Application of Homogeneity Tests: Problems and Solution

Boris Yu. Lemeshko(✉), Irina V. Veretelnikova, Stanislav B. Lemeshko, and Alena Yu. Novikova

Novosibirsk State Technical University, Novosibirsk, Russian Federation
Lemeshko@ami.nstu.ru

Abstract. The properties of the homogeneity tests of Smirnov, Lehmann-Rosenblatt, Anderson-Darling, k-sampling tests of Anderson-Darling and Zhang have been studied. Models of limiting distributions for k-sampling Anderson-Darling test under various numbers of compared samples have been presented. Power ratings have been obtained. Comparative analysis of the power of the homogeneity tests has been performed. The tests have been ordered in terms of power relative to various alternatives. Recommendations on the application of tests have been given.

Keywords: Homogeneity test · Statistical simulation
Smirnov's test · Zhang's tests · Anderson-Darling test
Lehmann-Rosenblatt test · Test power

1 Introduction

Statistician constantly encounter with problems of testing hypotheses about the belonging of two (or more) random variables samples to the same general population (homogeneity check) in various applications. In this case, there are problems of correct application and choice of the most preferable test.

With limited sample sizes, the statistics of the tests can differ significantly from the limiting (asymptotic) distributions of these statistics. For some tests of homogeneity, the distributions of statistics are unknown. In such situations, it is possible to assess the achieved significance level p_{value} and ensure the correctness of statistical conclusions only through the use of computer technologies to study statistical distributions involving the intensive use of statistical simulating methods [1]. Without the use of computer technology, one can not obtain reliable knowledge of the power of the test.

The problem of checking the homogeneity of samples is formulated as follows. Let x_{ij} be the j observation of the i sampling $j = \overline{1, n_i}, i = \overline{1, k}$. Let's pretend that $F_i(x)$ corresponds to i sample. It is necessary to test the hypothesis

$$H_0 : F_1(x) = F_2(x) = \cdots = F_k(x)$$

V. V. Rykov et al. (Eds.): ACMPT 2017, LNCS 10684, pp. 461–475, 2017.
https://doi.org/10.1007/978-3-319-71504-9_38

for any x without specifying the common for them distribution law. The empirical distribution function corresponding to i sample is designated as $F_{in_i}(x)$.

In practice, two-sampling test of Smirnov [2] and Lehmann-Rosenblatt are most often used [2–4]. Significantly less mention is made of the use of the Anderson-Darling test [5] (Anderson-Darling-Petit) or its k-sampling [6], and even more rarely of the k-sampling variants of the Smirnov or Lehmann- Rosenblatt test [7–9] application. It is practically not said about the use of Zhang's homogeneity test [10,11].

The main goal of this paper, which is the development of [12], is to study the distributions of statistics and the homogeneity test power for limited sample sizes, to refine the sample sizes, from which one can use the limiting distributions, to clarify the nature of the alternatives concerning which tests have power advantage. In carrying out the research, computer simulation and analysis of statistical regularities methodology was used, which has proved itself in analogous works [1,13–18], based mainly on the statistical modeling method.

2 The Tests Under Consideration

2.1 The Smirnov Test

The Smirnov homogeneity test is proposed in [19]. It is assumed that the distribution functions $F_1(x)$ and $F_2(x)$ are continuous. The Smirnov test statistics measure the distance between the empirical distribution functions constructed from the samples

$$D_{n_1,n_2} = \sup_x \mid F_{1,n_1}(x) - F_{2,n_2}(x) \mid.$$

In practical use of the test of statistics D_{n_1,n_2} is calculated in accordance with the relations [2]:

$$D^+_{n_1,n_2} = \max_{1\le r\le n_1} [\frac{r}{n_1} - F_{2,n_2}(x_1 r)] = \max_{1\le s\le n_2} [F_{1,n_1}(x_2 s) - \frac{s-1}{n_2}],$$

$$D^-_{n_1,n_2} = \max_{1\le r\le n_1} [F_{2,n_2}(x_1 r) - \frac{r-1}{n_1}] = \max_{1\le s\le n_2} [\frac{s}{n_2} - F_{1,n_1}(x_2 s)],$$

$$D_{n_1,n_2} = \max(D^+_{n_1,n_2}, D^-_{n_1,n_2}).$$

If the hypothesis is valid statistics of the Smirnov test

$$S_C = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_{n_1,n_2} \tag{1}$$

in the limit belongs to the Kolmogorov distribution $K(S)$ [2].

However, for limited values n_1 and n_2 random variable D_{n_1,n_2} is discrete, and the number of its possible values is the smallest common multiple of n_1 and n_2 [2]. The stepwiseness of the conditional distribution $G(S_C \mid H_0)$ of statistics

S_C with equal n_1 and n_2 remains even with $n_i = 1000$. Therefore, it is preferable to apply the test when the sample sizes n_1 and n_2 are not equal and are in fact the prime numbers. Another drawback of the test with statistics (1) is that the distributions $G(S_C \mid H_0)$ with n_1 and n_2 and growth slowly approach the limiting distribution on the left and with bounded n_1 and n_2 substantially differ from $K(s)$ (see Fig. 1). Thereby, simple modification of the statistics (1) was proposed in [12]:

$$S_C M = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}(D_{n,m} + \frac{n_1 + n_2}{4.6 n_1 n_2}),$$

which practically does not have the drawback mentioned above.

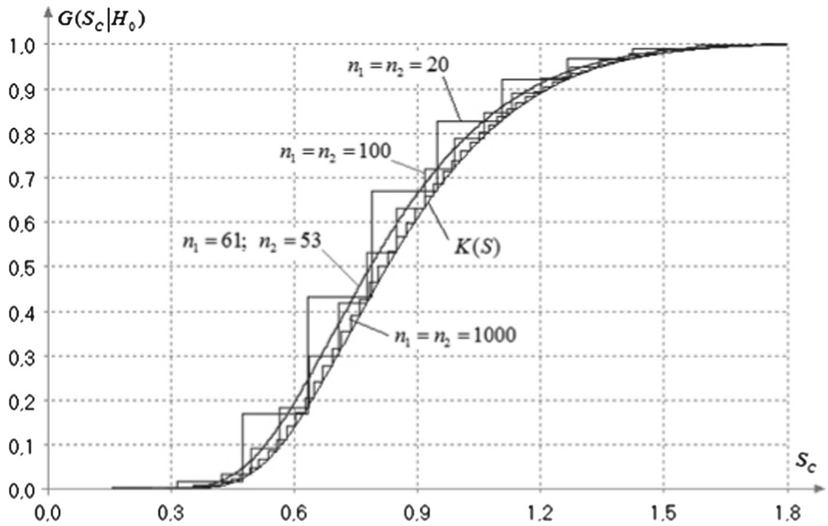

Fig. 1. The distributions of statistics (1) with validity H_0 as a function of n_1 and n_2

2.2 The Lehmann-Rosenblatt Test

The Lehmann-Rosenblatt homogeneity test is a ω^2 type test. The test was proposed in [3] and was investigated in [4]. Statistics of the testIs used in the form [2]

$$T = \frac{1}{n_1 n_2 (n_1 + n_2)}(n_2 \sum_{i=1}^{n_2} (r_i - i)^2 + n_1 \sum_{j=1}^{n_1} (s_j - j)^2) - \frac{4 n_1 n_2 - 1}{6(n_1 + n_2)}, \quad (2)$$

where r_i is ordinal number (rank) of x_{2i}; s_j is ordinal number (rank) x_{1j} in the combined variational series. It was shown in [4] that the statistics (2) in the limit is distributed as $a1(t)$ [2].

In contrast to Smirnov's test, the distribution of Lehman-Rosenblatt statistics converges rapidly to the limiting $a1(T)$. When $n_1 = n_2 = 100$ distribution visually coincides with $a1(T)$, while in practice deviation $G(T \mid H_0)$ from $a1(T)$ when $n_1, n_2 \geq 45$ can be neglected.

2.3 The Anderson-Darling Test

The two-sampling Anderson-Darling test (test for homogeneity) was considered in [5]. The statistics of the applied test is determined by the expression

$$A^2 = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1+n_2-1} \frac{(M_i(n_1+n_2) - n_1 i)^2}{i(n_1+n_2-i)}, \tag{3}$$

where M_i is the number of elements in the first sample that are less than or equal to i element of the variation series of the combined sample.

The limiting distribution of the statistics (3) with the validity of the hypothesis being tested H_0 is the same distribution $a2(t)$ [5], which is the limiting for Anderson-Darling's consent statistics.

Convergence of distribution $G(A^2 \mid H_0)$ of statistics (3) $a2(A^2)$ with limited sample volumes was investigated in [20], where it was shown that when $n_1, n_2 \geq 45$ deviation of the distribution function $G(A^2 \mid H_0)$ from $a2(A^2)$ does not exceed 0.01.

2.4 The *k*-Sampling Anderson-Darling Test

The k-sampling variant of the Anderson-Darling's consent test was proposed in [6]. Assuming continuity $F_i(x)$ the sample is built on the base of analyzed samples and generalized total volume $n = \sum_{i=1}^{k} n_i$ and ordered $X_1 \leq X_2 \leq \cdots \leq X_n$. The statistics of the test has the form [6]:

$$A_{kn}^2 = \frac{1}{n} \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{n-1} \frac{(nM_{ij} - jn_i)^2}{j(n-j)}, \tag{4}$$

where M_{ij} is number of elements in i sample, which are not greater than X_j. The hypothesis to be tested H_0 deviates at large values of the statistics (4).

In [6], the table of upper percentage points is not presented for statistics (4), but for statistics of the form:

$$T_{kn} = \frac{A_{kn}^2 - (k-1)}{\sqrt{D[A_{kn}^2]}}. \tag{5}$$

The parameter of the scale of statistics A_{kn}^2 is given by [6]

$$D[A_{kn}^2] = \frac{an^3 + bn^2 + cn + d}{(n-1)(n-2)(n-3)}$$

$$at\ a = (4g-6)(k-1) + (10-6g)H,$$

$$b = (2g-4)k^2 + 8hk + (2g-14h-4)H - 8h + 4g - 6,$$

$$c = (6h+2g-2)k^2 + (4h-4g+6)k + (2h-6)H + 4h,$$

$$d = (2h+6)k^2 - 4hk,$$

where

$$H = \sum_{i=1}^{k} \frac{1}{n_i}, h = \sum_{i=1}^{n-1} \frac{1}{i}, g = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \frac{1}{(n-i)j}.$$

Dependence of the limiting distributions of statistics (5) on the number of compared samples is k illustrates in Fig. 2. The distribution of statistics slowly converges to the standard normal law with increasing number of compared samples.

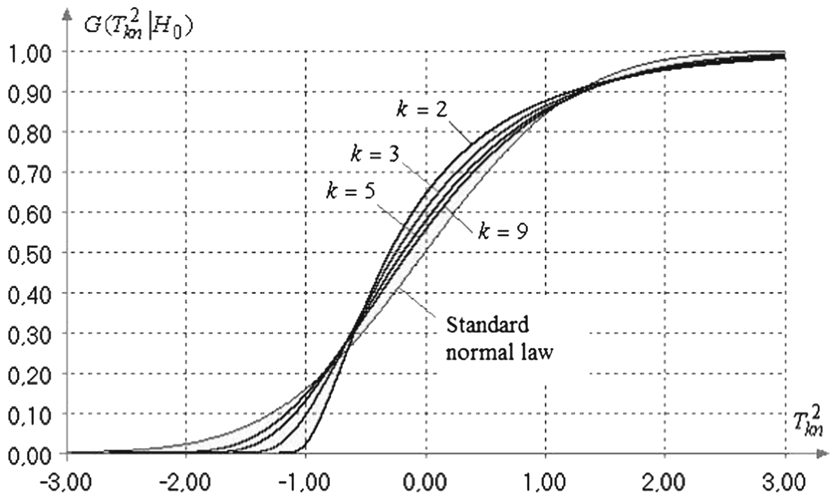

Fig. 2. Dependence distributions of statistics (5) on the number of samples being compared

The study of statistical distributions by methods of statistical modeling showed that when using test, the difference between the distributions of statistics from the corresponding limiting ones does not have practical significance for $n_i \geq 30$.

The table of upper percentage points of limiting distributions of statistic (5) is presented in [6]. Also interpolation polynomials are constructed there, allowing to find critical values $T_{kn}^2(\alpha)$ for the number of samples being compared k, absent in the table. As a result of studies of statistical distributions (5), (statistical modeling $n_i = 1000$ and the number of simulation experiments $N = 10^6$) we have somewhat refined and expanded the Table 1 of critical values.

Simultaneously, for the limiting distributions of statistics (5), approximate models of laws (for $k = 2 \div 11$) were built. Good models were [21] laws of the family of beta distributions of the third kind with density

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

Table 1. Refined upper critical values $T^2_{kn}(\alpha)$ and models of limiting distributions of statistics (5)

k	$1-\alpha$					Model
	0.75	0.90	0.95	0.975	0.99	
2	0.325	1.228	1.966	2.731	3.784	$B_{III}(3.1575, 2.8730, 18.1238, 15.0000, -1.1600)$
3	0.439	1.300	1.944	2.592	3.429	$B_{III}(3.5907, 4.5984, 7.8040, 14.1310, -1.5000)$
4	0.491	1.321	1.925	2.511	3.277	$B_{III}(4.2657, 5.7035, 5.3533, 12.8243, -1.7500)$
5	0.523	1.331	1.900	2.453	3.153	$B_{III}(6.2992, 6.5558, 5.6833, 13.010, -2.0640)$
6	0.543	1.333	1.885	2.410	3.078	$B_{III}(6.7446, 7.1047, 5.0450, 12.8562, -2.2000)$
7	0.557	1.337	1.870	2.372	3.017	$B_{III}(6.7615, 7.4823, 4.0083, 11.800, -2.3150)$
8	0.567	1.335	1.853	2.344	2.970	$B_{III}(5.8057, 7.8755, 2.9244, 10.900, -2.3100)$
9	0.577	1.334	1.847	2.323	2.927	$B_{III}(9.0736, 7.4112, 4.1072, 10.800, -2.6310)$
10	0.582	1.3345	1.838	2.306	2.899	$B_{III}(10.2571, 7.9758, 4.1383, 11.186, -2.7988)$
11	0.589	1.332	1.827	2.290	2.867	$B_{III}(10.6848, 7.5950, 4.2041, 10.734, -2.8400)$
∞	0.674	1.282	1.645	1.960	2.326	N(0.0, 1.0)

Values of the parameters of the law $B_{III}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ were found on the basis of the statistics samples obtained as a result of modeling $N = 10^6$.

The models $B_{III}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ presented in Table 1 with the given parametersvalues, allow to find p_{value} with an appropriate number k of compared samples from the statistics values calculated as (5).

In the case of $k = 2$, the test with statistics (5) is equivalent in power to the two-sample Anderson-Darling test with statistics (3).

2.5 Test for the Homogeneity of Zhang

The tests of homogeneity proposed by Zhang [10,11] are the of the Smirnov, Lehmann-Rosenblatt and Anderson-Darling tests development enabling us to compare $k \geq 2$ samples. Zhang's goodness-of-fit test [10] shows some advantage in power compared to the Kramer-Mises-Smirnov and Anderson-Darling goodness-of-fittests [22], but the drawback that limits the use of Zhang's test is the dependence of statistical distributions on sample sizes. The same drawback is possessed by variants of Zhang's test for checking the homogeneity of laws. To overcome this disadvantage, the author [10] proposes to use the Monte Carlo method for p_{value} estimation. The problem of modeling distributions of the Zhang homogeneity test statistics,is much simpler in comparison with a similar problem for the goodness-of-fit, since it is necessary to model the distributions of statistics $G(S \mid H_0)$ in the case of analyzed samples belonging to the uniform law.

Let $x_{i1}, x_{i2}, \cdots, x_{in_i}$ be ordered samples of continuous random variables with distribution functions $F_i(x)$, $(i = \overline{1,k})$ and combined ordered sample $X_1 < X_2 < \cdots < X_n$, $n = \sum_{i=1}^{k} n_i$. Rank j of the ordered x_{ij} observation of sample i in the combined sample is denoted as R_{ij}. Let $X_0 = -\infty, X_{n+1} = +\infty$, and ranks $R_{i,0} = 1, R_{i,n_i+1} = n + 1$.

The modification of the empirical distribution function $\hat{F}(t)$ is used in the tests, which is equal $\hat{F}(X_m) = (m - 0.5)/n$ [10] at break points $X_m, m = \overline{1,n}$.

Z_k the Zhang homogeneity test has the form [10]:

$$Z_K = \max_{1 \le m \le n} \sum_{i=1}^{k} [F_{i,m} \ln \frac{F_{i,m}}{F_m} + (1 - F_{i,m}) \ln \frac{1 - F_{i,m}}{1 - F_m}], \tag{6}$$

where $F_m = \hat{F}(X_m)$, so that $F_m = (m - 0.5)/n$, and the calculation $F_{i,m} = \hat{F}_i(X_m)$ is carried out as follows. At the initial moment the values are $j_i = 0, i = \overline{1,k}$. If $R_{i,j_i+1} = m$, then $j_i := j_i + 1$ and $F_{i,m} = (j_i - 0.5)/n_i$, otherwise, if $R_{i,j_i} < m < R_{i,j_i+1}$, then $F_{i,m} = j_i/n_i$.

Right-hand test indicates that hypothesis H_0 deviates at **large values** of the statistics (6). The distributions of statistics depend on n_i and k. Decision-making is influenced by the discreteness of statistics, which, with growth of k becomes less pronounced (see Fig. 3).

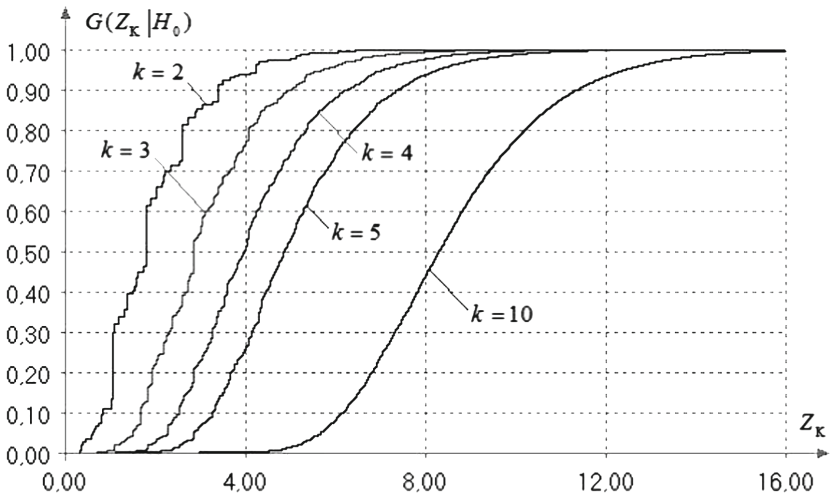

Fig. 3. Dependence of the distributions of statistics (6) on k where $n_i = 20$

Statistics Z_A of the Zhang homogeneity testis is determined by the expression [10]:

$$Z_A = -\sum_{m=1}^{n} \sum_{i=1}^{k} n_i \frac{F_{i,m} \ln F_{i,m} + (1 - F_{i,m}) \ln(1 - F_{i,m})}{(m - 0.5)(n - m + 0.5)}, \tag{7}$$

where F_m and $F_{i,m}$ are calculated as defined above.

Left-sided test indicates that verifiable hypothesis H_0 deviates for **small values** of the statistics (7). The distributions of statistics depend on n_i and k.

Statistics Z_C the test for homogeneity of samples is calculated in accordance with expression [10]:

$$Z_C = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln(\frac{n_i}{j-0.5}-1)\ln(\frac{n}{R_{i,j}-0.5}-1). \quad (8)$$

The test is also left-handed: the hypothesis being tested H_0 deviates at **small values** of the statistics (8). The distributions of statistics depend on n_i and k, this dependence is shown in Fig. 4.

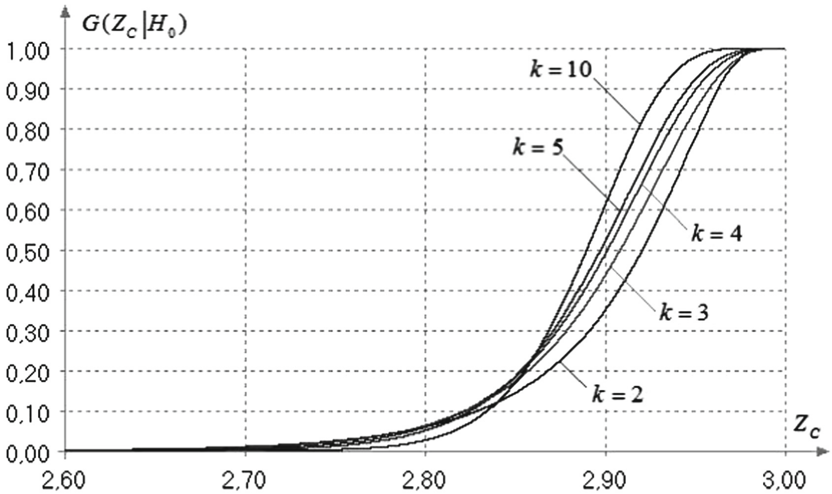

Fig. 4. Dependence of the distributions of statistics (8) on k where $n_i = 20$

The lack of information on the distribution laws of statistics and tables of critical values in modern conditions is not a serious disadvantage of Zhang's test, since in software supporting the application of test it is not difficult to organize the calculation of the achieved significance levels p_{value}, using methods of statistical modeling.

3 Comparative Analysis of the Test Power

The power of homogeneity test has been investigated with respect to the number of pairs of competing hypotheses. For definiteness, the hypothesis tested H_0 corresponded to the samples with same standard normal distribution law with density

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

and the shift parameters $\theta_0 = 0$ and scale $\theta_1 = 1$. With all alternatives, the first sample always corresponded to the standard normal law, and the second sample to some other one. In particular, with a shift alternative, in the case of competing hypothesis H_1, the second compilation corresponded to the normal law with the shift parameter $\theta_0 = 0.1$ and scale parameter $\theta_1 = 1$, in the case of competing hypothesis H_2 - normal law with parameters $\theta_0 = 0.5$ and $\theta_1 = 1$. When the scale is changed, in the case of competing hypothesis H_3, the second assembly corresponds to the normal law with parameters $\theta_0 = 0$ and $\theta_1 = 1.1$, in the case of competing hypothesis H_4 - normal law with parameters $\theta_0 = 0$ and $\theta_1 = 1.5$. In the case of competing hypothesis H_5 the second assembly corresponded to the logistic law with density

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

and parameters $\theta_0 = 0$ and $\theta_1 = 1$. Normal and logistic laws are really close and difficult to distinguish using the goodness-of-fit test.

The obtained power estimates of the considered test for equal n_i when $k = 2$ with respect to competing hypotheses $H_1 - H_5$ - are presented in the Table 2, where the test are ordered in descending order with respect to the corresponding H_i. Power ratings of k-sampling tests where $k = 4$ with respect to competing hypotheses H_1, H_3, H_5 are given in the Table 3.

Naturally, with the increase in the number of compared samples of the same volumes, the power of the test decreases. For example, it is more difficult to single out the situation and give preference to the competing hypothesis, when only one of the samples analyzed belongs to some other law. This can be seen by comparing the corresponding power ratings in Tables 2 and 3.

Analysis of the obtained power estimates allows us to draw the following conclusions. Concerning competing hypotheses corresponding to some changes in the shift parameter, Smirnov's (Sm), Lehmann-Rosenblatt (LR), Anderson-Darling-Petite (AD) test and Zhang's test with statisticians Z_K, Z_A, Z_C in descending order are in the following order:

$$AD \succ LR \succ Z_C \succ Z_A \succ Sm \succ Z_K.$$

Concerning competing hypotheses corresponding to some changes in the scale parameter, the test are already arranged in a following order:

$$Z_A \succ Z_C \succ Z_K \succ AD \succ LR \succ Sm.$$

However, the difference in the power of tests with statistic Z_A and Z_C is small. Again, with relatively close alternatives for small sample sizes, the Smirnov test is more preferable when the Lehmann-Rosenblatt test. In the case of one sample belongs to the normal law and the second to the logistic one, the test are ordered in terms of power as follows:

$$Z_K \succ Z_A \succ Z_C \succ AD \succ Sm \succ LR.$$

Table 2. Estimates of the power of tests to alternatives $H_1 - H_5$ where $k = 2$ with equal n_i

Test	$n_i = 20$	$n_i = 50$	$n_i = 100$	$n_i = 300$	$n_i = 500$	$n_i = 1000$	$n_i = 2000$
Concerning the alternative H_1							
Sm	0.111	0.132	0.164	0.280	0.381	0.617	0.869
LR	0.115	0.136	0.173	0.313	0.438	0.678	0.910
AD	0.114	0.137	0.175	0.319	0.447	0.691	0.919
Z_K	0.111	0.126	0.152	0.238	0.333	0.526	0.798
Z_A	0.113	0.133	0.162	0.272	0.374	0.583	0.851
Z_C	0.114	0.134	0.164	0.278	0.382	0.600	0.859
Concerning the alternative H_2							
Sm	0.365	0.703	0.910	1	1	1	1
LR	0.430	0.757	0.954	1	1	1	1
AD	0.435	0.768	0.959	1	1	1	1
Z_K	0.344	0.650	0.906	1	1	1	1
Z_A	0.419	0.733	0.941	1	1	1	1
Z_C	0.425	0.743	0.946	1	1	1	1
Concerning the alternative H_3							
Sm	0.105	0.108	0.120	0.150	0.186	0.297	0.551
LR	0.103	0.107	0.114	0.149	0.1908	0.324	0.624
AD	0.104	0.112	0.128	0.202	0.290	0.528	0.861
Z_K	0.107	0.127	0.154	0.268	0.390	0.624	0.892
Z_A	0.108	0.128	0.164	0.318	0.464	0.745	0.958
Z_C	0.107	0.127	0.163	0.320	0.468	0.748	0.961
Concerning the alternative H_4							
Sm	0.152	0.288	0.510	0.964	0.999	1	1
LR	0.154	0.280	0.548	0.989	1	1	1
AD	0.185	0.424	0.777	1	1	1	1
Z_K	0.248	0.552	0.849	1	1	1	1
Z_A	0.267	0.651	0.937	1	1	1	1
Z_C	0.256	0.640	0.936	1	1	1	1
Concerning the alternative H_5							
Sm	0.104	0.110	0.121	0.159	0.198	0.319	0.564
LR	0.103	0.106	0.113	0.142	0.178	0.288	0.547
AD	0.103	0.108	0.117	0.156	0.203	0.343	0.640
Z_K	0.105	0.110	0.122	0.179	0.266	0.429	0.759
Z_A	0.104	0.108	0.115	0.177	0.275	0.563	0.916
Z_C	0.104	0.108	0.116	0.1721	0.265	0.556	0.913

Table 3. Estimates of power of the k-sampling homogeneity test to alternatives H_1, H_3, H_5 where $k = 4$ with equal n_i

Test	$n_i = 20$	$n_i = 50$	$n_i = 100$	$n_i = 300$	$n_i = 500$	$n_i = 1000$
Concerning the alternative H_1						
AD	0.112	0.131	0.164	0.301	0.433	0.701
Z_K	0.109	0.121	0.141	0.219	0.300	0.502
Z_A	0.111	0.127	0.153	0.255	0.360	0.579
Z_C	0.111	0.126	0.155	0.260	0.368	0.595
Concerning the alternative H_3						
AD	0.104	0.110	0.123	0.180	0.254	0.474
Z_K	0.106	0.120	0.145	0.249	0.367	0.606
Z_A	0.107	0.124	0.158	0.305	0.463	0.745
Z_C	0.106	0.122	0.158	0.306	0.468	0.761
Concerning the alternative H_5						
AD	0.102	0.106	0.113	0.143	0.179	0.291
Z_K	0.103	0.107	0.114	0.161	0.222	0.410
Z_A	0.103	0.107	0.116	0.179	0.274	0.566
Z_C	0.103	0.107	0.115	0.173	0.257	0.555

When k sample are compared, the order of preference is maintained for k-sampling variants of the Anderson-Darling and Zhang test. In particular, with respect to change the shift parameter, the order of preference is:

$$AD \succ Z_C \succ Z_A \succ Z_K.$$

Regarding the change in the scale parameter, the order of tests can be written as follows:

$$Z_C \succ Z_A \succ Z_K \succ AD.$$

In this case, the test with statistics Z_A and Z_C are practically equivalent in power, and the Anderson-Darling test is noticeably inferior to all. Regarding the situation when the three samples belong to the normal law, and the fourth to the logistic one, the test are arranged according to the power in the following order:

$$Z_A \succ Z_C \succ Z_K \succ AD.$$

It should be noted that the Zhang test have an advantage in power relative to the alternatives associated with changing scale characteristics, and are inferior in power under shift alternatives.

4 Application Examples

The application of the tests considered in the section for checking the homogeneity of laws is considered by analyzing the three samples below, each with a volume of 40 observations:

0.321	0.359	−0.341	1.016	0.207	1.115	1.163	0.900	−0.629	−0.524
−0.528	−0.177	1.213	−0.158	−2.002	0.632	−1.211	0.834	−0.591	−1.975
−2.680	−1.042	−0.872	0.118	−1.282	0.766	0.582	0.323	0.291	1.387
−0.481	−1.366	0.351	0.292	0.550	0.207	0.389	1.259	−0.461	−0.283

0.890	−0.700	0.825	1.212	1.046	0.260	0.473	0.481	0.417	1.825
1.841	2.154	−0.101	1.093	−1.099	0.334	1.089	0.876	2.304	1.126
−1.134	2.405	0.755	−1.014	2.459	1.135	0.626	1.283	0.645	1.100
2.212	0.135	0.173	−0.243	−1.203	−0.017	0.259	0.702	1.531	0.289

0.390	0.346	1.108	0.352	0.837	1.748	−1.264	−0.952	0.455	−0.072
−0.054	−0.157	0.517	1.928	−1.158	−1.063	−0.540	−0.076	0.310	−0.237
−1.109	0.732	2.395	0.310	0.936	0.407	−0.327	1.264	−0.025	−0.007
0.164	0.396	−1.130	1.197	−0.221	−1.586	−0.933	−0.676	−0.443	−0.101

The empirical distributions corresponding to these samples are shown in Fig. 5.

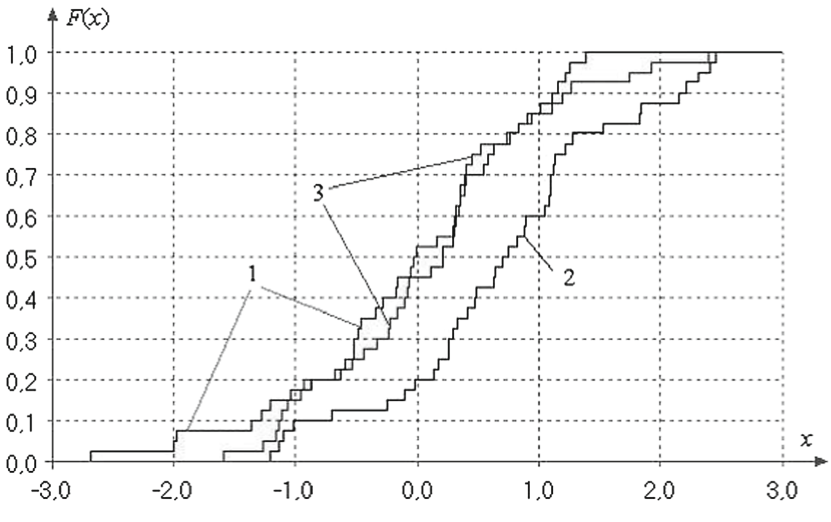

Fig. 5. Empirical distributions corresponding to the samples compared

Let us test the hypothesis of homogeneity of the 1st and 2nd samples [23]. Table 4 shows the results of the check: the values of the test statistics and the achieved significance levels p_{value}. Estimates of p_{value} were calculated from the value of statistics in accordance with the distribution $a2(A^2)$ for the Anderson-Darling test, in accordance with the distribution $a1(T)$ for the Lehmann - Rosenblatt test, in accordance with the distribution $K(S)$ for the Smirnov test, in accordance with the beta distribution of the third kind from Table 2 for $k = 2$, k-sampling Anderson-Darling test. The distributions of statistics (6), (7) and (8) of the Zhang test and estimates of p_{value} were obtained by modeling. It is obvious that the hypothesis of homogeneity should be rejected by all tests.

Table 4 also shows the results of testing the hypothesis of homogeneity of the first and third samples. Here the estimates of p_{value} by all test are really high, therefore the hypothesis of homogeneity to be tested should be not rejected.

Table 5 shows the results of testing the hypothesis of homogeneity of the three samples considered by k-sampling Anderson-Darling and the Zhang tests. In this case, the estimate of p_{value} for the Anderson-Darling test was calculated in accordance with the beta distribution of the third kind from Table 1 for $k = 3$, and for the Zhang test on the basis of statistical modeling carried out in interactive mode (number of simulation experiments $N = 10^6$). The results show that the hypothesis tested should be rejected.

In this case, the results of the test were fairly predictable. The first and third samples were modeled in accordance with the standard normal law, and the resulting pseudorandom values were rounded to 3 significant digits after the decimal point. The second sample was obtained in accordance with the normal law with a shift parameter of 0.5 and a standard deviation of 1.1.

Table 4. The results of testing the homogeneity of two samples

Tests	Of the 1st and 2nd		Of the 1st and 3rd	
	Statistics	p_{value}	Statistics	p_{value}
Anderson-Darling	5.19801	0.002314	0.49354	0.753415
k-sampling Anderson-Darling	5.66112	0.003259	−0.68252	0.767730
Lehmann-Rosenblatt	0.9650	0.002973	0.0500	0.876281
Smirnov	1.5625	0.015101	0.447214	0.989261
Smirnov's modified	1.61111	0.011129	0.495824	0.966553
Zhang Z_A	2.99412	0.0007	3.1998	0.332
Zhang Z_C	2.87333	0.0008	3.07077	0.384
Zhang Z_K	5.58723	0.0150	1.7732	0.531

Table 5. The results of testing the homogeneity of 3 samples

Tests	Statistics	p_{value}
k-sampling Anderson-Darling	4.73219	0.0028
Zhang Z_A	3.02845	0.0015
Zhang Z_C	2.92222	0.0017
Zhang Z_K	7.00231	0.0217

5 Conclusion and Outlook

Use of limiting distribution of the statistics (4) and of the statistics of the Lehmann-Rosenblatt test is correct for small volumes of compared samples since these distributions converges rapidly to limiting distribution. The same can be said about the convergence of the distribution of statistics (3) of the Anderson-Darling homogeneity test to the distribution $a2(t)$.

The models of limited distributions of statistics (5) constructed in this paper using k-sampling homogeneity Anderson-Darling test for analysis k compared samples ($k = 2 \div 11$) give an opportunity to find estimates of p_{value}, which will undoubtedly make the statistical conclusion more informative and substantiated.

In the case of the Smirnov test, due to the stepped nature of the statistics distribution (1) (especially, for equal sample sizes), the use of the Kolmogorov distribution $K(S)$ will be associated with approximate knowledge of the actual level of significance (the probability of error of the first kind) and the corresponding critical value. In the case of constructing the procedures for testing homogeneity by the Smirnov test, it is recommended: (1) to choose $n_1 \neq n_2$ so that they are relatively prime numbers, and their least common multiple k was maximal and equal $n_1 n_2$; (2) to use a modification of Smirnov's statistics. Then the application of the Kolmogorov distribution as the distribution of the modified Smirnov test statistic will be correct for relatively small n_1 and n_2.

The Zhang test with statistics Z_K, Z_A and Z_C with respect to some alternatives have a noticeable advantage in power. The drawback that limits their use is the dependence of the distributions of statistics on sample volumes. This disadvantage is easily overcome by using the Monte Carlo method to construct empirical distributions $G_N(Z \mid H_0)$ for statistics Z_K, Z_A and Z_C at specific sample sizes with subsequent evaluation of the values of p_{value}. This procedure is easily realized, since in the construction $G_N(Z \mid H_0)$ comparable samples are modeled according to the uniform law on the interval $[0, 1]$.

Acknowledgments. The studies were carried out with the support of the Ministry of Education and Science of the Russian Federation in the framework of the state work "Ensuring the conduct of scientific research" (No. 1.4574.2017/6.7) and the design part of the state task (No. 1.1009.2017/4.6).

References

1. Lemeshko, B.Y., Postovalov, S.N., Chimitova, E.V.: Statistical Data Analysis, Simulation and Study of Probability Regularities. Computer Approach: Monograph. NSTU Publisher, Novosibirsk (2011)
2. Bol'shev, L.N., Smirnov, N.V.: Tables for Mathematical Statistics. Nauka, Moscow (1983). (in Russian)
3. Lehmann, E.L.: Consistency and unbiasedness of certain nonparametric tests. Ann. Math. Statist. **22**(1), 165–179 (1951)
4. Rosenblatt, M.: Limit theorems associated with variants of the von Mises statistic. Ann. Math. Statist. **23**, 617–623 (1952)

5. Pettitt, A.N.: A two-sample Anderson-Darling rank statistic. Biometrika **63**(1), 161–168 (1976)
6. Scholz, F.W., Stephens, M.A.: K-sample Anderson-Darling tests. J. Am. Stat. Assoc. **82**(399), 918–924 (1987)
7. Kiefer, J.: K-sample analogues of the Kolmogorov-Smirnov and Cramer-V. Mises Tests. Ann. Math. Stat. **30**(2), 420–447 (1959)
8. Conover, W.J.: Several k-sample Kolmogorov-Smirnov tests. Ann. Math. Stat. **36**(3), 1019–1026 (1965)
9. Conover, W.J.: Practical Nonparametric Statistics, 3rd edn. Wiley, New York (1999)
10. Zhang, J.: Powerful goodness-of-fit and multi-sample tests. Ph.D. thesis, York University, Toronto (2001)
11. Zhang, J.: Powerful two-sample tests based on the likelihood ratio. Technometrics **48**(1), 95–103 (2006)
12. Lemeshko, B.Y., Lemeshko, S.B.: Statistical distribution convergence and homogeneity test power for Smirnov and Lehmann-Rosenblatt tests. Measur. Tech. **48**(12), 1159–1166 (2005)
13. Lemeshko, B.Y., Lemeshko, S.B.: Power and robustness of test used to verify the homogeneity of means. Measur. Tech. **51**(9), 950–959 (2008)
14. Lemeshko, B., Mirkin, E.: Bartlett and Cochran tests in measurements with probability laws different from normal. Measur. Tech. **47**(10), 960–968 (2004)
15. Lemeshko, B.Y., Lemeshko, S.B., Gorbunova, A.A.: Application and power of test for testing the homogeneity of variances. Part I. Parametric test. Measur. Tech. **53**(3), 237–246 (2010)
16. Lemeshko, B.Y., Lemeshko, S.B., Gorbunova, A.A.: Application and power of test for testing the homogeneity of variances. Part II. Nonparametric test. Measur. Tech. **53**(5), 476–486 (2010)
17. Lemeshko, B.Y., Sataeva, T.S.: Application and power of parametric test for testing the homogeneity of variances. Part III. Measur. Tech. **60**(1), 7–14 (2017)
18. Lemeshko, B.Y., Sataeva, T.S.: Application and power of parametric test for testing the homogeneity of variances. Part IV. Measur. Tech. **60**(5), 425–431 (2017)
19. Smirnov, N.V.: Estimation of the discrepancy between the empirical distribution curves in two independent samples. Byul. MGU, Seriya A **2**(2), 3–14 (1939)
20. Postovalov, S.N.: Using of computer modeling for expanding application of classical methods of statistics hypothesis checking. DEA thesis. NSTU Publisher, Novosibirsk (2013)
21. Lemeshko, B.Y.: Tests for homogeneity. Guide on the application. INFRA-M, Moscow (2017)
22. Lemeshko, B.Y.: Nonparametric goodness-of-fit tests. Guide on the application. INFRA-M, Moscow (2014)
23. ISW-Program System for Statistical Analysis of Univariate Random Quantities. http://ami.nstu.ru/headrd/ISW.htm. Accessed 23 July 2017