DOI: 10.1515/bile-2015-0008

DE GRUYTER OPEN

Biometrical Letters
Vol. 52 (2015), No. 2, 85 - 93

Normalization of the Kolmogorov–Smirnov and Shapiro–Wilk tests of normality

Zofia Hanusz, Joanna Tarasińska

Department of Applied Mathematics and Computer Science, University of Life Sciences in Lublin, Głęboka 28, 20-612 Lublin, Poland, e-mail: zofia.hanusz@up.lublin.pl, joanna.tarasinska@up.lublin.pl

SUMMARY

Two very well-known tests for normality, the Kolmogorov–Smirnov and the Shapiro–Wilk tests, are considered. Both of them may be normalized using Johnson's (1949) S_B distribution. In this paper, functions for normalizing constants, dependent on the sample size, are given. These functions eliminate the need to use non-standard statistical tables with normalizing constants, and make it easy to obtain *p*-values for testing normality.

Key words: Shapiro–Wilk *W* test, Kolmogorov–Smirnov test, normality, Johnson's S_B transformation

1. Introduction

Let us consider the following problem. For a given random sample $\left(X_1, X_2, \ldots, X_n\right)$ we wish to test the null hypothesis of data normality:

H_0: The sample comes from a normal distribution.

A review of techniques for solving such problems can be found, for example, in Thode (2002).

In this paper we consider, after Stevens (1974), three special cases of the null hypothesis H_0:

Case 1: both parameters μ and σ^2 of the normal distribution are known;

Case 2: μ is known and σ^2 is estimated by $S_1^2 = n^{-1}\sum_{i=1}^{n}\left(X_i - \mu\right)^2$;

Case 3: both μ and σ^2 are unknown and are estimated by $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $S_2^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ respectively.

The most interesting cases, from a practical point of view, are Cases 2 and 3. Case 2 arises, for example, when we consider model residuals which are theoretically normal with null mean, or when we test for no difference between two means of normal distributions.

We focus here on the Shapiro–Wilk test (for Case 3) and the Kolmogorov–Smirnov test (for Cases 1, 2 and 3). The Shapiro–Wilk test for Case 2 was considered in Hanusz and Tarasińska (2014). A normalization of the test statistics via Johnson's S_B transformation is used. Moreover, functions for normalizing constants, dependent on the sample size *n*, are proposed. Thus, *p*-values for these tests are easily obtainable, as is shown in the 'Illustration' section.

Let us recall some facts concerning the Shapiro–Wilk and Kolmogorov–Smirnov tests. The Shapiro–Wilk test (1965) is based on the statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2},$$

where $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ are the ordered values of the sample and a_i are tabulated constants. Normality is rejected for small values of *W*. Critical points are tabulated in Shapiro and Wilk (1965) and are reproduced in other papers. The *W* test is considered very powerful for the hypothesis that a random variable *X* is normally distributed with unknown mean μ and variance σ^2. Shapiro and Wilk (1968) normalized the distribution of the *W* statistic using Johnson's S_B transformation (Johnson, 1949) and gave tables for three normalizing constants γ, δ and ε. In this paper we are interested in the relationships of these constants with the sample size.

We should add that another idea for the normalization of *W* was given by Royston (1992). This idea is implemented, for example, in the procedure 'shapiro.test{stats}' in the R environment.

The Kolmogorov–Smirnov statistic for testing normality belongs to the subclass of goodness-of-fit statistics, so-called EDF (Empirical Distribution Function) statistics, which are based on comparison of the population cumulative distribution function $F(x)$ with the empirical cumulative distribution function $F_n(x)$.

The Kolmogorov–Smirnov *D* statistic is defined as follows

$$D = \max_{1 \le i \le n} \{ | F(X_{(i)}) - \frac{i-1}{n} |, | F(X_{(i)}) - \frac{i}{n} | \},$$

where:

$$F(X_{(i)}) = \Phi\left(\frac{X_{(i)} - \mu}{\sigma}\right) \text{ for Case 1,}$$

$$F(X_{(i)}) = \Phi\left(\frac{X_{(i)} - \mu}{S_1}\right) \text{ for Case 2,}$$

$$F(X_{(i)}) = \Phi\left(\frac{X_{(i)} - \overline{X}}{S_2}\right) \text{ for Case 3,}$$

and Φ is the cumulative distribution function of the standard normal distribution. Normality is rejected for large values of *D*, and critical values (percentage points of the *D* distribution) are required for different sample sizes.

To eliminate the need to use tables of critical values, some efforts have been made to find the distribution of the *D* statistic. For Case 1, Marsaglia et al. (2003) gave such a distribution for a null hypothesis of any specified distribution, not only normal. This test is implemented in the R environment with the procedure 'ks.test{stats}'. Earlier, only the asymptotic distribution had been known (Kolmogorov, 1933; Miller, 1956; Birnbaum, 1952).

For Cases 2 and 3, Stephens (1974, Table 1) identified modified statistics of *D* and sample size *n* and gave asymptotic percentage points (15, 10, 5, 2.5 and 1) for them, which made the test more widely applicable. For example, in Case 3 Stephen's modified statistic is $D(\sqrt{n} - 0.01 + 0.85/\sqrt{n})$.

In this paper we propose normalization of the *D* statistics via Johnson's S_B transformation for Cases 1–3, and give the relationships of the normalizing constants with the sample size.

2. Normalization constants and their relationship with sample size for the Shapiro–Wilk test

For a bounded test statistic *T*, Johnson's S_B transformation (Johnson, 1949) can be used to obtain a normal approximation of *T* (S_B is an abbreviation for Bounded System):

$$Z = \gamma + \delta \ln\left(\frac{T-\varepsilon}{\lambda - T}\right)$$

where *Z* has approximately a standard normal distribution. The parameters ε and λ are respectively the minimum and maximum attainable values of the statistic *T*. The values of γ and δ may be evaluated by a Monte Carlo method.

Shapiro and Wilk (1968) transformed the *W* statistic by Johnson's S_B system, as $\varepsilon < W < 1$ with the minimal value $\varepsilon = (n-1)^{-1} n a_1$. For each *n* from 3 to 50 they obtained the least squares regression of the empirical sampling value of

$$u(p) = \ln \frac{W(p) - \varepsilon}{1 - W(p)}$$

on the *p*-th quantile of the standard normal distribution z_p, where $W(p)$ was the *p*-th empirical sampling quantile of the statistic *W*. They took the following values of *p*:

$$p = 0.01, 0.02, 0.05\,(0.05)\,0.25\,(0.25)\,0.75\,(0.05)\,0.95, 0.98, 0.99\ .$$

The regression gave estimates of $-\gamma\delta^{-1}$ and δ^{-1}, from which values for γ and δ were obtained. The normality hypothesis is rejected for small values of *Z*. Shapiro and Wilk (1968, Figure 1) plotted $-\gamma\delta^{-1}$ and δ^{-1} as a function of sample size *n*. However, they did not try to fit any analytical functions to the points; they merely gave a table with values of $\varepsilon, \gamma, \delta$ for different *n*.

Here we shall propose such functions for all three constants ε, γ and δ. Figure 1 shows the points (n,ε), (n,γ), (n,δ) and the lines with determination coefficient R-squared:

$$\varepsilon = 1.415 \cdot n^{-0.591} \text{ with } R^2 \approx 0.9987,$$

$$\gamma = -2.600 \cdot \ln(n) + 2.635 \text{ with } R^2 \approx 0.9978,$$

$$\delta = 1.364 \cdot \ln(\ln(n)) + 0.305 \text{ with } R^2 \approx 0.9961.$$

The null hypothesis of normality is rejected for the values of $Z = \gamma + \delta \ln((W - \varepsilon)(1 - W)^{-1})$ in the lower tail of the standard normal distribution.

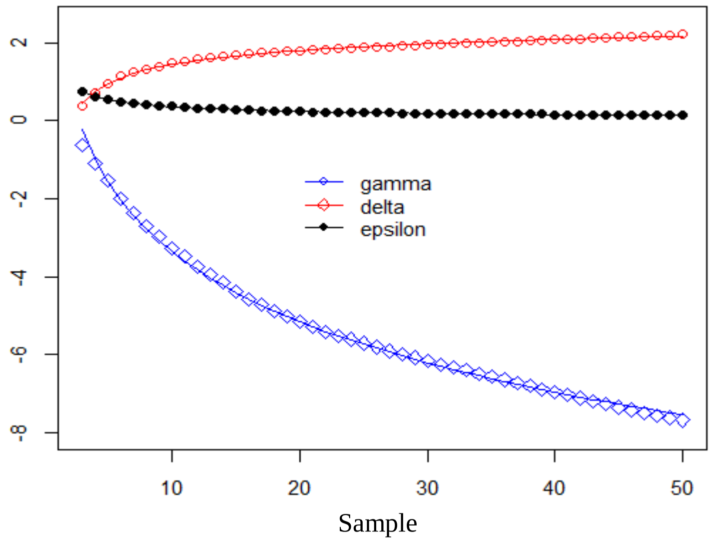

Figure 1. Scatterplots and regression lines in Johnson's S_B transformation of the Shapiro–Wilk W statistic

3. Normalization constants and their relationship with sample size for the Kolmogorov–Smirnov test

The Kolmogorov–Smirnov D statistic is also bounded ($0 < D < 1$), and therefore Johnson's S_B transformation $Z = \gamma + \delta \ln(D(1 - D)^{-1})$ can be applied. For Cases 1, 2 and 3, for each n from five to one hundred, 100,000 random samples from a standard normal distribution were generated. A least squares regression of the empirical sampling values of

$$u(p) = \ln \frac{D(p)}{1 - D(p)}$$

on the *p*-th quantile of the standard normal distribution was obtained, with the same values of $p = 0.01, 0.02, 0.05\,(0.05)\,0.25\,(0.25)\,0.75\,(0.05)\,0.95, 0.98, 0.99$ as in Shapiro and Wilk (1968).

By this means the estimates of γ and δ were calculated. Next the functions of γ and δ dependent on sample size *n* were found with *R*-squared near to one. The results are given in Table 1. The null hypothesis of normality is rejected for values of *Z* in the upper tail of the standard normal distribution.

Table 1. Regression functions on sample size and *R*-squared for coefficients for the Kolmogorov–Smirnov statistic *D*

Case	γ	δ
1	$\gamma = -2.132 + 2.130 \cdot \ln(n)$;	$\delta = 1.623 + 0.985 \cdot \ln(\ln(n))$;
	$R^2 \approx 0.9999$	$R^2 \approx 0.9978$
2	$\gamma = -2.148 + 2.083 \cdot \ln(n)$;	$\delta = 1.597 + 0.951 \cdot \ln(\ln(n))$;
	$R^2 \approx 0.9999$	$R^2 \approx 0.9970$
3	$\gamma = -0.521 + 2.525 \cdot \ln(n)$;	$\delta = 2.821 + 0.815 \cdot \ln(\ln(n))$;
	$R^2 \approx 0.9994$	$R^2 \approx 0.9952$

In order to check whether the tests for normality based on the statistic $Z = \gamma + \delta \ln(D(1-D)^{-1})$, with γ and δ given in Table 1, preserve the significance level α, a simulation study was performed. For sample sizes *n* from five to one hundred, 10,000 pseudorandom normal samples were generated, and the rates of rejection of normality at the significance level $\alpha = 0.05$ were evaluated. The results are shown in Figure 2. It can be seen that for Cases 2 and 3 the test preserves the nominal significance level 0.05 quite well. In Case 1 the empirical significance level is a little too large, but for this case the exact distribution of the *D* statistic is known (Marsaglia et al., 2003).

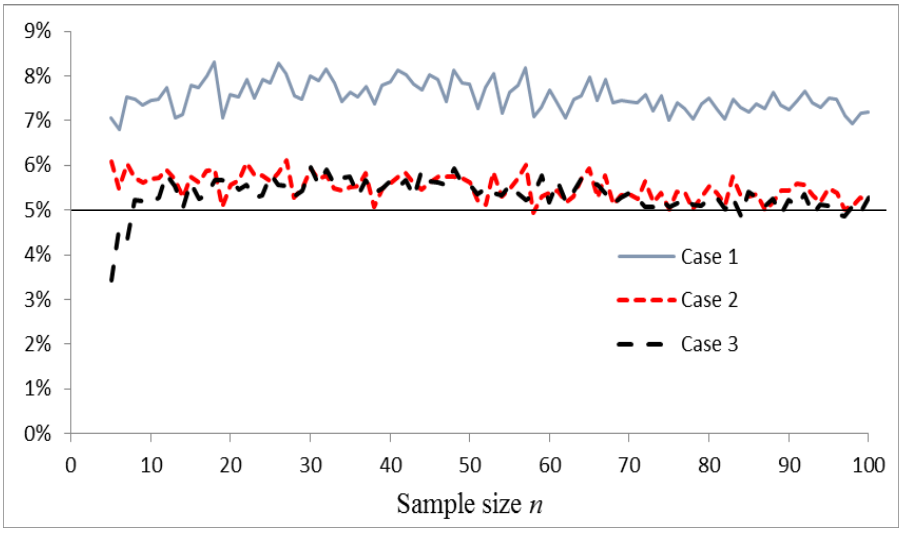

Figure 2. Empirical significance level for test statistics Z at significance level $\alpha = 0.05$

4. Illustration

As the normalization given in this paper for the Shapiro–Wilk test in Case 3 and the Kolmogorov–Smirnov test in Case 1 is not very interesting, because there exist other procedures giving *p*-values (Royston, 1992 for the Shapiro–Wilk and Marsaglia et al., 2003 for the Kolmogorov–Smirnov test), we give only an illustration of the Kolmogorov–Smirnov test for Cases 2 and 3.

Illustration 1.

Let us consider a set of the data giving the weights in centigrams of cork borings for the north and south sides of the trunk for 28 trees (Rao, 1948). Let us assume that we are interested in verifying the null hypothesis that the difference between the north and south weights are normally distributed with null mean. Thus we consider Case 2 with the null hypothesis $H_0 : X_N - X_S \sim N\left(0, \sigma^2\right)$. We obtain $D \approx 0.163$, $\gamma \approx 4.79$, $\delta \approx 2.74$, $Z \approx 0.314$, and the *p*-value equals 0.377. Thus the null hypothesis is not rejected.

Using the table of percentage points given by Stephens (1974), we can only conclude that the *p*-value is larger than 0.15.

The alternative approach to the data in Case 2 is to check normality with unknown parameters, e.g. by the Shapiro–Wilk test, and then (if normality is not rejected) apply the *t*-test for population mean. For our data, we obtain the *p*-value 0.1009 for the *W* statistic (by Royston's procedure) or 0.1180 (as the result of our test described in Section 2), so normality is not rejected. Next we obtain a *p*-value equal to 0.574 for the *t*-test of hypothesis $H_0 : \mu = 0$. The test based on $Z = \gamma + \delta \ln(D(1-D)^{-1})$ proposed in Section 3, giving one *p*-value, is more refined and is to be recommended.

Illustration 2.

To illustrate Case 3, let us take the following values of men's weights in pounds: 148, 154, 158, 160, 161, 162, 166, 170, 182, 195, 236, which were used by Shapiro and Wilk (1965) after Snedecor as an illustration of their test for normality. For the data we get $D \approx 0.2592$, $\gamma \approx 5.53$, $\delta \approx 3.53$, $Z \approx 1.823$, and a *p*-value equal to 0.034.

Using the table of percentage points given by Stephens (1975) we know only that the *p*-value is larger than 0.025 and less than 0.05. In this case, application of our test based on Johnson's S_B transformation of the *D* statistic is more advantageous.

5. Concluding remarks

For testing a null hypothesis of normality with unknown parameters (Case 3) and normality with known mean (Case 2), a test based on a normalizing transformation of the Kolmogorov–Smirnov *D* statistic, i.e. the test based on $Z = \gamma + \delta \ln(D(1-D)^{-1})$, can be used. There is no need for tables with the coefficients γ and δ, as there are well-fitting regression lines on sample size, given in Table 1. In both Case2 and Case 3, the test preserves the significance level of 0.05 quite well for sample sizes in a range from 5 to 100. The test provides the possibility of obtaining a *p*-value which is in the upper tail of the standard normal distribution.

We recommend the use of such a test especially in Case 2, instead of the combined Shapiro–Wilk test (for normality) and Student's *t*-test (for mean).

In the case of the Shapiro–Wilk test with normalizing constants, based on $Z = \gamma + \delta \ln((W - \varepsilon)(1 - W)^{-1})$, there is no need for the tables given in Shapiro and Wilk (1968), as there are well-fitting functions $\varepsilon = 1.415 \cdot n^{-0.591}$, $\gamma = -2.600 \cdot \ln(n) + 2.635$ and $\delta = 1.364 \cdot \ln(\ln(n)) + 0.305$.

REFERENCES

Birnbaum Z.W. (1952): Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. J. Amer. Statist. Assoc. 47: 425-44.

Johnson N.L. (1949): Systems of frequency curves generated by methods of translation. Biometrika 36: 149-176.

Hanusz Z., Tarasińska J. (2014): Normalization of Shapiro-Wilk test with known mean. Colloquium Biometricum 44: 35-41.

Kolmogorov A. (1933): Sulla derminazione empirica di una legge di distribuzione. Ist. Ital. Attuari G,4: 1-11.

Marsaglia G., Tsang W.W., Wang J. (2003): Evaluating Kolmogorov's distribution. Journal of Statistical Software 8/18.

Miller L.H. (1956): Table of percentage points of Kolmogorov statistics. J. Amer. Statist. Assoc. 51: 111-121.

R Development Core Team (2008): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rao C.R. (1948): Tests of significance in multivariate analysis. Biometrika 35: 58-79.

Royston J.P. (1992): Approximating the Shapiro-Wilk W-test for non-normality. Statistics and Computing 2: 117-119.

Shapiro S.S., Wilk M.B. (1965): An analysis of variance test for normality (complete samples). Biometrika 52: 591-611.

Shapiro S.S., Wilk M.B. (1968): Approximations for the null distribution of the W statistic. Technometrics 10: 861-866.

Stephens M.A. (1974): EDF statistics for goodness of fit and some comparisons. JASA, 69(347): 730-737.

Thode H.C. (2002): Testing for normality. Marcel Dekker Inc.