Sri Lankan Journal of Applied Statistics,Vol 26 (Special Issue)
Selected Papers from International Statistics Conference (ISC) 2024

Test for Normality Based on an Edgeworth Expansion

K. P. Hapuarachchi[1], K. G. H. S. Peiris[2,*], and D. C. Wickramarachchi[1]

[1]Department of Statistics, Faculty of Applied Sciences,
University of Sri Jayewardenepura, Sri Lanka

[2]Department of Statistics and Actuarial Science, Faculty of Science,
Simon Fraser University, Canada

*Corresponding author: hashan_peiris@sfu.ca

Received: 19th June, 2025/ Revised: 10th November, 2025/ Published: 30th November, 2025

ABSTRACT

Many statistical procedures require the assumption that the observations in a random sample are drawn from a normal distribution. Several statistical techniques, mostly based on either population moments or empirical distribution functions, are currently available to test whether the observations in a random sample are normally distributed. In this study, we use the Edgeworth series expansion to model deviations from normality, depending on the skewness and excess kurtosis of a normal population. We develop a test statistic based on these two measures to test for normality. However, the sampling distribution of the test statistic is not mathematically tractable. Therefore, we conducted a simulation study by generating the sampling distribution of the proposed test statistic for different sample sizes when the data are normally distributed. The critical values were also calculated at different levels of significance. The power of the proposed test was empirically compared with the Shapiro-Wilk (SW), Shapiro-Francia (SF), Jarque-Bera (JB), Cramer-von Mises (CM), Lilliefors (LF), Kolmogorov-Smirnov (KS), and Anderson-Darling (AD) tests. The proposed test demonstrates competitive power against JB, CM, LF, KS and AD tests for small samples under specific alternatives with low skewness.

Keywords: Normality tests, Edgeworth expansion, Power comparison, Skewness, Kurtosis

1 Introduction

Generally, the assumption that the observations of a random sample are normally distributed is a prerequisite for many statistical procedures. Widely used tests such as the F, Z, and t tests, as well as the error term in the classical linear model, rely on normality. When this assumption is violated, the resulting test statistics and associated inferences may be misleading. It is therefore standard practice to examine the validity of the normality assumption before performing statistical analyzes.

A large number of tests for normality have been proposed, and they can be broadly classified as parametric (often moment-based or regression-based) and non-parametric (typically empirical distribution function (EDF)-based) procedures. The early parametric work by Fisher (1930b,a), Pearson (1930, 1931, 1963, 1965), and Williams (1935) introduced the standardized third and fourth moments, β_1 and β_2, as test statistics for normality. These moment-based tests focus on detecting departures from normality attributable to skewness and kurtosis. Subsequently, Geary (1935, 1936, 1947) highlighted the ratio of the mean deviation to the standard deviation as a further test statistic, and David et al. (1954) emphasized the use of the relative range (the ratio of the range to the standard deviation). These results were further refined by Pearson and Stephens (1964).

Another important class of parametric tests is based on regressions or correlations designed to detect departures from linearity in normal probability plots. The Shapiro–Wilk family of tests (Shapiro and Wilk 1965; Shapiro et al. 1968; Shapiro and Francia 1972) and the procedures proposed by D'Agostino (1973, 1971) fall into this category. Filliben (1975) introduced the normal probability plot correlation coefficient as a test statistic for the composite hypothesis of normality, and Verrill and Johnson (1988) proposed a modification of the Shapiro–Wilk correlation statistic.

Non-parametric and EDF-based procedures provide another major class of normality tests. The Pearson χ^2 goodness-of-fit test is a well-known example for grouped or nominal data. For continuous distributions, Kolmogorov (1933) proposed a goodness-of-fit test based on the supremum distance between the empirical and hypothesized distribution functions, and Smirnov (1939) extended this idea to the two-sample case. These are now known as the Kolmogorov–Smirnov one-sample and two-sample tests. EDF-based tests have been further developed and modified over time: Friedman et al. (1988) introduced a modified Kolmogorov–Smirnov-type test for normality, while Sinclair et al. (1990) discussed two modifications of the Anderson–Darling test statistic. Additional EDF-type and related statistics for normality have been proposed by Birnbaum (1952), Miller (1956), Darling (1957), Lilliefors (1967), Srinivasan (1970), Guo et al. (2010), Torabi et al. (2016), and Bayoud (2021).

Further extensions of normality testing include multi-sample tests, such as the procedure of **?**, which is based on the generalized distance between the standardized order statistics vectors across samples and their expected values under normality, as well as tests for multivariate normality discussed by Székely and Rizzo (2005) and Hanusz and Tarasinska (2014). A substantial simulation literature has compared the power of many of these tests with a wide range of alternatives; see, for example, Dufour et al. (1998), Öztuna et al. (2006), Bayoud (2016, 2021), Kahya (1991), Seier (2002), Mendes and Pala (2003), **?**, Razali and Wah (2011), and Romao et al. (2009). These studies consistently indicate that no single test is uniformly most powerful and that performance often depends on the pattern of departure from normality (e.g., primarily skewness, primarily kurtosis, or a combination of both).

Despite this extensive literature, most existing tests either (i) emphasize separate assessments of skewness and kurtosis via low-order moments, or (ii) provide omnibus EDF- or regression-based procedures whose connection to specific features of the underlying distribution (such as the joint effect of skewness and kurtosis) is less explicit. In particular, there is a relative lack of tests that are explicitly constructed from an Edgeworth expansion, using the standardized third and fourth cumulants to build a single, interpretable measure of deviation from normality that simultaneously captures asymmetry and tail behavior.

In this study, we address this gap by deriving a test statistic based on the Edgeworth series expansion of the standardized density. Following the approach of Gayen (1949), who used Edgeworth expansions to approximate Student's t-distribution for non-normal populations, we consider a third-order approximation of the standard normal density of the form

$$f(z) = \phi(z) - \frac{\lambda_3}{3!}\,\phi^{(3)}(z) + \frac{\lambda_4}{4!}\,\phi^{(4)}(z) + \frac{10\lambda_3^2}{6!}\,\phi^{(6)}(z), \tag{1}$$

where $\phi(z)$ is the standard normal density, $\phi^{(r)}(z) = (\frac{d}{dz})^r \phi(z)$, and λ_3 and λ_4 measure the skewness and excess kurtosis of the distribution, respectively. For the normal distribution, $\lambda_3 = \lambda_4 = 0$, so that $f(z) = \phi(z)$.

By varying λ_3 and λ_4, we obtain a wide class of symmetric and asymmetric marginal distributions that represent a broad spectrum of deviations from normality. Our proposed test statistic combines information on both λ_3 and λ_4 into a single measure and is therefore specifically designed to be sensitive to departures characterized by simultaneous changes in symmetry and tail heaviness.

Because the exact sampling distribution of the proposed Edgeworth-based test statistic is not analytically tractable, we employ a large-scale Monte Carlo simulation study to obtain critical values at a few selected levels of significance and prepare a table to refer to test the null hypothesis. Then, we conduct another simulation study to investigate its empirical power at 5% level of significance using the statistical software `R` via `RStudio` and the packages `graphics`, `stats`, `tseries`, and `nortest`, where we compare the performance of our test with several existing moment-based (especially, the Jarque–

Bera test), regression-based, and EDF-based tests across a range of sample sizes and alternative distributions.

Consistent with earlier simulation studies, our results confirm that no test dominates in all scenarios, but show that the proposed Edgeworth-based procedure offers competitive, and in some cases superior, power for alternatives involving joint deviations in skewness and kurtosis.

The remainder of this article is organized as follows. In Section 2, we present the methodological framework: we first develop the proposed Edgeworth based test statistic, then introduce the penalized test statistic, derive the sampling distribution, and obtain critical value tables. After that, we describe the design of the power comparison with seven existing normality tests. Section 3 reports and discusses the main empirical findings, including the behaviour of the sampling distribution of the test statistic and the comparative power performance of the proposed test against the seven competitors under selected alternatives and sample sizes. Section 4 concludes with a summary of the main contributions, practical implications for normality assessment, and possible directions for future research.

2 Methodology

2.1 Development of the test statistic

In many classical statistical inferences, we typically assume that we have selected a sample from a normal population. Suppose, however, that the process of generating the data does not follow a normal distribution. We will still assume that the mean and the standard deviation are finite and denoted by μ and σ respectively. Let us introduce two additional parameters λ_3, the coefficient of skewness, and λ_4, a measure of excess (from kurtosis). We can then model the process through an Edgeworth expansion given below.

$$\begin{aligned} f\Big(\frac{x-\mu}{\sigma}|\mu,\sigma,\lambda_3,\lambda_4\Big) &= \phi\Big(\frac{x-\mu}{\sigma}\Big) - \frac{\lambda_3}{3!}\phi^{(3)}\Big(\frac{x-\mu}{\sigma}\Big) \\ &\quad + \frac{\lambda_4}{4!}\phi^{(4)}\Big(\frac{x-\mu}{\sigma}\Big) + \frac{10\lambda_3^2}{6!}\phi^{(6)}\Big(\frac{x-\mu}{\sigma}\Big) \end{aligned} \tag{2}$$

For a single observation of this distribution; $E\big(\frac{X-\mu}{\sigma}\big) = 0$, $E\big(\frac{X-\mu}{\sigma}\big)^2 = 1$, $E\big(\frac{X-\mu}{\sigma}\big)^3 = \lambda_3$ and $E\big(\frac{X-\mu}{\sigma}\big)^4 = 3 + \lambda_4$

When making statistical inferences about the mean based on a sample drawn from a normal population with mean μ and variance σ^2, we generally compare the value of the sample mean, $\bar{X}$ with the value of the sample standard deviation, s, as in the test statistic, $t = \sqrt{n}\big(\frac{\bar{X}-\mu}{s}\big)$, or the interval estimate, $(\bar{X} - ks, \bar{X} + ks)$.

What happens when ($\bar{X}$,s) are computed from a non-normal population? Consider sampling from a non-normal population, still with the mean μ and the variance σ^2, but with some skewness and excess of kurtosis quantified by means of the parameters λ_3 and λ_4, respectively. Thus, we are retaining the same description of the location and dispersion of the distribution by adding additional parameters to account for the way in which the mass is dispersed, namely a measure for the direction of the spread (skewness, λ_3) and the flatness/ peakedness (excess of kurtosis, λ_4).

We can then approximate the true distribution of the sample by an Edgeworth expansion, as in equation 2, to obtain the approximate joint distribution of ($\bar{X}$,s) for this non-normal population. Using the simplified joint distribution of ($\bar{X}$,s) given by Gayen (1949), a marginal distribution of the sample mean, $\bar{X}$, is obtained as

$$g(\bar{x}) = \Big(\frac{ne^{n\bar{x}^2}}{2\pi}\Big)^{1/2}\Big\{1 + \frac{\lambda_3}{6}\bar{x}(n\bar{x}^2 - 3) + \frac{\lambda_4}{24n}(n^2\bar{x}^2 - 6n\bar{x}^2 + 3) + \frac{\lambda_3^2}{72n}(n^3\bar{x}^6 - 15n^2\bar{x}^4 + 45n\bar{x}^2 - 15)\Big\} \tag{3}$$

here, $E\big(\frac{\bar{X}-\mu}{\sigma}\big) = 0\,, E\big(\frac{\bar{X}-\mu}{\sigma}\big)^2 = \frac{1}{n}\,, E\big(\frac{\bar{X}-\mu}{\sigma}\big)^3 = \frac{\lambda_3}{n^2}$ and $E\big(\frac{\bar{X}-\mu}{\sigma}\big)^4 = \frac{3}{n^2} + \frac{\lambda_4}{n^3}$.

Note that,$\sqrt{n}(\frac{\bar{X}-\mu}{\sigma}) \to N(0,1)$ as $n \to \infty$, which is a direct consequence of the Central Limit Theorem. Hence, the distribution of the sample mean is apparently unaffected by the additional parameters λ_3 and λ_4.

Now, let us consider the distribution of the sample standard deviation for this non-normal population. The distribution of the sample standard deviation, s, for the Edgeworth expansion in equation 2, as given by Gayen (1949) can be written as,

$$g(s) = \frac{2s^{n-2}}{\Gamma(\frac{n-1}{2})\big(\frac{2e^{s^2}}{n-1}\big)^{(n-1)/2}}\Big[1 + \frac{\lambda_4}{8n}(n-1)^2\Big\{\frac{(n-1)}{(n+1)}s^2 - 2s + 1\Big\} + \frac{\lambda_3^2}{12n}(n-2)(n-1)\Big\{\frac{(n-1)^2}{(n+3)(n+1)}s^6 - 3\frac{(n-1)}{(n+1)}s^4 + 3s^2 - 1\Big\}\Big] \tag{4}$$

Approximating this distribution by a normal with the same first two moments, as depicted in the appendix, we compute the moments as $E(s) = \gamma c\sigma$ and $V(s) = (1-\gamma^2c^2)\sigma^2$ using the distribution in 4, where cσ is the mean under normality, and γ is a correction factor given by

$$\gamma = \Big\{1 - \frac{\lambda_4}{8n}\frac{(n-1)}{(n+1)} + \frac{\lambda_3^2}{4n}\frac{(n-2)}{(n+3)(n+1)}\Big\} \tag{5}$$

Now suppose that we have taken a sample of size n from a non-normal population. In this case, we could compute the sample equivalent of γ, the correction factor.

$$\hat{\gamma} = \left\{ 1 - \frac{\hat{\lambda}_4}{8n} \frac{(n-1)}{(n+1)} + \frac{\hat{\lambda}_3^2}{4n} \frac{(n-2)}{(n+3)(n+1)} \right\} \tag{6}$$

where $\hat{\lambda}_3 = \hat{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\hat{E}(X-\mu)^3}{\hat{\sigma}^3}$ and $\hat{\lambda}_4 = \hat{E}\left(\frac{X-\mu}{\sigma}\right)^4 - 3 = \frac{\hat{E}(X-\mu)^4}{\hat{\sigma}^4} - 3$. Using the maximum likelihood estimators for the central moments,

$$\hat{\mu}_j = \hat{E}(X-\mu)^j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j, \quad j = 2, 3, 4$$

and, so $\hat{\lambda}_3 = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}}$ and $\hat{\lambda}_4 = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3$.

When sampling from a normal distribution, skewness and excess of kurtosis are equal to 0. Thus, γ is equal to 1 in the normal case. Hence, the null and alternative hypotheses can be written as,

H_0: Data are normally distributed ($\gamma = 1$) vs

H_1: Data are not normally distributed ($\gamma \neq 1$)

One can identify that the proposed statistic γ can be viewed as an Edgeworth-based analogue of the Jarque–Bera (JB) statistic (Jarque and Bera (1987)), but with a different scaling and weighting structure that gives it a distinct behaviour. The JB statistic, which is derived as a Lagrange multiplier test, aggregates the squared skewness and squared excess kurtosis through

$$\mathrm{JB} = \frac{n}{6}\left(\lambda_3^2 + \frac{\lambda_4^2}{4}\right),$$

so that departures from normality are measured by a growing (order n) chi-square type quantity in which λ_3^2 and λ_4^2 enter with fixed weights proportional to sample-size. In practice, the JB test relies on the asymptotic χ_2^2 approximation, which is known to provide a poor fit in small samples and to yield inflated Type I error rates relative to the nominal level.

In contrast, our statistic in equation 5 is constructed directly from a third-order Edgeworth expansion, so that excess kurtosis (λ_4) and squared skewness (λ_3^2) appear as finite-sample correction terms to the benchmark value $\gamma = 1$ under normality. This yields a bounded, interpretable index of normality (with nearby values corresponding to small Edgeworth corrections), in which skewness and kurtosis are combined through sample-size–dependent coefficients derived from the expansion.

Moreover, instead of relying on an asymptotic chi-square approximation, critical values for γ are obtained from its finite-sample sampling distribution by Monte Carlo simulation, leading to improved control of the Type I error rate at the nominal significance level. The novelty of γ therefore lies in providing a single, Edgeworth-based measure that simultaneously reflects skewness and kurtosis, incorporates explicit finite-sample adjustments, and is calibrated via its non-standard sampling distribution rather than via asymptotic χ^2 theory.

2.2 The $\hat{\gamma}$ statistic with penalty

The distribution of $\hat{\gamma}$ is suspected to be biased and skewed, especially for small samples, as it depends on the sample size. Ideally, as with many test statistics, with unknown sampling distributions, we would hope that $\hat{\gamma}$ or some simple transformation of $\hat{\gamma}$ will have an approximate normal distribution as the sample becomes large. However, even as *n* becomes large, the convergence to an unbiased, symmetric, normal-like distribution may be quite slow. This is a conjecture that needs to be tested.

The bias in $\hat{\gamma}$ is accounted for, in that, by calculating the maximum likelihood estimates of λ_3 and λ_4. The ratios of the sample moments are being taken which is unlikely to be unbiased estimates of λ_3^2 and λ_4. One possible remedy is to introduce a penalty factor, *k*, and estimate the sample moments by dividing the quadratic forms by (*n-k*) rather than *n*. A similar kind of method is used in the calculation of the sample variance where a penalty of $k = 1$ produces an unbiased estimator. Thus, we can define the penalized gamma statistic $\hat{\gamma}_k$.

$$\hat{\gamma}_k = \left\{1 - \frac{\hat{\lambda}_{4,k}}{8n}\frac{(n-1)}{(n+1)} + \frac{\hat{\lambda}_{3,k}^2}{4n}\frac{(n-2)}{(n+3)(n+1)}\right\} \tag{7}$$

where, $\hat{\lambda}_{3,k} = \frac{\hat{\mu}_{3,k}}{\hat{\mu}_{2,k}^{3/2}}$ and $\hat{\lambda}_{4,k} = \frac{\hat{\mu}_{4,k}}{\hat{\mu}_{2,k}^2} - 3$. and so,

$$\hat{\mu}_{j,k} = \frac{1}{(n-k)}\sum_{i=1}^{n}(x_i - \bar{x})^j, \quad j = 2, 3, 4$$

Then, to correct the bias only, a penalty between +2 and +3 is required for various sample sizes. A sampling distribution requires a penalty of approximately 2.923 for $n = 4$, a penalty of 2.8 for $n = 5$, a penalty of 2.3 for $n = 20$, and a penalty of 2 for large *n*. Thus it is suggested that a penalty of $k = 2$ be used for all sample sizes as this will give a statistic that is almost unbiased for all sample sizes and with a sampling distribution that converges more rapidly to γ as *n* becomes large. We anticipate that the sampling distribution of $\hat{\gamma}_2$ will also stabilize to a symmetric normal-like distribution more rapidly than the non-penalized $\hat{\gamma}$. Here calculating critical values of the penalized $\hat{\gamma}_2$ and drawing histograms have to be done following the same concepts discussed in the following sub-sections.

We name this test the '**HPW**' ('Hapuarachchi-Peiris-Wickramarachchi') test. The test statistic without penalty is given in equation 6 is named as

'**HPW1**' and the test statistic with penalty is given in equation 7 is named as '**HPW2**'. We intend to use these abbreviations throughout the paper.

2.3 The sampling distribution of $\hat{\gamma}$s and critical value tables

We need to obtain or approximate the distribution of $\hat{\gamma}$ when sampling from a normal population to use $\hat{\gamma}$ as a test statistic to test the normality of samples. The distribution of $\hat{\gamma}$ does not appear to be readily available or mathematically tractable as it involves both λ_3 and λ_4, combined non-normality. Instead, let us generate the empirical distribution for different sample sizes by simulating samples from a normal population similar to Bayoud (2016, 2021). According to Glass et al. (1972) simulation size and the level of significance are important factors in building critical values for the test and conducting an inferential power study.

First, 150000 samples were drawn from the standard normal distribution for each sample size from 3 to 200. Then HPW1 and HPW2 were calculated using those samples that eventually formed two empirical distributions for each sample size. After that, histograms were drawn for each of these distributions to identify changes in the shape of the distributions with the various sample sizes. A summary of these histograms is given in section 3.1. Furthermore, the minimum and maximum of each sampling distribution were calculated to recognize changes in the range of the distribution for different sample sizes. It allowed us to get an idea about the density of the empirical distribution. Then we determined the $(\frac{\alpha}{2} \times 100)$th and $((1 - \frac{\alpha}{2}) \times 100)$th percentiles from each of these distributions as critical values for different levels of significance (α = $\{1\%, 2.5\%, 5\%, 10\%\}$) to test normality. The critical values were observed to be almost identical up to 3 decimal places for sample sizes greater than 30. Reflecting changes in critical values, we selected a set of large sample sizes. Thus, we finalized user friendly critical value tables to use the test with {3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200} sample sizes which are listed in the Appendix under Tables 4, 5, 6 and 7.

2.4 Power comparison with seven other normality tests

We compared the power of HPWs for normality of observations with Shapiro–Wilks (SW), Shapiro–Francia (SF), Lilliefors (LF), Kolmogorov–Smirnov (KS), Cramer–von Mises (CM), Anderson–Darling (AD), and Jarque–Bera (JB) tests. These tests are from the categories of ”Tests based on the empirical distribution function”, ”Regression and correlation tests” and ”Tests based on measures of the moments”, according to Romao et al. (2009). Thus, we used 9 tests in the power comparison.

To calculate the power of a test, we first drew a data set from a selected alternative distribution with a size listed in the Table 1. Then we calculated the power at the 5% level of significance. This method was repeated for all sample sizes and alternatives in table 1. As described in section 3.1, empirical distributions of sample sizes from 3 to 6 had contradictions with the condition in the null hypothesis. Also, considering that the minimum sample size for the AD test is 8, we arbitrarily selected these sample sizes for the power comparison. In addition, we followed such categorization for the convenience of comparing and interpreting the results of power comparison, motivated by the work of Romao et al. (2009) and Torabi et al. (2016).

Table 1: Summary of sample sizes and alternative distributions used in the power comparison study.

Component	Category	Values
Sample size	Small	$n \in \{8, 9, 10, 12, 13, 15, 20\}$
	Moderate	$n \in \{25, 30, 35, 40, 45\}$
	Large	$n \in \{50, 75, 100, 150, 200\}$
Alternative distributions	Symmetric	Cauchy$(0, 1)$, Cauchy$(0, 3)$, $t(1)$, $t(30)$, Unif$(0, 1)$, Unif$(0, 5)$
	Skewed	$F(10, 1)$, $F(5, 10)$, $\chi^2(3)$, $\chi^2(30)$, Exp(1.5), Beta$(2, 5)$, Beta$(5, 1)$, Exp(0.5), Log-normal$(0, 0.25)$, Log-normal$(0, 1)$, Weibull$(1, 1.5)$, Weibull$(1.5, 1)$, Gamma$(2, 2)$, Gamma$(2, 3)$

Finally, we summarize the final results of the power values calculated for the 9 selected tests under 20 alternatives and 19 sample sizes. And it is organized according to the categories that appear in Table 1. In addition, Table 2 contains extracted power values from the final results of the power analysis for the median sample sizes of each category in Table 1 and 10 alternative distributions. It shows the power of each test under limited conditions.

3 Results and Discussion

3.1 Sampling distribution of $\hat{\gamma}$s (HPWs)

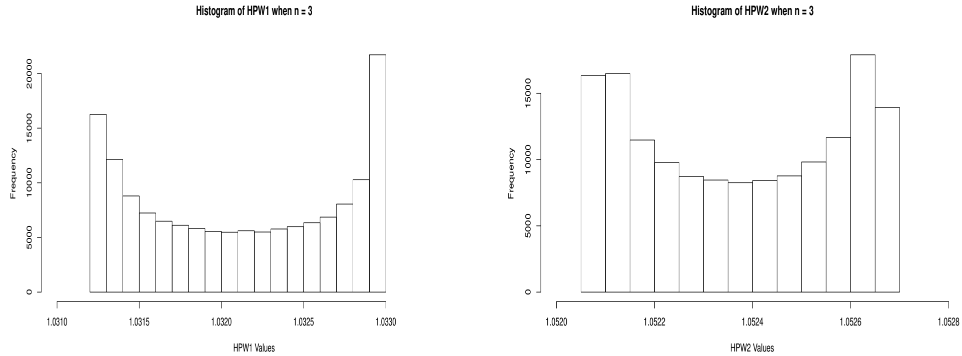
Histograms for various sampling distributions illustrated three types of shapes. The first shape was approximately symmetric and bimodal, which occurred when $n = 3, 5$, in figure 1 and 3. The second was a positively skewed shape, which occurred when $n = 4$, in figures 2. The third was a negatively skewed shape, occurred when $n > 5$, in figures 4. Thus we only list four figures, as for increasing *n*, histograms got narrower around 1 and the mode moved towards 1 while maintaining a similar shape in 4.

According to the form of both test statistics in equations 6 and 7, the value of 'HPW' is a summation of three components. Hence, we can expect extreme values of these test statistics on two occasions as given below.

A negative $\hat{\lambda}_4$ adds a positive amount to the sum. Hence a positive extreme value of 'HPW' can be expected if $\hat{\lambda}_3^2$ is high and $\hat{\lambda}_4$ is negatively low. Thus, a skewed and light-tailed distribution can give a positive extreme value for ‘HPWs’.

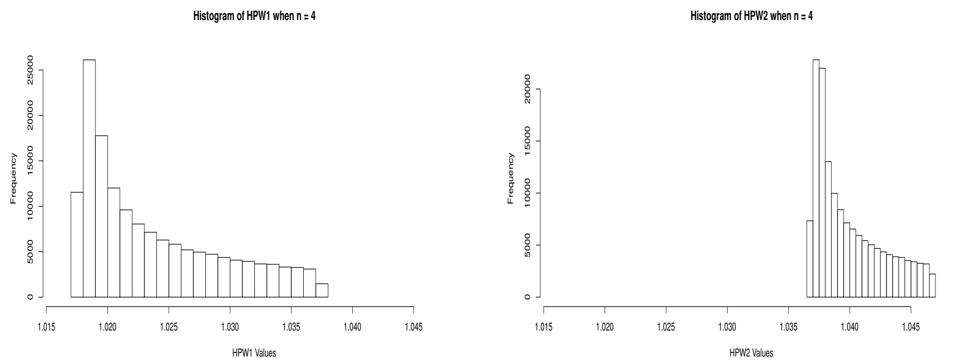
A positive $\hat{\lambda}_4$ adds a negative amount to the sum. Hence a negative extreme value of ‘HPW’ can be expected, if $\hat{\lambda}_3^2$ is close to 0 and $\hat{\lambda}_4$ is positively high. Thus, a symmetric or a least skewed and heavy-tailed distribution can give a negative extreme value for ‘HPWs’.

In addition, we can observe that when n increases, the effects of the second and third components in these equations get closer to 0. Thus we can expect that these tests can recognize non-normality for small samples from a skewed and light-tailed distribution or for small samples from a symmetric/least-skewed and heavy-tailed distribution.



(a) Histogram of ‘HPW1’. (b) Histogram of ‘HPW2’.

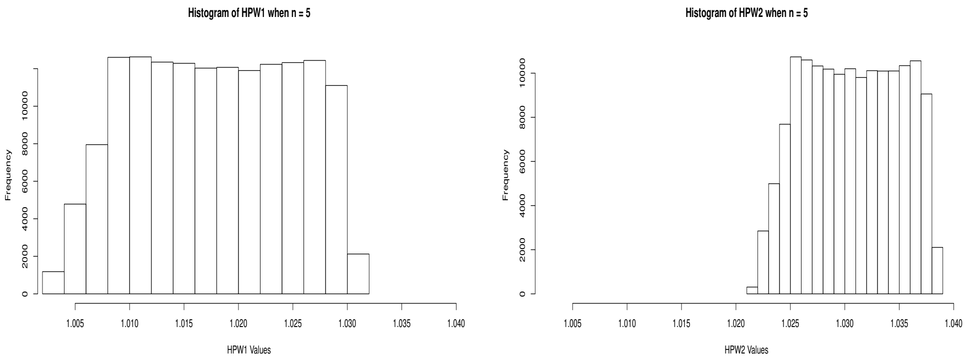
Fig. 1: Histograms of the sampling distributions of two test statistics (n=3).



(a) Histogram of ‘HPW1’. (b) Histogram of ‘HPW2’.

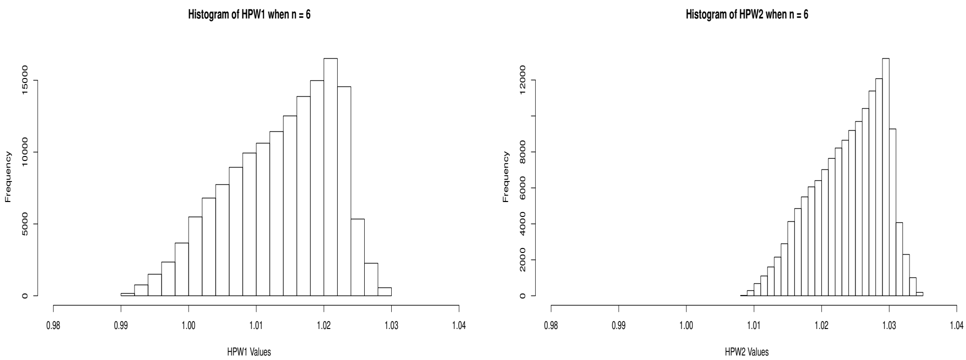
Fig. 2: Histograms of the sampling distributions of two test statistics (n=4).

The ranges of test statistics for each sample size are less than 0.065. Due to this, the critical values should be calculated at least up to five decimal places



(a) Histogram of 'HPW1'. (b) Histogram of 'HPW2'.

Fig. 3: Histograms of the sampling distributions of two test statistics (n=5).



(a) Histogram of 'HPW1'. (b) Histogram of 'HPW2'.

Fig. 4: Histograms of the sampling distributions of two test statistics (n=6).

in order to provide a good decision by minimizing the effect of truncation and round-off errors of the critical values. The minimum of the sampling distribution of 'HPW1' is greater than 1 for $n = 3, 4, 5$. The respective values of 'HPW2' are greater than 1 for $n = 3, 4, 5, 6$. Consequently, on each of these occasions, the values of test statistics are not equal to 1.

Based on the characteristics of the empirical distributions and the critical value tables of HPWs, we can suggest that HPW1 can be used at each of the four levels of significance for samples larger than 5. In comparison, HPW2 can be used at each of four levels of significance for samples greater than 9 as all critical value intervals include 1 for these sample sizes. According to ranges of empirical distributions of 'HPWs', we cannot expect critical values to be monotonically increasing or decreasing. Based on histograms, 'HPW2' has a lower skewness than 'HPW1'.

3.2 Power comparison with seven other normality tests

Considering the categories of sample size and alternative distributions in Table 1, we can draw several insights from the calculated power values.

As reported in Tables 2 and 3, it is difficult to declare one of the two HPW tests is uniformly superior, as their power values are very similar across the scenarios considered. However, both HPW tests outperform the KS test under every condition examined. According to the summary of power values in Table 3, the proposed tests exhibit competitive power relative to the JB test for all sample sizes under symmetric distributions and for small samples from asymmetric distributions.

Table 3 further indicates that HPW tests are more powerful than the LF test for moderate and large samples from symmetric distributions. They perform comparably to the JB test for small and moderate sample sizes and remain on a par with the JB test for large samples under symmetric alternatives. The HPW tests are also broadly comparable to the AD, SW, SF, LF, and CM tests for moderate and large samples from symmetric distributions. Nevertheless, the Shapiro–Wilk and Shapiro–Francia tests tend to show higher power than all other procedures in almost all situations considered in this study.

Table 2: Power of Normality Tests for Different Distributions and Sample Sizes

Distribution	**n**	**SW**	**KS**	**LF**	**AD**	**CM**	**SF**	**JB**	**HPW1**	**HPW2**
F(5,10)	12	0.5107	0.0172	0.3695	0.4819	0.4544	0.5087	0.2644	0.2834	0.2836
	35	0.9673	0.2363	0.8268	0.9414	0.9166	0.9559	0.8468	0.6378	0.6364
	100	1.0000	0.8844	0.9992	1.0000	0.9999	1.0000	0.9999	0.9550	0.9551
Chi-square(3)	12	0.3840	0.0033	0.2553	0.3543	0.3252	0.3780	0.1536	0.1901	0.1887
	35	0.9229	0.0664	0.6609	0.8604	0.8063	0.8937	0.6660	0.4014	0.3969
	100	1.0000	0.5620	0.9921	0.9998	0.9992	0.9999	0.9987	0.7544	0.7518
Exp(1.5)	12	0.5389	0.0084	0.3583	0.5019	0.4660	0.5226	0.2195	0.2428	0.2425
	35	0.9879	0.1724	0.8486	0.9672	0.9410	0.9772	0.8129	0.5190	0.5177
	100	1.0000	0.9206	0.9999	1.0000	0.9999	1.0000	0.9999	0.8828	0.8833
Log-normal(0,1)	12	0.7092	0.0523	0.5448	0.6820	0.6520	0.6976	0.3929	0.3852	0.3858
	35	0.9976	0.5465	0.9626	0.9941	0.9889	0.9959	0.9545	0.7995	0.7987
	100	1.0000	0.9969	1.0000	1.0000	1.0000	1.0000	1.0000	0.9939	0.9940
Beta(2,5)	12	0.0991	0.0001	0.0808	0.0961	0.0911	0.0962	0.0227	0.0751	0.0731
	35	0.3275	0.0018	0.1806	0.2644	0.2302	0.2604	0.0991	0.0994	0.0930
	100	0.8963	0.0185	0.5063	0.7563	0.6527	0.8234	0.5051	0.1060	0.1052
Weibull(1.5,1)	12	0.2137	0.0007	0.1445	0.1959	0.1774	0.2129	0.0758	0.1238	0.1225
	35	0.6987	0.0118	0.3800	0.5818	0.5096	0.6394	0.3862	0.2194	0.2153
	100	0.9981	0.1377	0.8528	0.9824	0.9527	0.9950	0.9520	0.3955	0.3906
Gamma(2,3)	12	0.2953	0.0017	0.2003	0.2731	0.2505	0.2972	0.1213	0.1614	0.1579
	35	0.8269	0.0340	0.5292	0.7414	0.6757	0.7878	0.5606	0.3304	0.3278
	100	0.9997	0.3366	0.9544	0.9977	0.9916	0.9995	0.9913	0.6419	0.6402
Cauchy(0,1)	12	0.6683	0.2498	0.6518	0.6923	0.6909	0.7120	0.5399	0.6061	0.6071
	35	0.9776	0.8049	0.9662	0.9812	0.9801	0.9840	0.9673	0.9681	0.9679
	100	1.0000	0.9986	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000	1.0000
t(30)	12	0.0537	0.0001	0.0536	0.0558	0.0544	0.0623	0.0184	0.0559	0.0541
	35	0.0672	0.0001	0.0545	0.0607	0.0594	0.0776	0.0575	0.0687	0.0638
	100	0.0809	0.0002	0.0590	0.0645	0.0619	0.1004	0.0966	0.0838	0.0817
Unif(0,1)	12	0.0933	0.0001	0.0681	0.0917	0.0829	0.0520	0.0019	0.1599	0.1562
	35	0.4851	0.0005	0.1707	0.3678	0.2855	0.2382	0.0002	0.6770	0.6741
	100	0.9959	0.0092	0.5919	0.9481	0.8411	0.9676	0.5616	0.9984	0.9985

Table 3: Results of Power Comparison for both Symmetric and Asymmetric Distributions

Sample size	Results
Symmetric Distributions	
Sample size	**Results**
Small	'HPWs' have power higher than KS and JB tests for all 6 distributions. 'HPWs' have power higher than AD, SW, SF, LF and CM tests for t(30), Unif(0,1) and Unif(0,5) distributions.
Medium	'HPWs' have power higher than KS and JB tests for all 6 distributions. 'HPWs' have power higher than LF test for t(30), Unif(0,1) and Unif(0,5) distributions. And those two have power higher than LF test for t(1), Cauchy(0,1) and Cauchy(0,3) when sample size is greater than 35. 'HPWs' have power higher than AD, SW, SF and CM tests for t(30), Unif(0,1) and Unif(0,5) distributions.
Large	'HPWs' have power higher than KS and LF tests for all 6 distributions. 'HPWs' have power closer to JB test for t(1) and Cauchy(0,3), while power is higher than JB test for Cauchy(0,1), Unif(0,1) and Unif(0,5). 'HPWs' have power higher than AD and CM tests for t(30), Cauchy(0,1), Unif(0,1) and Unif(0,5) distributions. 'HPWs' have power higher than SW and SF tests for t(30), Unif(0,1) and Unif(0,5) distributions.
Asymmetric Distributions	
Sample size	**Results**
Any sample	'HPWs' have power values higher than KS test for all 14 distributions.
Small	'HPWs' have power values closer to the power values of JB test for all 14 distributions while showing higher power values than JB test when sample size is less than 14.

4 Conclusion

The simulation study indicates that the proposed HPW tests perform particularly well when the data are generated from symmetric distributions, in line with the expectations set out in Section 3.1. Under the sample sizes and alternative distributions considered in Table 1, the HPW tests generally achieve power values comparable to those of the Kolmogorov–Smirnov, Jarque–Bera, and Lilliefors tests. For skewed alternatives, the HPW tests outperform the Kolmogorov–Smirnov test across all sample sizes and the Jarque–Bera test for small samples, and they exhibit power close to that of the competing procedures for sample sizes between 8 and 20.

At the same time, HPW tests do not uniformly dominate the classical procedures. In particular, they do not consistently outperform the Shapiro–Wilk or Shapiro–Francia tests, which often retain the highest power among the methods considered. Thus, our tests should be viewed as providing a complementary Edgeworth-based perspective on normality rather than a universal replacement for existing procedures.

In practice, the HPW tests appear most useful when the sample exhibits an approximately symmetric shape or when the practitioner is especially concerned with symmetric, heavy-tailed alternatives and small sample sizes. In such settings, HPW tests offer competitive or superior power while maintaining reasonable control of the Type I error rate. For highly skewed distributions, their main advantage is observed for small samples, as anticipated in Section 3.1. Further work on the behaviour of the sampling distributions for very small samples (e.g., $n < 6$) would be valuable.

The necessary calculations are demonstrated below to perform new tests for a given sample. Let x = {82, 84, 85, 88, 92, 93, 94, 95, 98, 100, 102, 107, 110, 116, 116} be a data set of 15 patients whose mean arterial pressure was measured by Mishra et al. (2019). Since the sample mean is 97.46 we can calculate HPW1 as follows.

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{15}(x_i - 97.46)^2}{15} = 113.049, \;\; \hat{\mu}_3 = \frac{\sum_{i=1}^{15}(x_i - 97.46)^3}{15} = 429.297,$$

$$\hat{\mu}_4 = \frac{\sum_{i=1}^{15}(x_i - 97.46)^4}{15} = 26208.95, \;\; \hat{\lambda}_3 = 0.35715, \;\; \hat{\lambda}_4 = -0.94923$$

$$HPW1 = \left\{1 - \frac{-0.94923}{8*15}\Big(\frac{14}{16}\Big) + \frac{0.35715^2}{4*15}\Big(\frac{13}{(18)(16)}\Big)\right\} = 1.007017$$

At 5% level of significance and $n = 15$, the critical values are given in the Appendix Table 4.

Decision rule: We reject H_0, if HPW1 is less than 0.988826 or greater than 1.010006 at 5% level of significance.

Decision: Since HPW1 (= 1.007017) is in between than 0.988826 and 1.010006, we cannot reject the null hypothesis.

Similarly, we can perform the proposed test for this sample data set using HPW2 and its related calculations. For more clarification, we performed the JB, SW, SF, LF, and AD tests with this data set. The P-values of these tests are 0.6434, 0.4541, 0.64, 0.7928 and 0.644. That is clear to see that all these tests give the same decision as the proposed test.

Furthermore, let $x = \{8, 9, 8, 15, 9, 10, 14, 16, 18, 19, 21, 18, 19, 10, 15, 21, 16, 19, 9, 18\}$ be a random sample of size 20. Since the sample mean is 14.6 we can calculate HPW1 as above and test statistics value becomes 1.008253. When $n = 20$, we reject H_0, if HPW1 is less than 0.991129 or greater than

1.007138 at 5% level of significance. Since HPW1 (= 1.008253) is greater than 1.007138, we reject the null hypothesis of normality.

We perform JB, SW, SF, LF, and AD tests with this data set. The P-values of these two tests are 0.382, 0.022, 0.053, 0.054, and 0.02 respectively. Thus, it is clear that the AD and SW tests suggest rejecting the null hypothesis of normality, while the JB, SF and LF tests suggest vice-versa at 5% level of significance. In conclusion, these two examples show how well the proposed test can test the normality assumption for small samples.

To facilitate practical use, we will work on an R package implementing the HPW test statistics, together with critical values for a range of sample sizes in the appendix. This implementation allows practitioners to apply the proposed tests directly in empirical work and is intended to encourage further exploration and comparison of HPW-based procedures in applied settings.

Acknowledgment

The second author wishes to express his gratitude to Dr. K.P. Hapuarachchi and Dr. D.C. Wickramarachchi, senior lecturers in the Department of Statistics, University of Sri Jayewardenepura for suggesting this topic, their guidance and support. Authors wish to thank University for providing facilities to conduct this research successfully.

References

Bayoud, H. A. (2016) Testing the similarity of two normal populations with application to the bioequivalence problem, *Journal of Applied Statistics*, 43(7):1322–1334.

Bayoud, H. A. (2021) Tests of normality: new test and comparative study, *Communications in Statistics-Simulation and Computation*, 50(12):4442–4463.

Birnbaum, Z. W. (1952) Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size, *Journal of the American Statistical Association*, 47(259):425–441.

D'Agostino, R. B. (1971) An omnibus test of normality for moderate and large size samples, *Biometrika*, 58(2):341–348.

D'Agostino, R. B. (1973) Monte Carlo power comparison of the W' and D tests of normality for n = 100, *Communications in Statistics*, 1(6):545–551.

Darling, D. A. (1957) The Kolmogorov–Smirnov, Cramér–von Mises tests, *The Annals of Mathematical Statistics*, 28(4):823–838.

David, H. A., Hartley, H. O., and Pearson, E. S. (1954) The distribution of the ratio, in a single normal sample, of range to standard deviation, *Biometrika*, 41(3/4):482–493.

Dufour, J.-M., Farhat, A., Gardiol, L., and Khalaf, L. (1998) Simulation-based finite sample normality tests in linear regressions, *The Econometrics Journal*, 1(1):154–173.

Filliben, J. J. (1975) The probability plot correlation coefficient test for normality, *Technometrics*, 17(1):111–117.

Fisher, R. A. (1930a) The moments of the distribution for normal samples of measures of departure from normality, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 130(812):16–28.

Fisher, R. A. (1930b) Moments and product moments of sampling distributions, *Proceedings of the London Mathematical Society*, 2(1):199–238.

Friedman, Y., Gelman, A., and Phadia, E. (1988) A modified Kolmogorov–Smirnov type goodness-of-fit test, *Communications in Statistics – Theory and Methods*, 17(12):4147–4162.

Gayen, A. K. (1949) The distribution of Student's t in random samples of any size drawn from non-normal universes, *Biometrika*, 36(3/4):353–369.

Geary, R. C. (1936) Moments of the ratio of the mean deviation to the standard deviation for normal samples, *Biometrika*, 28(3/4):295–307.

Geary, R. C. (1935) The ratio of the mean deviation to the standard deviation as a test of normality, *Biometrika*, 27(3/4):310–332.

Geary, R. C. (1947) Testing for normality, *Biometrika*, 34(3/4):209–242.

Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972) Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance, *Review of Educational Research*, 42(3):237–288.

Guo, J., Alemayehu, D., and Shao, Y. (2010) Tests for normality based on entropy divergences, *Statistics in Biopharmaceutical Research*, 2(3):408–418.

Hanusz, Z. and Tarasinska, J. (2014) On multivariate normality tests using skewness and kurtosis, *Colloquium Biometricum*, 44.

Jarque, C. M. and Bera, A. K. (1987) A test for normality of observations and regression residuals, *International Statistical Review / Revue Internationale de Statistique*, 55:163–172.

Kahya, G. (1991) *New modified Anderson–Darling and Cramér–von Mises goodness-of-fit tests for a normal distribution with specified parameters*, Air Force Institute of Technology, Wright–Patterson AFB, OH, School of Engineering. URL: https://apps.dtic.mil/sti/html/tr/ADA238630/ (5/11/2025)

Kolmogorov, A. (1933) Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari*, 4:461.

Lilliefors, H. W. (1967) On the Kolmogorov–Smirnov test for normality with mean and variance unknown, *Journal of the American Statistical Association*, 62(318):399–402.

Mendes, M. and Pala, A. (2003) Type I error rate and power of three normality tests, *Pakistan Journal of Information and Technology*, 2(2):135–139.

Miller, L. H. (1956) Table of percentage points of Kolmogorov statistics, *Journal of the American Statistical Association*, 51(273):111–121.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., and Keshri, A. (2019) Descriptive statistics and normality tests for statistical data, *Annals of Cardiac Anaesthesia*, 22(1):67–72.

Öztuna, D., Elhan, A. H., and Tüccar, E. (2006) Investigation of four different normality tests in terms of type I error rate and power under different distributions, *Turkish Journal of Medical Sciences*, 36(3):171–176.

Pearson, E. S. (1931) I. Note on tests for normality, *Biometrika*, 22(3–4):423–424.

Pearson, E. S. (1930) A further development of tests for normality, *Biometrika*, 22(1/2):239–249.

Pearson, E. S. (1963) Some problems arising in approximating to probability distributions, using moments, *Biometrika*, 50(1/2):95–112.

Pearson, E. S. (1965) Tables of percentage points of $\sqrt{b_1}$ and b_2 in normal samples; a rounding off, *Biometrika*, 52(1/2):282–285.

Pearson, E. S. and Stephens, M. A. (1964) The ratio of range to standard deviation in the same normal sample, *Biometrika*, 51(3/4):484–487.

Razali, N. M. and Wah, Y. B. (2011) Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests, *Journal of Statistical Modeling and Analytics*, 2(1):21–33.

Romao, X., Delgado, R., and Costa, A. (2009) An empirical power comparison of univariate goodness-of-fit tests for normality, *Journal of Statistical Computation and Simulation*, 80(5), 545–591.

Seier, E. (2002) Comparison of tests for univariate normality, *Interstat*, 1:1–17.

Shapiro, S. S. and Wilk, M. B. (1965) An analysis of variance test for normality (complete samples), *Biometrika*, 52(3/4):591–611.

Shapiro, S. S. and Francia, R. S. (1972) An approximate analysis of variance test for normality, *Journal of the American Statistical Association*, 67(337):215–216.

Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968) A comparative study of various tests for normality, *Journal of the American Statistical Association*, 63(324):1343–1372.

Sinclair, C. D., Spurr, B. D., and Ahmad, M. I. (1990) Modified Anderson–Darling test, *Communications in Statistics – Theory and Methods*, 19(10):3677–3686.

Smirnov, N. V. (1939) Sur les escarres de la contre de distribution empirique, *Recueil Mathématique*, 6(337):3–26.

Srinivasan, R. (1970) An approach to testing the goodness of fit of incompletely specified distributions, *Biometrika*, 57(3):605–611.

Székely, G. J. and Rizzo, M. L. (2005) A new test for multivariate normality, *Journal of Multivariate Analysis*, 93(1):58–80.

Torabi, H., Montazeri, N. H., and Grané, A. (2016) A test for normality based on the empirical distribution function, *SORT: Statistics and Operations Research Transactions*, 40(1):55–88.

Verrill, S. and Johnson, R. A. (1988) Tables and large-sample distribution theory for censored-data correlation statistics for testing normality, *Journal of the American Statistical Association*, 83(404):1192–1197.

Williams, P. (1935) Note on the sampling distribution of $\sqrt{\beta_1}$, where the population is normal, *Biometrika*, 27(1/2):269–271.

Appendix

A Derivation of the test statistic

Let $\phi(x)$ be the probability density function of standard normal distribution. And $\phi(x)^{(k)}$ be the k^{th} derivative of $\phi(x)$ with respect to x. Hence probability density test can be written as $\phi(x) = \frac{\exp^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$. Also we take $S_1 = n\bar{X}$ and $S_2 = (n-1)s^2$ as in Gayen (1949).

By the third order approximation of Edgeworth expansion where $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
f\Big(\frac{x-\mu}{\sigma}|\mu,\sigma,\lambda_3,\lambda_4\Big) &= \phi\Bigg(\frac{x-\mu}{\sigma}\Bigg) - \frac{\lambda_3}{3!}\phi^{(3)}\Bigg(\frac{x-\mu}{\sigma}\Bigg) + \frac{\lambda_4}{4!}\phi^{(4)}\Bigg(\frac{x-\mu}{\sigma}\Bigg) \\
&+ \frac{10\lambda_3^2}{6!}\phi^{(6)}\Bigg(\frac{x-\mu}{\sigma}\Bigg) \\
&= \phi\Bigg(\frac{x-\mu}{\sigma}\Bigg)\Bigg[1 + \frac{\lambda_3}{3!}\big[(\frac{x-\mu}{\sigma})^3 - 3(\frac{x-\mu}{\sigma})\big] \\
&+ \frac{\lambda_4}{4!}\big[(\frac{x-\mu}{\sigma})^4 - 6(\frac{x-\mu}{\sigma})^2 + 3\big] \\
&+ \frac{10\lambda_3^2}{6!}\big[(\frac{x-\mu}{\sigma})^6 + 15(\frac{x-\mu}{\sigma})^4 + 45(\frac{x-\mu}{\sigma})^2 - 15\big]\Bigg]
\end{aligned}
$$

Now consider the following mathematical relationships

$$
\begin{aligned}
E\Bigg(\frac{x-\mu}{\sigma}\Bigg) &= 0 \\
E\Bigg(\frac{x-\mu}{\sigma}\Bigg)^2 &= 1 + \frac{\lambda_3}{3!}\big[0 - 3(0)\big] + \frac{\lambda_4}{4!}\big[15 - 6(3) + 3\big] \\
&+ \frac{10\lambda_3^2}{6!}\big[105 + 15(15) + 45(3) - 15\big] = 1
\end{aligned}
$$

That is $Var(x-\mu)^2 = \sigma^2$.

$$
\begin{aligned}
E\Bigg(\frac{x-\mu}{\sigma}\Bigg)^3 &= 0 + \frac{\lambda_3}{3!}\big[15 - 3(3)\big] + 0 = \lambda_3 \\
E\Bigg(\frac{x-\mu}{\sigma}\Bigg)^4 &= 3 + \frac{\lambda_3}{3!}\big[0 - 3(0)\big] + \frac{\lambda_4}{4!}\big[105 - 6(15) + 3(3)\big] \\
&+ \frac{10\lambda_3^2}{6!}\big[945 + 15(105) + 45(15) - 15\big] = 3 + \lambda_4
\end{aligned}
$$

$$E(S_1) = \int_{-\infty}^{\infty} \frac{\exp^{\frac{-s_1^2}{2n}}}{\sqrt{2n\pi}} \left[s_1 + \frac{\lambda_3}{6n^2}\left(s_1^4 - 3ns_1^2\right) + \frac{\lambda_4}{24n^3}\left(s_1^5 - 6ns_1^3 + 3n^2 s_1\right)\right.$$
$$\left. + \frac{10\lambda_3^2}{72n^4}\left(s_1^7 + 15ns_1^5 + 45n^2 s_1^3 - 15n^3 s_1\right)\right] ds_1$$
$$= 0 + \frac{\lambda_3}{6n^2}\left(n^2 - 3n(n^2)\right) + \frac{\lambda_4}{24n^3}\left(0 - 6n(0) + 3n^2(0)\right)$$
$$+ \frac{10\lambda_3^2}{72n^4}\left(0 + 15n(0) + 45n^2(0) - 15n^3(0)\right) = 0$$

$$V(S_1) = E(S_1^2) = \int_{-\infty}^{\infty} \frac{\exp^{\frac{-s_1^2}{2n}}}{\sqrt{2n\pi}} \left[s_1^2 + \frac{\lambda_3}{6n^2}\left(s_1^5 - 3ns_1^3\right) + \frac{\lambda_4}{24n^3}\left(s_1^6 - 6ns_1^4\right.\right.$$
$$\left.\left. + 3n^2 s_1^2\right) + \frac{10\lambda_3^2}{72n^4}\left(s_1^8 + 15ns_1^6 + 45n^2 s_1^4 - 15n^3 s_1^2\right)\right] ds_1$$
$$= n + \frac{\lambda_3}{6n^2}\left(0 - 3n(0)\right) + \frac{\lambda_4}{24n^3}\left(15n^3 - 6n(3n^2) + 3n^2(n)\right)$$
$$+ \frac{10\lambda_3^2}{72n^4}\left(105n^4 + 15n(15n^3) + 45n^2(3n^2) - 15n^3(n)\right) = n$$
$$\therefore \bar{X} \sim (0, \frac{1}{n})$$

$$E(S_1^3) = \int_{-\infty}^{\infty} \frac{\exp^{\frac{-s_1^2}{2n}}}{\sqrt{2n\pi}} \left[s_1^3 + \frac{\lambda_3}{6n^2}\left(s_1^6 - 3ns_1^4\right) + \frac{\lambda_4}{24n^3}\left(s_1^7 - 6ns_1^5 + 3n^2 s_1^3\right)\right.$$
$$\left. + \frac{10\lambda_3^2}{72n^4}\left(s_1^9 + 15ns_1^7 + 45n^2 s_1^5 - 15n^3 s_1^3\right)\right] ds_1$$
$$\therefore E(S_1^3) = \frac{\lambda_3}{6n^2}\left(15n^3 - 3n(3n^2)\right) = n\lambda_3$$
$$\frac{E(S_1^3)}{E(S_1^2)^{\frac{3}{2}}} = \frac{n\lambda_3}{n^{\frac{3}{2}}} = \frac{\lambda_3}{\sqrt{n}}$$
$$E(S_1^4) = 3n^2 + \frac{\lambda_4}{24n^3}\left(105n^4 - 6n(15n^3) + 3n^2(3n^2)\right)$$
$$+ \frac{10\lambda_3^2}{72n^4}\left(945n^5 + 15n(105n^4) + 45n^2(15n^2) - 15n^3(3n^2)\right)$$
$$\therefore E(S_1^4) = 3n^2 + n\lambda_4$$

If $T = \sqrt{n}\bar{X}$ *then* $E(T) = 0$ *and* $V(T) = 1$, then $T \sim N(0, 1)$

$$E(T^3) = \frac{\lambda_3}{\sqrt{n}} \quad \text{and} \quad E(T^4) = 3 + \frac{\lambda_4}{n}$$

$$S_2 = (n-1)\frac{s^2}{\sigma^2} \to E\big(s^2\big) = \sigma^2\big[\frac{S_2}{(n-1)}\big]$$

$$J_k = \int S_2^k \frac{S_2^{\frac{n-3}{2}}}{2^{\frac{n-1}{2}}} \frac{\exp^{\frac{-S_2}{2}}}{\Gamma\big(\frac{n-2}{2}\big)} = \frac{2^{\frac{(n+2k-1)}{2}}\Gamma\big(\frac{n+2k-1}{2}\big)}{2^{\frac{(n-1)}{2}}\Gamma\big(\frac{n-1}{2}\big)} = \frac{2^k\Gamma\big(\frac{n+2k-1}{2}\big)}{\Gamma\big(\frac{n-1}{2}\big)}$$

Using the above equation we can derive the following quantities,

$$J_{1/2} = \big[\sqrt{2}\Gamma\big(\frac{n}{2}\big)\big]/\Gamma\big(\frac{n-1}{2}\big) = \alpha \qquad J_1 = \big[2\Gamma\big(\frac{n+1}{2}\big)\big]/\Gamma\big(\frac{n-1}{2}\big) = n-1$$

$$J_{3/2} = \big[2\sqrt{2}\Gamma\big(\frac{n+2}{2}\big)\big]/\Gamma\big(\frac{n-1}{2}\big) = n\alpha \qquad J_2 = (n-1)(n+1)$$

$$J_{5/2} = n(n+1)\alpha \qquad J_3 = (n-1)(n+1)(n+3)$$

$$J_{7/2} = n(n+2)(n+4)\alpha \qquad ect.$$

$$\begin{aligned}
E\big(S_2\big) &= J_1 + \frac{\lambda_4(n-1)}{8n(n+1)}\big(J_3 - 2(n+1)J_2 + (n^2-1)J_1\big) \\
&+ \frac{\lambda_3^2(n-2)}{12n(n+3)(n+1)}\big[J_4 - 3(n+3)J_3 + 3(n+3)(n+1)J_2 \\
&- (n+3)(n+1)(n-1)J_1\big] \\
&= (n-1) \\
E\big(\sqrt{S_2}\big) &= J_{1/2} + \frac{\lambda_4(n-1)}{8n(n+1)}\big(J_{5/2} - 2(n+1)J_{3/2} + (n^2-1)J_{1/2}\big) \\
&+ \frac{\lambda_3^2(n-2)}{12n(n+3)(n+1)}\big[J_{7/2} - 3(n+3)J_{5/2} + 3(n+3)(n+1)J_{3/2} \\
&- (n+3)(n+1)(n-1)J_{1/2}\big] \\
&= \frac{\sqrt{2}\Gamma\big(n/2\big)}{\Gamma\big((n-1)/2\big)}\left[1 - \frac{\lambda_4(n-1)}{8n(n+1)} + \frac{\lambda_3^2(n-2)}{12n(n+3)(n+1)}\right] \\
&= \frac{\sqrt{2}\Gamma\big(n/2\big)}{\Gamma\big((n-1)/2\big)}\gamma = \alpha\gamma
\end{aligned}$$

Thus, we get the for of the test statistic from above.

B Critical value tables

These critical value tables are made considering 4 different levels of significance and 49 sample sizes. Under each level of significance, there are 2 critical values that are used to test the normality.

Table 4: Table of critical values of sampling distribution of HPW1

Sample size	α=0.01		α=0.025		α=0.05		α=0.1	
6	0.993641	1.027717	0.995405	1.026629	0.997245	1.025526	0.999583	1.024166
7	0.987008	1.023986	0.989830	1.023241	0.992501	1.022439	0.995578	1.021303
8	0.983220	1.021130	0.986623	1.020359	0.989781	1.019658	0.993398	1.018711
9	0.980783	1.018947	0.985030	1.018145	0.988465	1.017401	0.992458	1.016502
10	0.979525	1.017058	0.983983	1.016342	0.987850	1.015647	0.991906	1.014769
11	0.979305	1.015469	0.984038	1.014798	0.987818	1.014167	0.991787	1.013330
12	0.979053	1.014132	0.983939	1.013473	0.987748	1.012840	0.991778	1.012059
13	0.979687	1.012973	0.984376	1.012355	0.988183	1.011787	0.991980	1.011061
14	0.979822	1.011956	0.984788	1.011378	0.988383	1.010826	0.992065	1.010148
15	0.980759	1.011085	0.985288	1.010531	0.988826	1.010006	0.992381	1.009342
16	0.981710	1.010304	0.986157	1.009791	0.989394	1.009296	0.992773	1.008677
17	0.982142	1.009649	0.986495	1.009139	0.989710	1.008670	0.992958	1.008080
18	0.982932	1.009015	0.987066	1.008535	0.990147	1.008085	0.993220	1.007511
19	0.983588	1.008482	0.987738	1.008025	0.990715	1.007594	0.993560	1.007064
20	0.984570	1.007979	0.988383	1.007561	0.991129	1.007138	0.993850	1.006634
21	0.985349	1.007555	0.988796	1.007128	0.991452	1.006736	0.994015	1.006240
22	0.985882	1.007164	0.989248	1.006755	0.991801	1.006375	0.994265	1.005900
23	0.986397	1.006768	0.989650	1.006395	0.992123	1.006030	0.994551	1.005586
24	0.986966	1.006433	0.990158	1.006075	0.992496	1.005736	0.994755	1.005302
25	0.987471	1.006180	0.990519	1.005804	0.992754	1.005460	0.994873	1.005027
26	0.987714	1.005869	0.990760	1.005522	0.992962	1.005193	0.995033	1.004790
27	0.988328	1.005628	0.991265	1.005276	0.993319	1.004958	0.995255	1.004569
28	0.988737	1.005363	0.991546	1.005048	0.993546	1.004746	0.995457	1.004369
29	0.989411	1.005122	0.991931	1.004818	0.993804	1.004527	0.995587	1.004171
30-34	0.989569	1.004937	0.992246	1.004634	0.994013	1.004347	0.995759	1.003990

Table 5: Table of critical values of sampling distribution of HPW1

Sample size	α=0.01		α=0.025		α=0.05		α=0.1	
35-39	0.991494	1.004092	0.993570	1.003828	0.995018	1.003582	0.996420	1.003276
40-44	0.992834	1.003476	0.994534	1.003239	0.995765	1.003023	0.996975	1.002754
45-49	0.993670	1.002995	0.995281	1.002780	0.996330	1.002589	0.997331	1.002357
50-54	0.994586	1.002615	0.995845	1.002437	0.996775	1.002270	0.997657	1.002056
55-59	0.995150	1.002313	0.996335	1.002152	0.997126	1.001993	0.997900	1.001805
60-64	0.995767	1.002065	0.996769	1.001906	0.997464	1.001769	0.998141	1.001605
65-69	0.996187	1.001865	0.997102	1.001726	0.997738	1.001597	0.998330	1.001442
70-74	0.996514	1.001683	0.997317	1.001557	0.997913	1.001445	0.998472	1.001303
75-79	0.996912	1.001554	0.997614	1.001431	0.998130	1.001318	0.998622	1.001191
80-84	0.997189	1.001423	0.997823	1.001314	0.998287	1.001211	0.998725	1.001089
85-89	0.997363	1.001316	0.998002	1.001209	0.998420	1.001112	0.998826	1.000999
90-94	0.997662	1.001217	0.998159	1.001120	0.998536	1.001030	0.998911	1.000924
95-99	0.997831	1.001132	0.998305	1.001041	0.998648	1.000958	0.998990	1.000859
100-109	0.997949	1.001059	0.998419	1.000972	0.998748	1.000897	0.999058	1.000802
110-119	0.998239	1.000937	0.998634	1.000857	0.998914	1.000787	0.999176	1.000702
120-129	0.998481	1.000830	0.998802	1.000758	0.999037	1.000696	0.999273	1.000620
130-139	0.998623	1.000743	0.998931	1.000682	0.999143	1.000626	0.999348	1.000558
140-49	0.998772	1.000675	0.999039	1.000615	0.999222	1.000564	0.999405	1.000503
150-159	0.998892	1.000615	0.999135	1.000562	0.999301	1.000513	0.999463	1.000455
160-169	0.998996	1.000561	0.999208	1.000513	0.999359	1.000469	0.999513	1.000417
170-179	0.999090	1.000517	0.999286	1.000470	0.999422	1.000430	0.999555	1.000381
180-189	0.999152	1.000479	0.999332	1.000436	0.999456	1.000398	0.999584	1.000352
190-199	0.999220	1.000445	0.999385	1.000405	0.999505	1.000369	0.999622	1.000326
200-	0.999298	1.000414	0.999434	1.000378	0.999543	1.000343	0.999645	1.000303

Table 6: Table of critical values of sampling distribution of HPW2

Sample size	α=0.01		α=0.025		α=0.05		α=0.1	
6	1.010733	1.033352	1.011835	1.032651	1.013040	1.031915	1.014641	1.031009
7	1.002282	1.028639	1.004193	1.028116	1.006043	1.027520	1.008252	1.026704
8	0.996547	1.024948	0.999190	1.024395	1.001566	1.023878	1.004286	1.023175
9	0.992313	1.022166	0.995549	1.021537	0.998464	1.020934	1.001451	1.020228
10	0.989986	1.019772	0.993527	1.019218	0.996490	1.018680	0.999697	1.017945
11	0.988318	1.017805	0.992159	1.017244	0.995161	1.016723	0.998448	1.016041
12	0.987030	1.016159	0.991018	1.015631	0.994139	1.015119	0.997530	1.014479
13	0.986527	1.014813	0.990571	1.014279	0.993706	1.013786	0.996964	1.013144
14	0.986318	1.013592	0.990269	1.013093	0.993408	1.012602	0.996586	1.012002
15	0.986210	1.012509	0.990356	1.012044	0.993315	1.011605	0.996376	1.011028
16	0.986486	1.011623	0.990223	1.011155	0.993235	1.010726	0.996222	1.010194
17	0.986624	1.010805	0.990514	1.010366	0.993350	1.009931	0.996061	1.009401
18	0.986916	1.010076	0.990648	1.009659	0.993490	1.009264	0.996131	1.008763
19	0.987051	1.009470	0.990710	1.009051	0.993452	1.008657	0.996046	1.008184
20	0.987616	1.008893	0.991034	1.008492	0.993532	1.008130	0.996069	1.007657
21	0.987842	1.008393	0.991290	1.008016	0.993749	1.007651	0.996094	1.007192
22	0.988381	1.007943	0.991644	1.007558	0.993886	1.007213	0.996193	1.006774
23	0.988640	1.007491	0.991797	1.007130	0.994021	1.006807	0.996250	1.006392
24	0.989409	1.007132	0.992231	1.006785	0.994291	1.006463	0.996346	1.006057
25	0.989596	1.006776	0.992507	1.006431	0.994567	1.006122	0.996531	1.005738
26	0.989963	1.006448	0.992624	1.006126	0.994524	1.005823	0.996472	1.005448
27	0.990284	1.006161	0.992891	1.005842	0.994759	1.005547	0.996612	1.005188
28	0.990444	1.005867	0.993058	1.005571	0.994936	1.005289	0.996670	1.004933
29	0.990861	1.005623	0.993347	1.005329	0.995032	1.005050	0.996728	1.004711
30-34	0.991132	1.005390	0.993560	1.005102	0.995285	1.004839	0.996845	1.004511

Table 7: Table of critical values of sampling distribution of HPW2

Sample size	α=0.01		α=0.025		α=0.05		α=0.1	
35-39	0.992519	1.004414	0.994480	1.004174	0.995866	1.003942	0.997201	1.003658
40-44	0.993602	1.003737	0.995207	1.003521	0.996381	1.003313	0.997525	1.003058
45-49	0.994411	1.003205	0.995878	1.003013	0.996875	1.002834	0.997824	1.002610
50-54	0.995084	1.002804	0.996305	1.002618	0.997187	1.002454	0.998051	1.002259
55-59	0.995630	1.002467	0.996716	1.002308	0.997506	1.002164	0.998229	1.001981
60-64	0.996093	1.002195	0.997059	1.002048	0.997739	1.001917	0.998403	1.001752
65-69	0.996503	1.001990	0.997326	1.001848	0.997949	1.001724	0.998549	1.001573
70-74	0.996872	1.001791	0.997600	1.001665	0.998146	1.001557	0.998671	1.001419
75-79	0.997107	1.001636	0.997803	1.001520	0.998314	1.001415	0.998797	1.001286
80-84	0.997367	1.001505	0.997996	1.001392	0.998442	1.001292	0.998868	1.001173
85-89	0.997625	1.001382	0.998170	1.001279	0.998578	1.001190	0.998965	1.001079
90-94	0.997773	1.001280	0.998285	1.001184	0.998660	1.001098	0.999026	1.000994
95-99	0.997944	1.001191	0.998425	1.001099	0.998761	1.001020	0.999091	1.000922
100-109	0.998115	1.001115	0.998528	1.001029	0.998854	1.000952	0.999155	1.000859
110-119	0.998339	1.000976	0.998719	1.000899	0.998980	1.000832	0.999244	1.000748
120-129	0.998549	1.000869	0.998875	1.000799	0.999112	1.000738	0.999336	1.000664
130-139	0.998675	1.000778	0.998992	1.000716	0.999195	1.000658	0.999401	1.000590
140-149	0.998819	1.000703	0.999079	1.000646	0.999270	1.000594	0.999450	1.000532
150-159	0.998930	1.000641	0.999174	1.000584	0.999343	1.000539	0.999507	1.000482
160-169	0.999041	1.000585	0.999249	1.000536	0.999396	1.000492	0.999545	1.000441
170-179	0.999120	1.000541	0.999309	1.000493	0.999449	1.000452	0.999585	1.000403
180-189	0.999194	1.000498	0.999367	1.000455	0.999493	1.000417	0.999616	1.000370
190-199	0.999276	1.000461	0.999422	1.000421	0.999531	1.000385	0.999643	1.000344
200-	0.999317	1.000430	0.999463	1.000393	0.999567	1.000360	0.999670	1.000320