ISSN 8756-6990, Optoelectronics, Instrumentation and Data Processing, 2016, Vol. 52, No. 2, pp. 128–140. © Allerton Press, Inc., 2016.
Original Russian Text © B.Yu. Lemeshko, P.Yu. Blinov, S.B. Lemeshko, 2016, published in Avtometriya, 2016, Vol. 52, No. 2, pp. 28–42.

ANALYSIS AND SYNTHESIS
OF SIGNALS AND IMAGES

Goodness-of-Fit Tests for Uniformity of Probability Distribution Law

B. Yu. Lemeshko, P. Yu. Blinov, and S. B. Lemeshko

Novosibirsk State Technical University,
pr. Karla Marksa 20, Novosibirsk, 630073 Russia
E-mail: Lemeshko@ami.nstu.ru

Received April 7, 2015

Abstract—This paper describes a set of tests for uniformity of observations which are ranked by power. It is shown that most tests commonly used for the hypothesis of uniformity are biased relative to a certain kind of competing hypotheses. It is emphasized that specific tests meant only for uniformity have no obvious advantages over nonparametric goodness-of-fit tests used for the same purpose.

Keywords: uniformity, testing hypotheses, statistical test, test power.

DOI: 10.3103/S8756699016020047

INTRODUCTION

For several reasons, the probability distribution law is prominent in the problems of statistical analysis. Uniformity is often used as a model for describing the measurement errors of some devices or systems, which is not least due to the lack of information. Naturally, its unjustified use can cause problems.

The hypothesis of uniformity of random variables (measurement errors) can be subjected to different statistical tests of a fairly long list that can be divided into two subsets. These include general goodness-of-fit tests used for uniformity and special tests meant only for the hypothesis of uniformity of the sample $X_1, X_2, \ldots, X_n$.

There are multiple tests available for professionals which makes it difficult for them to choose as the data given in the publications do not allow one to opt for a certain test, and every professional wants a chosen test (or tests) to be correct and statistical conclusions to be of high quality (reliable).

A plurality of considered tests can be used for a simple hypothesis of uniformity of random variable X on the interval [0, 1] or on the interval $[a, b]$ with known a and b, or for a composite hypothesis where a and b are unknown.

Usually a simple testable hypothesis of uniformity of the sample $X_1, X_2, \ldots, X_n$ of independent observations of a random value of X has the form: H_0: $F(x) = x$, $x \in [0, 1]$.

Most of the tests for the hypothesis of uniformity on the interval [0, 1] are based on the estimates of order statistics of a random value X (on the elements $x_{(i)}$ of the ordered series $0 < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < 1$ constructed according to the sample $X_1, X_2, \ldots, X_n$). Further, the following denotations are used in the statistic tests: $U_i = x_{(i)}$, $i = \overline{1, n}$, $U_0 = 0$, and $U_{n+1} = 1$.

As a rule, tests are aimed at a simple hypothesis H_0 on the interval [0, 1]. If it is necessary to test the simple hypothesis of uniformity of the sample $X_1, X_2, \ldots, X_n$ on the interval $[a, b]$ (with shift parameter a and scale parameter $b-a$), all uniformity tests can be used by converting the elements $x_{(i)}$ of the ordered series $a < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < b$ constructed according to the sample $X_1, X_2, \ldots, X_n$ into corresponding (required in the tests) order statistics the following way: $U_i = (x_{(i)} - a)/(b-a)$, $i = \overline{1, n}$, $U_0 = 0$, $U_{n+1} = 1$. The rest of the procedure of applying uniformity tests remains unchanged (as on the interval [0, 1]).

When testing the composite hypothesis of uniformity of the form H_0: $F(x) = (x-a)/(b-a)$, $x \in [a, b]$, where a and b are unknown and should be found according to the same sample, we use the elements of the ordered series $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ constructed according to the sample $X_1, X_2, \ldots, X_n$ to find the parameter estimates:

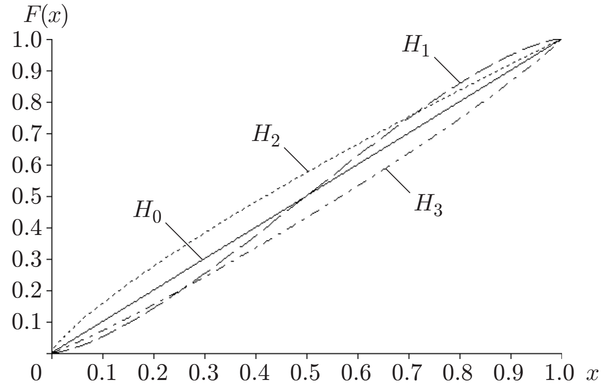

Fig. 1.

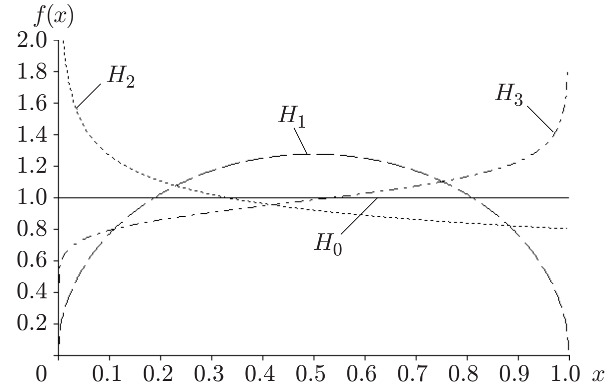

Fig. 2.

$$\hat{a} = x_{(1)} - \frac{x_{(n)} - x_{(1)}}{n-1}, \quad \hat{b} = x_{(n)} + \frac{x_{(n)} - x_{(1)}}{n-1}.$$

It is evident that testing the composite hypothesis of uniformity of the sample $X_1, X_2, \ldots, X_n$ on the interval $[\hat{a}, \hat{b}]$ obtained according to this sample is equivalent to testing the simple hypothesis of uniformity of part of the sample of size $n-2$ (a certain part of the ordered series $x_{(2)} < x_{(3)} < \ldots < x_{(n-1)}$) on the interval $[x_{(1)}, x_{(n)}]$ that corresponds to the scope of the sample. Any tests considered in this paper can be applied to this hypothesis if the required values of the order statistics are determined according to expressions $U_{i-1} = (x_{(i)} - x_{(1)})/(x_{(n)} - x_{(1)})$, $i = \overline{2, (n-1)}$, $U_0 = 0$, $U_{n-1} = 1$.

Note that generally the use of nonparametric goodness-of-fit tests for composite hypotheses with regard to different parametric models of probability distribution laws is seriously complicated due to the dependence of test statistic distribution on a number of factors [1]. Obviously, such a problem does not arise in the case of nonparametric tests used for composite hypotheses of uniformity.

Note. Sometimes testing composite hypotheses that a sample belongs to some parametric law comes down to testing a hypothesis of uniformity on the interval [0, 1]. Similarly, solving statistical analysis problems usually leads to testing uniformity on the interval [0, 1], e.g., when it comes to testing the adequacy of constructed wait-and-see models and reliability models. In these situations, using the above-described approach to testing complex hypotheses of uniformity would be incorrect.

The purpose of studies whose results are described in this paper is to show the features of application of tests for the hypotheses of uniformity of the sample under analysis, to compare the power of considered tests with regard to some close competing hypotheses, and to match the power of goodness-of-fit tests with the power of special tests solely focused on uniformity.

COMPETING HYPOTHESES UNDER CONSIDERATION

The results of testing hypotheses are associated with errors of two types: type I error is rejection of hypothesis H_0 when it is correct and type II error is acceptance (not rejection) of hypothesis H_0 as correct while the competing hypothesis H_1 is correct. Significance level α defines the probability of type I error.

Typically, a specific competing hypothesis is not considered when using tests for hypotheses. In this case, when hypotheses are tested for a type of law, it can be assumed that the competing hypothesis has the form H_1: $F(x) \neq F(x, \theta_0)$; here $F(x, \theta_0)$ corresponds to the tested hypothesis H_0. If the hypothesis H_1 has, for example, the form H_1 $F(x) = F_1(x, \theta)$, then setting the value of α for the used test also determines the type II error probability β. Test power is $1 - \beta$. Obviously, the higher the test power for a given value of α, the better it distinguishes the hypotheses H_0 and H_1.

Naturally, the most interesting is the ability of tests to distinguish close competing hypotheses. It is the analysis of close alternatives that helps find the subtle aspects that characterize the real properties of tests and to identify the key shortcomings and advantages.

In this work, the power of all the considered tests was investigated with respect to three competing hypotheses that correspond to an observed random variable that belongs to the family of beta distributions of type I with density function

$$f(x) = \frac{1}{\theta_2 \mathrm{B}(\theta_0, \theta_1)} \Big(\frac{x-\theta_3}{\theta_2}\Big)^{\theta_0 - 1} \Big(1 - \frac{x-\theta_3}{\theta_2}\Big)^{\theta_1 - 1},$$

where $\mathrm{B}(\theta_0, \theta_1) = \Gamma(\theta_0)\Gamma(\theta_1)/\Gamma(\theta_0 + \theta_1)$ is the beta function; $\theta_0, \theta_1 \in (0, \infty)$ is the shape parameter; $\theta_2 \in (0, \infty)$ is the scale parameter; $\theta_3 \in (-\infty, \infty)$ is the shift parameter; $x \in [0, \theta_2]$.

Let the beta distribution function of type I with specific parameter values be denoted as $B_I(\theta_0, \theta_1, \theta_2, \theta_3)$. Then the three considered competing hypotheses H_0, H_1, and H_2 that are quite close to H_3 take the following form:

$$H_1\colon F(x) = B_I(1.5; 1.5; 1; 0), \quad x \in [0, 1];$$

$$H_2\colon F(x) = B_I(0.8; 1; 1; 0), \quad x \in [0, 1];$$

$$H_3\colon F(x) = B_I(1.1; 0.9; 1; 0), \quad x \in [0, 1].$$

The probability distribution functions corresponding to the considered hypotheses are shown in Fig. 1, and the density distributions are in Fig. 2. As can be seen, the distribution functions of the laws for the hypotheses H_1, H_2, and H_3 are not so different from the uniformity distribution function, but the densities of the laws differ significantly.

It should be noted that competing hypothesis H_1 corresponds to the law whose distribution function intersects that of uniformity; in the case of H_2 and H_3, the law distribution functions lie above or below the function of the uniform law. The ability of tests to differentiate hypotheses H_0 and H_1, H_0 and H_2, or H_0 and H_3 turn out to be different.

It should be especially emphasized that the analysis of the test power with regard to hypothesis H_1 reveals the inability of most nonparametric goodness-of-fit tests to distinguish H_1 from H_0 with small sampling volumes n and low significance levels α, i.e., it shows that some tests are biased (the power $1 - \beta$ is smaller than α). Moreover, this drawback appears to be peculiar not only for most nonparametric goodness-of-fit tests, but also for most special uniformity tests.

APPLICATION OF GOODNESS-OF-FIT TESTS

Uniformity can be tested by using all the goodness-of-fit tests without exception.

When it comes to the Kolmogorov test [2], we test the hypothesis H_0 of uniformity of the sample by using a statistic with correction in the form [3]

$$S_K = \sqrt{n} D_n + 1/6\sqrt{n}, \tag{1}$$

where $D_n = \max(D_n^+, D_n^-)$, $D_n^+ = \max\limits_{1 \le i \le n} \{i/n - U_i\}$, $D_n^- = \max\limits_{1 \le i \le n} \{U_i - (i-1)/n\}$. If the simple testable hypothesis H_0 is valid, the limiting statistic distribution (1) is the Kolmogorov distribution with a function

$$K(s) = \sum_{k=-\infty}^{\infty} (-1)^k \mathrm{e}^{-2k^2 s^2}.$$

In the Kuiper test [4, 5], the distance between the empirical and theoretical law is a variable calculated from the expression $V_n = D_n^+ + D_n^-$.

The dependence of the distribution of the statistic $\sqrt{n} V_n$ used in the test on n can be reduced by applying the statistic versions [6] or [7], respectively:

$$V = V_n(\sqrt{n} + 0.155 + 0.24/\sqrt{n}), \tag{2}$$

$$V_n^{\mathrm{mod}} = \sqrt{n}\, V_n + 1/3\sqrt{n}. \tag{3}$$

When testing the simple hypothesis H_0, the limiting distribution of statistics (2) and (3) is the distribution [4, 5]

$$\mathrm{Kuiper}(s) = 1 - \sum_{m=1}^{\infty} 2(4m^2 s^2 - 1)\mathrm{e}^{-2m^2 s^2}.$$

After testing for uniformity, the statistic of the Cramer — von Mises — Smirnov test ω^2 takes the form

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \Big\{ U_i - \frac{2i-1}{2n} \Big\}^2. \tag{4}$$

If the simple hypothesis H_0 is valid, statistic (4) in the limit obeys the law with the distribution function $a1(s)$ of the form [3]

$$a1(s) = \frac{1}{\sqrt{2s}} \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)\sqrt{4j+1}}{\Gamma(1/2)\Gamma(j+1)} \exp\Big\{ -\frac{(4j+1)^2}{16s} \Big\} \Big\{ I_{-1/4}\Big[\frac{(4j+1)^2}{16s}\Big] - I_{1/4}\Big[\frac{(4j+1)^2}{16s}\Big] \Big\},$$

where $I_{-1/4}(\cdot)$ and $I_{1/4}(\cdot)$ are the modified Bessel functions of the form

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \quad |z| < \infty, \quad |\arg z| < \pi.$$

In the case of testing uniformity, the Watson test statistic [8, 9] is given by the following expression:

$$U_n^2 = \sum_{i=1}^{n} \Big(U_i - \frac{i-1/2}{n} \Big)^2 - n\Big(\frac{1}{n} \sum_{i=1}^{n} U_i - \frac{1}{2} \Big)^2 + \frac{1}{12n}. \tag{5}$$

If H_0 is valid, statistic (5) in the limit obeys the law with the distribution function [8, 9]

$$\mathrm{Watson}(s) = 1 - 2 \sum_{m=1}^{\infty} (-1)^{m-1} \mathrm{e}^{-2m^2\pi^2 s}.$$

In the case of testing uniformity, the statistic of the Anderson — Darling goodness-of-fit test Ω^2 [10, 11] takes the form

$$S_\Omega = -n - 2\sum_{i=1}^{n} \Big\{ \frac{2i-1}{2n} \ln U_i + \Big(1 - \frac{2i-1}{2n}\Big) \ln(1-U_i) \Big\}. \tag{6}$$

If the simple testable hypothesis H_0 is valid, statistic (6) in the limit obeys the law with the distribution function [3]

$$a2(s) = \frac{\sqrt{2\pi}}{s} \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma(j+1/2)(4j+1)}{\Gamma(1/2)\Gamma(j+1)} \exp\Big\{ -\frac{(4j+1)^2\pi^2}{8s} \Big\} \times$$
$$\times \int_0^\infty \exp\Big\{ \frac{s}{8(y^2+1)} - \frac{(4j+1)^2\pi^2 y^2}{8s} \Big\} dy.$$

Zhang [12] describes the nonparametric goodness-of-fit tests whose statistics take the following form in the case of testing the simple hypothesis of uniformity of the analyzed sample on the interval [0, 1]:

$$Z_A = -\sum_{i=1}^{n} \Big[\frac{\ln U_i}{n-i+1/2} + \frac{\ln\{1-U_i\}}{i-1/2} \Big], \tag{7}$$

$$Z_C = \sum_{i=1}^{n} \Big[\ln\Big\{ \frac{U_i^{-1}-1}{(n-1/2)/(i-3/4)-1} \Big\} \Big]^2, \tag{8}$$

$$Z_K = \max_{1 \le i \le n} \Big(\Big(i - \frac{1}{2}\Big) \ln\Big\{ \frac{i-1/2}{nU_i} \Big\} + \Big(n - i + \frac{1}{2}\Big) \ln\Big[\frac{n-i+1/2}{n(1-U_i)} \Big] \Big). \tag{9}$$

The Zhang tests are the development of the Anderson — Darling, Cramer — von Mises — Smirnov, and Kolmogorov tests, respectively. Application of the tests with statistics (7)–(9) is complicated by the strong dependence of the statistic distributions on the sample volume n.

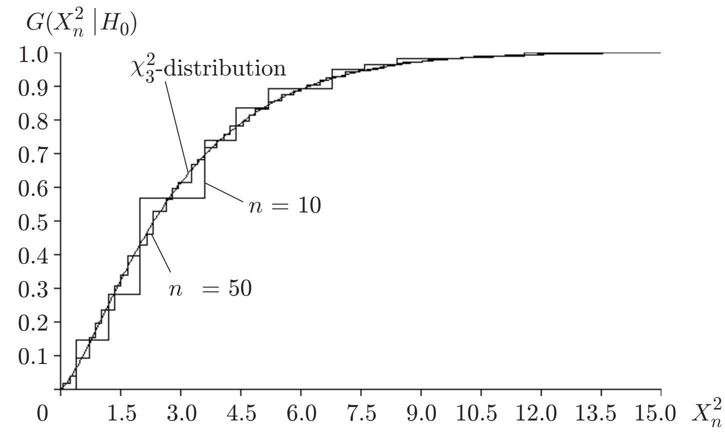

Fig. 3.

One of the factors supporting the application of nonparametric goodness-of-fit tests for uniformity is the known limiting statistic distribution that can be used to computate the achieved significance level usually for the sample volumes $n \geq 25$. Exceptions are the Zhang tests where the statistic distributions depend on n because of which hypotheses have to be tested with the help of the tables of percentage points.

When using the Pearson test χ^2, the range of the random variable is divided into k non-overlapping intervals by boundary points and the number of observations n_i fallen into the ith interval and the probability P_i of falling into the interval are counted. In this case, $n = \sum_{i=1}^{k} n_i$, $\sum_{i=1}^{k} P_i(\theta) = 1$. The test statistic is calculated from the expression

$$X_n^2 = n \sum_{i=1}^{k} \frac{(n_i/n - P_i(\theta))^2}{P_i(\theta)}. \tag{10}$$

If the simple testable hypothesis H_0 is valid, statistic (10) asymptotically obeys the χ^2_{k-1} distribution.

It is noteworthy that statistic (10) is a discrete random variable and its actual distribution $G(X_n^2\,|\,H_0)$ may differ significantly from the asymptotic χ^2_{k-1} distribution in the case of validity of the tested hypothesis H_0 and bounded n. For example, Fig. 3 shows the dependence of the test statistic distribution (if H_0) is valid) on the sample volume n with the range divided into intervals of equal probability (number of intervals $k = 4$). The use of equiprobable grouping in testing uniformity is quite logical. As the actual statistic distribution is discrete, the estimate of the achieved significance level calculated in accordance with the χ^2_{k-1} distribution is erroneous.

As the number of discrete intervals increases, the discrete distribution of the statistic rapidly converges to the continuous χ^2_{k-1} distribution, but this does not imply an increase in the test power. The test power χ^2 depends on the considered alternative (laws that correspond to testable and competing hypotheses) and on the method of dividing into intervals and their number [13, 14].

For example, Fig. 4 shows the dependence of the Pearson test power relative to H_1, H_2, and H_3 on the number of intervals k (division of the range into equiprobable grouping intervals). The results are shown for the sample volume $n = 100$ and the given probability of the type I error (significance level) $\alpha = 0.1$.

GOODNESS-OF-FIT TEST POWER ANALYSIS

Based on the power estimates derived in [15], the goodness-of-fit tests used for uniformity can be arranged in the order of descending power with regard to the competing hypothesis H_1 (in the case of intersection of the distribution law functions corresponding to H_0 and H_1)):

$$Z_A \text{ Zhang} \succ Z_C \text{ Zhang} \succ \text{Watson} \succ \text{Kuiper} \succ Z_K \text{ Zhang} \succ \chi^2 \text{ Pearson} \succ$$
$$\succ \text{Anderson — Darling} \succ \text{Cramer — von Mises — Smirnov} \succ \text{Kolmogorov}.$$

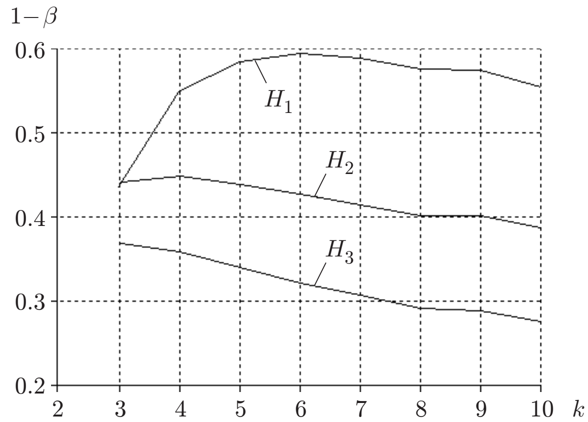

Fig. 4.

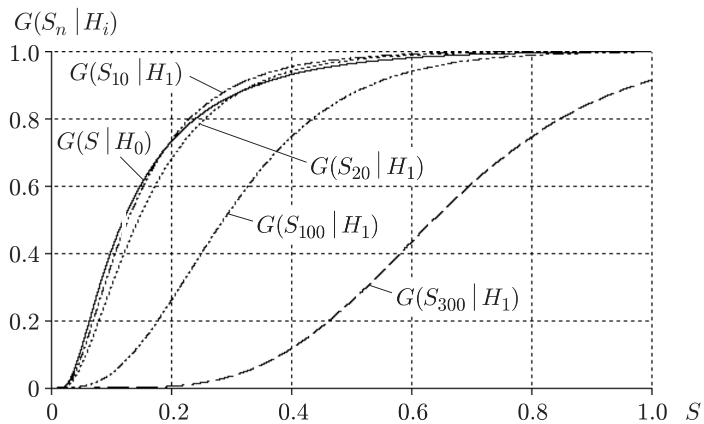

Fig. 5.

It should be noted that the bias of the Kolmogorov, Cramer — von Mises — Smirnov, and Anderson — Darling tests (relative to hypotheses such as H_1)) for small sample volumes n and low significance level α is mentioned for the first time in [15]. The Zhang tests with statistics Z_K and Z_C are biased as well, and the test with statistic Z_A is biased to a smaller extent. The fact that the goodness-of-fit tests with statistics Z_A and Z_C are biased relative to some competing hypotheses is found when testing the hypotheses that the samples belong to the normal law [16].

The bias is illustrated in Fig. 5 that shows the distribution of the Cramer — von Mises test statistic $G(S \,|\, H_0)$ with validity of the testable hypothesis H_0 and the distribution $G(S_n \,|\, H_1)$ of this statistic with validity of the competing hypothesis H_1 (sample volume $n = 10, 20, 100, 300$). Obviously, the statistic distribution $G(S_n \,|\, H_1)$ for $n = 10, 20$ intersect $G(S \,|\, H_0)$, which explains why the power $1 - \beta$ is lower than α.

In the figure, the distribution $G(S \,|\, H_0)$ is shown only for $n = 10$. If $n \geq 20$, the distributions $G(S_n \,|\, H_0)$ are not visually different from $G(S_{10} \,|\, H_0)$ and practically coincide with the limiting distribution $a1(s)$ of the Cramer — von Mises — Smirnov test statistic when testing simple hypotheses.

The considered set of goodness-of-fit tests can be arranged by descending power relative to the competing hypothesis H_2 (without intersection of the distribution functions that correspond to H_0 and H_2)): It is located in a different order:

$$\text{Anderson — Darling} \succ Z_C \text{ Zhang} \succ \text{Cramer — von Mises — Smirnov} \succ$$
$$\succ Z_A \text{ Zhang} \approx Z_K \text{ Zhang} \succ \text{Kolmogorov} succ \chi^2 \text{ Pearson} \succ \text{Kuiper} \succ \text{Watson}.$$

The same tests listed by the order of descending power relative to the competing hypothesis H_3 (also without intersection of the law distribution functions corresponding to H_0 and H_3)) are arranged approximately in the same order:

$$\text{Anderson — Darling} \succ \text{Cramer — von Mises — Smirnov} \succ Z_C \text{ Zhang} \succ$$
$$\succ Z_A \text{ Zhang} \succ \text{Kolmogorov} \succ Z_K \text{ Zhang} \succ \chi^2 \text{ Pearson} \succ \text{Kuiper} \succ \text{Watson}.$$

Note [15] that the Kuiper and Watson tests not known for being biased are greater in power as compared to the Kolmogorov — von Mises — Smirnov tests in the case of an alternative case with intersection of the distribution laws (e. g., the situation with H_0 and H_1)) and are essentially smaller in power in the case of an alternative without intersection (the situation with H_2 and H_3)).

In general, the preference should be given to the Anderson — Darling and Zhang tests with statistics Z_C and Z_A and to the Cramer — von Mises — Smirnov tests. However, it should be taken into account that there might be a situation when some competing laws would not be distinguished by tests with small n and α.

SPECIAL TESTS FOR UNIFORMITY

There are three groups in the subset of specific tests of uniformity. The test statistics of the first group provide the use of the differences of successive values of the ordered series $U_i - U_{i-1}$, where $i = \overline{1, (n+1)}$, $U_0 = 0$, $U_{n+1} = 1$.

The second group includes the tests using the differences of estimates of order statistics obtained from the analyzed sample and, for example, from the mathematical expectations of these order statistics.

The third group is the so-called entropy tests based on various entropy estimates.

The first group of tests using the difference of the elements of an ordered series include the Sherman [17, 18], Kimball [19], Moran 1 [20], Moran 2 [21], and Young [22] tests, as well as the Cressy tests with expressions of the statistics $S_n^{(m)}$ and $L_n^{(m)}$ [15] clarified as compared to [23], the Pardo [24] tests, and the Schwartz [25] tests.

The second group where the deviations of order statistics from their mathematical expectations are considered (on medians, etc.) include the Hegazy — Green tests with statistics T_1 and T_2 [26], the Frosini [27], Cheng — Spiring [28], Greenwood [29], Greenwood — Quesenberry — Miller [30] tests, as well as the Neyman — Barton tests with statistics N_2, N_3, and N_4 [31].

The third group consists of the Dudewics — van der Meulen entropy tests [32] and two versions in whose statistics other entropy estimates are used [33].

In this work that is the development of [34, 35], we used statistical simulation methods [14] to study the statistic distribution of all the above-mentioned tests and expand the tables of percentage points. We tested how well the distribution of normalized statistics are described by the corresponding asymptotic laws (in those cases where the literature had references to such results) and studied the test power relative to various competing hypotheses, particularly H_1, H_2, and H_3 [15]. Unfortunately, most of the considered tests (the Sherman, Kimball, Moran 1, and Moran 2 tests; one of the Cressie tests; the Hegazy — Green, Frosini, Young, Greenwood, Greenwood — Quesenberry — Miller, and Neyman — Barton tests) are biased relative to the competing hypothesis H_1.

Other practically general disadvantages of the most special tests are the dependence of the statistic distribution on the sample volume n and the need to used the tables of critical values (percentage points). The exceptions are the Neyman — Barton tests where the distributions of three statistics for $n > 20$ are well approximated by the χ^2_2-, χ^2_3-, and χ^2_4 distributions, as well as the Moran 2 and Young tests. However, the approximations by the χ^2 distribution and by the normal law of the corresponding versions of the Moran 2 test statistics are essentially different from the actual distributions of these versions, and the test itself has a very low power relative to the competing hypotheses H_1, H_2, and H_3. The normalized Young test statistic on the contrary is well approximated by the standard normal law, but the test has such a low power that recommending its usage would be unreasonable.

Expressions for the statistics of the considered specific tests of uniformity are shown in Table 1.

All the tests of uniformity in the columns of Table 2 are ordered by descending power relative to the competing hypotheses H_1, H_2, and H_3 (by the power $1 - \beta$ manifested for $n = 100$ and for the significance level $\alpha = 0.1$)).

In the column with ordering by power relative to hypothesis H_1, the tests that are obviously biased for small n relative to hypothesis H_1 are marked in bold-faced type. The Neyman — Barton tests with

Table 1. Statistics of special tests for uniformity

No.	Test	Statistics
1	Sherman	$\omega_n = \frac{1}{2}\sum_{i=1}^{n+1}\left\lvert U_i - U_{i-1} - \frac{1}{n+1}\right\rvert$
2	Kimball	$A = \sum_{i=1}^{n+1}\left(U_i - U_{i-1} - \frac{1}{n+1}\right)^2$
3	Moran 1, 2	$B = \sum_{i=1}^{n+1}(U_i - U_{i-1})^2$; $M_n = -\sum_{i=1}^{n+1}\ln[(n+1)(U_i - U_{i-1})]$
4	Young	$M = \sum_{i=1}^{n}\min(D_i, D_{i+1})$, $D_1 = U_1$, $D_i = U_i - U_{i-1}$, $D_{n+1} = 1 - U_n$
5	Greenwood	$G = (n+1)\sum_{i=1}^{n+1}(U_i - U_{i-1})^2$
6	Greenwood — Quesenberry — Miller	$Q = \sum_{i=1}^{n+1}(U_i - U_{i-1})^2 + \sum_{i=1}^{n}(U_{i+1} - U_i)(U_i - U_{i-1})$
7	Swartz	$A_n^* = \frac{n}{2}\sum_{i=1}^{n}\left(\frac{U_{i+1} - U_{i-1}}{2} - \frac{1}{n}\right)^2$, $U_0 = -U_1$, $U_{n+1} = 2 - U_n$
8	Pardo	$E_{m,n} = \frac{1}{n}\sum_{i=1}^{n}\frac{2m}{n(U_{i+m} - U_{i-m})}$
9	Cressie 1, 2	$S_n^{(m)} = \sum_{i=0}^{n+1-m}\left(U_{i+m} - U_i - \frac{m}{n+1}\right)^2$; $L_n^{(m)} = -\sum_{i=0}^{n+1-m}\ln\left[\frac{n+1}{m}(U_{i+m} - U_i)\right]$
10	Cheng — Spiring	$W_p = \left[(U_n - U_1)\frac{n+1}{n-1}\right]^2 \Big/ \sum_{i=1}^{n}(U_i - \bar{U})^2$
11	Hegazy — Green T_1, T_1^*	$T_1 = \frac{1}{n}\sum_{i=1}^{n}\left\lvert U_i - \frac{i}{n+1}\right\rvert$; $T_1^* = \frac{1}{n}\sum_{i=1}^{n}\left\lvert U_i - \frac{i-1}{n-1}\right\rvert$
12	Hegazy — Green T_2, T_2^*	$T_2 = \frac{1}{n}\sum_{i=1}^{n}\left(U_i - \frac{i}{n+1}\right)^2$; $T_2^* = \frac{1}{n}\sum_{i=1}^{n}\left(U_i - \frac{i-1}{n-1}\right)^2$
13	Frosini	$B_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\lvert U_i - \frac{i-0.5}{n}\right\rvert$
14	Neyman — Barton N_2	$N_2 = \sum_{j=1}^{2}V_j^2$, $V_j = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_j(U_i - 0.5)$, $\pi_1(y) = 2\sqrt{3}y$, $\pi_2(y) = \sqrt{5}(6y^2 - 0.5)$
15	Neyman — Barton N_3	$N_3 = \sum_{j=1}^{3}V_j^2$, $V_j = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_j(U_i - 0.5)$, $\pi_3(y) = \sqrt{7}(20y^3 - 3y)$
16	Neyman — Barton N_4	$N_4 = \sum_{j=1}^{4}V_j^2$, $V_j = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_j(U_i - 0.5)$, $\pi_4(y) = 3(70y^4 - 15y^2 + 0.375)$
17	Dudewics — van der Meulen	$H(m,n) = -\frac{1}{n}\sum_{i=1}^{n}\ln\left\{\frac{n}{2m}(U_{i+m} - U_{i-m})\right\}$, m — unit and $m \le n/2$; if $i + m \ge n$, then $U_{i+m} = U_n$; if $i - m \le 1$, then $U_{i-m} = U_1$
18	Entropy test version 1	$HY_1 = -\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{U_{i+m} - U_{i-m}}{\hat{F}(U_{i+m}) - \hat{F}(U_{i-m})}\right)$, $\hat{F}(U_i) = \frac{n-1}{n(n+1)}\left(i + \frac{1}{n-1} + \frac{U_i - U_{i-1}}{U_{i+1} - U_{i-1}}\right)$, $i = \overline{2, (n-1)}$, $\hat{F}(U_1) = 1 - \hat{F}(U_n) = 1/(n+1)$
19	Entropy test version 2	$HY_2 = -\sum_{i=1}^{n}\ln\left(\frac{U_{i+m} - U_{i-m}}{\hat{F}(U_{i+m}) - \hat{F}(U_{i-m})}\right)\left(\frac{\hat{F}(U_{i+m}) - \hat{F}(U_{i-m})}{\sum_{j=1}^{n}(\hat{F}(U_{j+m}) - \hat{F}(U_{j-m}))}\right)$

Table 2. Order of uniformity tests by power

No.	Relative to H_1	$1-\beta$	Relative to H_2	$1-\beta$	Relative to H_3	$1-\beta$
1	Entropy test version 2	0.883	Anderson — Darling	0.648	Anderson — Darling	0.526
2	**Zhang Z_A**	0.850	Hegazy — Green T_1	0.610	Hegazy — Green T_1	0.522
3	**Neyman — Barton N_2**	0.837	Zhang Z_C	0.606	Frosini	0.522
4	Cressie 2	0.820	Frosini	0.603	Hegazy — Green T_1^*	0.520
5	**Zhang Z_C**	0.819	Hegazy — Green T_2	0.602	Hegazy — Green T_2	0.508
6	Dudewics — van der Meulen	0.790	Neyman — Barton N_2	0.597	Cramer — von Mises — Smirnov	0.507
7	Entropy test version 1	0.789	Cramer — von Mises — Smirnov	0.595	Hegazy — Green T_2^*	0.506
8	Watson	0.779	Hegazy — Green T_1^*	0.595	Zhang Z_C	0.463
9	**Neyman — Barton N_3**	0.766	Z_K	0.590	Zhang Z_A	0.459
10	**Neyman — Barton N_4**	0.739	Hegazy — Green T_2^*	0.585	Kolmogorov	0.450
11	Kuiper	0.736	Neyman — Barton N_3	0.577	Neyman — Barton N_2	0.447
12	Cheng — Spiring	0.722	Zhang Z_A	0.574	Zhang Z_K	0.438
13	**Zhang Z_K**	0.617	Neyman — Barton N_4	0.557	Neyman — Barton N_3	0.416
14	χ^2 Pearson	0.593	Kolmogorov	0.542	Neyman — Barton N_4	0.381
15	Swartz	0.583	Pardo	0.463	χ^2 Pearson	0.374
16	**Anderson — Darling**	0.505	χ^2 Pearson	0.448	Pardo	0.291
17	**Hegazy — Green T_1^***	0.443	Kuiper	0.364	Dudewics — van der Meulen	0.275
18	**Hegazy — Green T_2^***	0.409	Zhang Z_A	0.356	Entropy test version 1	0.275
19	Pardo	0.408	**Entropy test version 1**	0.328	Entropy test version 2	0.267
20	**Frosini**	0.384	**Dudewics — van der Meulen**	0.327	Watson	0.257

Table 2 (cont.)

No.	Relatively to H_1	$1-\beta$	Relatively to H_2	$1-\beta$	Relatively to H_3	$1-\beta$
21	**Cramer — von Mises — Smirnov**	0.358	Cressie 1	0.314	Kuiper	0.254
22	**Hegazy — Green T_1**	0.322	**Entropy test version 2**	0.266	Cressie 2	0.226
23	**Kolmogorov**	0.322	Greenwood — Quesenberry — Miller	0.244	Cressie 1	0.218
24	**Hegazy — Green T_2**	0.308	Schwartz	0.226	Schwartz	0.206
25	**Greenwood — Quesenberry — Miller**	0.290	Cressie 2	0.217	Greenwood — Quesenberry — Miller	0.186
26	**Kimball**	0.279	Sherman	0.204	Kimball	0.165
27	**Moran 1**	0.279	Kimball	0.201	Moran 1	0.165
28	**Greenwood**	0.279	Moran 1	0.201	Greenwood	0.165
29	**Sherman**	0.215	Greenwood	0.201	Sherman	0.154
30	**Cressie 1**	0.187	Moran 2	0.193	Moran 2	0.143
31	**Moran 2**	0.187	Cheng — Spiring	0.168	Cheng — Spiring	0.106
32	**Young**	0.115	Young	0.108	Young	0.104

statistics N_2 and N_3 are biased relative to H_1 to a smaller extent. This disadvantage is not observed only for some tests: the Kuiper — Watson test and the Dudewics — van der Meulen entropy test with its versions, as well as the Cheng — Spiring, Swartz, Pardo, Cressie 2, and χ^2 Pearson tests.

All modifications of the tests that use various estimates of entropy [32, 33] as statistics demonstrate a high power relative to the competing hypothesis H_1. At the same time, the estimates of power these tests are more modest relative to hypotheses H_2 and H_3. Only these tests for small n manifest bias relative to hypothesis H_2. It should be noted that power of these tests and of the Cressie — Pardo tests depends on the "window size" choice m [15].

The Neyman — Barton test with statistic N_2 shows a high power with respect to H_1 and comparatively great results relative to H_2 and H_3.

A consistently good ability to distinguish between competing hypotheses and uniformity is demonstrated by the Hegazy — Green and Frosini tests.

Low power is demonstrated by the tests where modules or squares of differences $U_i - U_{i-1}$ of values of successive order statistics (the Sherman, Kimball, Moran, Greenwood, and Greenwood — Quesenberry — Miller tests).

The Cheng — Spiring test demonstrates comparatively high power relatively to H_1, but it shows low power relatively to H_2 and H_3. The power is especially low for all three considered hypotheses because of the Young test [22], which indicates a very unsuccessful attempt to use the corresponding statistic in the test for uniformity.

Based on the study of the properties of the set of tests used for uniformity, a manual is prepared [15].

CONCLUSION

If the hypothesis that a sample under analysis belongs to a certain distribution law is tested by developing a set of special tests, then this set would probably include tests whose application is more preferable with limited sample volumes due to obviously higher power as compared, for example, with general goodness-of-fit

Table 3. Minimal sample volumes $\boldsymbol{n}$, required to distinguish hypotheses $\boldsymbol{H_0}$ and $\boldsymbol{H_i}$ with given probabilities of type I and II $\boldsymbol{\alpha = 0.1}$ and $\boldsymbol{\beta = 0.1}$

No.	Test	n		
		from H_1	from H_2	from H_3
1	Anderson — Darling	200	210	295
2	Cramer — von Mises — Smirnov	274	245	315
3	Zhang Z_C	127	240	353
4	Zhang Z_A	120	253	355
5	Zhang Z_K	180	250	390
6	Watson	145	455	730
7	Kuiper	156	410	595
8	Kolmogorov	335	275	370
9	χ^2 Pearson	210	335	435
10	Neyman — Barton N_2	122	235	350
11	Dudewics — van der Meulen	150	485	805
12	Hegazy — Green T_1	277	235	300
13	Frosini	264	240	300
14	Pardo	280	375	830
15	Swartz	300	2600	4000
16	Sherman, Kimball, Moran 2	617	4350	6950

tests. In testing uniformity, this difference is not observed when it comes to nonparametric goodness-of-fit tests: the Zhang tests with statistics Z_A and Z_C and the Anderson — Darling tests demonstrate very poor results.

It follows from the analysis of the properties of the entire set of tests that can be used for hypotheses of uniformity of a sample that the correct use of any single test for the formation of a reliable statistical conclusion may often be insufficient. For greater objectivity of statistical conclusions, it is preferable to use some number of tests having certain advantages. The use of not one, but a set of uniformity tests supported by various measures of deviation of empirical distribution from theoretical distribution improves the quality of statistical conclusions.

In the analysis of the measurement results, there might be a problem concerning the sample volume sufficient for correctly using a certain test for a hypothesis. This problem can be solved only in terms of probabilities of type I and II errors after clarifying the information about the competing hypothesis (competing law). For example, for some of the tests considered in this paper and hypotheses H_1, H_2, and H_3, Table 3 shows sample volumes n required for a situation where the probability of the type II error does not exceed $\alpha = 0.1$ for the specified probability of the type I error $\beta = 0.1$.

Making a decision about the results of testing hypothesis H_0 on the basis of the achieved significance level is always more justified than in the case of comparing the obtained value of the statistic with a given critical value derived from the corresponding table of percentage points. In the latter case, it remains unclear how different is the true distribution to which the analyzed sample belongs (and which always remains unknown) from uniformity.

Unfortunately, the distributions of the majority of special tests for uniformity are essentially dependent on the sample volume, so it is necessary to rely on the tables of percentage points. A similar problem arises with the use of the Zhang nonparametric goodness-of-fit tests with statistics Z_A, Z_C, and Z_K whose distributions depend on n.

What can be done to improve the quality of statistical conclusions? The use of computer data analysis and the study of the unknown distribution of applied test statistics (for a given sample volume n) in real-time testing (online) of a hypothesis [1, 16, 17, 36]. For example, one can interactively explore the unknown distribution of the statistic of any test of uniformity that depends on the sample volume for the value of n which corresponds to the sample under analysis and use the empirical distribution of the statistic to estimate the achieved significance level.

This work was supported by the Ministry of Education and Science of the Russian Federation (Grant No. 2.541.2014/K).

REFERENCES

1. B. Yu. Lemeshko, *Nonparametric Goodness-of-Fit Tests: Application Guide* (INFRA-M, 2014) [in Russian].
2. A. N. Kolmogoroff, "Sulla determinazione empirica di una legge di distribuzione," G. Ist. Ital. attuar. **4** (1), 83–91 (1933).
3. L. N. Bol'shev and N. V. Smirnov, *Tables of Mathematical Statistics* (Nauka, Moscow, 1983) [in Russian].
4. N. H. Kuiper, "Tests Concerning Random Points on a Circle," Proc. Koninklijke Nederlandse Akademie van Wetenschappen. Ser. A. **63**, 38–47 (1960).
5. M. A. Stephens, "Use of Kolmogorov — Smirnov, Cramer — von Mises and Related Statistics — without Extensive Table," J. Royal. Stat. Soc. Ser. B **32** (1), 115–122 (1970).
6. M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," J. Amer. Stat. Assoc. **69** (347), 730–737 (1974).
7. B. Yu. Lemeshko and A. A. Gorbunova, "Application and Power of the Nonparametric Kuiper, Watson, and Zhang Tests of Goodness-of-Fit," Izmeritel'naya Tekhnika **56** (5), 3–9 (2013) [Measur. Techn. **56** (5), 465–475 (2013)].
8. G. S. Watson, "Goodness-of-Fit Tests on a Circle. I," Biometrika **48** (1–2), 109–114 (1961).
9. G. S. Watson, "Goodness-of-Fit Tests on a Circle. II," Biometrika **49** (1–2), 57–63 (1962).
10. T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes," Ann. Math. Stat. **23** (2), 193–212 (1952).
11. T. W. Anderson and D. A. Darling, "A Test of Goodness of Fit," J. Amer. Stat. Assoc. **49** (268), 765–769 (1954).
12. J. Zhang, *Powerful Goodness-of-Fit and Multi-Sample Tests: PhD Thesis* (York University, Toronto, 2001), http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf.
13. B. Yu. Lemeshko and E. V. Chimitova, "Maximizing the Strength of χ^2 Tests," Dokl. Sib. Otd. Akad. Nauk Vyssh. Shk., No. 2, 53–61 (2000).
14. B. Yu. Lemeshko, S. B. Lemeshko, S. N. Postovalov, and E. V. Chimitova, *Statistical Data Analysis, Simulation, and Study of the Laws of Probability. Computer Approach* (Izd. Nov. Gos. Tekh. Univ., Novosibirsk, 2011) [in Russian].
15. B. Yu. Lemeshko and P. Yu. Blinov, *Criteria for Testing the Distribution Deviation from the Uniform Law: Application Guide* (INFRA-M, Moscow, 2016) [in Russian].
16. B. Yu. Lemeshko, *Criteria for Testing the Distribution Deviation from the Normal Law: Application Guide* (INFRA-M, Moscow, 2015) [in Russian].
17. B. Sherman, "A Random Variable Related to the Spacing of Sample Values," Ann. Math. Stat. **21** (3), 339–361 (1950).
18. B. Sherman, "Percentiles of the Statistic," Ann. Math. Stat. **28** (1), 257–261 (1957).
19. B. F. Kimball, "Some Basic Theorems for Developing Tests of Fit for the Case of the Non-parametric Probability Distribution Function," Ann. Math. Stat. **18** (1), 540–548 (1947).
20. P. A. P. Moran, "The Random Division of an Intervals," J. Royal. Stat. Soc. Ser. B **9** (1), 92–98 (1947).
21. P. A. P. Moran, "The Random Division of an Intervals II," J. Royal. Stat. Soc. Ser. B **13** (2), 147–150 (1951).
22. D. L. Young, "The Linear Nearest Neighbour Statistic," Biometrika **69** (2), 477–480 (1982).
23. N. Cressie, "An Optimal Statistic Based on Higher Order Gaps," Biometrika **66** (3), 619–627 (1979).
24. M. C. Pardo, "A Test for Uniformity Based on Informational Energy," Statist. Papers **44** (4), 521–534 (2003).
25. T. Swartz, "Goodness-of-Fit Tests Using Kullback — Leibler Information," Communs Statist. Theory and Methods **21** (3), 711–729 (1992).
26. Y. A. Hegazy and J. R. S. Green, "Some New Goodness-of-Fit Tests Using Order Statistics," Appl. Statist. **24** (3), 299–308 (1975).
27. B. V. Frosini, "On the Distribution and Power of Goodness-of-Fit Statistic with Parametric and Nonparametric Applications," *Goodness-of-Fit*, Ed. by P. Revesz, K. Sarkadi, and P. K. Sen (North-Holland Publ. Comp., Amsterdam, Oxford, New York, 1987).

28. S. W. Cheng and F. A. Spiring, "A Test to Identify the Uniform Distribution, with Applications to Probability Plotting and Other Distributions," IEEE Trans. Reliability **R-36** (1), 98–105 (1987).

29. M. Greenwood, "The Statistical Study of Infection Disease," J. Royal. Stat. Soc. Ser. A **109** (2), 85–110 (1946).

30. C. P. Quesenberry and F. L. Miller, "Power Studies of Some Tests for Uniformity," J. Statist. Computat. Simulat. **5** (3), 169–191 (1977).

31. J. Neyman, "Smooth Tests for Goodness-of-Fit," Scandinavisk Aktuarietidskrift **20** (3–4), 149–199 (1937).

32. E. J. Dudewics and E. C. van der Meulen, "Entropy-Based Test of Uniformity," J. Amer. Stat. Assoc. **76** (376), 967–974 (1981).

33. E. Zamanzade, "Testing Uniformity Based on New Entropy Estimators," J. Statist. Computat. Simulat. **82** (11), 1701–1713 (2014).

34. P. Yu. Blinov and B. Yu. Lemeshko, "A Review of the Properties of Tests for Uniformity," in *Proc. of the 12th Iintern. Conf. Actual Problems of Electronics Instrument Engineering (APEIE)* (NSTU Publisher, Novosibirsk, 2014), Vol. 1, pp. 540–547.

35. P. Yu. Blinov and B. Yu. Lemeshko, "The Strength of Tests for the Hypotheses of Uniformity of Samples," in *Proceedings of the Russian Scientific and Engineering Conference "Processing of Information Signals and Mathematical Modeling"* (Siberian State University of Telecommunications and Information Sciences, Novosibirsk, 2013) [in Russian].

36. B. Yu. Lemeshko, A. A. Gorbunov, S. B. Lemeshko, and A. P. Rogozhnikov, " Solving Problems of Using Some Nonparametric Goodness-of-Fit Tests," Avtometriya **50** (1), 26–43 (2014) [Optoelectron., Instrum. Data Process. **50** (1), 21–35 (2014)].