Measurement Techniques, Vol. 56, No. 5, August, 2013

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

APPLICATION AND POWER OF THE NONPARAMETRIC KUIPER, WATSON, AND ZHANG TESTS OF GOODNESS-OF-FIT

B. Yu. Lemeshko and A. A. Gorbunova UDC 519.24

The power of the Kuiper, Watson, and three Zhang tests of goodness-of-fit with different statistics are estimated relative to some pairs of competing distributions for testing simple and composite hypotheses. The powers of these tests are compared with those of the Kolmogorov, Cramer–Mises–Smirnov, and Anderson–Darling tests.
Keywords: *nonparametric tests of goodness-of-fit, Kuiper, Watson, Zhang tests, power, simple and composite hypotheses.*

The purpose of this paper is to draw attention to some nonparametric tests of goodness-of-fit that are rarely, if ever, used by Russian specialists for the statistical analysis of experimental data. These include, for example, the Kuiper [1] and Watson [2, 3] tests because they appear to be extensions and analogs of the Kolmogorov and Cramer–Mises–Smirnov tests and, therefore, have no clear advantages over the latter, as well as others (proposed in [4–7]) because of limited source availability and a lack of independent recommendations for their use.

The testing of goodness-of-fit differs for simple and composite hypotheses. A simple testable hypothesis has the form H_0: $F(x) = F(x, \theta)$, where $F(x, \theta)$ is a known theoretical probability distribution function with a known scalar or vector parameter θ. For testing simple hypotheses, nonparametric tests of goodness-of-fit are distribution-free; i.e., when the tested hypothesis is true, the distribution of the test statistics, $G(S|H_0)$, is independent of the form of the distribution $F(x, \theta)$ with which the goodness-of-fit is being tested.

When testing composite hypotheses of the form H_0: $F(x) \in \{F(x, \theta), \theta \in \Theta\}$, where an estimate $\hat{\theta}$ of the scalar or vector parameter of the distribution $F(x, \theta)$ is calculated over a given sample, nonparametric tests of goodness-of-fit lose the distribution-free property. Then the conditional distribution of the statistics $G(S|H_0)$ depends on a number of factors: the form of the observed distribution $F(x, \theta)$ corresponding to a true hypothesis H_0 to be tested; the type and number of parameters to be estimated; in some cases on the specific value of the parameter (e.g., in the case of families of gamma- and beta-distributions); and the method for estimating the parameters. The differences in the distributions for a given statistic during testing of simple and composite hypotheses are so large that this cannot be ignored in any case.

We emphasize that the available classical results for the tests discussed in this paper (distributions of the test statistics or tables of percentage points) apply only to testing of simple hypotheses.

The Kuiper Test. Kuiper proposed [1] an extended Kolmogorov type test for testing the hypothesis that a random sample belongs to a distribution law with a continuous distribution function $F(x, \theta)$. The statistic V_n for the test is defined as

$$V_n = \sup_{-\infty<x<\infty} \{F_n(x) - F(x,\theta)\} - \inf_{-\infty<x<\infty} \{F_n(x) - F(x,\theta)\},$$

Novosibirsk State Technical University, Novosibirsk, Russia; e-mail: lemeshko@fpm.ami.nstu.ru. Translated from Izmeritel'naya Tekhnika, No. 5, pp. 3–9, May, 2013. Original article submitted January 14, 2013.

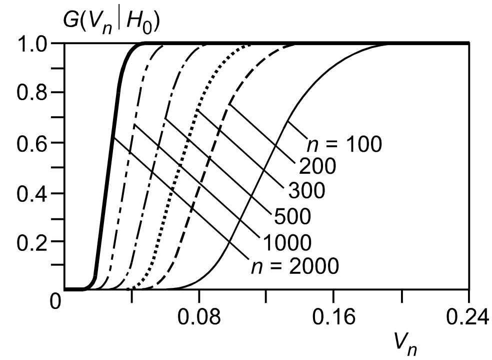

Fig. 1. Distributions of the (1) statistic for the Kuiper test $G(V_n|H_0)$ as a function of the sample size n for testing a simple hypothesis.

where $F_n(x)$ is the empirical distribution function, and is used in the form

$$V_n = D_n^+ + D_n^-, \tag{1}$$

where

$$D_n^+ = \max\{i/n - F(x_i, \theta)\}; \quad D_n^- = \max\{F(x_i, \theta) - (i-1)/n\},$$

with $i = \overline{1, n}$, n is the volume of the sample, and here and in the following the x_i are the elements of a variation series constructed from the sample (in increasing order).

The major deficiency of the test employing the (1) statistic is that its distribution depends on the sample size n. This kind of dependence of the distributions $G(V_n|H_0)$ of the statistic when the simple tested hypothesis H_0 is true is illustrated by the simulations shown in Fig. 1.

Tables of the percentage points for tests of simple hypothesis using the test statistic of (1) can be found in [8, 9]. As a limiting distribution $G(\sqrt{n}V_n|H_0)$ for the statistic $\sqrt{n}V_n$, Kuiper [1] gives the following distribution function [9]:

$$G(s|H_0) = 1 - \sum_{m=1}^{\infty} 2(4m^2s^2 - 1)e^{-2m^2s^2}.$$

Percentage points are tabulated in [10] for the modified statistic:

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

for which the distribution is no longer as strongly dependent on n. (Here and in the following α is the probability of a type I error.)

α	0.15	0.10	0.05	0.01
V_n^{mod}	1.537	1.620	1.747	2.001

The dependence of the distribution for the (2) statistic on the sample size can be neglected for $n \geq 20$, since the deviation of its real distribution from the limit is negligible and essentially has no effect on the statistical conclusions.

We propose the use of the modified statistic

$$V_n^{\text{mod}} = \sqrt{n}(D_n^+ + D_n^-) + 1/(3\sqrt{n}) \tag{3}$$

in the Kuiper test, where the validity of using this correction follows naturally from the expression for the Smirnov goodness-of-fit test (p. 81 of [11]). The dependence of the distribution of statistic (3) on the sample size can be neglected for $n \geq 30$. Percentage points for this statistic have also been tabulated and are essentially the same as those for statistic (2).

α	0.15	0.10	0.05	0.01
V_n^{mod}	1.537	1.619	1.747	2.000

Statistics (2) and (3) have the same limiting distribution. For small n, the difference between their distributions is quite substantial. However, for $n \geq 20$, in the decision making region (for values of $G(V | H_0) > 0.9$ and $G(V_n^{\text{mod}} | H_0) > 0.9$) these distributions are essentially the same.

As a model for the limiting distribution of statistic (3), we can use a beta distribution of the third kind with density

$$f(x) = \frac{\theta_2^{\theta_0}}{\theta_3 \mathrm{B}(\theta_0, \theta_1)} \left(\frac{x - \theta_4}{\theta_3} \right)^{\theta_0 - 1} \left(1 - \frac{x - \theta_4}{\theta_3} \right)^{\theta_1 - 1} \left[1 + (\theta_2 - 1) \frac{x - \theta_4}{\theta_3} \right]^{-\theta_0 - \theta_1}$$

and the parameter vector $\boldsymbol{\theta} = (7.8624; 7.6629; 2.6927; 2.6373; 0.495)^{\mathrm{T}}$. This distribution corresponds to the following tabulated percentage points:

α	0.15	0.10	0.05	0.01
V_n^{mod}	1.537	1.620	1.747	1.993

This model describes the distribution of the statistic over its entire domain of definition; it, as well as the limiting distribution, can be used for calculating the attained level of significance $P\{S > S^* | H_0\}$: the probability that, when the tested hypothesis H_0 is true, the statistic S for the test exceeds the value S^* of the statistic calculated from the sample.

The Watson Test. The statistic for the Watson test [2, 3] is

$$U_n^2 = n \int_{-\infty}^{\infty} \left\{ F_n(x) - F(x, \theta) - \int_{-\infty}^{\infty} (F_n(y) - F(y, \theta)) dF(y, \theta) \right\}^2 F(x, \theta)$$

and is used in the computationally convenient form:

$$U_n^2 = \sum_{i=1}^{n} \left(F(x_i, \theta) - \frac{i - 0.5}{n} \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^{n} F(x_i, \theta) - 0.5 \right)^2 + \frac{1}{12n}. \quad (4)$$

The percentage points of the statistic U_n^2 for test of a simple hypothesis can be found in [2, 12]. The limiting distribution $G(U_n^2 | H_0)$ for the statistic U_n^2 is [2, 3]

$$G(s | H_0) = 1 - 2 \sum_{m=1}^{\infty} (-1)^{m-1} \mathrm{e}^{-2m^2 \pi^2 s}.$$

Modifications of the Kuiper and Watson tests have been discussed in [13] and of the Watson test, in [14]. The percentage points for the distributions of the modified statistics are given in [13]. In particular, the upper percentage points for the modified Watson statistic of the form

$$U_n^{2*} = (U_n^2 - 0.1/n + 0.1/n^2)(1 + 0.8/n)$$

are given by [13]

α	0.15	0.10	0.05	0.01
U_n^{2*}	0.131	0.152	0.187	0.267

The upper percentage points for the distribution of statistic (4) are essentially the same:

α	0.15	0.10	0.05	0.01
U_n^{2*}	0.131	0.151	0.187	0.268

It should be emphasized that the distribution of statistic (4) has a weak dependence on the sample size. For $n \geq 20$, the difference between the distribution of statistic (4) and the limiting distribution can be neglected.

The limiting distribution for statistic (4) over the entire domain of definition can be approximated using a model of an inverse gaussian law with density

$$f(x) = \frac{1}{\theta_2}\left(\theta_0 \Bigg/ \left[2\pi\left(\frac{x-\theta_3}{\theta_2}\right)^3\right]\right)^{1/2} \exp\left(-\theta_0\left(\left(\frac{x-\theta_3}{\theta_2}\right)-\theta_1\right)^2 \Bigg/ \left[2\theta_1^2\left(\frac{x-\theta_3}{\theta_2}\right)\right]\right)$$

with a parameter vector $\boldsymbol{\theta} = (0.2044;\ 0.08344;\ 1.0;\ 0.0)^{\mathrm{T}}$. This distribution (along with the limiting distribution) can be used for calculating the attained level of significance.

The asymptotic effectiveness of the Watson test has been studied elsewhere [15].

The Zhang tests. Nonparametric tests of goodness-of-fit with the following statistics have been proposed by Zhang [4, 5–7]:

$$Z_K = \max_{1 \leq i \leq n}\left((i-0.5)\log\left\{\frac{i-0.5}{nF(x_i,\theta)}\right\} + (n-i+0.5)\log\left[\frac{n-i+0.5}{n\{1-F(x_i,\theta)\}}\right]\right); \tag{5}$$

$$Z_A = -\sum_{i=1}^{n}\left[\frac{\log\{F(x_i,\theta)\}}{n-i+0.5} + \frac{\log\{1-F(x_i,\theta)\}}{i-0.5}\right]; \tag{6}$$

$$Z_C = \sum_{i=1}^{n}\left[\log\left\{\frac{[F(x_i,\theta)]^{-1}-1}{(n-0.5)/(i-0.75)-1}\right\}\right]^2. \tag{7}$$

Zhang's claim that these tests have greater powers than the Kolmogorov, Cramer–Mises–Smirnov, and Anderson–Darling tests has been confirmed recently [16]. However, a recommendation for wider use of these tests was denied because of the strong dependence of the distributions of statistics (5)–(7) on the sample size n. As an example, this dependence for the statistic Z_A is illustrated by the simulated distributions of the statistic in Fig. 2. This dependence makes it more difficult to use the tests. Naturally, the dependence on n is retained when testing composite hypotheses.

A Comparative Analysis of the Power of the Tests. It is essentially impossible to study the power of the tests without using computer technology and statistical modelling of the distributions of the statistics for the tests. For studying the distributions of the statistics for the validity of the tested $G(S|H_0)$ and competing $G(S|H_1)$ hypotheses, here we have used our earlier approach [17]. The statistical modelling provided an accuracy in constructing the distributions of the statistics $G(S|H_i)$, $i = \overline{0, 1}$ on the order of $\pm 10^{-3}$ with a confidence coefficient of 0.9. This quantity determines the maximum length of the confidence interval covering the true value of the distribution function at a point. It approaches this value in the region of the median.

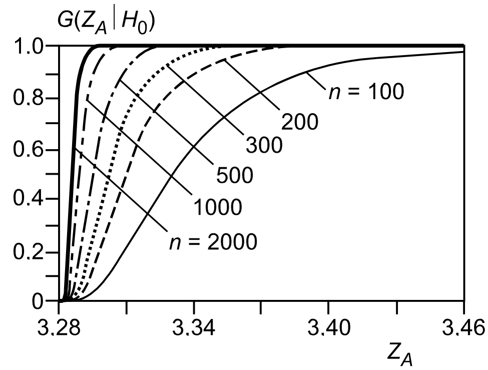

Fig. 2. Distributions $G_n(Z_A \mid H_0)$ of statistic (5) as functions of sample volume n for simple hypothesis testing.

In order to compare the power of these distributions with the Kolmogorov (K), Cramer–Mises–Smirnov (CMS), and Anderson–Darling (AD) tests, we show the results of studies for the same two pairs of competing distributions as in [18–20]. The first pair consists of a normal and a logistic distribution: the hypothesis H_0 to be tested corresponds to a normal law with density

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

and the competing hypothesis H_1, to a logistic distribution with density

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

and parameters $\theta_0 = 1$ and $\theta_1 = 0$. In the case of a simple hypothesis H_0, the parameters of the normal distribution have the same values. These two distributions are close and difficult to distinguish using goodness-of-fit criteria.

The second pair has H_0 as a Weibull distribution with a density

$$f(x) = \frac{\theta_0(x-\theta_2)^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x-\theta_2}{\theta_1}\right)^{\theta_0}\right\}$$

and parameters $\theta_0 = 2$, $\theta_1 = 2$, and $\theta_2 = 0$, and H_1 as a gamma distribution with a density

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

and parameters $\theta_0 = 3.12154$, $\theta_1 = 0.557706$, and $\theta_2 = 0$, for which the gamma distribution is closest to this Weibull distribution.

The power of these distributions was studied for testing of simple and composite hypotheses H_0 against a simple competing hypothesis H_1.

Table 1 lists the estimated power of the tests for testing a simple hypothesis H_0 (normal distribution) against hypothesis H_1 (logistic) for different values of the probability α of a type I error and different sample sizes. Table 2 does the same for testing a simple hypothesis H_0 (Weibull distribution with parameters 2, 2, 0) against hypothesis H_1 (gamma distribution with parameters 3.12154, 0.557706, 0).

TABLE 1. Power of Goodness-of-Fit Tests for Testing a Simple Hypothesis H_0 (normal distribution) Against Hypothesis H_1 (logistic)

α	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = 2000$
Zhang test (Z_C)							
0.150	0.214	0.258	0.330	0.410	0.638	0.792	0.997
0.100	0.155	0.195	0.265	0.344	0.574	0.740	0.997
0.050	0.095	0.125	0.187	0.260	0.481	0.654	0.995
0.025	0.063	0.086	0.138	0.201	0.406	0.577	0.991
0.010	0.040	0.057	0.096	0.148	0.327	0.487	0.982
Zhang test (Z_A)							
0.150	0.205	0.243	0.297	0.360	0.582	0.757	0.999
0.100	0.143	0.175	0.221	0.274	0.485	0.675	0.999
0.050	0.075	0.097	0.129	0.167	0.341	0.532	0.996
0.025	0.039	0.052	0.074	0.099	0.230	0.401	0.989
0.010	0.016	0.022	0.034	0.048	0.129	0.259	0.970
Zhang test (Z_K)							
0.150	0.169	0.194	0.246	0.314	0.529	0.693	0.996
0.100	0.115	0.134	0.178	0.235	0.434	0.601	0.991
0.050	0.060	0.072	0.102	0.144	0.303	0.458	0.974
0.025	0.032	0.039	0.059	0.088	0.209	0.340	0.941
0.010	0.015	0.018	0.029	0.047	0.127	0.224	0.872
Watson test							
0.150	0.163	0.175	0.214	0.278	0.506	0.680	0.995
0.100	0.111	0.120	0.153	0.208	0.421	0.602	0.992
0.050	0.057	0.064	0.086	0.126	0.301	0.477	0.981
0.025	0.029	0.033	0.048	0.075	0.211	0.368	0.964
0.010	0.012	0.014	0.022	0.037	0.128	0.250	0.929
Kuiper test							
0.150	0.163	0.174	0.209	0.268	0.482	0.652	0.993
0.100	0.110	0.119	0.149	0.199	0.396	0.570	0.987
0.050	0.057	0.062	0.082	0.118	0.279	0.443	0.972
0.025	0.029	0.032	0.045	0.070	0.192	0.335	0.948
0.010	0.012	0.014	0.020	0.035	0.113	0.223	0.904

Similarly, the estimated powers for the same pairs of competing distributions are shown in Tables 3 and 4 for testing of composite hypotheses. A maximum likelihood method was used for estimating the parameters of the distribution when testing the composite hypotheses.

TABLE 2. Power of Foodness-of-Fit Tests for Testing a Simple Hypothesis H_0 (Weibull distribution with parameters 2, 2, 0) Against Hypothesis H_1 (gamma distribution with parameters 3.12154, 0.557706, 0)

α	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = 2000$
Zhang test (Z_C)							
0.150	0.194	0.224	0.305	0.427	0.786	0.938	1.000
0.100	0.140	0.168	0.239	0.350	0.718	0.906	1.000
0.050	0.084	0.106	0.163	0.254	0.603	0.837	1.000
0.025	0.053	0.070	0.116	0.188	0.499	0.756	1.000
0.010	0.031	0.044	0.077	0.131	0.384	0.641	1.000
Zhang test (Z_A)							
0.150	0.183	0.204	0.272	0.394	0.774	0.935	1.000
0.100	0.127	0.142	0.196	0.300	0.693	0.898	1.000
0.050	0.068	0.076	0.107	0.180	0.549	0.815	1.000
0.025	0.036	0.040	0.057	0.103	0.413	0.711	1.000
0.010	0.015	0.017	0.024	0.047	0.264	0.558	1.000
Zhang test (Z_K)							
0.150	0.183	0.205	0.266	0.364	0.684	0.868	1.000
0.100	0.129	0.146	0.198	0.282	0.593	0.805	1.000
0.050	0.070	0.082	0.118	0.182	0.451	0.684	1.000
0.025	0.039	0.047	0.071	0.116	0.335	0.561	0.999
0.010	0.018	0.022	0.037	0.065	0.222	0.415	0.996
Watson test							
0.150	0.171	0.190	0.251	0.350	0.661	0.842	1.000
0.100	0.117	0.132	0.185	0.273	0.581	0.787	1.000
0.050	0.061	0.072	0.108	0.175	0.455	0.685	0.999
0.025	0.032	0.038	0.063	0.111	0.346	0.581	0.998
0.010	0.013	0.017	0.030	0.059	0.235	0.448	0.995
Kuiper test							
0.150	0.170	0.187	0.243	0.335	0.633	0.819	1.000
0.100	0.116	0.130	0.178	0.258	0.550	0.759	1.000
0.050	0.060	0.069	0.103	0.163	0.423	0.649	0.999
0.025	0.031	0.037	0.058	0.102	0.317	0.541	0.997
0.010	0.013	0.016	0.028	0.054	0.208	0.407	0.992

On comparing the estimates of the power of these tests and of the Kolmogorov, Cramer–Mises–Smirnov, and Anderson–Darling tests [19, 20], we can order the tests in terms of their power for testing as follows:

- simple hypotheses with respect to the normal distribution – logistic distribution pair $Z_C > Z_A > Z_K > U_n^2 > V_n >$ $> AD > K > CMS$;

TABLE 3. Power of Goodness-of-Fit Tests for Testing a Composite Hypothesis H_0 (normal distribution) Against Hypothesis H_1 (logistic)

α	n = 10	n = 20	n = 50	n = 100	n = 300	n = 500	n = 2000
Zhang test (Z_A)							
0.150	0.210	0.259	0.340	0.434	0.706	0.865	1.000
0.100	0.153	0.198	0.272	0.358	0.633	0.815	1.000
0.050	0.090	0.126	0.185	0.256	0.512	0.718	0.999
0.025	0.052	0.080	0.126	0.180	0.403	0.615	0.998
0.010	0.025	0.045	0.075	0.112	0.282	0.478	0.995
Zhang test (Z_C)							
0.150	0.195	0.236	0.321	0.427	0.711	0.866	0.998
0.100	0.142	0.186	0.270	0.374	0.662	0.831	0.998
0.050	0.086	0.126	0.205	0.302	0.584	0.770	0.998
0.025	0.052	0.086	0.156	0.243	0.510	0.705	0.997
0.010	0.026	0.051	0.105	0.176	0.411	0.609	0.996
Zhang test (Z_K)							
0.150	0.176	0.221	0.316	0.424	0.692	0.839	0.999
0.100	0.123	0.162	0.249	0.349	0.614	0.779	0.998
0.050	0.068	0.097	0.167	0.248	0.492	0.668	0.995
0.025	0.038	0.058	0.111	0.176	0.386	0.557	0.988
0.010	0.018	0.029	0.066	0.112	0.271	0.424	0.966
Watson test							
0.150	0.177	0.201	0.268	0.367	0.673	0.848	1.000
0.100	0.124	0.145	0.204	0.294	0.599	0.798	1.000
0.050	0.068	0.084	0.128	0.200	0.481	0.704	0.999
0.025	0.038	0.049	0.081	0.135	0.380	0.608	0.999
0.010	0.018	0.025	0.044	0.080	0.270	0.486	0.996
Kuiper test							
0.150	0.171	0.194	0.256	0.346	0.633	0.812	1.000
0.100	0.119	0.139	0.192	0.273	0.554	0.752	0.999
0.050	0.064	0.079	0.118	0.181	0.433	0.646	0.998
0.025	0.035	0.045	0.073	0.119	0.333	0.544	0.996
0.010	0.016	0.022	0.039	0.069	0.228	0.416	0.989

- simple hypotheses with respect to the Weibull distribution – gamma distribution pair $Z_C > Z_A > Z_K > U_n^2 > V_n > > AD > CMS > K$;
- composite hypotheses with respect to the normal distribution – logistic distribution pair $Z_A \approx Z_C > Z_K > AD > > CMS > U_n^2 > V_n > K$; and

TABLE 4. Power of Goodness-of-Fit Tests for Testing a Composite Hypothesis H_0 (Weibull distribution with parameters 2, 2, 0) Against Hypothesis H_1 (gamma distribution with parameters 3.12154, 0.557706, 0)

α	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = 2000$
Zhang test (Z_A)							
0.150	0.165	0.219	0.358	0.533	0.880	0.974	1.000
0.100	0.112	0.161	0.287	0.456	0.837	0.960	1.000
0.050	0.059	0.095	0.196	0.345	0.754	0.925	1.000
0.025	0.030	0.056	0.132	0.256	0.664	0.879	1.000
0.010	0.013	0.028	0.077	0.169	0.540	0.799	1.000
Zhang test (Z_C)							
0.150	0.170	0.215	0.341	0.509	0.867	0.971	1.000
0.100	0.113	0.152	0.264	0.426	0.818	0.954	1.000
0.050	0.055	0.081	0.166	0.303	0.722	0.913	1.000
0.025	0.024	0.041	0.098	0.203	0.610	0.853	1.000
0.010	0.007	0.014	0.044	0.107	0.440	0.732	1.000
Zhang test (Z_K)							
0.150	0.147	0.173	0.277	0.436	0.814	0.947	1.000
0.100	0.097	0.117	0.206	0.352	0.747	0.916	1.000
0.050	0.048	0.060	0.121	0.236	0.623	0.844	1.000
0.025	0.023	0.030	0.070	0.151	0.499	0.751	1.000
0.010	0.009	0.012	0.032	0.081	0.350	0.609	1.000
Watson test							
0.150	0.169	0.195	0.267	0.377	0.710	0.885	1.000
0.100	0.116	0.138	0.200	0.299	0.634	0.838	1.000
0.050	0.061	0.077	0.122	0.199	0.511	0.748	1.000
0.025	0.033	0.043	0.075	0.131	0.401	0.650	1.000
0.010	0.015	0.020	0.039	0.075	0.284	0.523	0.999
Kuiper test							
0.150	0.167	0.189	0.248	0.343	0.661	0.852	1.000
0.100	0.114	0.132	0.183	0.266	0.579	0.797	1.000
0.050	0.060	0.072	0.108	0.171	0.450	0.691	1.000
0.025	0.032	0.040	0.064	0.109	0.341	0.583	0.999
0.010	0.014	0.018	0.032	0.059	0.230	0.449	0.998

• composite hypotheses with respect to the Weibull distribution – gamma distribution pair $Z_A > Z_C > AD > Z_K >$ $> CMS > U_n^2 > V_n > K$.

Comparing these results with the power of the χ^2 type tests, we find that for testing of simple hypotheses [19] the Pearson χ^2 test ends up in the third position if asymptotic optimum grouping [17] is used along with a choice of the number of intervals such that the test will have a maximum power [17, 21]. However, for testing of composite hypotheses [20] the positions of the Pearson χ^2 test and of the Nikulin–Rao–Robson χ^2 type test [22–24] are inferior: they end up in the 7th or 8th place in the overall series of tests ranked by decreasing power. We note, however, that the power of these tests can be maximized relative to a given competing hypothesis through optimal choice of the limits and umber of grouping intervals [17, 21].

This paper can be summarized as follows: the Kuiper and Watson tests are best for testing simple hypotheses, since then they have an advantage in power over the Kolmogorov, Cramer–Mises–Smirnov, and Anderson–Darling tests. There are no difficulties in applying these tests to simple hypotheses.

In testing of composite hypotheses the Kuiper and Watson tests lose their advantage over the Cramer–Mises–Smirnov and Anderson–Darling tests. However, this does not mean that they should not be used in this case. At present, the obstacle is the lack of knowledge about the distributions (i.e., their percentage points) of the statistics for testing the corresponding composite hypotheses. Later it is planned to provide some results (models of limiting distributions and tables of percentage points) that, as do [17, 25–27], will make it possible to apply the Kuiper and Watson tests to composite hypotheses regarding various parametric models for the probability distributions.

The Zhang tests, especially with the statistics Z_C and Z_A, have an indisputable advantage in power compared to the other tests that is more noticeable in the testing of simple hypotheses. These tests have certain difficulties associated with the strong dependence of the distributions of the statistics on sample size, but these difficulties are not fundamental.

In the testing of composite hypotheses, when the dependence of the distributions of the statistics on n is supplemented by their dependence on various factors which determine the complexity of the hypothesis, these difficulties become fundamental since it is impossible to find distributions for the test statistics in advance for an infinite variety of composite hypotheses. However, there is also a way out that involves changing the technology for testing a composite hypothesis with a particular criterion [28]. The corresponding approach assumes a study of the required behavior in an interactive mode. In this situation this means that the distribution of the statistic for the test being used, which is unknown at the start of solving the statistical analysis problem (since all the factors determining the character of the composite hypothesis are unknown), must be found in real time for this analysis and used in the decision making stage (to accept or reject the test hypothesis based on a calculated value of the test statistic and computer derived knowledge of the distribution of the statistic). This approach has been used successfully with multiprocessor computing [28].

In conclusion, we note that further development of the apparatus of applied mathematical statistics and success in the use of statistical methods in science, technology, and the economy are inseparable from extensive use of computer simulations and studies of statistical and probabilistic behavior and the development of suitable scientifically based computer programs.

This work was supported by the Ministry of Education and Science of the Russian Federation (Project No. 8.1274.2011) and the Federal Targeted Program on Scientific and Teaching Staff for an Innovative Russia (agreement No. 14.V37.21.0860).

REFERENCES

1. N. H. Kuiper, “Tests concerning random points on a circle,” *Proc. Konikl. Nederl. Akad. van Wetenschappen*, Ser A, **63**, 38–47 (1960).
2. G. S. Watson, “Goodness-of-fit tests on a circle. I,” *Biometrika*, **48**, No. 1–2, 109–114 (1961).
3. G. S. Watson, “Goodness-of-fit tests on a circle. I,” *Biometrika*, **49**, No. 1–2, 57–63 (1962).
4. J. Zhang, *Powerful Goodness-of-Fit and Multi-Sample Tests: PhD Thesis*, York University, Toronto (2001).
5. J. Zhang, “Powerful goodness-of-fit tests based on the likelikhood ratio,” *J. R. Statist. Soc. B*, **64**, Part 2, 281–294, (2002).
6. J. Zhang and Y. Wu, “Likelihood-ratio tests for normality,” *Comp. Statist. Data Analysis*, **49**, No. 3, 709–721 (2005).
7. J. Zhang, “Powerful two-sample tests based on the likelihood ratio,” *Technometrics*, **48**, No. 1, 95–103 (2006).
8. H. Arsham, “Kuiper’s P-value as a measuring tool and decision procedure for the goodness-of-fit test,” *J. Appl. Statist.*, **15**, 131–135 (1988).
9. M. A. Stephens, “The goodness-of-fit statistic VN: distribution and significance points,” *Biometrika*, **52**, No. 3–4, 309–321 (1965).

10. M. A. Stephens, "EDF statistics for goodness-of-fit and some comparisons," *J. Am. Statist. Assoc.*, **69**, No. 347, 730–737 (1974).
11. L. N. Bol'shev and N. V. Smirnov, *Mathematical Statistics Tables* [in Russian], Nauka, Moscow (1983).
12. E. Pearson and H. O. Hartley, "Tables for statisticians," *Biometrika* (1972), Vol. 2.
13. M. A. Stephens, "Use of the Kolmogorov–Smirnov, Cramer–von Mises, and related statistics without extensive tables," *J. Royal Stat. Soc. Ser. B (methodological)*, **32**, No. 1, 115–122 (1970).
14. J. A. Koziol, "A modification of Watson's statistic for goodness-of-fit," *Comm. in Statistics – Theory and Methods*, **18**, No. 10, 3739–3747 (1989).
15. Ya. Yu. Nikitin, "Bahadur effectiveness of Watson–Darling goodness-of-fit tests," *J. Math. Sci.*, **43**, No. 6, 2833–2838 (1988).
16. A. V. Kozlova and B. Yu. Lemeshko, "A study of the distributions of statistics and power of the nonparametric goodness-of-fit tests proposed by Jin Zhang," *Proc. Russ. Sci.-Techn. Conf. Information and Telecommunications Problems*, Novosibirsk (2001), Vol. 1, pp. 136–139.
17. B. U. Lemeshko et al., *Statistical Analysis of Data, Modelling and Study of Probability Behavior. A Computerized Approach* [in Russian], Izd. NGTU, Novosibirsk (2011).
18. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "The power of goodness-of-fit tests for close alternatives," *Izmer. Tekhn.*, No. 2, 22–27 (2007); *Measur. Techn.*, **50**, No. 2, 132–141 (2007).
19. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "Comparative analysis of the power of goodness-of-fit tests for close competing hypotheses. I. Testing simple hypotheses," *Sibir. Zh. Industr. Matem.*, **11**, No. 2 (34), 96–111 (2008); *J. Appl. Industr. Math.*, **3**, No. 4, 462–475 (2009).
20. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "Comparative analysis of the power of goodness-of-fit tests for close competing hypotheses. II. Testing composite hypotheses," *Sibir. Zh. Industr. Matem.*, **11**, No. 4 (36), 78–93 (2008); *J. Appl. Industr. Math.*, **4**, No. 1, 70–93 (2010).
21. B. Yu. Lemeshko and E. V. Chimitova, "Choice of the number of intervals in χ^2 type goodness-of-fit tests," *Zavodsk. Lab. Diagn. Mater.*, **69**, No. 1, 61–67 (2003).
22. M. S. Nikulin, "A chi-squared test for continuous distributions," *Teor. Veroyatn. Ee Primen.*, **18**, No. 3, 675–676 (1973).
23. M. S. Nikulin, "Chi-squared test for continuous distributions with shift and scale parameters," *ibid.*, pp. 583–591.
24. K. C. Rao and D. S. Robson, "A chi-squared statistic for goodness-of-fit tests within the exponential family," *Comm. Statist.*, **3**, 1139–1153 (1974).
25. B. Yu. Lemeshko and S. B. Lemeshko, "Models for distributions of the statistics for nonparametric tests of goodness-of-fit testing of composite hypotheses using maximum likelihood estimates. Part I," *Izmer. Tekhn.*, No. 6, 3–11 (2009); *Measur. Techn.*, **52**, No. 6, 555–565 (2009).
26. B. Yu. Lemeshko and S. B. Lemeshko, "Models for distributions of the statistics for nonparametric tests of goodness-of-fit testing of composite hypotheses using maximum likelihood estimates. Part II," *Izmer. Tekhn.*, No. 8, 17–26 (2009); *Measur. Techn.*, **52**, No. 8, 799–812 (2009).
27. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "Statistic distributions models for some nonparametric goodness-of-fit tests in testing composite hypotheses," *Comm. in Statistics – Theory and Methods*, **39**, No. 3, 460–471 (2010).
28. B. Yu. Lemeshko, S. B. Lemeshko, and A. P. Rogozhnikov, "Real-time study of statistic distributions of nonparametric goodness-of-fit tests when testing complex hypotheses," *Proc. Int. Workshop Applied Methods of Statistical Analysis. Simulations and Statistical Inference, AMSA'2011,* Novosibirsk, Russia, Sept. 20–22, 2011, pp. 19–27.