigma}^3}, \quad K = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X})^4}{\hat{\sigma}^4},$$

where, as usual, we have the following:

$$\hat{\sigma} = \left(\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X})^2 \right)^{\frac{1}{2}}, \quad \overline{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i.$$

The main result is as follows:

Theorem 1. *Let $X, X_1, \ldots, X_n, n \in \mathbb{N}$ be i.i.d. random variables. Then, we have the following:*
(1) If X has a non-degenerate normal distribution and $\mathbf{E}X = a$, then

$$JB_a = n \left(\frac{S_a^2}{15} + \frac{(K_a - 3)^2}{24} \right) \xrightarrow[n \to \infty]{d} Y \sim \chi^2(2),$$

where

$$S_a = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - a)^3}{\hat{\sigma}_a^3}, \quad K_a = \frac{\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - a)^4}{\hat{\sigma}_a^4}, \quad \hat{\sigma}_a = \left(\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - a)^2 \right)^{\frac{1}{2}};$$

(2) If X has a normal distribution and $\mathbf{D}X = \sigma^2 > 0$, then

$$JB_{\sigma^2} = n \left(\frac{S_{\sigma^2}^2}{6} + \frac{(K_{\sigma^2} - 3)^2}{96} \right) \xrightarrow[n \to \infty]{d} Y \sim \chi^2(2),$$

where

$$S_{\sigma^2} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^3}{\sigma^3}, \quad K_{\sigma^2} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^4}{\sigma^4};$$

(3) If X has a normal distribution with $\mathbf{E}X = a$ *and* $\mathbf{D}X = \sigma^2 > 0$, *then*

$$JB_{a,\sigma^2} = n\left(\frac{S_{a,\sigma^2}^2}{15} + \frac{(K_{a,\sigma^2} - 3)^2}{96}\right) \xrightarrow[n\to\infty]{d} Y \sim \chi^2(2),$$

where

$$S_{a,\sigma^2} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^3}{\sigma^3}, \quad K_{a,\sigma^2} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^4}{\sigma^4}.$$

Theorem 1 yields the following asymptotic tests. Let α be the significance level. Recall that $\chi^2_{1-\alpha}(2)$ denotes the $1 - \alpha$ quantile of the $\chi^2(2)$ distribution.

1. If we test the null hypothesis $H_0 : X_i \sim N(a, \sigma^2)$, $1 \leq i \leq n$, (where a is known and σ^2 is unknown) against the alternative hypothesis $H_1 : X_i$ follows a distribution from the Pearson family that is not normal but has a mean equal to a. Then, from statement (1) of Theorem 1, for a sufficiently large sample size, the following rule can be used: if $JB_a < \chi^2_{1-\alpha}(2)$, then the null hypothesis H_0 is accepted; otherwise, it is rejected.

2. When we test the null hypothesis $H_0 : X_i \sim N(a, \sigma^2)$, $1 \leq i \leq n$, (where a is unknown and σ^2 is known) against the alternative $H_1 : X_i$ follows a distribution from the Pearson family that is not normal with variance equal to σ^2. Then, from statement (2) of Theorem 1, for a sufficiently large sample size, the following rule can be used: if $JB_{\sigma^2} < \chi^2_{1-\alpha}(2)$, then the null hypothesis H_0 is accepted; otherwise, it is rejected.

3. When the null hypothesis $H_0 : X_i \sim N(a, \sigma^2)$, $1 \leq i \leq n$, (a and σ^2 are known) is tested against the alternative $H_1 : X_i$ follows a distribution from the Pearson family that is not $N(a, \sigma^2)$. Then, from statement (3) of Theorem 1, for a sufficiently large sample size, the following testing rule can be applied: if $JB_{a,\sigma^2} < \chi^2_{1-\alpha}(2)$, then the null hypothesis H_0 is accepted; otherwise, it is rejected.

It should also be noted that the above tests are goodness-of-fit tests, i.e., if the alternative hypothesis H_1 holds for the sample elements, then the values of the corresponding statistics JB_a, JB_{σ^2}, and JB_{a,σ^2} converge in probability to ∞ as $n \to \infty$.

3. Simulation Study

In this section, we compare the power of various tests for normality using Monte Carlo simulations of alternative hypotheses. The simulations were performed in R software, version 4.2.3. We used the following sample sizes (small, moderate, and large): $n = 25, 50, 75, 100, 150, 200, 250, 500$, and 1000. The null hypothesis is $N(0, 1)$ in almost all cases; we specify separately where this is not the case. As alternative hypotheses, we considered normal, log-normal, mixed normal, Student, gamma, and uniform distributions. Note that the log-normal and mixed normal distributions do not belong to the Pearson family of distributions and uniform distribution is the limit of the Pearson type I distribution. All codes are written in R and available at https://github.com/KhrushchevSergey/Modified-Jarque-Bera-test, accessed on 1 June 2024.

Here, we consider the following tests for normality:

Kolmogorov–Smirnov (KS) test. The test statistic measures the maximum deviation between the theoretical cumulative distribution function and the empirical cumulative distribution function. When the parameters of the normal distribution are unknown, they are estimated from the sample and used in the test.

Anderson–Darling (AD) test. The Anderson–Darling test assesses whether a sample comes from a specific distribution, often the normal distribution. It gives more weight

to the tails of the distribution compared to other tests, making it sensitive to deviations from normality in those areas.

Cramér–von Mises (CVM) test. The Cramér–von Mises test, like the KS test, is based on the distance between the empirical and specified theoretical distributions. It measures the cumulative squared differences between the empirical and theoretical cumulative distribution functions, providing a robust assessment of overall fit.

Lilliefors (LF) Kolmogorov–Smirnov test. The Lilliefors test is based on the Kolmogorov–Smirnov test. It tests the null hypothesis that data come from a normally distributed population without specifying the parameters.

Shapiro–Wilk (SW) test. The Shapiro–Wilk test is one of the most popular tests with good power. It is based on a correlation between given observations and associated normal scores.

We estimate the power in the following way. For the given n, we generate 1000 samples with the sample size n according to the alternative hypothesis. The empirical power is the ratio of the number of rejections of the null hypothesis to 1000. We categorize our findings based on the following cases of the alternative hypothesis distribution.

3.1. Normal versus Normal

Different variances and the same means. We start by comparing the test powers when the alternative distribution is the normal distribution with a zero mean and a variance different from one, specifically, $N(0, 2)$. See Table A1. Since we considered two normal distributions with different variances but the same mean, we added the column with the power of the Fisher test. The Fisher test checks the hypothesis if the variance is equal to one. We expect that, in this situation, the modified Jarque–Bera statistic would exhibit the highest power. The KS and CVM statistics have demonstrated similarly lower power, while the power of the AD statistic falls between that of the KS, CVM, and modified Jarque–Bera statistics.

Different means and the same variances. Here, we compare the test powers when the alternative distribution is a normal distribution with a mean of one and a variance of one, $N(1, 1)$. See Table A2. Since two normal distributions with different means but the same variance are considered, we added two additional columns with the test powers of the Student and Welch tests, respectively. All statistics perform similarly well, except for the modified Jarque–Bera with a known mean for the small sample sizes.

3.2. Normal versus Student's t with Degrees of Freedom 1 (Cauchy), 5, and 9

Cauchy. Here, consider the case where the alternative distribution is the Cauchy distribution. See Table A3. Since the alternative distribution is not normal, we did not conduct additional tests such as the Student or Fisher tests. In this case, the J_{a,σ^2} test provided the best power, but all tests performed similarly well, except for the Kolmogorov–Smirnov and the Cramér–von Mises tests. Since the normal distribution differs significantly from the Cauchy distribution, almost all tests provided good power.

Student's t-distribution with 5 and 9 degrees of freedom. Here, we consider the comparison between the test powers when the alternative distribution is the t_ν distribution with $\nu = 5$. In contrast to the Cauchy distribution, the Student's t distribution is more similar to the normal distribution, so the expected values of the powers are smaller than in the Cauchy case. Moreover, the J_{a,σ^2} statistic provided significantly better power. See Table A4. Since the Student's t distribution with $\nu = 9$ is even more similar to the standard normal distribution, the power will be smaller with similar relationships between different tests. Therefore, we omitted the entire table of powers. To give an idea about the magnitude of the changes in power, for the J_{a,σ^2} statistic with $n = 25$ and $\alpha = 0.05$, the power changed from 0.56 to 0.34. Note that the KS test exhibited the worst power. Additionally, observe that the performance of J_a is worse than those of J_{a,σ^2} and J_{σ^2}. This, of course, is expected because the null and alternative

distributions have the same zero mean. Note that the AD_c, CVM_c, and SW statistics exhibited power lower than but comparable to those of all the Jarque–Bera statistics. For this case, for Student's t with 5 degrees of freedom, we plotted the test power for both α values to provide a more detailed breakdown of these power comparisons. See Figures A1 and A2 in Appendix A.

3.3. Normal versus Non-Symmetric and Non-Pearson Type Distributions

For non-symmetric alternative distributions, (i) the gamma (2,1), $\gamma(2,1)$, (ii) log-normal (0,1), and $LN(0,1)$ were considered; for non-Pearson type alternative distributions, we considered (iii) the uniform on interval $[-3,3]$, $U[-3,3]$, and (iv) the mixture of standard normal $N(0,1)$ and normal $N(0,9)$ distributions with the same mixture weights, denoted by Mix. Since all alternative distributions are "significantly different" from the standard normal distribution, all tables "are similar" to that of the Cauchy distribution. See Tables A5–A8. It is expected that the JB_a statistic will lose power under symmetric alternative distributions. Moreover, the AD statistic is likely to perform comparably well to the Jarque–Bera statistics JB_{a,σ^2} and JB_{σ^2}. See Tables A5 and A8. In the non-symmetric case (Tables A6 and A7), the KS, AD, CVM, and SW statistics exhibit similar high power to the modified Jarque–Bera statistic, JB_{a,σ^2}.

3.4. Normal versus Gamma Distribution with the Same Mean Value

Finally, we decided to perform the comparison between test powers when the null hypothesis was not a standard normal distribution. Here, we test the normal distribution $N(2,2)$ versus gamma distribution $\gamma(2,1)$, which has a mean value of 2. The results are presented in Table A9. In this case, as before, the modified JB_{a,σ^2}, JB_{σ^2}, and AD statistics exhibit the highest power, while the KS, CVM, and SW statistics show a loss in power.

3.5. Robustness in the Presence of Outliers

To evaluate the performance of the modified Jarque–Bera tests in the presence of outliers, we generated data from a mixture of a standard normal distribution (weight 0.9) and a sum of independent random variables—one with a standard normal distribution and the other with a Poisson distribution with a mean of 5 (weight 0.1). This type of mixture is rarely used in simulation studies, but such discrete-value outliers can occur due to failures in production machines, for example; see Table A10. In this case, the KS and CVM tests showed the lowest power performance. The modified Jarque–Bera tests showed the best power, while the other tests (JB, LF, AD, AD_c, CVM_c, and SW) had lower but similar power. We also refer the readers to [7] for robust modifications to the Jarque–Bera statistic.

3.6. Application to Real Data

We tested the hypothesis that the mass of penguins, based on their species and gender, follows a normal distribution. We applied a modified Jarque–Bera test, using the sample mean and variance as known values. The observations were taken from a popular dataset featuring penguin characteristics from the study [8], where sexual size dimorphism (SSD), i.e., ecological sexual dimorphism, was studied in penguin populations. The normal variability of penguin mass is well-accepted, thus, accepting the null hypothesis is anticipated for this dataset. The corresponding p-values are provided in the table below.

Species	Male	Female
Adelie	0.7787	0.7361
Chinstrap	0.9194	0.5876
Gentoo	0.9598	0.7687

In the next section, we provide the detailed proof of our main results.

4. Proof of Theorem 1

Let us prove proposition (1) of the theorem. Consider the sequence of random variables, as follows:

$$\overrightarrow{Z}_n := (Z_{1,n}, Z_{2,n}) := \left(\frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} (X_i - a)^3}{\hat{\sigma}_a^3}, \frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} ((X_i - a)^4 - 3\hat{\sigma}_a^4)}{\hat{\sigma}_a^4} \right).$$

From the convergence, i.e.,

$$\lim_{n \to \infty} \hat{\sigma}_a = \sigma > 0 \text{ a.s.}$$

and Slutsky's theorem [9], it follows that the limiting distribution of the sequence $\overrightarrow{Z}_n$ coincides with the limiting distribution of the sequence

$$\overrightarrow{Z'}_n := (Z'_{1,n}, Z'_{2,n}) := \left(\frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} (X_i - a)^3}{\sigma^3}, \frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} ((X_i - a)^4 - 3\hat{\sigma}_a^4)}{\sigma^4} \right).$$

It is easy to see the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - a)^4 - 3\hat{\sigma}_a^4) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 - 3 \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right)^2 \pm 3\sigma^4 \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 - 3\sigma^4 + 3 \left(\sigma^2 - \frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right) \left(\sigma^2 + \frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right) \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 - 3\sigma^4 + 3 \left(\sigma^2 - \frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right) \left(2\sigma^2 + \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 - \sigma^2 \right) \right) \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 - 3\sigma^4 + 6\sigma^2 \left(\sigma^2 - \frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right) \right) - \frac{3}{\sqrt{n}} \sum_{i=1}^{n} \left(\sigma^2 - \frac{1}{n} \sum_{j=1}^{n} (X_j - a)^2 \right)^2$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 + 3\sigma^4 - 6\sigma^2 (X_i - a)^2 \right) - 3\sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} (\sigma^2 - (X_j - a)^2) \right)^2$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a)^4 + 3\sigma^4 - 6\sigma^2 (X_i - a)^2 \right) - 3 \left(\frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (\sigma^2 - (X_j - a)^2) \right)^2.$$

The law of the iterative logarithm yields the following:

$$\lim_{n \to \infty} \frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (\sigma^2 - (X_j - a)^2) = 0 \text{ a.s.}$$

Therefore, applying Slutsky's theorem, we can conclude that the limiting distribution of the sequence $\overrightarrow{Z'}_n$ coincides with the limiting distribution of the sequence, as follows:

$$\overrightarrow{Z''}_n := (Z'_{1,n}, Z''_{2,n}) := \left(\frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} (X_i - a)^3}{\sigma^3}, \frac{\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} ((X_i - a)^4 - 6\sigma^2 (X_i - a)^2 + 3\sigma^4)}{\sigma^4} \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{(X_i - a)^3}{\sigma^3}, \frac{(X_i - a)^4 - 6\sigma^2 (X_i - a)^2 + 3\sigma^4}{\sigma^4} \right).$$

By the central limit theorem, the sequence $\overrightarrow{Z''}_n$ converges to the random vector (W_1, W_2), whose coordinates have a joint normal distribution. Therefore, it suffices to show that these coordinates are uncorrelated and have the required variances.

It is easy to see that $\mathbf{D}(X_i - a)^3 = 15\sigma^6$ (thus, $\mathbf{D}W_1 = 15$); $\mathbf{D}((X_i - a)^4 - 6\sigma^2(X_i - a)^2 + 3\sigma^4) = 24\sigma^8$ (thus $\mathbf{D}W_2 = 24$). From the fact that the odd moments of a centered normally distributed random variable are equal to zero, it follows that

$$\mathbf{E}\left(\frac{(X_i - a)^3}{\sigma^3} \cdot \frac{(X_i - a)^4 - 6\sigma^2(X_i - a)^2 + 3\sigma^4}{\sigma^4} \right) = 0$$

(therefore, $\mathbf{E}W_1 W_2 = 0$).

Therefore, we have shown that (W_1, W_2) has a joint normal distribution with a covariance matrix, as follows:

$$\Sigma = \begin{pmatrix} 15 & 0 \\ 0 & 24 \end{pmatrix}.$$

Thus,

$$\frac{W_1^2}{15} + \frac{W_2^2}{24} \sim \chi^2(2).$$

Let us prove statement (2). Consider the following sequence of random vectors:

$$\overrightarrow{Z}_n := (Z_{1,n}, Z_{2,n}) := \left(\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X})^3}{\sigma^3}, \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - \overline{X})^4 - 3\sigma^4)}{\sigma^4} \right).$$

It is easy to see the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X})^3 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - a) - (\overline{X} - a))^3 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a) - \frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^3$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^3 - \frac{3}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^2 \cdot \frac{1}{n} \sum_{j=1}^{n} (X_j - a)$$

$$+ \frac{3}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a) \cdot \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^2 - \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^3$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^3 - \left(\frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2 \right) \cdot \frac{3}{\sqrt{n}} \sum_{j=1}^{n} (X_j - a)$$

$$+ \frac{3}{n} \sum_{i=1}^{n} (X_i - a) \cdot \left(\frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (X_j - a) \right)^2 - \left(\frac{1}{n^{\frac{5}{6}}} \sum_{j=1}^{n} (X_j - a) \right)^3$$

Finally, introducing $g_{1,n}, g_{1,n}$ and $g_{3,n}$ we have the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X})^3 =: \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^3 - g_{1,n} \cdot \frac{3}{\sqrt{n}} \sum_{j=1}^{n} (X_j - a) + g_{2,n} - g_{3,n}. \tag{1}$$

The law of large numbers and the law of the iterated logarithm yield the following:

$$\lim_{n \to \infty} g_{1,n} = \sigma^2 \text{ a.s. } \text{ and } \lim_{n \to \infty} g_{3,n} = 0 \text{ a.s.} \tag{2}$$

The sequences $\frac{3}{n} \sum_{i=1}^{n} (X_i - a)$ and $\frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (X_j - a)$ converge almost surely to zero as $n \to \infty$, due to the law of large numbers and the law of the iterated logarithm, respectively; therefore, we have the following:

$$\lim_{n\to\infty} g_{2,n} = 0 \text{ a.s.} \tag{3}$$

Let us consider the numerator of the second coordinate of the random vector as follows:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X})^4 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - a) - (\overline{X} - a))^4 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left((X_i - a) - \frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^4$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^4 - \frac{4}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^3 \cdot \frac{1}{n} \sum_{j=1}^{n} (X_j - a) + \frac{6}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^2 \cdot \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^2$$

$$- \frac{4}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a) \cdot \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^3 + \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} (X_j - a) \right)^4$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^4 - \frac{4}{n^{\frac{3}{4}}} \sum_{i=1}^{n} (X_i - a)^3 \cdot \frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (X_j - a) + \frac{6}{n} \sum_{i=1}^{n} (X_i - a)^2 \cdot \left(\frac{1}{n^{\frac{3}{4}}} \sum_{j=1}^{n} (X_j - a) \right)^2$$

$$- \frac{4}{n} \sum_{i=1}^{n} (X_i - a) \cdot \left(\frac{1}{n^{\frac{5}{6}}} \sum_{j=1}^{n} (X_j - a) \right)^3 + \left(\frac{1}{n^{\frac{7}{8}}} \sum_{j=1}^{n} (X_j - a) \right)^4 .$$

Denoting $g_{4,n}$, the four last terms, we have the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \overline{X})^4 =: \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - a)^4 + g_{4,n}. \tag{4}$$

From the law of large numbers and the law of the iterated logarithm, we have the following:

$$\lim_{n\to\infty} g_{4,n} = 0 \text{ a.s.} \tag{5}$$

From (1)–(5) and Slutsky's theorem, it follows that the limiting distribution of the sequence $\overrightarrow{Z}_n$ coincides with the limiting distribution of the sequence, as follows:

$$\overrightarrow{Z'}_n := (Z'_{1,n}, Z'_{2,n}) := \left(\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - a)^3 - 3\sigma^2(X_i - a))}{\sigma^3}, \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} ((X_i - a)^4 - 3\sigma^4)}{\sigma^4} \right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{(X_i - a)^3 - 3\sigma^2(X_i - a)}{\sigma^3}, \frac{(X_i - a)^4 - 3\sigma^4}{\sigma^4} \right).$$

By the central limit theorem, the sequence $\overrightarrow{Z'}_n$ converges to the random vector (W_1, W_2), whose coordinates have a joint normal distribution with a covariance matrix, as follows:

$$\Sigma = \begin{pmatrix} \mathbf{D}W_1 & \mathbf{E}W_1 W_2 \\ \mathbf{E}W_1 W_2 & \mathbf{D}W_2 \end{pmatrix}.$$

It is easy to see that $\mathbf{D}((X_i - a)^3 - 3\sigma^2(X_i - a)) = 6\sigma^6$ (thus, $\mathbf{D}W_1 = 6$); $\mathbf{D}((X_i - a)^4 - 3\sigma^4) = 96\sigma^8$ (thus, $\mathbf{D}W_2 = 96$). From the fact that the odd moments of a centered normally distributed random variable are equal to zero, we derive the following:

$$\mathbf{E}\left(\frac{(X_i - a)^3 - 3\sigma^2(X_i - a)}{\sigma^3} \cdot \frac{(X_i - a)^4 - 3\sigma^4}{\sigma^4} \right) = 0$$

(and, thus, $\mathbf{E}W_1 W_2 = 0$). Therefore,

$$\frac{W_1^2}{6} + \frac{W_2^2}{96} \sim \chi^2(2).$$

Proof. The proof of statement (3) is similar (just simpler since a and σ^2 are known), so we omit it. □

5. Conclusions

In this paper, new modifications to the Jarque–Bera statistics are proposed. Detailed proofs are provided, which are simple and accessible even to undergraduate students in probability and statistics.

A Monte Carlo study showed that the Jarque–Bera statistic and its new modifications perform well on the class of Pearson distributions. When the alternative distribution does not belong to the Pearson family, Jarque–Bera and its modifications perform well alongside other statistics such as Anderson–Darling, Cramér–von Mises, and Shapiro–Wilk. Like any specific test, the Jarque–Bera test and its modifications have natural limitations in their application. Despite the test performing well on classical distributions, a significant drawback is that it cannot distinguish between symmetric distributions with a kurtosis of 3.

Our goal was not to explore and compare all existing tests; therefore, we limited our comparison to the most widely used tests for normality. Comparative studies on a broader class of normality tests can be found in [10,11]. Note that the findings of [10,11] are aligned with our simulation results. For more comparative studies, we also refer to [12].

In this paper, we are limited to the univariate case. Multivariate normality tests represent a curious and interesting area of research. For discussions on existing tests and the possibility of multivariate extensions of some known statistics, including Jarque–Bera we refer the readers to [12,13], and the references therein.

Author Contributions: Conceptualization, V.G. and A.L.; methodology, A.L. and S.K.; software, data curation and validation, S.K., Y.I., O.L., L.S. and K.Z.; writing—original draft preparation, A.L. and A.Y.; writing—review and editing, Y.I., O.L., L.S. and K.Z.; visualization, S.K. and A.Y.; project administration, V.G. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

Appendix A. Tables and Figures

In this section, we report the power of various tests for normality using Monte Carlo simulations under alternative hypotheses. The simulations were performed in R. The sample sizes were small, moderate, and large, with n = 25, 50, 75, 100, 150, 200, 250, 500, and 1000. Although simulations were conducted for all stated sample sizes, the table includes only rows up to the first row where all criteria have a power of 1, to maintain shorter tables.

The null hypothesis is $N(0, 1)$ in almost all cases; exceptions are specified separately. As alternative hypotheses, we considered normal, log-normal, mixed normal, Student's t, gamma, and uniform distributions.

Recall that the following procedure to estimate the power was used: 1000 samples with a given sample size were generated from the alternative hypothesis with specific parameters, and the ratio of the number of rejections of the null hypothesis to 1000 was calculated.

We used the notations AD_c and CVM_c for Anderson–Darling and Cramér–von Mises tests respectively, where the parameters were replaced with their estimates (i.e., modifications for testing composite hypotheses).

Table A1. The estimated power is reported when the null hypothesis $H_0 : X \sim N(0,1)$ is tested against samples simulated from the normal distribution $N(0,2)$. The last column contains the test power of the Fisher test for the null hypothesis where the variance is equal to 1.

	n	JB_{a,σ^2}	JB_{σ^2}	KS	AD	CVM	Fisher
$\alpha = 0.05$	25	0.739	0.673	0.175	0.418	0.173	0.392
	50	0.907	0.882	0.265	0.655	0.276	0.680
	75	0.975	0.970	0.391	0.832	0.433	0.835
	100	0.991	0.991	0.510	0.924	0.607	0.929
	150	1	0.999	0.732	0.993	0.822	0.994
	200	1	1	0.857	1	0.933	1
	250	1	1	0.951	1	0.978	1
	500	1	1	1	1	1	1
$\alpha = 0.01$	25	0.655	0.576	0.042	0.179	0.039	0.164
	50	0.852	0.831	0.077	0.348	0.074	0.401
	75	0.953	0.940	0.137	0.529	0.124	0.631
	100	0.988	0.986	0.191	0.752	0.204	0.801
	150	1	1	0.385	0.940	0.454	0.943
	200	1	1	0.539	0.982	0.669	0.991
	250	1	1	0.730	0.997	0.841	0.998
	500	1	1	0.996	1	1	1
	1000	1	1	1	1	1	1

Table A2. The estimated power is reported when the null hypothesis $H_0 : X \sim N(0,1)$ is tested against samples simulated from the normal distribution $N(1,1)$. The two last columns contain the test powers of the Student and Welch statistics used to test whether the difference between the two means is statistically significant or not.

	n	JB_{a,σ^2}	JB_a	KS	AD	CVM	Student	Welch
$\alpha = 0.05$	25	0.946	0.015	0.993	0.997	0.997	0.929	0.929
	50	1	0.961	1	1	1	0.999	0.999
	75	1	0.999	1	1	1	1	1
	100	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.889	0.002	0.949	0.987	0.978	0.803	0.803
	50	0.997	0.030	0.999	1	1	0.988	0.988
	75	1	0.967	1	1	1	0.999	0.999
	100	1	1	1	1	1	1	1

Table A3. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from the Cauchy distribution.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	0.999	0.900	0.997	0.896	0.262	0.909	0.971	0.927	0.248	0.927	0.939
	50	1	0.994	1	0.993	0.478	0.992	0.999	0.995	0.478	0.997	0.997
	75	1	1	1	1	0.720	1	1	1	0.676	1	1
	100	1	1	1	1	0.864	1	1	1	0.853	1	1
	150	1	1	1	1	0.984	1	1	1	0.971	1	1
	200	1	1	1	1	0.999	1	1	1	1	1	1
	250	1	1	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.995	0.854	0.994	0.854	0.072	0.816	0.945	0.882	0.071	0.880	0.894
	50	1	0.988	1	0.986	0.186	0.986	0.996	0.995	0.161	0.996	0.994
	75	1	1	1	1	0.329	1	1	1	0.283	1	1
	100	1	1	1	1	0.512	1	1	1	0.463	1	1
	150	1	1	1	1	0.867	1	1	1	0.806	1	1
	200	1	1	1	1	0.971	1	1	1	0.935	1	1
	250	1	1	1	1	0.997	1	1	1	0.994	1	1
	500	1	1	1	1	1	1	1	1	1	1	1

Table A4. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from the Student's t distribution with 5 degrees of freedom.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	0.560	0.215	0.531	0.216	0.069	0.135	0.168	0.200	0.078	0.178	0.265
	50	0.742	0.371	0.715	0.385	0.069	0.173	0.186	0.269	0.065	0.230	0.411
	75	0.857	0.523	0.828	0.527	0.065	0.27	0.254	0.387	0.078	0.341	0.531
	100	0.910	0.628	0.911	0.624	0.06	0.326	0.297	0.473	0.070	0.422	0.624
	150	0.976	0.764	0.974	0.758	0.073	0.443	0.420	0.605	0.086	0.555	0.750
	200	0.992	0.861	0.991	0.856	0.082	0.540	0.510	0.736	0.112	0.685	0.861
	250	0.996	0.904	0.997	0.914	0.102	0.634	0.625	0.829	0.124	0.771	0.907
	500	1	0.990	1	0.991	0.174	0.905	0.936	0.976	0.205	0.965	0.987
	1000	1	1	1	1	0.480	0.998	1	1	0.497	1	1
$\alpha = 0.01$	25	0.506	0.138	0.459	0.148	0.012	0.057	0.053	0.097	0.016	0.081	0.136
	50	0.66	0.285	0.647	0.288	0.011	0.078	0.066	0.153	0.013	0.129	0.235
	75	0.803	0.421	0.803	0.442	0.007	0.124	0.077	0.247	0.011	0.191	0.385
	100	0.882	0.504	0.875	0.517	0.008	0.158	0.094	0.287	0.013	0.248	0.450
	150	0.96	0.702	0.95	0.698	0.013	0.264	0.163	0.462	0.024	0.407	0.644
	200	0.986	0.789	0.983	0.793	0.020	0.346	0.240	0.575	0.023	0.512	0.744
	250	0.991	0.865	0.991	0.873	0.022	0.388	0.313	0.672	0.026	0.596	0.840
	500	1	0.990	1	0.992	0.035	0.750	0.716	0.940	0.040	0.902	0.986
	1000	1	1	1	1	0.128	0.970	0.990	1	0.122	0.997	1

Table A5. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from $U[-3,3]$.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	0.989	0	0.976	0.001	0.677	0.125	0.972	0.230	0.708	0.178	0.128
	50	1	0	0.999	0.001	0.953	0.252	1	0.567	0.976	0.414	0.446
	75	1	0.112	1	0.084	0.995	0.435	1	0.844	1	0.681	0.837
	100	1	0.602	1	0.543	1	0.574	1	0.941	1	0.832	0.964
	150	1	0.987	1	0.985	1	0.839	1	0.997	1	0.974	1
	200	1	1	1	1	1	0.947	1	1	1	0.996	1
	250	1	1	1	1	1	0.989	1	1	1	1	1
	500	1	1	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.977	0	0.957	0	0.339	0.024	0.874	0.058	0.323	0.043	0.014
	50	1	0	0.999	0	0.782	0.074	0.998	0.271	0.823	0.179	0.126
	75	1	0	1	0	0.959	0.147	1	0.553	0.982	0.373	0.444
	100	1	0.005	1	0.002	0.999	0.276	1	0.799	1	0.576	0.777
	150	1	0.531	1	0.483	1	0.524	1	0.976	1	0.887	0.990
	200	1	0.975	1	0.970	1	0.725	1	0.998	1	0.972	0.999
	250	1	0.999	1	0.999	1	0.891	1	1	1	0.996	1
	500	1	1	1	1	1	1	1	1	1	1	1

Table A6. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from $\gamma(2,1)$.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	1	0.128	0.760	0.380	1	0.409	1	0.576	1	0.512	0.623
	50	1	1	0.950	0.772	1	0.730	1	0.896	1	0.843	0.934
	75	1	1	0.993	0.944	1	0.880	1	0.975	1	0.954	0.984
	100	1	1	0.999	0.992	1	0.948	1	0.999	1	0.994	0.999
	150	1	1	1	1	1	0.994	1	1	1	0.998	1
	200	1	1	1	1	1	1	1	1	1	1	1

Table A6. *Cont.*

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	*JB*	*KS*	*LF*	*AD*	AD_c	*CVM*	CVM_c	*SW*
$\alpha = 0.01$	25	1	0.046	0.676	0.244	1	0.186	1	0.332	1	0.287	0.349
	50	1	0.307	0.930	0.640	1	0.444	1	0.755	1	0.676	0.805
	75	1	1	0.983	0.847	1	0.700	1	0.936	1	0.879	0.957
	100	1	1	0.994	0.935	1	0.825	1	0.982	1	0.955	0.994
	150	1	1	1	0.999	1	0.965	1	1	1	0.997	1
	200	1	1	1	1	1	0.995	1	1	1	1	1
	250	1	1	1	1	1	0.999	1	1	1	1	1
	500	1	1	1	1	1	1	1	1	1	1	1

Table A7. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from the log-normal distribution $LN(0,1)$.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	*JB*	*KS*	*LF*	*AD*	AD_c	*CVM*	CVM_c	*SW*
$\alpha = 0.05$	25	0.994	0.760	0.907	0.860	1	0.889	1	0.960	1	0.947	0.963
	50	1	1	0.994	0.994	1	0.998	1	1	1	1	1
	75	1	1	1	1	1	1	1	1	1	1	
$\alpha = 0.01$	25	0.996	0.574	0.850	0.737	1	0.727	1	0.881	1	0.853	0.888
	50	1	0.966	0.980	0.978	1	0.972	1	0.996	1	0.991	0.998
	75	1	1	0.998	0.999	1	1	1	1	1	1	1
	100	1	1	0.999	1	1	1	1	1	1	1	1
	150	1	1	1	1	1	1	1	1	1	1	1 1

Table A8. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from the *Mix*, which is a mixture of the standard normal distribution $N(0,1)$ and the normal distribution $N(0,9)$ with equal mixture weights.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	*JB*	*KS*	*LF*	*AD*	AD_c	*CVM*	CVM_c	*SW*
$\alpha = 0.05$	25	0.997	0.228	0.997	0.212	0.277	0.271	0.955	0.339	0.292	0.345	0.366
	50	1	0.430	1	0.433	0.499	0.425	0.999	0.553	0.538	0.536	0.581
	75	1	0.599	1	0.575	0.791	0.626	1	0.756	0.831	0.742	0.737
	100	1	0.737	1	0.724	0.889	0.745	1	0.878	0.937	0.87	0.863
	150	1	0.884	1	0.889	0.991	0.893	1	0.972	0.996	0.967	0.97
	200	1	0.957	1	0.957	1	0.975	1	0.996	1	0.997	0.993
	250	1	0.988	1	0.986	1	0.991	1	0.999	1	0.999	0.999
	500	1	1	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.998	0.155	0.993	0.154	0.089	0.105	0.840	0.147	0.077	0.139	0.170
	50	1	0.349	1	0.334	0.196	0.235	0.983	0.374	0.202	0.350	0.363
	75	1	0.446	1	0.444	0.420	0.355	0.998	0.537	0.436	0.525	0.505
	100	1	0.591	1	0.581	0.564	0.485	1	0.721	0.613	0.699	0.662
	150	1	0.789	1	0.791	0.896	0.738	1	0.909	0.934	0.899	0.883
	200	1	0.893	1	0.893	0.986	0.892	1	0.983	0.994	0.981	0.961
	250	1	0.958	1	0.952	1	0.961	1	0.996	1	0.995	0.991
	500	1	1	1	1	1	1	1	1	1	1	1

Table A9. The null hypothesis $H_0 : X \sim N(2,2)$ is tested against data sampled from the gamma $(2,1)$ distribution.

	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	0.292	0.274	0.353	0.410	0.136	0.400	0.134	0.558	0.131	0.505	0.620
	50	0.421	0.477	0.592	0.761	0.269	0.732	0.307	0.885	0.263	0.837	0.913
	75	0.523	0.652	0.765	0.942	0.370	0.856	0.482	0.975	0.382	0.949	0.991
	100	0.643	0.805	0.948	0.995	0.421	0.952	0.658	0.999	0.465	0.990	1
	150	0.815	0.935	1	1	0.619	0.996	0.913	1	0.679	1	1
	200	0.923	0.971	1	1	0.831	1	0.992	1	0.837	1	1
	250	0.989	0.994	1	1	0.979	1	1	1	0.928	1	1
	500	1	1	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.219	0.176	0.260	0.272	0.045	0.171	0.034	0.316	0.040	0.270	0.360
	50	0.363	0.412	0.488	0.617	0.098	0.459	0.084	0.757	0.092	0.675	0.790
	75	0.451	0.581	0.665	0.832	0.173	0.671	0.179	0.928	0.169	0.872	0.962
	100	0.563	0.701	0.800	0.945	0.197	0.818	0.250	0.979	0.197	0.946	0.991
	150	0.684	0.882	0.967	0.997	0.371	0.969	0.571	1	0.407	0.997	1
	200	0.807	0.951	0.999	1	0.507	0.996	0.797	1	0.549	1	1
	250	0.904	0.99	1	1	0.667	1	0.951	1	0.724	1	1
	500	1	1	1	1	1	1	1	1	0.992	1	1
	1000	1	1	1	1	1	1	1	1	1	1	1

Table A10. The null hypothesis $H_0 : X \sim N(0,1)$ is tested against data sampled from a mixture of a standard normal distribution (weight 0.9) and a sum of independent random variables—with a standard normal distribution and Poisson distribution with $\lambda = 5$ (weight 0.1).

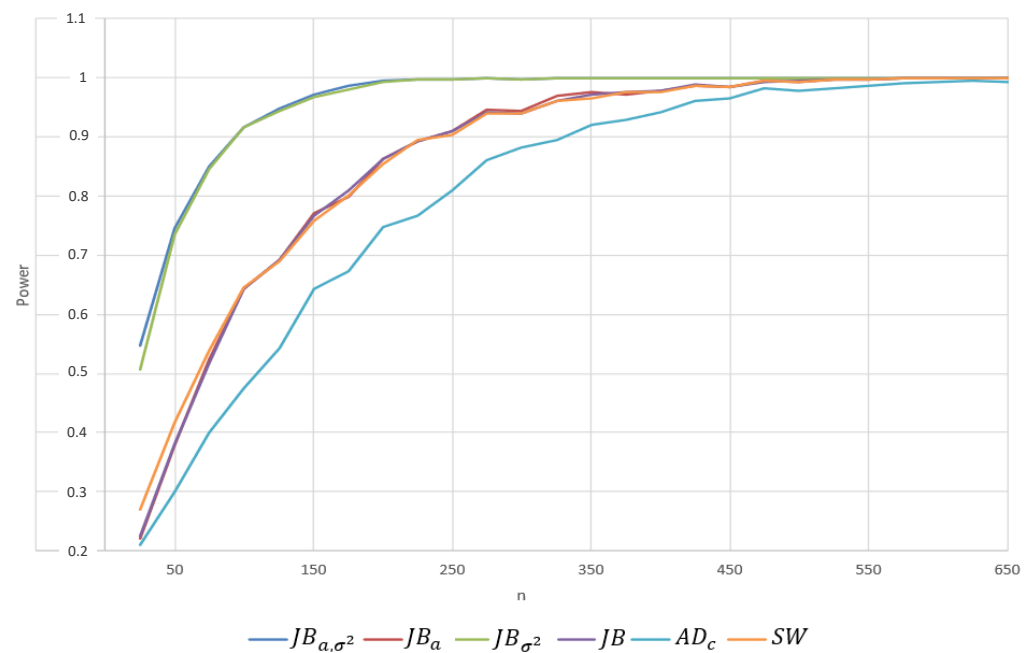|	n	JB_{a,σ^2}	JB_a	JB_{σ^2}	JB	KS	LF	AD	AD_c	CVM	CVM_c	SW
$\alpha = 0.05$	25	0.994	0.923	0.985	0.884	0.111	0.808	0.878	0.890	0.160	0.878	0.920
	50	1	0.999	1	0.990	0.134	0.952	0.966	0.990	0.194	0.981	0.995
	75	1	1	1	1	0.235	0.991	0.999	0.999	0.319	0.999	1
	100	1	1	1	1	0.310	0.998	1	1	0.398	1	1
	150	1	1	1	1	0.570	1	1	1	0.609	1	1
	200	1	1	1	1	0.823	1	1	1	0.770	1	1
	250	1	1	1	1	0.974	1	1	1	0.887	1	1
	500	1	1	1	1	1	1	1	1	0.999	1	1
	1000	1	1	1	1	1	1	1	1	1	1	1
$\alpha = 0.01$	25	0.996	0.861	0.992	0.796	0.033	0.647	0.671	0.791	0.049	0.754	0.842
	50	1	0.992	1	0.985	0.043	0.884	0.854	0.966	0.065	0.951	0.984
	75	1	1	1	0.998	0.085	0.979	0.990	0.994	0.131	0.989	0.999
	100	1	1	1	1	0.108	0.994	0.998	1	0.169	0.998	1
	150	1	1	1	1	0.207	1	1	1	0.275	1	1
	200	1	1	1	1	0.379	1	1	1	0.424	1	1
	250	1	1	1	1	0.622	1	1	1	0.617	1	1
	500	1	1	1	1	1	1	1	1	0.972	1	1
	1000	1	1	1	1	1	1	1	1	1	1	1

Figure A1. The power for $\alpha = 0.05$ depending on the sample size n ($H_0 : X \sim N(0,1)$ is tested against data sampled from the Student's t distribution with 5 degrees of freedom. Figure created by S. Khrushchev and A. Yambartsev.
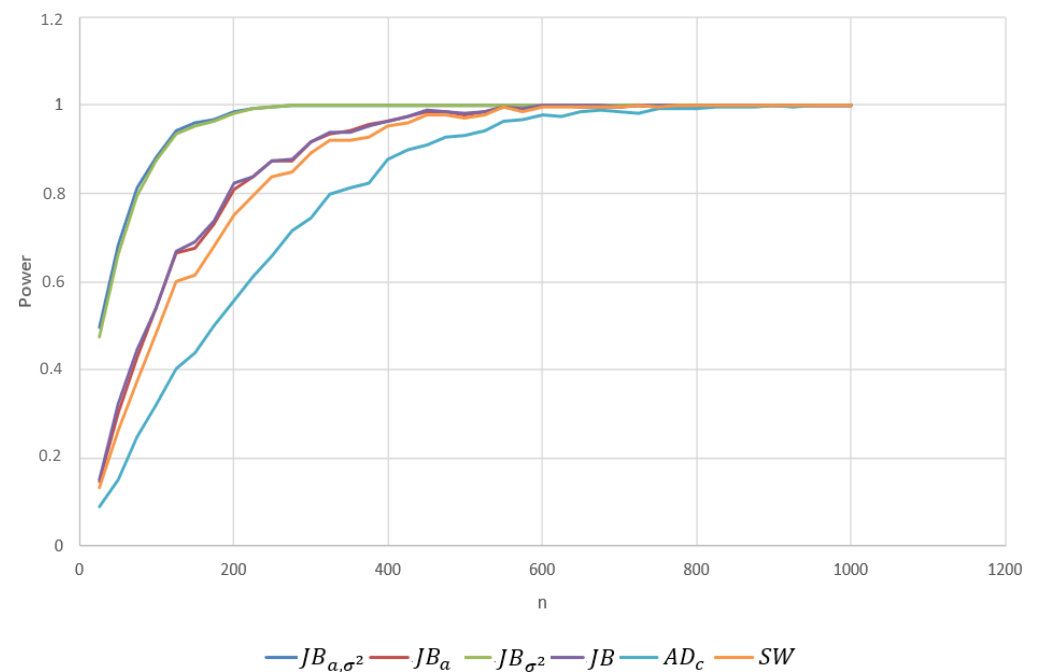

Figure A2. The power for $\alpha = 0.01$ depending on the sample size n ($H_0 : X \sim N(0,1)$ is tested against data sampled from the Student's t distribution with 5 degrees of freedom. Figure created by S. Khrushchev and A. Yambartsev.

References

1. Jarque, C.M.; Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* **1980**, *6*, 255–259. [CrossRef]
2. Jarque, C.M.; Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Econ. Lett.* **1981**, *7*, 313–318. [CrossRef]
3. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **1987**, *55*, 163–172. [CrossRef]

4. Pearson, K. Mathematical contributions to the theory of evolution, XIX: Second supplement to a memoir on skew variation. *Philos. Trans. R. Soc. A* **1916**, *216*, 429–457.
5. Searls, D.T. The utilization of a known coefficient of variation in the estimation procedure. *J. Am. Stat. Assoc.* **1964**, *59*, 1225–1226. [CrossRef]
6. Fu, Y.; Wang, H.; Wong, A. Inference for the normal mean with known coefficient of variation. *Open J. Stat.* **2013**, *3*, 41368. [CrossRef]
7. Rana, S.; Eshita, N.N.; Al Mamun, A.S.M. Robust normality test in the presence of outliers. *J. Phys. Conf. Ser.* **2021**, *1863*, 012009. [CrossRef]
8. Gorman, K.B.; Williams, T.D.; Fraser, W.R. Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus Pygoscelis). *PLoS ONE* **2014**, *9*, e90081. [CrossRef] [PubMed]
9. Slutsky, E. Über stochastische Asymptoten und Grenzwerte. *Metron* **1925**, *5*, 3–89.
10. Yap, B.W.; Sim, C.H. Comparisons of various types of normality tests. *J. Stat. Comput. Simul.* **2011**, *81*, 2141–2155. [CrossRef]
11. Razali, N.M.; Wah, Y.B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33.
12. Khatun, N. Applications of normality test in statistical analysis. *Open J. Stat.* **2021**, *11*, 113. [CrossRef]
13. Chen, W.; Genton, M.G. Are you all normal? It depends! *Int. Stat. Rev.* **2023**, *91*, 114–139. [CrossRef]