Journal of Physics: Conference Series

PAPER • OPEN ACCESS

About the effect of rounding on the properties of tests for testing statistical hypotheses

To cite this article: B Yu Lemeshko and S B Lemeshko 2021 *J. Phys.: Conf. Ser.* **1715** 012063

About the effect of rounding on the properties of tests for testing statistical hypotheses

B Yu Lemeshko[1] **and S B Lemeshko**[1]

[1] Faculty of Applied Mathematics and informatics, Novosibirsk state technical university, 630073, Novosibirsk, Russia

E-mail: `lemeshko@ami.nstu.ru`

Abstract. The paper shows the results of the influence of rounding errors on the distribution of statistics for testing various statistical hypotheses. It is shown that when rounding errors are commensurable with the standard deviation, measurement errors in the distribution of statistical tests (goodness-of-fit tests, homogeneity tests, normality tests, etc.) can significantly deviate from those that occur in the absence of rounding. Recommendations are given for applying the tests in such situations.

1. Introduction

Most of the existing tests are designed to test statistical hypotheses about continuous random variables. This is a standard assumption, which is rarely emphasized, and which determines whether the appropriate tests are applied correctly. Under this assumption, the samples cannot have duplicate values.

In real situations, this assumption is often violated due to rounding errors, and then repeated observations are found in the samples. This is typical for medical and biological experiments, where rounding errors are very significant due to their specificity. These can be samples with the results of high-precision measurements, in which the changes concern, for example, only the last decimal place, which, as a rule, is determined by the resolution of the measuring system. In automated data processing systems usually also have the rounded results of the measurements coming from different sensors.

Any measurements are accompanied by some rounding error, depending on the resolution of the measuring system, including the characteristics of the sensors and analog-to-digital converters.

Obviously, the presence of rounding errors somehow affects the results of applying statistical methods, and in some situations can lead to incorrect statistical conclusions.

In the works [1, 2, 3] it was shown that the main problem of applying various tests for testing statistical hypotheses for the analysis of large samples is related to the presence of rounding errors of the analyzed data. With large sample sizes, due to the natural presence of rounding errors, the distributions of the statistic statistics differ significantly from those in the classical situation. As a result of this, the classical results concerning the distributions of statistics can be used only up to certain sample sizes $n \leq n_{max}$.

The impossibility of using classical results is encountered not only with large samples.

In this paper, based on the methods of statistical modeling, we study the effect of rounding errors on the distribution of statistics of various tests for testing statistical hypotheses. It is shown that, when rounding errors are commensurate with the standard deviation of the measurement errors, the distributions of the statistics of the tests used to analyze the corresponding data can deviate significantly from the classical ones with limited sample sizes. In such conditions, the application of tests using classical results may lead to incorrect conclusions. The effect of rounding errors on the distribution of statistic of goodness-of-fit tests, tests of homogeneity, normality, exponentiality and others is demonstrated.

2. Effect of rounding on goodness-of-fit tests

2.1. Basic equations

The applications are often faced with a situation where due to rounding in the analyzed samples is a lot of duplicate values. This can be regarded as a warning, indicating that the actual distributions of the $G(S_n|H_0)$ test statistics (for rounding error $\triangle$ and sample sizes n) can differ significantly from the limit distributions $G(S|H_0)$ or from $G(S_n|H_0)$, which occur in a situation without rounding measurements. When the value $\triangle$ turns out to be commensurate with the standard deviation σ of the distribution law of the measurement error, the distribution of statistics may not try to converge to the limit law, and with increasing n it will only move away from it.

In particular, in such a situation, the presence of rounding leads to a change in the distribution of statistics of all the tests of agreement most used in practice (Kolmogorov [4], Kuiper [5], Cramer-Mises-Smirnov [6], Watson [7, 8], Anderson-Darling [9, 10], Zhang [11], χ^2 Pearson, Nikulin-Rao-Robson [12, 13]) both when testing simple and complex hypotheses. By the example of the Anderson-Darling test [9, 10] with statistics

$$S_\Omega = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\Big\{\ln F(x_i,\theta) + \ln(1 - F(x_{n-i+1},\theta))\Big\}$$

we show the change in the distribution of $G(S_\Omega|H_0)$ depending on $\triangle$ for $n = 50$ in the case of testing the simple hypothesis that the samples belong to the normal law. Figure 1 shows the distribution of $G(S_\Omega|H_0)$ when $\triangle$ changes from 0 to σ of the observed law. As you can see, even with the rounding error $\triangle = 0.1\sigma$, the deviation of the distribution of statistics from the asymptotic $a2(S)$ [6] (for $\triangle = 0$) cannot be neglected. At the same time, with $\triangle = 0.01\sigma$ and $n = 50$, the distribution of statistics does not differ from the asymptotic one. With an increase in sample sizes, the picture presented in the figure 1 will change: the distributions will shift to the right of the asymptotic for the corresponding distributions.

Similarly, depending on $\triangle$, the distributions of the statistics of the Anderson-Darling test change when testing complex hypotheses. On the whole, a similar picture of the dependence on rounding errors of the distributions of statistics can be observed for other tests of agreement.

If you really competing hypothesis holds H_1 sample of belonging to some other law, the conditional distribution of $G(S_n|H_1)$, which occur when testing the hypothesis H_0, also depend on $\triangle$. Moreover, with an increase in the rounding error $\triangle$, the power of tests like χ^2 decreases noticeably, and the power of nonparametric tests for agreement can increase somewhat.

Relatively recently proposed goodness-of-fit tests, based on the evaluation of the Kullback-Leibler information [14], which has the form of statistics

$$S_{KL} = -\frac{1}{n}\sum_{i=1}^{n}\ln\Big\{\frac{n}{2m}[F(x_{i+m},\theta) - F(x_{i-m},\theta)]\Big\}.$$

where $m \le n/2$; $x_i = x_1$, if $i < 1$; $x_i = x_n$, if $i > n$. This test, as well as a number of special tests for checking normality and uniformity, for example, [15, 16, 17], the statistics of which use different estimates of entropy, as a rule, demonstrate high power. However, analysis of samples with the number of identical values of $l\ ge 2m + 1$ using similar tests is problematic.

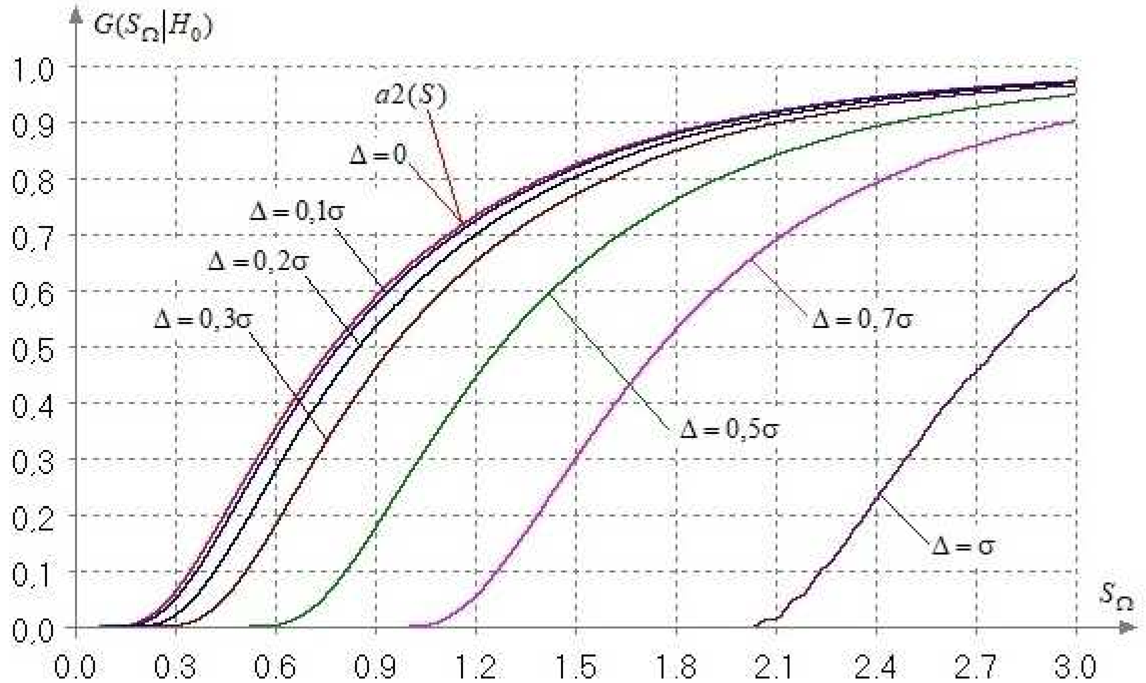

Figure 1. Distributions of statistics of the Anderson-Darling test when testing a simple hypothesis depending on $\triangle$ (for $n = 50$)

Changing the properties of the tests under the influence of rounding errors does not exclude the possibility of their correct application. You just need to know (or be able to find) the distribution of the test $G(S_n|H_0)$ statistics for the same rounding error $\triangle$ and the same sample size n that correspond to the analyzed sample. To do this, it is best to use the methods of statistical modeling and as a result of N experimentation to find the empirical distribution of statistic $G_N(S_n|H_0)$ test (for the same $\triangle$ and n), by which to calculate the estimate of p_{value}. This feature is implemented in the developed software system [18].

3. Effect of rounding on the normality tests

To test the hypothesis that the sample belongs to the normal law, in addition to the goodness-of-fit tests, more than 20 special normality tests can be used, the distributions of statistics also affected by rounding errors. We demonstrate this influence on the Frosini tests.

The statistics of the Frosini test [19], designed to check normality, has the form

$$B_n = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mid \Phi(z_i) - \frac{i-0,5}{n} \mid,$$

wherein sample elements x_i, $i = \overline{1,n}$, sorted ascending; $z_i = \frac{x_i - \bar{x}}{s}$; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$; $\Phi(z_i)$ - distribution function of the standard normal law.

The conditional distributions of the $G(B_n|H_0)$ statistics of the Frosini test for the validity of the tested hypothesis H_0 depend on the sample size n, but for $n > 100$ the distribution of the statistics does not change significantly. In [20] proposes a limit law model that can be used for $n > 50$.

Similarly to the previous example with the Anderson-Darling test, the figure 2 shows the distributions of $G(B_n|H_0)$ statistics of the Frosini test when $\triangle$ changes in the same interval for

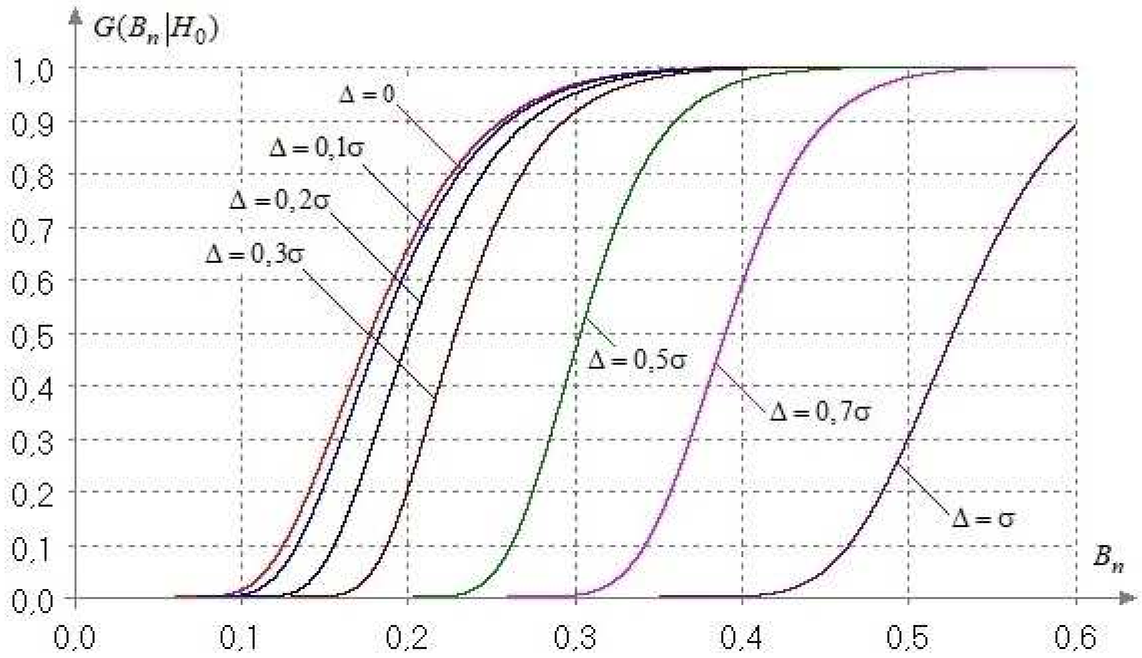

Figure 2. Frosini normality test statistics distributions depending on $\triangle$ (for $n = 50$)

$n = 50$. And in this case, when the rounding error is $\triangle = 0.1\sigma$, the deviation of the distribution of statistics from the one taking place with $\triangle = 0$ cannot be neglected either. And at $\triangle = 0.01\sigma$ and $\triangle = 0$ distribution statistics do not differ.

4. Effect of rounding on the exponentiality tests

To test the hypothesis that the samples belong to an exponential law can be used all goodness-of-fit tests and more than 3 dozen special tests. Naturally, the statistics distributions of these tests change under the influence of data rounding.

When checking the exponentiality, the statistics of the Frosini test [19] takes the form

$$B_n = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mid 1 - \exp\left(-\frac{x_i}{\bar{x}}\right) - \frac{i-0{,}5}{n} \mid .$$

Dependence of the distribution $G(B_n|H_0)$ on the value of the test statistic $\triangle$ for $n = 50$ illustrated in Figure 3. Naturally, these distributions differ significantly from the distributions of the test statistics when checking normality, since the tested complex hypotheses differ.

It should be noted that when rounding observations obeying the exponential law, values that coincide with the values of the shift parameter may appear in the samples, which cannot be in samples of a continuous random variable. This fact, as a rule, does not cause serious problems in estimating the parameters of the law, but leads to problems when using certain tests. In particular, the statistics of the Anderson-Darling test and the statistics of the three Zhang tests [11] due to the presence in the analyzed sample of values that coincide with the shift parameter, take the value $+\infty$, and $p_{value} = 0$, which is evidence of the rejection (even true) of the test hypothesis H_0. The same thing happens with a number of special tests for exponentiality.

If such a situation is detected, the possibility of the correct application of the corresponding test (without loss of generality with a zero shift parameter) can be realized by the following algorithm.

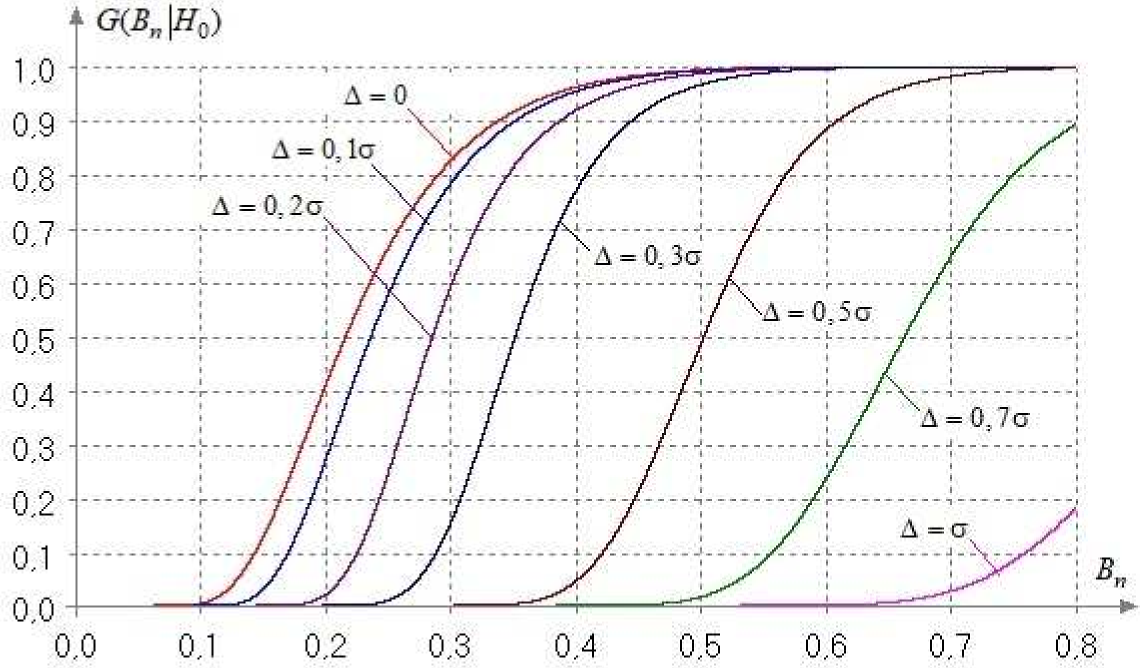

Figure 3. Frosini exponentiality test statistics distributions depending on $\triangle$ (at $n = 50$)

1. Using the estimate of the $\bar{x}$ scale parameter obtained from the sample, restore k the minimum sample elements that coincide with the shift parameter and appear in the sample as a result of rounding off observations that fall into the left interval of $0.5\triangle$, with the values $x_i = -\bar{x}\ln(1-\xi_i)$, $i = \overline{1,k}$, where ξ_i are uniform pseudorandom variables on the interval $[1, 1-\exp(-0.5\triangle/\bar{x})$.
2. Using the obtained sample, calculate the value of the test S^* statistics.
3. Simulate the distribution of $G_N(S_n|H_0)$ test statistics for given n and $\triangle$. In the process of generating pseudo-random samples $x_1, x_2, ..., x_n$ by law $F(x,\bar{x}) = 1-\exp(-x/\bar{x})$ value $x_i < 0.5\triangle$ should not be rounded.
4. Using the distribution of $G_N(S_n|H_0)$ and the value of S^*, evaluate the value of p_{value}.

An implementation of this approach is provided in [18].

5. Effect of rounding on the homogeneity tests

In the tests for the homogeneity of laws, a hypothesis of the form H_0 is verified: $F_1(x) = F_2(x) = ... = F_k(x)$. On the distribution of $G(S|H_0)$ statistics tests affect the degree of rounding and differences in the analyzed samples. The effect of rounding errors on the distributions of statistics of homogeneity tests is shown by the example of the two-sample Anderson-Darling-Petit test [21], whose statistics are determined by the expression

$$A^2 = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1+n_2-1}\frac{[M_i(n_1+n_2)-in_1]^2}{i(n_1+n_2-i)},$$

where M_i is the number of elements of the first sample that are less than or equal to the i -th element of the variational series of the combined sample. The limit distribution of statistics under the validity of the tested hypothesis H_0 is the same distribution of $a2(S)$ [6].

The influence of the rounding error on the distribution of statistics of the tests for the homogeneity of laws with justice H_0 without loss of generality will be considered if the analyzed samples belong to the standard normal law.

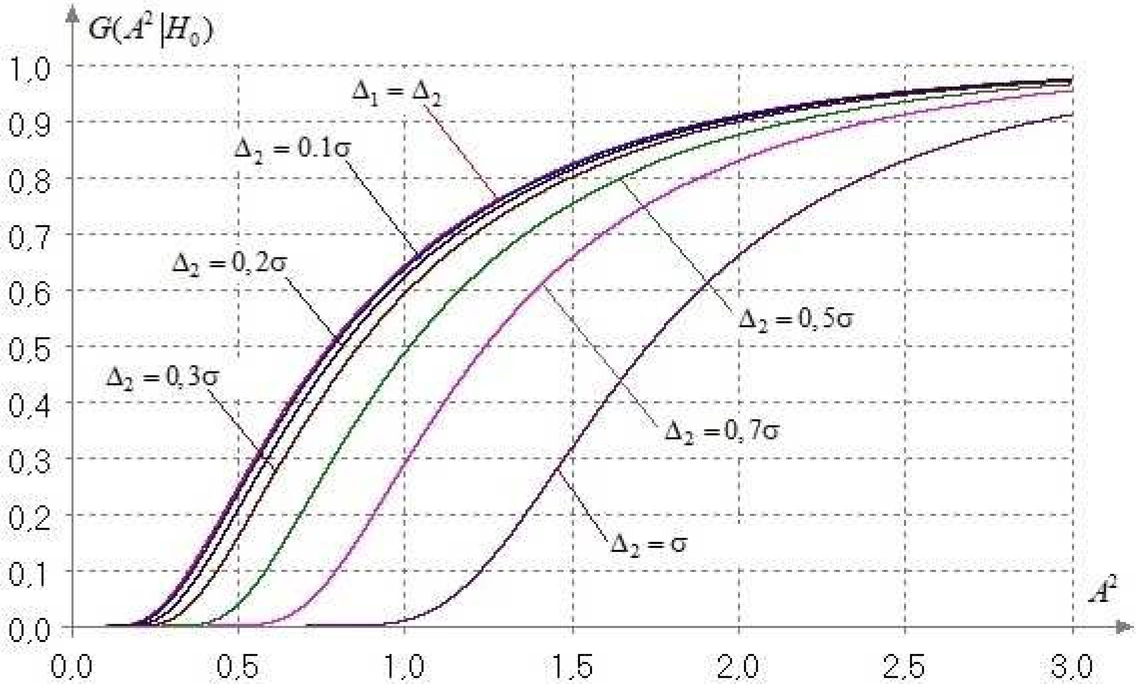

Figure 4. The dependence of the distributions of the statistics of the Anderson-Darling-Petit test on $\triangle_2$ for $\triangle_1 = 0.01\sigma$ and $n_i = 50$

With $\triangle_1 = \triangle_2$ and sample sizes $n_i = 50$ and $\triangle_i \leq 0, 0\sigma$, the distributions of $G(A^2|H_0)$ practically do not deviate from the distribution of $a2(S)$, but deviate for unequal $\triangle_i$. In the figure 4 the distributions of $G(A^2|H_0)$ test statistics for $n_i = 50$ are shown depending on $\triangle_2$ for $\triangle_1 = 0.01\sigma$. In this case, the deviation of the distribution of $G(A^2|H_0)$ from $a2(S)$ with $\triangle_1 = 0.1\sigma$ still has no practical value. For the same $\triangle_i$, as n_i grows, the deviation of $G(A^2|H_0)$ from $a2(S)$ increases.

In the same way, rounding errors affect the distributions of statistics of other two-sample ones (Smirnov [6], Leman-Rosenblatt [22, 23]) and k-sampled (Anderson-Darling [24] ,Zhang [25] and other [26]) tests for the uniformity of laws.

You can pay attention to the fact that the growth of rounding errors $\triangle_i$ does not lead to significant changes in the power of the tests for the homogeneity of laws.

Investigation of the distributions of statistics of 2 and k-sample parametric tests for the homogeneity of means used to test the H_0 hypothesis: $\mu_1 = \mu_2 = ... = \mu_k$, or the test used to test the H_0 hypothesis : $\mu = \mu_0$ on the equality of mathematical expectation to the nominal value, showed that the rounding errors of the measurement results do not have any noticeable effect on them.

In contrast, the statistics distributions of the parametric tests used to test similar hypotheses regarding the variances H_0: $\sigma_1 = \sigma_2 = ... = \sigma_k$ and H_0: $\sigma = \sigma_0$, rounding errors can have significant impact.

We show this on Bartlett's tests [27], whose statistics are calculated in accordance with the relation [6]

$$\chi^2 = M\Big\{1 + \tfrac{1}{3(k-1)}\Big\{\sum_{i=1}^{k} \tfrac{1}{v_i} - \tfrac{1}{N}\Big\}\Big\}^{-1},$$

where $M = N \ln\Big\{\frac{1}{N}\sum_{i=1}^{k} v_i s_i\Big\} - \sum_{i=1}^{k} v_i \ln s_i^2$; k-number of samples; n_i - sample sizes; $v_i = n_i$

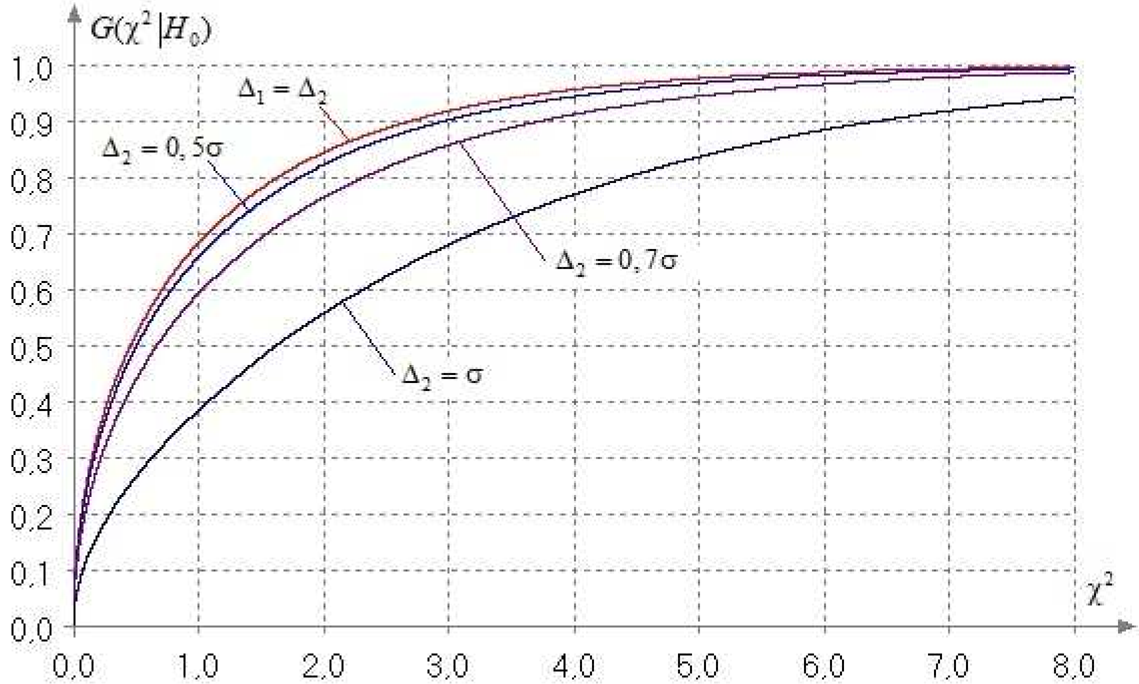

Figure 5. The dependence of the distribution of Bartlett statistics on $\triangle_2$ for $\triangle_1 = 0.01\sigma$ and $n_i = 1000$

if expectation is known, and $v_i = n_i - 1$ if not known; $N = \sum_{i=1}^{k} v_i$; s_i^2 - estimates of sample variances. When an unknown mathematical expectation used estimates $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ where x_{ij} is the j-th observation in the i-th sample.

The asymptotic distribution of statistics of the Bartlett test under the standard assumption of normality and the number of compared samples of k is the χ^2_{k-1} -distribution. Given rounding and equal $\triangle_i$, the real distributions of the $G(\chi^2|H_0)$ statistics of the test do not deviate from the χ^2_{k-1} -distribution. For significant and different $\triangle_i$, the real distributions of $G(\chi^2|H_0)$ deviate from the χ^2_{k-1} -distribution.

Figure 5 illustrates the dependence of the distribution of $G(\chi^2|H_0)$ statistics of the Bartlett test on the rounding error of $\triangle_2$ observations in the second sample with $\triangle_1 = 0.01\sigma$ and with sample sizes $n_i = 1000$ in the case of $k = 2$.

As you can see, with significantly different $\triangle_1$ and $\triangle_2$ the distributions of $G(\chi^2|H_0)$ significantly deviate from the χ^2_1 distribution. Note that for smaller sample sizes, for example, for $n_i = 100$, the deviation of the real distribution from the asymptotic distribution for $\triangle_2 = 0.5\sigma$ can still be neglected.

If the real distributions of the $G(\chi^2|H_0)$ statistics in the presence of rounding and equal to $\triangle_i$ do not deviate from the asymptotic ones, then the validity of competing hypotheses H_i of the distribution of $G(\chi^2|H_i)$ depend on rounding error. At the same time, the power of the tests decreases with increasing $\triangle_i$ even with equal $\triangle_i$ [2].

Studies have shown that in a similar way when unequal $\triangle_i$ with an increase in rounding errors are changing the distribution statistics of other parametric tests homogeneity dispersion discussed in [28].

In general, it can be stated that the distributions of the $G(S|H_0)$ statistics of the parametric tests for the dispersion homogeneity for the same degree of rounding $\triangle_i$ of the measurement results in the analyzed samples do not differ from the corresponding distributions in the absence of rounding ($\triangle_i$, $i = \overline{1,k}$). But the same distributions can differ significantly for different degrees of rounding (for $\triangle_i \neq \triangle_j$, $i \neq j$, $i, j = \overline{1,k}$) in analyzed samples. The foregoing concerns the set of all parametric tests for the variance homogeneity considered in [28].

Thus, all the homogeneity tests show the stability of the distributions of $G(S|H_0)$ statistics to the presence of rounding in the analyzed samples in the case of equality $\triangle_i$. In the case of unequal $\triangle_i$, the distributions of the statistics of the tests for the homogeneity of laws change to the greatest extent. To a lesser extent in the same situation changing distribution statistics parametric tests homogeneity dispersions. The greatest stability with respect to rounding errors is demonstrated by the tests for checking the homogeneity of averages.

Conclusion

When using specific tests in applications, as well as in automated data processing systems, it is necessary to take into account the possible effect of rounding errors on the distribution of tests statistics.

A signal of caution in using classical results with respect to the applied tests is the presence in the analyzed samples of too many repeating values. If not, one can rely on classical results.

Changing the properties of the test under the influence of rounding errors does not exclude the possibility of its correct application. To do this, it suffices to simulate the unknown distribution of statistics of the $G(S_n|H_0)$ test for the corresponding $\triangle_i$ and n_i, and as a result of N simulation experiments, find the empirical distribution of statistics $G_N(S_n|H_0)$. Further, having the distribution $G_N(S_n|H_0)$, from the calculated value of the test statistics S^*, we can find the estimate $p_{value} = 1 - G_N(S^*|H_0)$.

Such an approach to the application of the set of tests mentioned in the article in the presence of rounding errors is implemented in the system [18], using which the present research was carried out. The implementation of such software support in specialized data processing systems does not present fundamental difficulties.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation in the framework of state work The maintenance of scientific research (1.4574.2017/6.7) and the project part of the government assignment (1.1009.2017/4.6).

References

[1] Lemeshko B Yu, Lemeshko S B and Semenova M A 2018 To question of the statistical analysis of big data *Tomsk State University Journal of Control and Computer Science* **44** 40-49

[2] Lemeshko B, Lemeshko S and Semenova M 2019 Features of testing statistical hypotheses under big data analysis *Applied Methods of Statistical Analysis. Statistical Computation and Simulation - AMSA'2019, Novosibirsk, Russia, 18-20 September, 2019: Proceedings of the International Workshop* (Novosibirsk: NSTU publisher) pp 122-137

[3] Lemeshko B Yu, Lemeshko S B, Veretelnikova I V and Blinov P Yu 2019 Tests for testing statistical hypotheses in the analysis of large samples: problems and their solution *Marchukov Scientific Readings - 2019 : Proceedings of the International Conference "Actual Problems of Computational and Applied Mathematics" / Institute of Computational Mathematics and Mat. geophysics SB RAS. Novosibirsk, July 1-5, 2019* (Novosibirsk: IPC NSU) pp 277-283

[4] Kolmogoroff A N 1933 Sulla determinazione empirica di una legge di distribuzione *Giornale del Istituto Italiano degli Attuari* **4** (1) 83-91

[5] Kuiper N H 1960 Tests concerning random points on a circle *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series A* **63** 38-47

[6] Bolshev L N and Smirnov N V 1983 *Tables of Mathematical Statistics* (Moscow: Science) p 416

[7] Watson G S 1961 Goodness-of-fit tests on a circle. I *Biometrika* **48** (1-2) 109-114
[8] Watson G S 1962 Goodness-of-fit tests on a circle. II *Biometrika* **49** (1-2) 57- 63
[9] Anderson T W and Darling D A 1954 A test of goodness of fit *Journal of the American Statistical Association* **29** 765-769
[10] Anderson T W and Darling D A 1952 Asymptotic theory of certain Goodness of fit tests based on stochastic processes *The Annals of Mathematical Statistics* **23** 193-212
[11] Zhang J 2001 Powerful goodness-of-fit and multi-sample tests *PhD Thesis* (Toronto:York University) p 113 URL: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf (acc. 03.12.2019).
[12] Nikulin M S 1973 On a chi-square test for continuous distribution *Theory Probability and Its applications* **18** (3) 675-676
[13] Rao K C and Robson D S 1974 A chi-squared statistic for goodness-of-fit tests within the exponential family *Communications in Statistics - Theory and Methods* **3** 1139-1153
[14] Noughabi H A 2019 A new estimator of Kullback-Leibler information and its application in goodness of fit tests *Journal of Statistical Computation and Simulation* **89** (10) 1914-1934
[15] Vasicek O 1976 A test of normality based on sample entropy *Journal of the Royal Statistical Society: Series B* **38** (1) 54-59
[16] Zamanzade E and Arghami N R 2012 Testing normality based on new entropy estimators *Journal of Statistical Computation and Simulation* **82** (11) 1701-1713
[17] Zamanzade E 2014 Testing uniformity based on new entropy estimators *Journal of Statistical Computation and Simulation* **85** (16) 3191-3205
[18] ISW - Software System for the Statistical Analysis of One-Dimensional Observations,. https://ami.nstu.ru/~ headrd/ISW.htm. (acc. 03.27.2020)
[19] Frosini B V 1978 A survey of a class of goodness-of-fit statistics *Metron* **36** (1-2) 3-49
[20] Lemeshko B Yu 2015 *Tests for checking the deviation from normal distribution law. Guide on the application* (Moscow.: INFRA-M) p 160
[21] Pettitt A N 1976 A two-sample Anderson-Darling rank statistic *Biometrika* **63** (1) 161-168
[22] Lehmann E L 1951 Consistency and unbiasedness of certain nonparametric tests *Ann. Math. Statist.* **22** (1) 165-179
[23] Rosenblatt M 1952 Limit theorems associated with variants of the von Mises statistic *Ann. Math. Statist.* **23** 617-623
[24] Scholz F W and Stephens M A 1987 K-Sample Anderson-Darling Tests *Journal of the American Statistical Association* **82** (399) 918-924
[25] Zhang J and Wu Y 2007 k-Sample tests based on the likelihood ratio *Computational Statistics and Data Analysis* **51** (9) 4682-4691
[26] Lemeshko B Yu and Veretelnikova I V 2018 Power of k-sample tests aimed at checking the homogeneity of laws *Measurement Techniques* **61** (7) 647-654
[27] Bartlett M S 1937 Properties of sufficiency of statistical tests *Proceedings of the Royal Society A* **160** (901) 268-282
[28] Lemeshko B Yu 2017 *Tests for homogeneity. Guide on the application* (Moscow: INFRA-M) p 208