Simulation in Comparative Analysis of Several Tests for Normality

Boris Yu. Lemeshko
Faculty of Applied Mathematics and Informatics (FAMI)
Novosibirsk State Technical University
20, Karl Marx Avenue, Novosibirsk 630092
Russia
lemeshko@fpm.ami.nstu.ru

Andrey P. Rogozhnikov
FAMI
rogozhnikov.andrey@gmail.com

Abstract

Advantages and disadvantages are studied, and powers are estimated for different goodness-of-fit tests for the normal distribution (tests by Frosini, Hegazy-Green, Spiegelhalter, Geary and David-Hartley-Pearson).

1 Introduction

Due to objective reasons, testing for deviation from normal distribution is frequent procedure when conducting measurements, control, and tests. After *State Standard (2002)* have been released, *Lemeshko and Lemeshko (2005)* conducted a comparative analysis of a number of statistical tests supposed for testing for deviations from the normal distribution. A power was analyzed and shortcomings of particular tests were revealed that had not been mentioned in literature before.

In present work, analysis started by *Lemeshko et al. (2005)* is continued. A set of tests is extended by criteria proposed by *Frosini (1978,1987)*, *Hegazy and Green (1975)*, *Spiegelhaler (1977)*, *Geary(1935)*, and *David, Hartley, and Pearson (1964)*. Properties and powers of these tests were compared to the ones that had been analyzed by *Lemeshko et al. (2005)*. Recommendations on suitability of use of these tests are given.

In comparative analysis of tests power we considered the same competing hypotheses as in *Lemeshko et al. (2005)*. Hypothesis H_0 corresponds to normal law with density

$$f(x) = \frac{1}{\theta_1\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta_0)^2}{2\theta_1^2}\right\} \tag{1}$$

with the scale parameter $\theta_1 = 1$ and shift parameter $\theta_0 = 0$. The distribution of family

$$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \exp\left\{-\left(\frac{|x-\theta_0|}{\theta_1}\right)^{\theta_2}\right\} \tag{2}$$

with the shape parameter $\theta_2 = 4$, scale parameter $\theta_1 = 1$, and shift parameter $\theta_0 = 0$, is considered as competing hypothesis H_1; distribution of family (2) with shape parameter $\theta_2 = 1$ (Laplace distribution), scale parameter $\theta_1 = 1$, and shift parameter $\theta_0 = 0$ – as H_2; logistic distribution (3) with scale parameter $\theta_1 = 1$, and shift parameter $\theta_0 = 0$ – as H_3:

$$f(x) = \frac{\pi}{\theta_1\sqrt{3}} \exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\} \Bigg/ \left[1+\exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\}\right]^2 \tag{3}$$

To test deviations of an empirical distribution from the normal law it is possible to apply goodness-of-fit tests (non-parametric and χ^2-type). It seems natural to suppose that specially intended criteria should have certain advantages (sufficiently wide collection of such is given by *Kobzar (2006)*). Actually, such advantages are present when sample sizes are small, as rule.

But there are some complications. A simulation study in *Lemeshko et al. (2005)* showed that popular Shapiro-Wilk's and Epps-Pulley's tests, recommended by the *State Standard (2002)*, are biased under small sample sizes and small significance levels α (type I error probabilities), i.e. with respect to H_1 competing hypothesis (the power turns to be less than α). And, as we will see below, such serious shortcomings are typical to several other tests studied in the present work.

2 Tests under Consideration

Tests under consideration are based upon the following statistics.

Frosini:

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \Phi(z_i) - \frac{i-0.5}{n} \right|, \tag{4}$$

where

$$z_i = \frac{x_{(i)} - \overline{x}}{s}, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

$x_{(i)}$ are order statistics, $\Phi(z)$ is distribution function of standard normal law $N(0,1)$.

Hegazy-Green:

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} \left| z_i - \eta_i \right|, \tag{5}$$

$$T_2 = \frac{1}{n}\sum_{i=1}^{n} \left\{ z_i - \eta_i \right\}^2, \tag{6}$$

where z_i, $x_{(i)}$, and $\overline{x}$ are defined the same way as for Frosini's statistic,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

η_i – mathematical expectation of i-th order statistic of standard normal law, which can be found as

$$\eta_i = \Phi^{-1}\left(\frac{i}{n+1}\right).$$

Normality hypothesis is rejected under high values of the statistic.

Geary:

$$d = \frac{1}{ns}\sum_{i=1}^{n} \left| x_i - \overline{x} \right|, \tag{7}$$

where $\overline{x} = \sum_{i=1}^{n} x_i/n$, $s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2/n$. The test is two-sided, and normality hypothesis is accepted when

$$d\left(\frac{\alpha}{2}\right) \le d \le d\left(1 - \frac{\alpha}{2}\right)$$

where $d(\alpha)$ is quantile of d statistic's distribution.

David-Hartley-Pearson:

$$U = \frac{R}{s}, \tag{8}$$

where $R = x_{max} - x_{min}$ is range of sample, $s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2/n$ is unbiased variance estimator. Normality hypothesis is rejected when $U < U_1(\alpha)$ or $U > U_2(\alpha)$ (U_1 and U_2 are left and right quantiles of U statistic's distribution, respectively; α is significance level).

Spiegelhalter's statistic is a combination of Geary's and David-Hartley-Pearson's statistics:

$$T' = \left\{ (C_n U)^{-(n-1)} + g^{-(n-1)} \right\}^{\frac{1}{n-1}}, \tag{9}$$

where $C_n = \frac{1}{2n}(n!)^{\frac{1}{n-1}}$, U is statistic (8), $g = \frac{d}{\sqrt{(n-1)/n}}$, d is statistic (7). Normality hypothesis is rejected under high values of statistic T'.

3 Methodology of Study

In distributions study, percentage points calculation, and estimation of tests power with respect to different competing hypotheses, we used statistical simulation method by *Lemeshko et al. (2004)*. Distributions modeling was conducted by means of specially written module for the system ISW (*Lemeshko et al. 2004*). A count of trials (sizes of samples of statistics being studied) was choosen to $N = 10^6$ which allowed estimation of corresponding probabilities with error within $\pm 10^{-3}$.

4 Results of the Study

In course of research, the distributions of aforementioned tests were built under the assumptions that hypotheses H_i, $i = \overline{0,3}$, are true and $n = 10, 20, 40, 60, 80, 100, 200, 300$. For every sample size, tables of percentage points were calculated and tests powers were estimated with respect to competing hypotheses under consideration.

The common disadvantage of all six tests is that statistics distributions strongly depend on sample size and that their analytical distribution is unknown. Consequently, when deciding whether to accept a hypothesis or to reject it, one should follow the values of percentage points and can't estimate an achieved significance level, and it is hard to determine a degree of conformity or non-conformity of a given sample to the normal law.

Basing upon the research of tests properties and taking into account the powers that tests have shown with respect to competing hypotheses H_i, $i = \overline{1,3}$, these tests could be ranged as follows:

Geary's $\succ$ Spiegelhalter's[1] $\succ$ Hegazy-Green's $(T_2)^2$ $\succ$ Hegazy-Green's $(T_1)^3$ $\succ$ David-Hartley-Pearson's $\succ$ Frosini's.

But one should consider significant shortcomings of Spiegelhalter's and Hegazy-Green's tests:

- [1] Siegelhalter's test can't distinguish between hypotheses H_0 and H_1;
- [2] Hegazy-Green's test with statistic T_2, under small sample sizes, can't distinguish between H_0 and H_1 owing to bias;
- [3] Hegazy-Green's test with statistic T_1, under small sample sizes, is also somewhat biased as Shapiro-Wilk's and Epps-Pulley's tests (*Lemeshko et al. 2005*).

In the given row of preference, Epps-Pulley's test (*Epps and Pulley 1983*), that is included in the *State Standard (2002)*, should be placed after the Hegazy-Green's T_1 test due to it's power; Shapiro-Wilk's test (*Shapiro and Wilk 1965, Shapiro and Francia 1972*) follows right after David-Hartley-Pearson's test.

In *Lemeshko et al. (2005)* we gave the preference to a test with statistic z_2 (*D'Agostino, 1970*) which shown to be the most powerful with respect to competing hypotheses H_1 and H_3. In the row given above it asks to be placed to the first place, but it is worse than other tests with respect to the more distant hypothesis H_2.

It should be mentioned that, when testing a composite hypothesis, Anderson-Darling's Ω^2 and Nikulin's χ^2 goodness-of-fit tests are not much worse than tests with statistics z_2 (*Lemeshko et al. 2007, Lemeshko et al. 2008*), Hegazy-Green's T1 and T2, Spiegelhalter's, and Geary's tests; they are better than the rest of tests for normality, that we studied.

This research was partly supported by Russian Foundation for Basic Research (project #06-01-00059a) and by Federal Education Agency of Russian Ministry of Education in the Analytical departmental purposeful program framework "Development of Higher School Potential" (project #2.1.2/3970).

References

[1] D'Agostino R.B. (1970). *Transformation to normality of the null distribution of* g_1, Biometrika, V.57, pp. 679-681.

[2] David H.A., Hartley H.O., Pearson E.S. (1964) *The distribution of the ratio, in a single normal sample, of range to standard deviation*, Biometrika, V.51, #3-4, pp. 484-487.

[3] Epps T.W., Pulley L.B. (1983) *A test for normality based on the empirical characteristic function*, Biometrika. V.70, pp. 723-726.

[4] Frosini B.V. (1978). *A survey of a class of goodness-of-fit statistics* Metron. V.36, #1-2, pp. 3-49.

[5] Frosini B.V. (1987). *On the distribution and power of goodness-of-fit statistic with parametric and nonparametric applications*, "Goodness-of-fit" / Ed. by Revesz P., Sarkadi K., Sen P.K. - Amsterdam-Oxford-New York: North Holland Publ.Comp, 1987, pp. 133-154.

[6] Geary R.C. (1935). *The ratio of the mean deviation to the standard deviation as a test of normality*, Biometrika, V.27, pp. 310-322.

[7] Hegazy Y.A.S., Green J.R. (1975). *Some new goodness-of-fit tests using order statistics.* Applied Statistics, V.24, #3, pp. 299-308.

[8] Kobzar A.I. (2006). *Applied mathematical statistics. For engineers and scientists* (in Russian), Moscow: FIZMATLIT Publishing House.

[9] Lemeshko B.Yu., Postovalov S.N. (2004). *Computer technologies of data analysis and statistical regularities research: a tutorial* (in Russian), Novosibirsk: NSTU Publishing House.

[10] Lemeshko B.Yu., Lemeshko S.B. (2005). *A comparative analysis of tests for deviations of probabilities distribuion from normal distribution* (in Russian), Metrology, #2, pp. 3-24.

[11] Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N. (2007). *The power of goodness of fit tests for close alternatives*, Measurement Techniques, V.50, #2 pp. 132-141.

[12] Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N. (2008). *A comparative ananlysis of goodness-of-fit tests power under close competing hypotheses. II. Testing composite hypotheses* (in Russian), Siberian Journal of Industrial Mathematics, V.11, #4(36)

[13] Nikulin M.S. (1973a) Chi-squared test for continuous distributions with shift and scale parameters, Theory of probabilities and its applications, V.18, #3, pp. 559-568.

[14] Nikulin M.S. (1973b). *On a chi-squared test for continuous distributions*, Theory of probabilities and its applications, V.18, #3, p.638-639.

[15] Shapiro S.S., Francia R.S. (1972) *An approximate analysis of variance test fo normality*, J. Amer. Statist. Assoc., 337, pp. 215-216.

[16] Shapiro S.S., Wilk M.B. (1965) *An analysis of variance test for normality (complete samples)* Biometrika, V.52, pp. 591-611.

[17] Spiegelhalter D.J. (1977). *A test for normality against symmetric alternatives*, Biometrika, V.64, #2, pp. 415-418.

[18] State Standard GOST R ISO 5479-2002. *Statistical methods. Testing for deviations of probabilities distribuion from normal distribution*, Moscow: Standards Publishing House. 2002.