# On the application of k-samples tests for testing the homogeneity of distribution laws

Boris Yu. Lemeshko, Irina V. Veretelnikova

Novosibirsk State Technical University, Novosibirsk, Russian Federation
e-mail: Lemeshko@ami.nstu.ru

Models of limiting distributions of the k-samples Anderson–Darling homogeneity test have been made.

New k-samples homogeneity tests based on the Smirnov, Lehmann–Rosenblatt and Anderson–Darling two-sample tests have been proposed.

Models of the limiting distributions for the proposed test have been made.

Comparative analysis of the power of the set of k-samples tests, including the Zhang test, has been carried out.

# 1 Introduction

The necessity of solving the task of checking the hypotheses of two (or more) samples of random values belonging to the same universe estimates (the homogeneity test) may arise in different areas. For example, this task may arise naturally when checking the measurement means and trying to be certain that the random measurement errors distribution law has not undergone any serious changes within some time period.

The task of testing the homogeneity of k-samples can be stated as follows.

We have $x_{ij}$, where $j$ is the observation in the set of order statistics of $i$-sample $j = \overline{1, n_i}$, $i = \overline{1, k}$. Let us assume that the $i$-sample correlates with the continuous distribution function of $F_i(x)$.

It is required to test the hypothesis of

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

type without defining the common distribution law.

The general approach to the construction of *k*-sample homogeneity tests which are the counterparts of the two–sample Kolmogorov-Smirnov and Cramer–von Mises (Lehmann–Rosenblatt) tests, was considered in (**Kiefer, 1959)** [1]. Under this approach, the statistics of the criterion is a measure of deviation of empirical distributions corresponding to specific samples from the empirical distribution based on the totality of the analyzed samples. The *k*-samples variant of the Kolmogorov–Smirnov test based on this principle is mentioned in (**Conover**, **1965**), [2, 3]. The k-samples version of the Anderson–Darling test is proposed in (**Scholz & Stephens, 1987**) [4]. The homogeneity tests constructed by **Zhang** (**1999, 2006, 2007**) in [5, 6, 7] are the development of the homogeneity tests by Smirnov (**1939**) [8], Lehmann–Rosenblatt (**1951**, **1952**) [9, 10] and Anderson–Darling (**Pettitt, 1976**) [11] and allow us to analyze samples.

The application of *k*-samples tests in practice is constrained by the fact that, at best, only critical values of statistics for the relevant ones are known, as in the case of the Anderson–Darling test (**Scholz, 1987**) [4] or Kolmogorov-Smirnov tests (**Conover**, **1965**) [2], (**Wolf & Naus, 1972**) [12], and the possibility of using Zhang's criteria rests on the need to look for the distribution of test statistics (or estimation of the achieved significance level $p_{value}$) using statistical modeling in order to form a conclusion about the results of the hypothesis test.

The only exception is the homogeneity test $\chi^2$ for which the asymptotic distributions of statistics are known with the truth of $H_0$.

In the present work we illustrate the dependence of the distributions of statistics of the $k$-sample tests on the sample sizes and the number of $k$ compared samples. For the $k$-sample Anderson–Darling test [4] we give models of limit distributions of statistics constructed by us (**Lemeshko**, **2017**) [13, 14, 15]. Suggested variants of $k$-sample tests based on the use of 2-sample Smirnov test [8], Lehmann–Rosenblatt test [9, 10] and Anderson–Darling test [11], and present the constructed model for the limit distributions of the statistics of the proposed test for various $k$. The constructed models make it possible to carry out correct and informative conclusions with the calculation of $p_{value}$ with the usage of the corresponding criteria. In addition, we present estimates of the power of the test considered with respect to some competing hypotheses, which allows us to organize the $k$-sample tests by preference with respect to various alternatives.

The studies were based on the intensive use of the Monte Carlo method in the simulation of distributions of tests statistics.

# 2 k-samples homogeneity tests

## 2.1 Anderson-Darling test

The Anderson-Darling *k*-sample test is proposed in (**Scholz & Stephens, 1987**) [4]. Let us denote the empirical distribution function corresponding to the $i^{\text{th}}$ sample $F_{in_i}(x)$, and the empirical distribution function corresponding to the combined sample volume $n=\sum_{i=1}^{k} n_i$ as $H_n(x)$. Statistics of the Anderson-Darling sample test (AD) is defined by the expression

$$A_{kn}^2 = \sum_{i=1}^{k} n_i \int_{B_n} \frac{[F_{in_i}(x) - H_n(x)]^2}{\left(1-H_n(x)\right) H_n(x)} dH_n(x),$$

where $B_n = \{x \in R:\ H_n(x) < 1\}$. Under the assumption of continuity of $F_i(x)$ on the ordered combined sample $X_1 \le X_2 \le \ldots \le X_n$ in [4] this simple expression for the calculation of statistics is obtained:

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

where $M_{ij}$ is the number of elements in the $i^{\text{th}}$ sample which are not larger than $X_j$. The hypothesis $H_0$ being tested is rejected for large values of statistics.

The statistics acquires the following final form in [4]:

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

where the dispersion is determined by the following expression [4]

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

with

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

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

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

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

where

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

Asymptotic (limiting) distributions of statistics (1) depend on the $k$-number of samples compared and do not depend on $n_i$. With the growth of $k$ the distribution of statistics (1) slowly converges to the standard normal law (see Fig. 1).

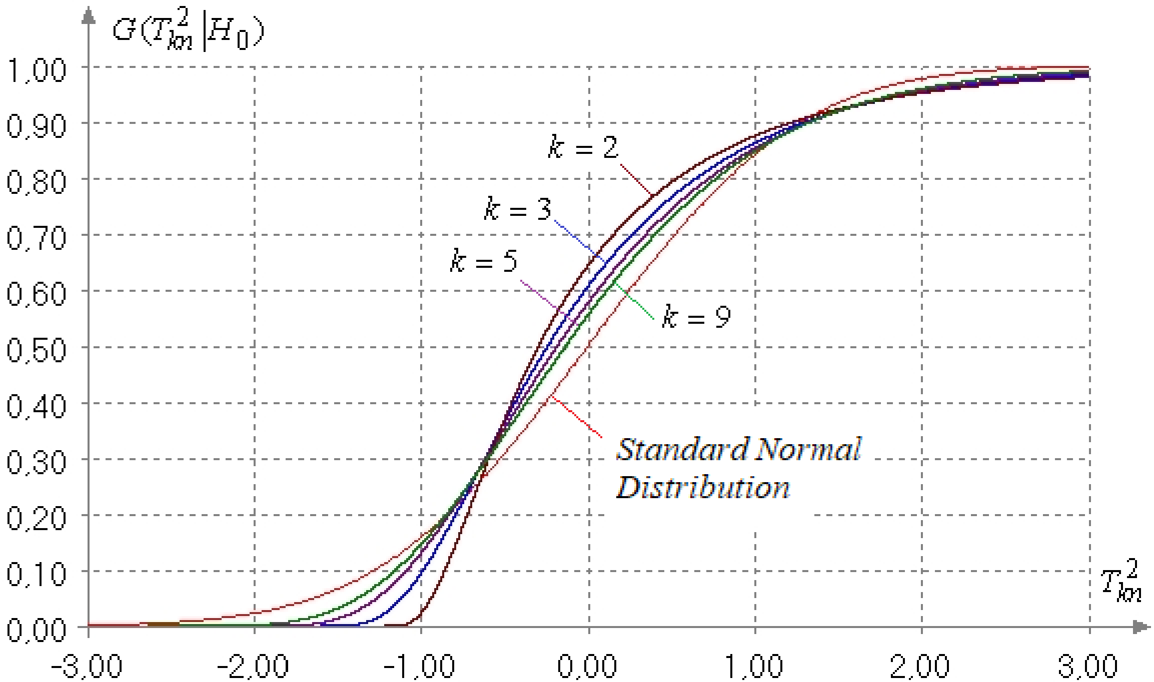


Fig. 1. Dependence of the limiting distributions of statistics (1) on $k$

In (**Scholz & Stephens, 1987)** [4] for statistics (1) the table of critical values has been constructed for a number of $k$ .

Based on the results of statistical modeling, we built models of limiting distributions of statistics (4) for $k = 2 \div 11$(**Lemeshko, 2017**) [13, 14, 15]. The laws of the family of beta distributions of the III type with density turned out to be good models when having the density of

$$f(x) = \frac{\theta_2^{\theta_0}}{\theta_3 \mathrm{B}(\theta_0, \theta_1)} \left( \frac{x - \theta_4}{\theta_3} \right)^{\theta_0 - 1} \left( 1 - \frac{x - \theta_4}{\theta_3} \right)^{\theta_1 - 1} \Bigg/ \left[ 1 + (\theta_2 - 1) \frac{x - \theta_4}{\theta_3} \right]^{\theta_0 + \theta_1}, \qquad (2)$$

as shown in Table 1 as $\mathrm{B}_{\mathrm{III}}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ having exact values for this law's parameters. These models are based on simulated samples of statistics with the number of simulation experiments $N = 10^6$ and $n_i = 10^3$.

Table 1: **Models of the limiting distributions of statistics (1)**

| $k$ | Model |
|---|---|
| 2 | $B_{III}$ (3.1575, 2.8730, 18.1238, 15.0000, –1.1600) |
| 3 | $B_{III}$ (3.5907, 4.5984, 7.8040, 14.1310, –1.5000) |
| 4 | $B_{III}$ (4.2657, 5.7035, 5.3533, 12.8243, –1.7500) |
| 5 | $B_{III}$ (6.2992, 6.5558, 5.6833, 13.010, –2.0640) |
| 6 | $B_{III}$ (6.7446, 7.1047, 5.0450, 12.8562, –2.2000) |
| 7 | $B_{III}$ (6.7615, 7.4823, 4.0083, 11.800, –2.3150) |
| 8 | $B_{III}$ (5.8057, 7.8755, 2.9244, 10.900, –2.3100) |
| 9 | $B_{III}$ (9.0736, 7.4112, 4.1072, 10.800, –2.6310) |
| 10 | $B_{III}$ (10.2571, 7.9758, 4.1383, 11.186, –2.7988) |
| 11 | $B_{III}$ (10.6848, 7.5950, 4.2041, 10.734, –2.8400) |
| $\infty$ | N(0.0, 1.0) |

## 2.2 Zhang tests

The Zhang tests [5, 6, 7] allow comparing $k \geq 2$ samples.

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

In the tests a modification of the empirical distribution function $\hat{F}(t)$ is used, having the values of $\hat{F}(X_m) = (m - 0.5)/n$ at break points $X_m$, $m = \overline{1,n}$ [5].

### 2.2.1 $Z_K$ Zhang Test

The $Z_K$ statistic of the Zhang homogeneity test is of the following form [5]:

$$Z_K = \max_{1\le m\le n}\left\{\sum_{i=1}^{k} n_i \left[F_{i,m}\ln\frac{F_{i,m}}{F_m} + \left(1-F_{i,m}\right)\ln\frac{1-F_{i,m}}{1-F_m}\right]\right\}, \qquad (3)$$

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

This is a *right-hand* test: the hypothesis being tested is rejected at *high* statistical values (3).

The dependence of statistics distributions $G(Z_K|H_0)$ on the volume of samples (with equal $n_i$ and $k=2$) is illustrated in Fig. 2, and the dependence on the *k* number of samples under comparison with $n_i = 20$, $i=\overline{2,k}$ is shown in Fig. 3.

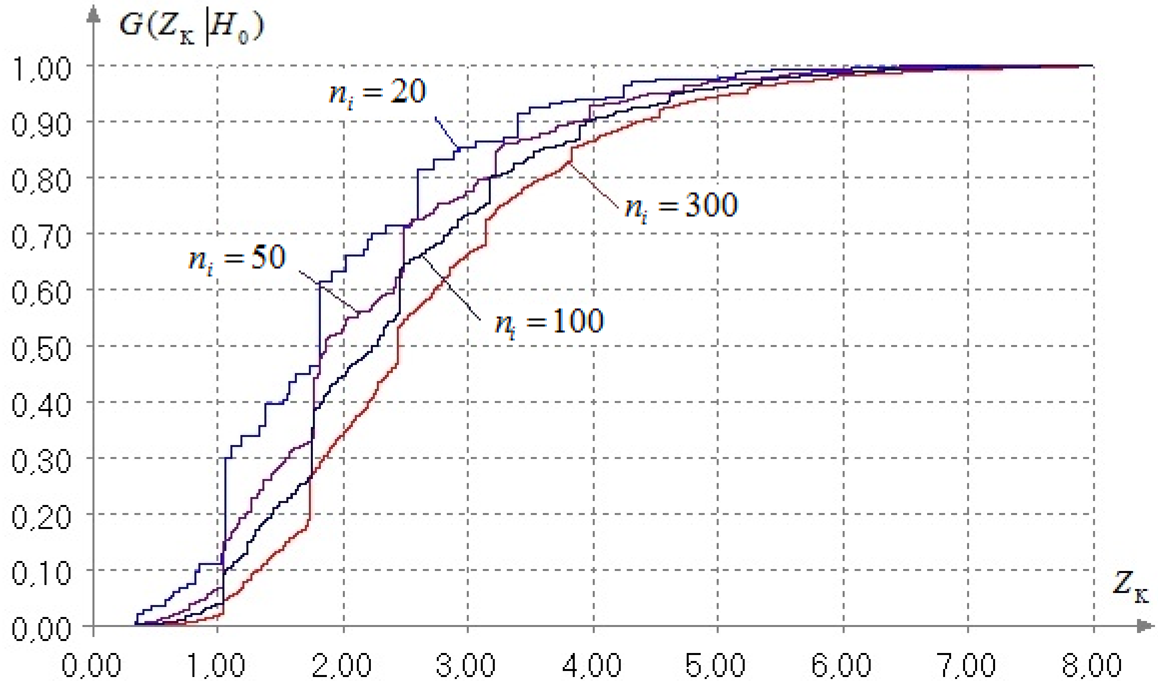


Fig. 2. The dependence of the distributions of statistics (3) on sample volume

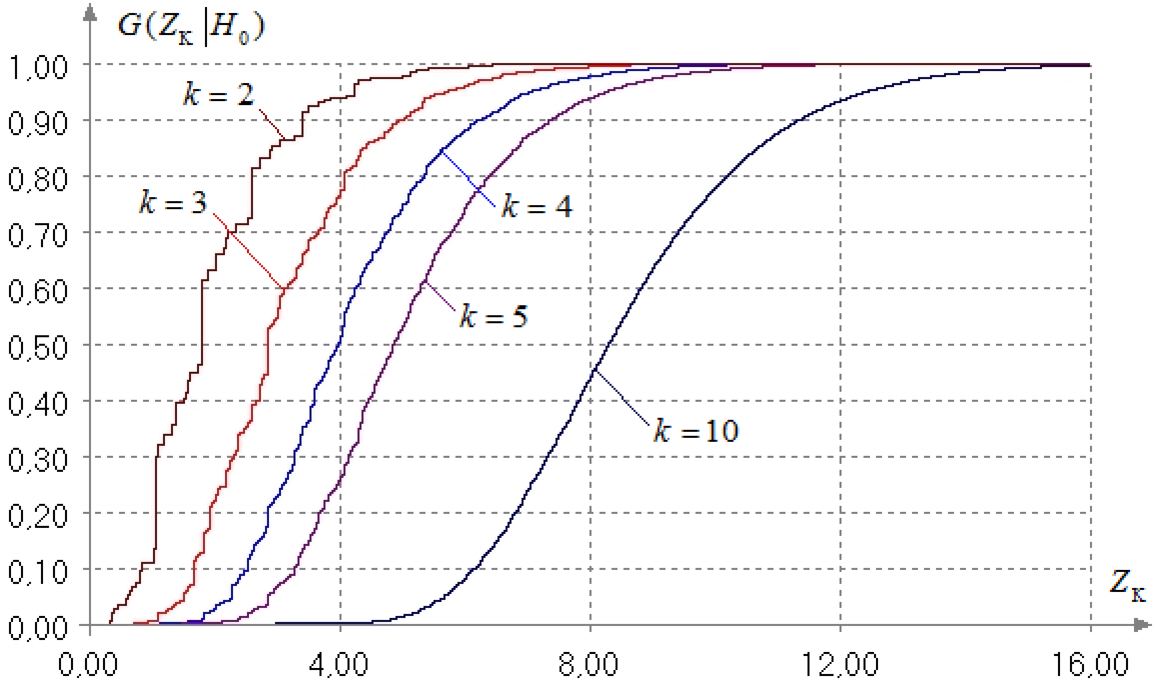


Fig. 3. The dependence of the distributions of statistic (3) on $k$

### 2.2.2 $Z_{\mathrm{A}}$ Zhang Test

Statistic $Z_{\mathrm{A}}$ of the homogeneity test of $k$ samples is defined by the following expression [5]:

$$Z_{\mathrm{A}} = -\sum_{m=1}^{n}\sum_{i=1}^{k} n_i \frac{F_{i,m}\ln F_{i,m} + \left(1-F_{i,m}\right)\ln\left(1-F_{i,m}\right)}{(m-0.5)(n-m+0.5)}, \tag{4}$$

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

This is a *left-side* test: the hypothesis $H_0$ being tested is rejected for *small* values of statistics (4).

Distributions of the statistic (4) depend on the sample volume and the number of samples compared as well. The dependence of distributions $G(Z_A | H_0)$ of statistic on the volume of the samples with equal $n_i$ and $k=2$ is shown in Fig. 4, and the dependence on the *k*-number of compared samples with $n_i = 20,\ i=\overline{2,k}$ is given in Fig. 5.

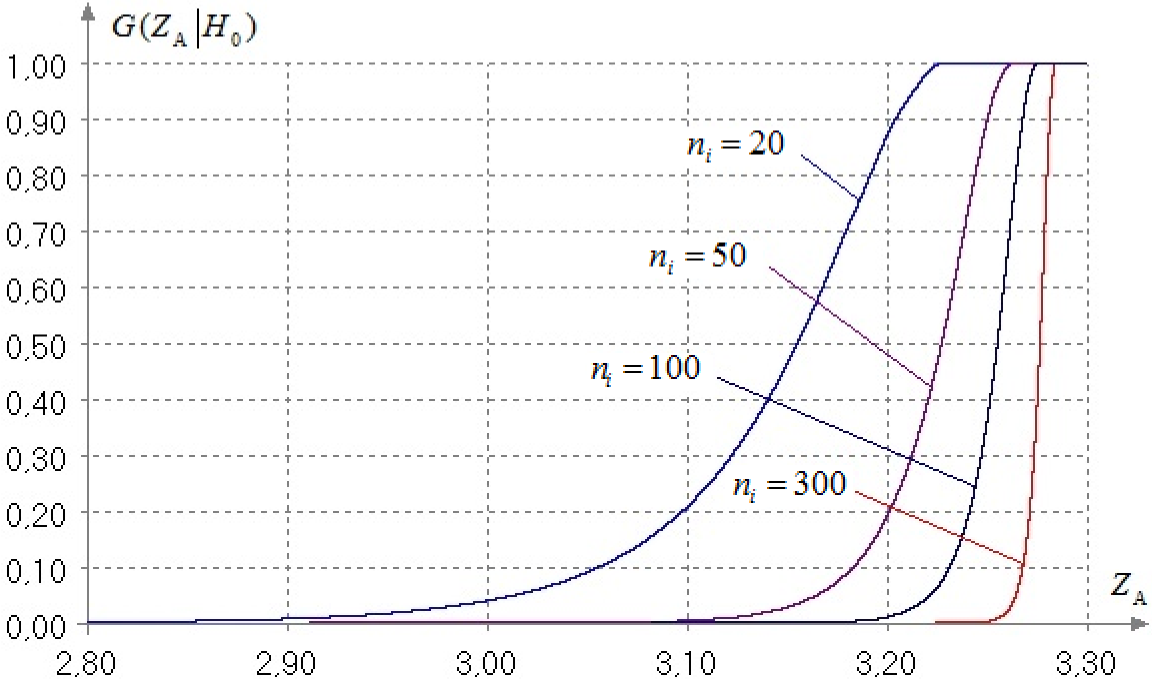


Fig. 4. The dependence of the distributions of statistics (4) on the sample volume, $k = 2$

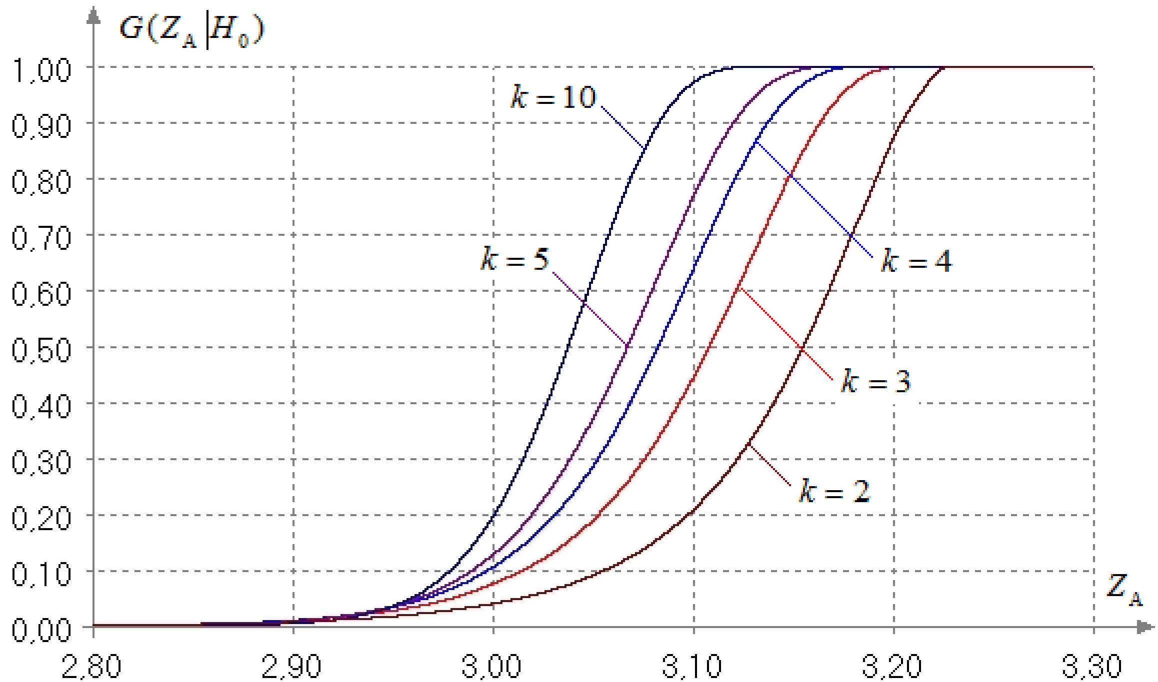


Fig. 5. The dependence of the distributions of statistics (4) on $k$ ($n_i = 20$)

### 2.2.3 $Z_C$ Zhang Test

**Statistic $Z_C$ of the homogeneity test** of $k$ samples is defined by the following expression [5]:

$$Z_C = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{n_i}{j-0.5}-1\right)\ln\left(\frac{n}{R_{i,j}-0.5}-1\right). \tag{5}$$

This is also a *left-hand* test: the tested hypothesis $H_0$ is rejected at *small* values of the statistic (5).The distributions $G(Z_C|H_0)$ of the statistic depend on the sample volume and the number of samples under analysis in the similar way.

The dependence of the distributions of statistics (3) – (5) of the sizes of the samples complicates the use of the Zhang test since there are problems with the calculation of the evaluation of $p_{value}$.

At the same time, the lack of information on the laws of distribution of statistics and tables of critical values in modern conditions is not a serious disadvantage of the tests as it is easy to calculate the achieved levels of significance of $p_{value}$ with the software that supports the application of the tests, merely using statistic modeling methods.

## 2.3 k-samples Tests Based on 2-sample Ones

In order to analyze the $k$-samples it is possible to apply a two-sample test with the $S$ statistic to each pair (totaling $(k-1)k/2$ pairs), and the decision on accepting or rejecting the $H_0$ hypothesis will be made on the strength of all results. The following statistic can be taken as a statistic of this $k$ -sample criterion (when having a right-hand two-sample criterion):

$$S_{\max} = \max_{\substack{1 \le i \le k \\ i < j \le k}} \{S_{i,j}\}, \tag{6}$$

where $S_{i,j}$ are the values of the statistics of the used two-sample criterion as calculated in the course of analysis of the $i$ th and the $j$ th samples.

The hypothesis $H_0$ to be tested will be rejected at **large** values of statistics $S_{\max}$ . The advantage of this kind of test is that as a result a pair of samples will be determined, the difference between them being the most significant from the standpoint of the two-sample test used.

Statistics of the two-sample Smirnov, Lehmann-Rosenblatt and Anderson-Darling tests can be used as $S_{i,j}$ . I n this case the distributions of the relevant statistics $S_{\max}$ converge to some limiting ones, models of which can be found on the results of statistical modeling.

### 2.3.1 Smirnov Maximum Test

The $D_{n_2,n_1}$ statistic used in the Smirnov test is calculated according to the following formulae [8]:

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

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

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

With the $H_0$ hypothesis being true and with unlimited increase of the number of samples the statistic

$$S_C = \sqrt{\frac{n_1 n_2}{n_1+n_2}} D_{n_2,n_1} \qquad (7)$$

will in the limit fall with the Kolmogorov arrangement of $K(S)$ [8] .

In case of using the $k$ -samples variant of the Smirnov test as $S_{i,j}$ in (6) it seems more preferable to use a modification of the Smirnov statistic

$$S_{\mathrm{mod}} = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left(D_{n_2,n_1} + \frac{n_1+n_2}{4.6 n_1 n_2}\right), \qquad (8)$$

its distribution being always closer to the limiting distribution of Kolmogorov $K(S)$ (**Lemeshko, 2005**) [16]. Statistic $S_{\max}$ will be defined as $S^{Sm}_{\max}$ in this case.

With equal volumes of samples under comparison the statistic distributions $S_{\max}^{Sm}$ will be of substantial discreteness (similar to the two-sample case (see Figure 6) and be different from the asymptotic (limiting) distributions (see Figure 7). If possible, it is preferable to use co-primes as $n_i$, then the distributions $G(S|H_0)$ of the $S_{\max}^{Sm}$ statistic will not be actually different from the asymptotic ones.

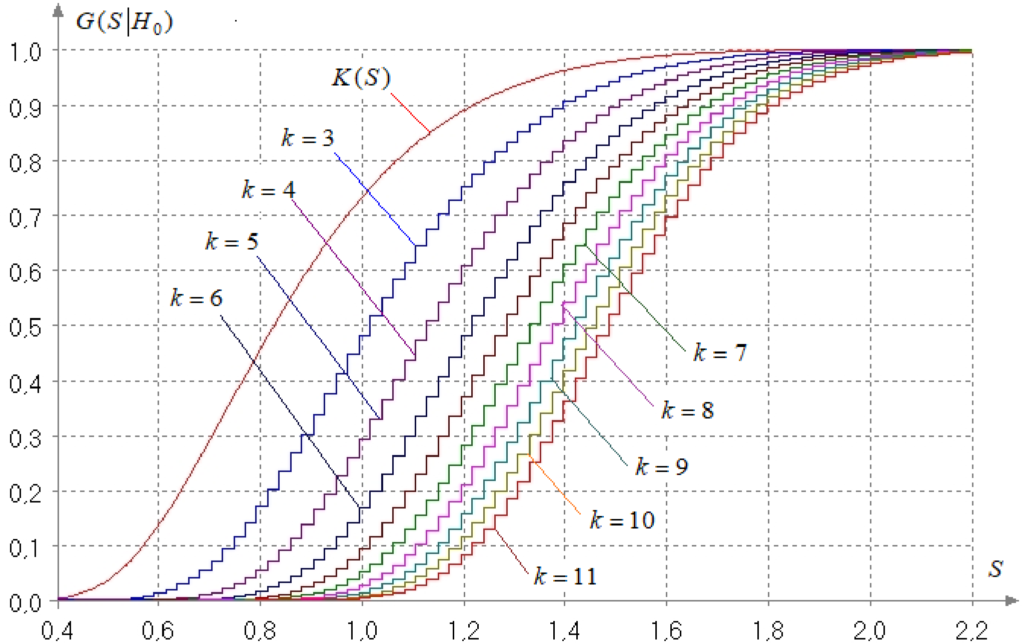


Fig. 6. Statistic distributions with $n_i = 1000, \quad i = \overline{1,k}$

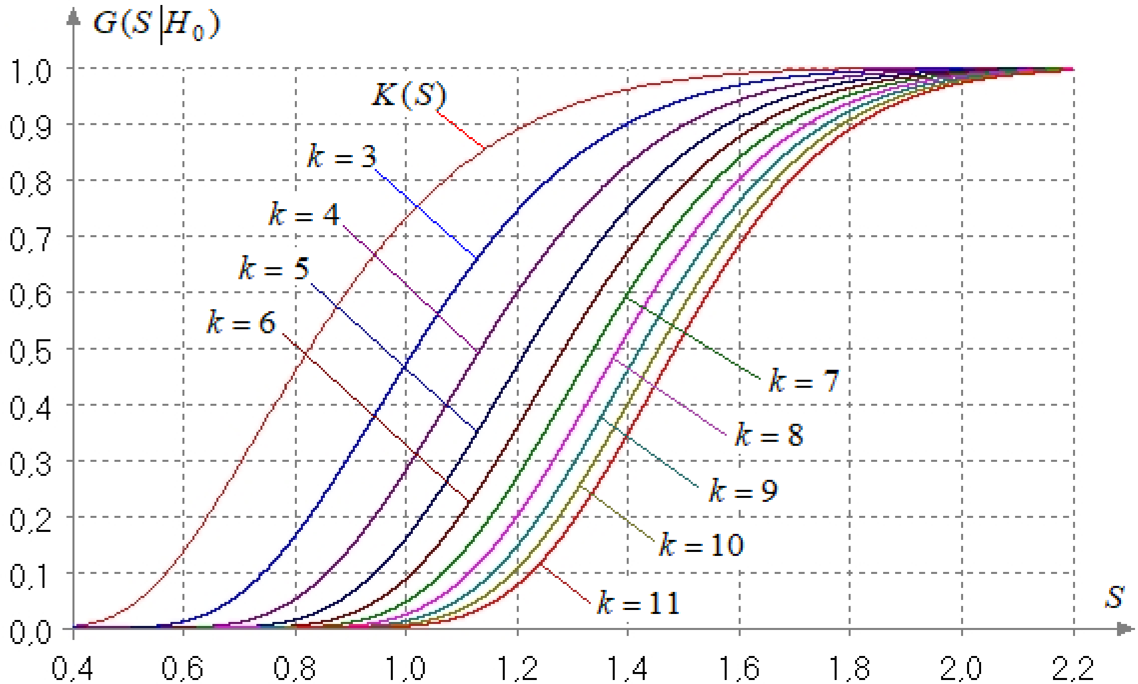


Figure 7. Asymptotic statistic distributions $S_{\max}^{Sm}$

Models of asymptotic $S_{\max}^{Sm}$ statistic distributions with $k = 3 \div 11$ in the form of beta distributions of the III type (2) $\mathrm{B}_{\mathrm{III}}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ having exact values of parameters and constructed in this paper based on the results of statistic modeling are shown in Table 2.

Table 2: **Models of limiting distributions of statistic** $S_{\max}^{Sm}$

| $k$ | Model |
|---|---|
| 2 | $K(S)$ |
| 3 | $B_{III}$ (6.3274, 6.6162, 2.8238, 2.4073, 0.4100) |
| 4 | $B_{III}$ (7.2729, 7.2061, 2.6170, 2.3775, 0.4740) |
| 5 | $B_{III}$ (7.1318, 7.3365, 2.4813, 2.3353, 0.5630) |
| 6 | $B_{III}$ (7.0755, 8.0449, 2.3163, 2.3818, 0.6320) |
| 7 | $B_{III}$ (7.7347, 8.6845, 2.3492, 2.4479, 0.6675) |
| 8 | $B_{III}$ (7.8162, 8.9073, 2.2688, 2.4161, 0.7120) |
| 9 | $B_{III}$ (7.8436, 8.8805, 2.1696, 2.3309, 0.7500) |
| 10 | $B_{III}$ (7.8756, 8.9051, 2.1977, 2.3280, 0.7900) |
| 11 | $B_{III}$ (7.9122, 9.0411, 2.1173, 2.2860, 0.8200) |

### 2.3.2 Lehman-Rosenblatt Maximum Test

Statistic of the two-sample Lehmann-Rosenblatt test as introduced in [9] is used in the following form [8]:

$$T = \frac{1}{(n_1+n_2)}\left[ n_2 \sum_{i=1}^{n_2} (r_i - i)^2 + n_1 \sum_{j=1}^{n_1} \left(s_j - j\right)^2 \right] - \frac{4n_1 n_2 - 1}{6(n_1+n_2)}, \qquad (9)$$

where $r_i$ is the numerical order (rank) of $x_{2i}$; $s_j$ is the numerical order (rank) of $x_{1j}$ in the unified ordered series. In [10] it was shown that the statistic (9) at the limit is distributed as $a1(t)$ [8].

In the case of using the $k$ -samples variant of the Lehman-Rosenblatt test as $S_{i,j}$ in the statistic $S_{\max}^{LR}$ of form (6) statistic (9) is used. Dependence of distributions of statistic $S_{\max}^{LR}$ on the number of samples with $H_0$ being true is illustrated in Fig. 8.

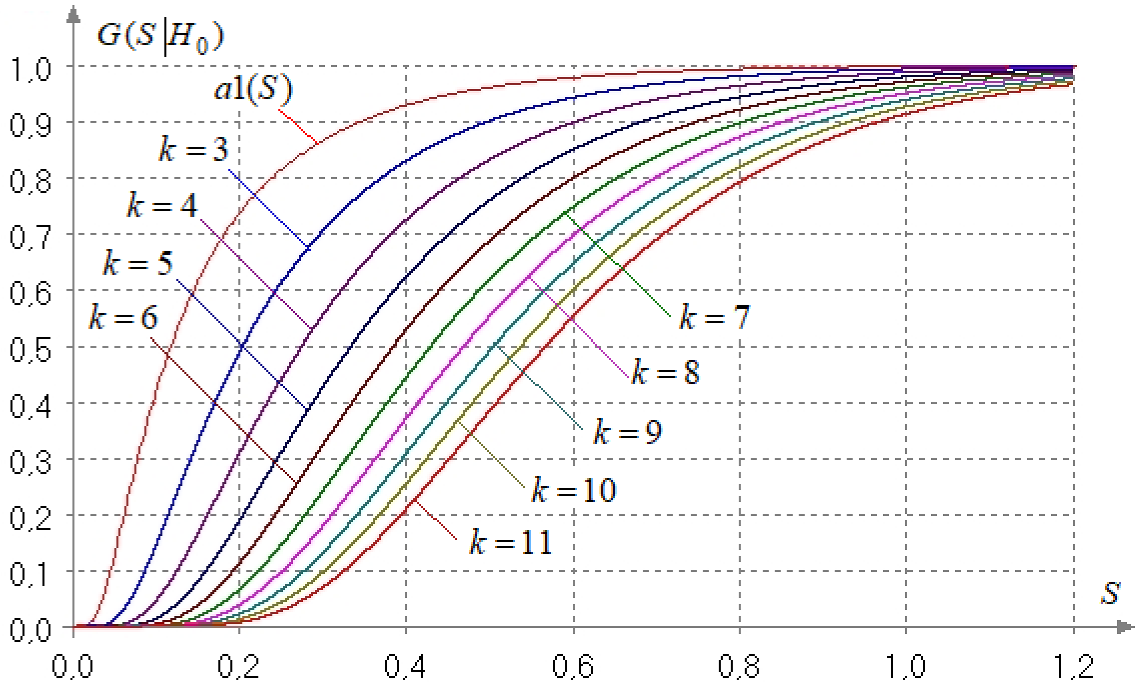


Fig. 8. Distributions of statistic $S_{\max}^{LR}$

The constructed models of asymptotic (limiting) distributions of statistic $S_{\max}^{LR}$ with the number of compared samples $k = 3 \div 11$ are shown in Table 3.

In this case the Sb-Johnson distributions proved to be the best with the density of

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

with exact values of this law's parameters, the law being shown in Table 3 as $Sb(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$. These represented models allow finding the estimates of $p_{value}$ by the values of statistic $S_{\max}^{LR}$ with corresponding $k$ number of samples under comparison.

Table 3: **Models of limiting distributions of statistic** $S_{\max}^{LR}$

| $k$ | Model |
|---|---|
| 2 | $a1(t)$ |
| 3 | Sb(3.2854, 1.2036, 3.0000, 0.0215) |
| 4 | Sb(2.5801, 1.2167, 2.2367, 0.0356) |
| 5 | Sb(3.1719, 1.4134, 3.1500, 0.0320) |
| 6 | Sb(2.9979, 1.4768, 2.9850, 0.0380) |
| 7 | Sb(3.2030, 1.5526, 3.4050, 0.0450) |
| 8 | Sb(3.2671, 1.6302, 3.5522, 0.0470) |
| 9 | Sb(3.4548, 1.7127, 3.8800, 0.0490) |
| 10 | Sb(3.4887, 1.7729, 3.9680, 0.0510) |
| 11 | Sb(3.4627, 1.8168, 3.9680, 0.0544) |

### 2.3.3 Anderson-Darling Maximum Test

The Anderson-Darling two-sample test was dealt with in [11]. This test's statistic is defined by the following expression:

$$A^2 = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1+n_2-1} \frac{\left(M_i(n_1+n_2) - n_1 i\right)^2}{i(n_1+n_2-i)}, \tag{10}$$

where $M_i$ is the number of elements of the first sample, smaller or equal to the $i$ th element of the variation set of the unified sample. Distribution $a2(t)$ will be the limiting distribution (10) with the tested hypothesis $H_0$ being true [8].

In the case of using the $k$ -samples variant of the Anderson-Darling test as $S_{i,j}$ in the $S_{\max}^{AD}$ statistic (6) statistic (10) will be used. Dependence of distributions of statistic $S_{\max}^{AD}$ on the number of samples with $H_0$ being true is shown in Fig. 9.

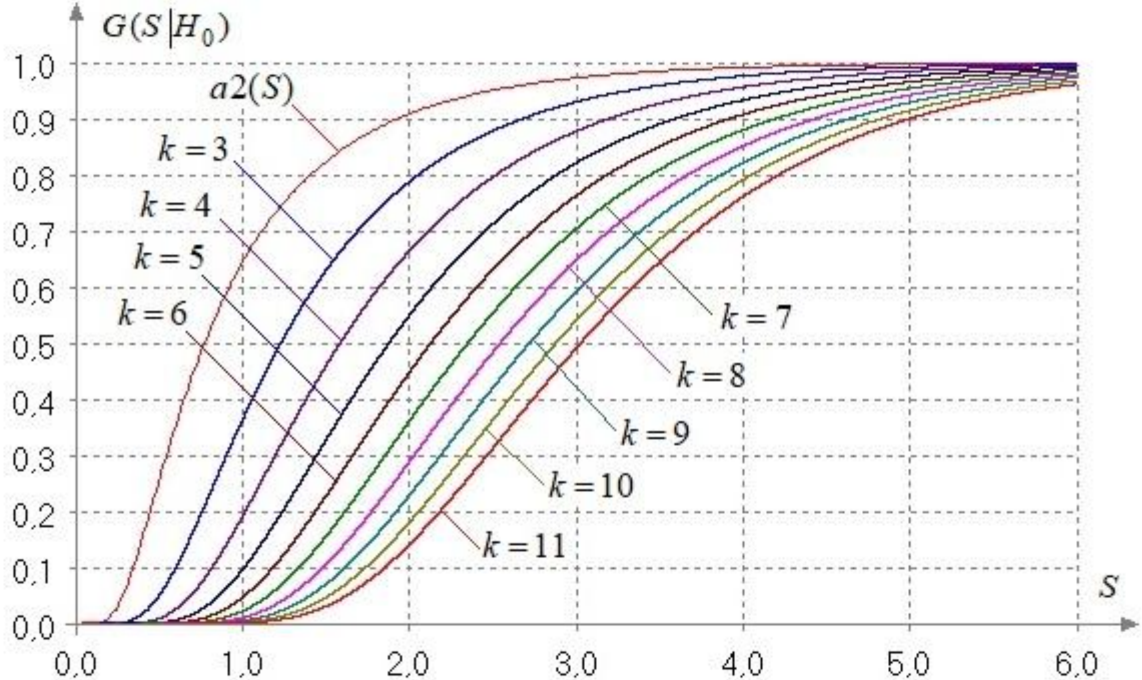


Fig. 9. Distributions of statistic $S_{\max}^{AD}$

Models of asymptotic (limiting) distributions of statistic $S_{\max}^{AD}$ for the $k$ number of samples under comparison $k = 3 \div 11$ have been constructed for distributions $G(S_{\max}^{AD} | H_0)$ and shown in Table 4. In this case the beta distributions of the III type proved to be the best (2) as shown as $\mathrm{B}_{\mathrm{III}}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ with exact values of parameters shown in Table 4; these can be used for estimating $p_{value}$ with the $k$ number of compared samples.

Table 4: **Models of limiting distributions of statistic $S_{\max}^{AD}$**

| $k$ | Model |
|---|---|
| 2 | $a2(t)$ |
| 3 | $B_{III}$ (4.4325, 2.7425, 12.1134, 8.500, 0.1850) |
| 4 | $B_{III}$ (5.2036, 3.2160, 10.7792, 10.000, 0.2320) |
| 5 | $B_{III}$ (5.7527, 3.3017, 9.7365, 10.000, 0.3000) |
| 6 | $B_{III}$ (5.5739, 3.4939, 7.7710, 10.000, 0.3750) |
| 7 | $B_{III}$ (6.4892, 3.6656, 8.0529, 10.500, 0.3920) |
| 8 | $B_{III}$ (6.3877, 3.8143, 7.3602, 10.800, 0.4800) |
| 9 | $B_{III}$ (6.7910, 3.9858, 7.1280, 11.100, 0.5150) |
| 10 | $B_{III}$ (6.7533, 4.2779, 6.6457, 11.700, 0.5800) |
| 11 | $B_{III}$ (7.1745, 4.3469, 6.6161, 11.800, 0.6100) |

## 2.4 Homogeneity Test $\chi^2$

The homogeneity test $\chi^2$ can successfully be used to analyze $k \geq 2$ samples. In this case the common area of the samples is split into $r$ intervals (groups). Let $\eta_{ij}$ be the number of elements of the $i^{\text{th}}$ sample of the $j^{\text{th}}$ interval, then $n_i = \sum_{j=1}^{r} \eta_{ij}$ .

The $\chi^2$ homogeneity test statistic will be of the following form:

$$\chi^2 = n \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{(\eta_{ij} - \nu_j n_i / n)^2}{\nu_j n_i} = n \left( \sum_{i=1}^{k} \sum_{j=1}^{r} \frac{\eta_{ij}^{\;2}}{\nu_j n_i} - 1 \right), \tag{11}$$

where $\nu_j = \sum_{l=1}^{k} \eta_{lj}$ is the total number of elements of all samples falling into the $j^{\text{th}}$ interval.

The $\chi^2$-distribution with the number of degrees of freedom $(k-1)(r-1)$ shall be the asymptotic distribution of statistic [17].

## 3 Comparative analysis of powers

One of the main characteristics of the statistical test is its power relative to a given competing hypothesis $H_1$. The power is the remainder of $1-\beta$, where $\beta$ is the possibility of type II error (accept hypothesis $H_0$ with $H_1$ being true) at specified probability $\alpha$ of type I error (reject $H_0$ when true).

The power of k-samples tests was investigated for various k and situations when the tested hypothesis $H_0$ was whether all samples belonged to the standard normal law, the competing hypothesis $H_1$ being if all samples but the last one belonged to the standard normal law and the last sample belonged to the normal law with the shift parameter $\theta_0 = 0.1$ and the scale parameter $\theta_1 = 1$; hypothesis $H_2$ being that the last sample belonged to the normal law with the shift parameter $\theta_0 = 0$ and the scale parameter $\theta_1 = 1.1$, the competing hypothesis $H_3$ being the last sample belonged to the logistic law with the density of

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

and parameters $\theta_0 = 0$ and $\theta_1 = 1$.

The power was evaluated on the results of modeling statistic distributions with the tested $G(S|H_0)$ being true, and competing hypotheses $G(S|H_1)$, $G(S|H_2)$ and $G(S|H_3)$ having equal volumes of $n_i$ compared samples. As an example, Tables 5 and 6 show evaluation of the power of tests with $\alpha = 0.1$ for $k = 3$ and $k = 4$ correspondingly. In the case of the homogeneity test $\chi^2$ the unified sample was split into $r = 10$ equifrequent intervals.

Table 5: **Assessment of the power of test against alternatives $H_1$, $H_2$ and $H_3$, $k=3$, $n_i=n$**

| Test | $n$ =20 | $n$ =50 | $n$ = 100 | $n$ = 300 | $n$ = 500 | $n$ = $10^3$ |
|---|---|---|---|---|---|---|
| Against alternative hypothesis $H_1$ | | | | | | |
| $S_{\max}^{AD}$ | 0.113 | 0.134 | 0.171 | 0.314 | 0.450 | 0.712 |
| AD | 0.113 | 0.134 | 0.171 | 0.313 | 0.449 | 0.711 |
| $S_{\max}^{LR}$ | 0.114 | 0.134 | 0.168 | 0.306 | 0.437 | 0.694 |
| $S_{\max}^{Sm}$ | 0.110 | 0.128 | 0.155 | 0.272 | 0.383 | 0.622 |
| $Z_C$ | 0.113 | 0.131 | 0.160 | 0.273 | 0.380 | 0.612 |
| $Z_A$ | 0.112 | 0.130 | 0.158 | 0.268 | 0.371 | 0.599 |
| $Z_K$ | 0.110 | 0.125 | 0.144 | 0.231 | 0.321 | 0.525 |
| $\chi^2$ | 0.100 | 0.108 | 0.120 | 0.173 | 0.226 | 0.385 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Against alternative hypothesis $H_2$ | | | | | | |
| $Z_C$ | 0.107 | 0.125 | 0.160 | 0.319 | 0.475 | 0.771 |
| $Z_{\mathrm{A}}$ | 0.107 | 0.126 | 0.162 | 0.319 | 0.470 | 0.767 |
| $Z_{\mathrm{K}}$ | 0.107 | 0.123 | 0.147 | 0.263 | 0.376 | 0.621 |
| AD | 0.104 | 0.111 | 0.124 | 0.191 | 0.273 | 0.509 |
| $\chi^2$ | 0.105 | 0.114 | 0.129 | 0.202 | 0.277 | 0.495 |
| $S_{\max}^{AD}$ | 0.102 | 0.107 | 0.114 | 0.165 | 0.231 | 0.446 |
| $S_{\max}^{Sm}$ | 0.103 | 0.104 | 0.114 | 0.136 | 0.164 | 0.253 |
| $S_{\max}^{LR}$ | 0.103 | 0.104 | 0.108 | 0.127 | 0.152 | 0.241 |
| Against alternative hypothesis $H_3$ | | | | | | |
| $Z_{\mathrm{A}}$ | 0.103 | 0.108 | 0.116 | 0.181 | 0.279 | 0.580 |
| $Z_C$ | 0.103 | 0.108 | 0.116 | 0.176 | 0.270 | 0.568 |
| $Z_{\mathrm{K}}$ | 0.104 | 0.110 | 0.117 | 0.170 | 0.233 | 0.423 |
| $\chi^2$ | 0.100 | 0.113 | 0.121 | 0.173 | 0.226 | 0.382 |
| AD | 0.103 | 0.107 | 0.114 | 0.148 | 0.189 | 0.315 |
| $S_{\max}^{Sm}$ | 0.102 | 0.105 | 0.111 | 0.148 | 0.183 | 0.288 |
| $S_{\max}^{AD}$ | 0.102 | 0.104 | 0.110 | 0.134 | 0.166 | 0.272 |
| $S_{\max}^{LR}$ | 0.103 | 0.104 | 0.107 | 0.124 | 0.145 | 0.218 |

Table 6: **Assessment of the power of test against alternatives $H_1$, $H_2$ and $H_3$, $k=4$, $n_i=n$**

| Test | $n$ =20 | $n$ =50 | $n$ = 100 | $n$ = 300 | $n$ = 500 | $n$ = $10^3$ |
|---|---|---|---|---|---|---|
| Against alternative hypothesis $H_1$ | | | | | | |
| $S_{\max}^{AD}$ | 0.112 | 0.131 | 0.165 | 0.302 | 0.438 | 0.706 |
| AD | 0.112 | 0.131 | 0.164 | 0.301 | 0.433 | 0.701 |
| $S_{\max}^{LR}$ | 0.113 | 0.130 | 0.162 | 0.293 | 0.425 | 0.686 |
| $S_{\max}^{Sm}$ | 0.111 | 0.125 | 0.151 | 0.261 | 0.366 | 0.605 |
| $Z_C$ | 0.111 | 0.126 | 0.155 | 0.260 | 0.368 | 0.595 |
| $Z_{\mathrm{A}}$ | 0.111 | 0.127 | 0.153 | 0.255 | 0.360 | 0.579 |
| $Z_{\mathrm{K}}$ | 0.109 | 0.121 | 0.141 | 0.219 | 0.300 | 0.502 |
| $\chi^2$ | 0.102 | 0.109 | 0.118 | 0.167 | 0.221 | 0.358 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Against alternative hypothesis $H_2$ | | | | | | |
| $Z_C$ | 0.106 | 0.122 | 0.158 | 0.306 | 0.468 | 0.761 |
| $Z_A$ | 0.107 | 0.124 | 0.158 | 0.305 | 0.463 | 0.745 |
| $Z_K$ | 0.106 | 0.120 | 0.145 | 0.249 | 0.367 | 0.606 |
| AD | 0.104 | 0.110 | 0.123 | 0.180 | 0.254 | 0.474 |
| $\chi^2$ | 0.107 | 0.113 | 0.127 | 0.189 | 0.271 | 0.458 |
| $S_{\max}^{AD}$ | 0.101 | 0.104 | 0.111 | 0.145 | 0.195 | 0.381 |
| $S_{\max}^{Sm}$ | 0.102 | 0.105 | 0.108 | 0.128 | 0.153 | 0.221 |
| $S_{\max}^{LR}$ | 0.102 | 0.103 | 0.105 | 0.118 | 0.135 | 0.197 |
| Against alternative hypothesis $H_3$ | | | | | | |
| $Z_A$ | 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 |
| $Z_K$ | 0.103 | 0.107 | 0.114 | 0.161 | 0.222 | 0.410 |
| $\chi^2$ | 0.102 | 0.110 | 0.116 | 0.164 | 0.218 | 0.357 |
| AD | 0.102 | 0.106 | 0.113 | 0.143 | 0.179 | 0.291 |
| $S_{\max}^{Sm}$ | 0.103 | 0.104 | 0.112 | 0.138 | 0.166 | 0.257 |
| $S_{\max}^{AD}$ | 0.101 | 0.103 | 0.107 | 0.124 | 0.147 | 0.229 |
| $S_{\max}^{LR}$ | 0.102 | 0.102 | 0.105 | 0.116 | 0.130 | 0.183 |

Thus-conducted power analysis of $k$ -samples tests allows making some conclusions.

The tests can be organized power-wise with respect to changes in the shift parameter in the following way:

$$S_{\max}^{AD} \succ \mathrm{AD} \succ S_{\max}^{LR} \succ S_{\max}^{Sm} \succ Z_C \succ Z_{\mathrm{A}} \succ Z_{\mathrm{K}} \succ \chi^2 .$$

With respect to changes in the scale parameter:

$$Z_C \succ Z_{\mathrm{A}} \succ Z_{\mathrm{K}} \succ \mathrm{AD} \succ \chi^2 \succ S_{\max}^{AD} \succ S_{\max}^{Sm} \quad S_{\max}^{LR} .$$

At that, the Zhang tests of $Z_{\mathrm{A}}$ and $Z_C$ statistics are almost equivalent power-wise, and the Anderson-Darling test is noticeably inferior to the Zhang tests.

The tests can be organized power-wise with respect to situations when all but one sample belongs to the normal law and the last one belongs to the logistic law, in the following way:

$$Z_{\mathrm{A}} \succ Z_C \succ Z_{\mathrm{K}} \succ \chi^2 \succ \mathrm{AD} \succ S_{\max}^{Sm} \succ S_{\max}^{AD} \succ S_{\max}^{LR} .$$

It can be noted that with the increase in the number of compared samples of the same volumes the power of the criterion relative to similar competing hypotheses decreases as a rule, which is absolutely natural. It is more difficult to single out a situation and to give preference to a competing hypothesis, when only one of the analyzed samples belongs to some other law.

We can't but mention that the Zhang tests with statistics of $Z_{\mathrm{K}}$, $Z_{\mathrm{A}}$, $Z_C$ possess quite substantial advantage in power with respect to some alternatives.

# 4. Application examples

Let's view the application of the homogeneity tests to analyze the 3 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 results of testing the hypothesis of homogeneity of the three samples under consideration are shown in table 7. The $p_{value}$ for the Anderson-Darling test was calculated in accordance with beta distribution of type III as taken from the Table for $k=3$. The $p_{value}$ values for the Zhang tests were found on the basis of statistic modeling conducted in interactive mode and with the number of simulation experiments $N=10^6$. For the test with $S_{\max}^{Sm}$ statistic the $p_{value}$ value at $k=3$ was calculated in accordance with beta distribution of type III from Table 2, for that with $S_{\max}^{LR}$ statistic – in accordance with the Sb-Johnson distribution from Table 3, and for that with $S_{\max}^{AD}$ statistic – in accordance with beta distribution of type III from Table 4.

It must be noted that the Anderson-Darling $S_{\max}^{AD}$ and the Lehmann-Rosenblatt $S_{\max}^{LR}$ tests noted the maximum deviation between the 1st and the 2nd samples, and the Smirnov $S_{\max}^{Sm}$ test noted that between the 2nd and the 3rd samples. The overall result is that the hypothesis of the homogeneity of the 3 samples should be rejected.

Table 7: **Results of the homogeneity of 3 samples**

| Test | Statistic | $p_{value}$ |
|---|---|---|
| $k$-sample 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 |
| $S_{\max}^{AD}$ Anderson-Darling | 5.19801 | 0.0064 |
| $S_{\max}^{LR}$ Lehmann-Rosenblatt | 0.9650 | 0.0094 |
| $S_{\max}^{Sm}$ Smirnov (modified) | 1.72566 | 0.0144 |
| $\chi^2$, $r=10$ | 25.556 | 0.1104 |
| $\chi^2$, $r=8$ | 19.200 | 0.1574 |
| $\chi^2$, $r=7$ | 21.627 | 0.0419 |

We may note that $\chi^2$ homogeneity tests results depend quite essentially on the number of intervals $r$ chosen.

In this case, the test results were quite predictable, as samples 1 and 3 were simulated in accordance with the standard normal law, and the obtained values of pseudo-random values were rounded to 3 significant digits after the decimal point. The second sample was obtained in accordance with the normal law with the shift value of 0.5 and the standard deviation of 1.1.

## 5 Conclusions

The constructed models of statistic limiting distributions for *k*-samples homogeneity tests (the Anderson-Darling ones and those proposed in this paper) allows obtaining correct and informational conclusions on and calculating the tests significance $p_{value}$. Software can is available for this purpose [18].

## 6 Acknowledgements

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. **Kiefer J.** K-samples Analogues of the Kolmogorov–Smirnov and Cramer-v. Mises Tests // Annals of Mathematical Statistics. 1959. Vol. 30. No. 2. – P. 420-447. URL: http://www.jstor.org/stable/2237091

2. **Conover W.J.** Several k-samples Kolmogorov–Smirnov tests // The Annals of Mathematical Statistics. 1965. Vol. 36, No. 3. – P.1019-1026. URL: http://www.jstor.org/stable/2238210

3. **Conover W.J.** Practical Nonparametric Statistics. – 3d ed. – Wiley, 1999. – 584 p.

4. **Scholz F.W., Stephens M.A**. K-samples Anderson–Darling Tests // Journal of the American Statistical Association. 1987. Vol. 82. No. 399. – P. 918-924. DOI: 10.1080/01621459.1987.10478517

5. **Zhang J.** Powerful goodness-of-fit and multi-sample tests // PhD Thesis. York University, Toronto. 2001. – 113 p. URL: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf (accessed 28.01.2013).

6. **Zhang J.** Powerful Two-Sample Tests Based on the Likelihood Ratio // Technometrics. 2006. V. 48. No. 1. – P.95-103. DOI: 10.1198/004017005000000328

7. **Zhang J., Wu Y.** k-samples tests based on the likelihood ratio // Computational Statistics & Data Analysis. 2007. V. 51. No. 9. – P. 4682-4691. DOI: 10.1016/j.csda.2006.08.029

8. **Bolshev L.N., Smirnov N.V.** Tables of mathematical statistic. – M.: Nauka, 1983. – 416 pp.

9. **Lehmann E. L.** Consistency and unbiasedness of certain nonparametric tests // Ann. Math. Statist. 1951. Vol. 22, № 1. – P. 165–179. URL: http://www.jstor.org/stable/2236420

10. Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Ann. Math. Statist. 1952. Vol. 23. – P. 617–623. URL: https://projecteuclid.org/euclid.aoms/1177729341

11. **Pettitt A.N.** A two-sample Anderson-Darling rank statistic // Biometrika. 1976. Vol. 63. No.1. P. 161-168. DOI: 10.1093/biomet/63.1.161

12. **Wolf E.H., Naus J.I.** Tables of critical values for a *k*-sample Kolmogorov–Smirnov test statistic // J. Amer. Statist. Assoc., 1973. Vol. 68. – P.994–997. DOI: 10.1080/01621459.1973.10481462

13. **Lemeshko B.Y.** Tests for homogeneity. Guide on the application. M: INFRA–M, 2017. – 207 pp. DOI: 10.12737/22368

14. Lemeshko B.Y. On application of distribution laws homogeneity tests / **B.Y. Lemeshko, S.B. Lemeshko, I.V. Veretelnikova** // Tomsk State University Journal of Control and Computer Science. 2017. No. 41. – P. 24-31. DOI: 10.17223/19988605/41/3

15. **Lemeshko B.Y., Veretelnikova I.V., Lemeshko S.B., Novikova A.Y**. (2017) Application of Homogeneity Tests: Problems and Solution. In: Rykov V., Singpurwalla N., Zubkov A. (eds) Analytical and Computational Methods in Probability Theory. ACMPT 2017. Lecture Notes in Computer Science, vol 10684. Springer, Cham. – P. 461-475. DOI: 10.1007/978-3-319-71504-9_38

16. **Lemeshko B.Yu., Lemeshko S.B.** Statistical distribution convergence and homogeneity test power for Smirnov and Lehmann–Rosenblatt tests // Measurement Techniques, 2005. V. 48, № 12. – P.1159-1166. DOI: 10.1007/s11018-006-0038-3

17. **Cramèr H.** Mathematical methods of statistics. – M.: Mir, 1975. – 648 pp.

18. ISW–Software for statistic analysis of one-dimensional observations. https://ami.nstu.ru/~headrd/ISW.htm. (accessed date 17.03.2018)

# Thank you for attention!

**Extended Abstract**

Present work illustrates the dependence of the distributions of $k$-sample tests statistics for the homogeneity of laws of distribution on the volume of samples and the number $k$ of compared samples.

For the $k$-sample Anderson–Darling test (Scholz, 1987), the authors construct a model of limit distributions of statistics for $k = 3,..., 11$.

Variants of $k$-sample tests based on the use of 2-sample Smirnov, Lehmann-Rosenblatt and Anderson–Darling tests for the analysis of all combinations of sample pairs are proposed. The maximum value of the statistics of the 2-sample test obtained during the analysis of combinations of pairs of samples is considered as a statistic of $k$-sample test. The constructed models for limit distributions of statistics of the proposed tests for $k$ =3,...,11 are given.

The constructed models make it possible to carry out correct and informative conclusions with the calculation of p-value using the corresponding tests.

Power estimates of the studied tests are presented in relation to some competing hypotheses, which allows to order $k$-sample tests by preference with respect to different alternatives.

The studies were based on the intensive use of the Monte-Carlo method in the simulation of distributions of tests statistics.