Measurement Techniques, Vol. 48, No. 12, 2005

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

STATISTICAL DISTRIBUTION CONVERGENCE AND HOMOGENEITY TEST POWER FOR SMIRNOV AND LEHMANN–ROSENBLATT TESTS

B. Yu. Lemeshko and S. B. Lemeshko

UDC 519.24

The distributions are examined for the statistics for the Smirnov and Lehmann–Rosenblatt homogeneity tests for restricted sample volumes and the convergence of them in the limit. A correction is proposed that provides closeness of the Smirnov statistic distribution to the limiting Kolmogorov one. The power of the tests is compared in relation to various alternatives.

Key words: *Smirnov and Lehmann–Rosenblatt homogeneity tests, test power.*

It is necessary to check whether two sets of random quantities belong to the same population in the analysis of the random errors of measuring instruments and in statistical management of process quality. The corresponding statistical test is just as important as the means of measurement.

There are three groups of statistical tests for homogeneity: for checking hypotheses on the means (mathematical expectations and medians), for checking hypotheses on the spread characteristics (variances and ranges), and for checking hypotheses on the distributions.

Although there are many publications, it remains uncertain when each test is preferable and the most powerful.

For example, for various tests on hypotheses concerning the means, one has information on the relative stability of the statistic distributions under deviations in the observed law from a gaussian distribution [1, 2].

On the other hand, tests on hypotheses about the variances are always sensitive to any deviations from the assumptions on which they were made [3, 4].

Tests for hypotheses on the distributions of two sets of data as a rule employ either the Smirnov test or the Lehmann–Rosenblatt one [3]. In [5], there is a discussion of the preference in checking homogeneity by means of those tests.

Here we examine the distributions of the statistics and the power of the Smirnov and Lehmann–Rosenblatt tests with restricted sample volumes. Computer simulation and statistic regularity analysis are used [1, 4], on the basis of Monte Carlo simulation and appropriate software.

One can check the homogeneity of two ranked sets of size $m \le n$ as represented by the variation series

$$x_1 < x_2 < \ldots < x_m \quad \text{and} \quad y_1 < y_2 < \ldots < y_n,$$

and testing hypothesis H_0 that the two samples derived from the same population, i.e., $F(x) = G(x)$ for any $x = y$.

Smirnov Test. This test [6] involves the assumption that the distributions $F(x)$ and $G(x)$ are continuous. The Smirnov test statistic characterizes the maximum difference between the empirical distributions:

$$D_{m,n} = \sup_x \left| G_m(x) - F_n(x) \right|.$$

Translated from Izmeritel'naya Tekhnika, No. 12, pp. 9–14, December, 2005. Original article submitted June 1, 2005.

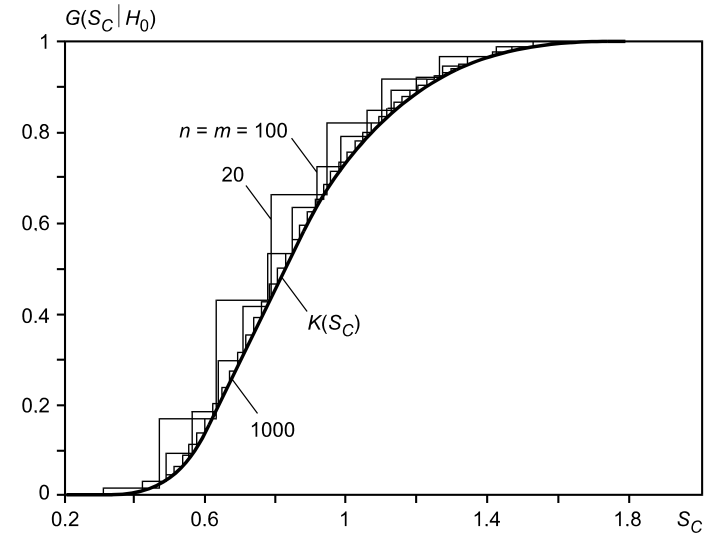

Fig. 1. Distributions of the statistic (1) with H_0 correct in accordance with m and n.

The actual values are calculated in accordance with the following [3]:

$$D_{m,n} = \max(D^+_{m,n}, D^-_{m,n});$$

$$D^+_{m,n} = \max_{1 \le r \le m}\left[\frac{r}{m} - F_n(x_r)\right] = \max_{1 \le s \le n}\left[G_m(y_s) - \frac{s-1}{n}\right];$$

$$D^-_{m,n} = \max_{1 \le r \le m}\left[F_n(x_r) - \frac{r-1}{m}\right] = \max_{1 \le s \le n}\left[\frac{s}{n} - G_m(y_s)\right].$$

If H_0 is correct, i.e., $\lim_{m\to\infty} P\left\{\sqrt{\frac{mn}{m+n}} D_{m,n} < s\right\} = K(S)$, and the statistic

$$S_C = \sqrt{\frac{mn}{m+n}} D_{m,n} \tag{1}$$

in the limit follows a Kolmogorov distribution $K(s)$ [3]. However, for restricted values of m and n, then statistics $D^+_{m,n}$ and $D^-_{m,n}$ are discrete, and the set of possible values constitutes a lattice with pitch $1/k$, in which k is the least common multiple of m and n. For m, $n \le 20$, tables of the percentage points of $D_{m,n}$ have been given in [3]. The conditional distribution $G(S_C|H_0)$ of statistic S_C if H_0 is correct converges slowly to $K(S)$ and differs substantially from it for not very large m and n. Asymptotic formulas for $D^+_{m,n}$ and $D^-_{m,n}$ have been considered in [7–9].

Figure 1 shows the conditional distributions for the (1) statistic when H_0 is correct as functions of m and n (for $m = n$). Even with $m = 1000$ and $n = 1000$, the stepped structure of $G(S_C|H_0)$ persists. Another deficiency of using (1) is (see Fig. 1) that the distribution $G(S_C|H_0)$ approximates to the limiting distribution $K(S)$ from the left as m and n increase.

The smoothness of the statistic distribution is dependent on k, so it is best to use the test when m and n are not equal and constitute mutually prime numbers. In that case, the least common multiple of m and n is maximal and equal to $k = mn$, and the distribution more closely resembles a continuous one. Then for small and moderate values of m and n, there is a substantial difference between $G(S_C|H_0)$ and the limiting $K(S)$, since $G(S_C|H_0)$ is appreciably shifted to the left from $K(S)$.

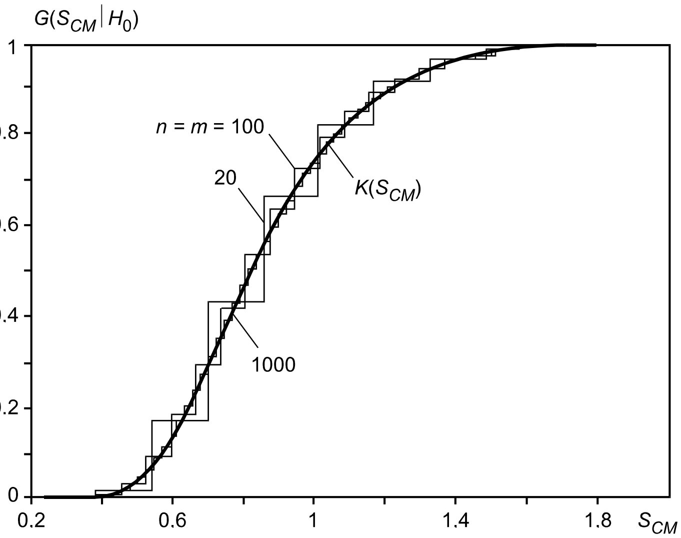

Fig. 2. Distributions of the statistic (2) with H_0 correct in accordance with m and n.

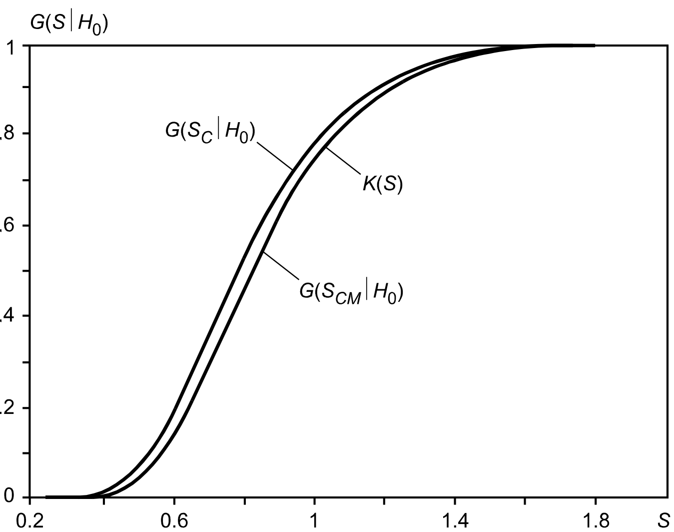

Fig. 3. Distributions of the (1) and (2) statistics with H_0 correct; m = 61 and n = 53.

In that connection, one proposes the following simple model for the (1) statistic:

$$S_{CM} = \sqrt{\frac{mn}{m+n}}\left(D_{m,n} + \frac{m+n}{4.6mn}\right), \tag{2}$$

in which that last deficiency is almost absent, as Fig. 2 shows.

The smoothness is dependent on k. Figure 3 gives the limiting Kolmogorov distribution $K(S)$ and the empirical distributions $G(S_C|H_0)$ of the (1) statistic and $G(S_{CM}|H_0)$ of the (2) statistic with m = 61 and n = 53 as obtained by simulation. The (1) statistic differs substantially from $K(S)$, while the distribution of the (2) statistic is visually fitting. The sample volumes for the simulated statistics in this case, as in all others in this paper, constitute 10000. When one checks the empir-

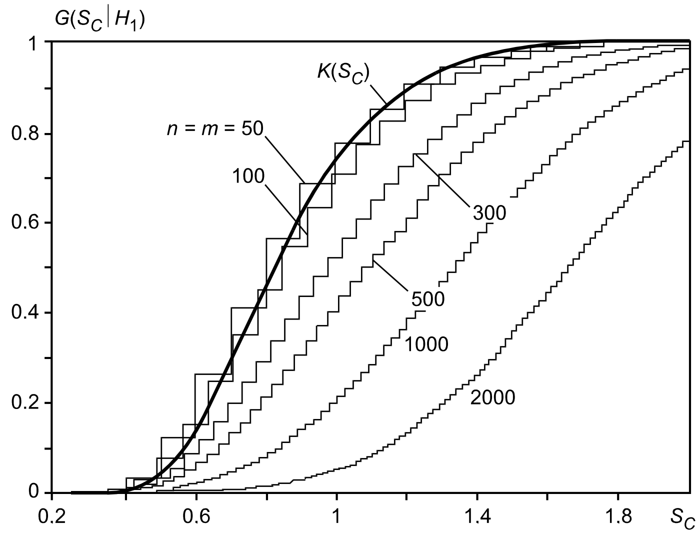

Fig. 4. Distributions of the (1) statistic with H_1 correct.

ical distribution for the (2) statistic against the Kolmogorov distribution, the significance levels attained on the corresponding tests are as follows: 0.72 on the χ^2 test (with 10 equally probable intervals), 0.83 on the Kolmogorov test, 0.97 on the Kramer–Mises–Smirnov ω^2 test, and 0.94 on the Ω^2 Anderson–Darling test.

Using mutually prime m and n in the Smirnov test with the (2) statistic makes it better to calculate the attained significance level $P\{S_{CM} > S_{CM}^*\} = 1 - G(S_{CM}^* | H_0)$, in which S_{CM}^* is the value of the (2) statistic found on testing hypothesis H_0 with the particular sets in accordance with the Kolmogorov distribution: $P\{S_{CM} > S_{CM}^*\} \approx 1 - K(S_{CM}^*)$. Correspondingly, one is more justified in using the percentage points (quantiles) of the Kolmogorov distribution in the test. This cannot be said about the Smirnov test with the (1) statistic because in that case the physical values as determined from the Kolmogorov distribution are larger than the true ones. Consequently, the hypothesis may be adopted without justification (not rejected).

The coefficient 4.6 in the (2) statistic has been selected empirically for sample volumes up to $m = n = 1000$, but for large values of the least common multiple, m and n are mutually prime numbers, e.g., with $m = 641$ and $n = 643$, the 4.6 should be replaced by 3.4

In what follows, the power of the Smirnov test is considered for the distribution of the (1) statistic. However, all conclusions on the power apply also for tests with the (2) statistic, since all the distributions with identical sample volumes are shifted by the same amount, as (2) shows.

One can check the simulation results against the available theoretical ones [3], and one finds complete coincidence for the simulation critical values with the exact ones for the statistic $D_{m,n}$ in [3].

Here the power of the homogeneity tests is examined for a series of alternatives. For definiteness, H_0 corresponded to the samples belonging to the same gaussian standard distribution with density

$$f(x) = \frac{1}{\theta_2\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2^2}\right\}, \quad \theta_1 = 0, \quad \theta_2 = 1. \tag{3}$$

In all the alternatives, the first set always corresponds to the (3) distribution with parameters $\theta_1 = 0$ and $\theta_2 = 1$, while the second corresponds to something different: in the case of hypothesis H_1, a gaussian distribution with $\theta_1 = 0.1$ and $\theta_2 = 1$; in the case of H_2, with $\theta_1 = 0.5$ and $\theta_2 = 1$; in the case of H_3, with $\theta_1 = 0$ and $\theta_2 = 1.1$; and in the case of H_4, with $\theta_1 = 0$ and $\theta_2 = 1.5$. In the case of hypothesis H_5, the second set corresponded to the logit distribution with density

TABLE 1. Power of the Smirnov Homogeneity Test with Respect to the Alternatives H_1–H_5 in Relation to Sample Volumes ($m = n$)

Significance level α	Power for n of						
	20	50	100	300	500	1000	2000
			Relative to alternative H_1				
0.1	0.0937	0.1480	0.1766	0.2775	0.3806	0.6171	0.8688
0.05	0.0410	0.0569	0.0944	0.1883	0.2682	0.4899	0.7762
0.025	0.0410	0.0344	0.0505	0.1163	0.1829	0.3859	0.6737
			Relative to alternative H_2				
0.1	0.3457	0.7200	0.9332	1	1	1	1
0.05	0.2202	0.5341	0.8722	0.9996	1	1	1
0.025	0.2202	0.4328	0.7842	0.9992	1	1	1
			Relative to alternative H_3				
0.1	0.0884	0.1229	0.1257	0.1466	0.1856	0.2967	0.5508
0.05	0.0352	0.0458	0.0630	0.0789	0.1024	0.1677	0.3520
0.025	0.0352	0.0257	0.0280	0.0410	0.0518	0.0967	0.2098
			Relative to alternative H_4				
0.1	0.1396	0.2986	0.5213	0.9609	0.9989	1	1
0.05	0.0570	0.1268	0.3161	0.8977	0.9952	1	1
0.025	0.0570	0.0763	0.1689	0.7738	0.9786	1	1
			Relative to alternative H_5				
0.1	0.0836	0.1209	0.1308	0.1568	0.1976	0.3191	0.5639
0.05	0.0341	0.0455	0.0673	0.0891	0.1158	0.1879	0.3754
0.025	0.0341	0.0258	0.0316	0.0471	0.0618	0.1119	0.2390

TABLE 2. Actual Significance Levels of the Smirnov Homogeneity Test Corresponding to the $1 - \alpha$ Quantiles of the Kolmogorov Distribution in Accordance with the Sample Volumes ($m = n$)

Given significance level α	Actual significance level α for n of						
	20	50	100	300	500	1000	2000
0.1	0.0835	0.1120	0.1085	0.0927	0.0970	0.0980	0.1041
0.05	0.0334	0.0410	0.0543	0.0496	0.0514	0.0471	0.0480
0.025	0.0334	0.0240	0.0252	0.0254	0.0238	0.0259	0.0245

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

and parameters $\theta_1 = 0$ and $\theta_2 = 1$. A gaussian distribution and a logit one are very similar and difficult to distinguish by means of fishing tests. Figure 4 shows the conditional distributions obtained by simulation with the $G(S_C | H_1)$ statistic with H_1 correct, from which one can estimate the values of the power for various sample volumes m and n.

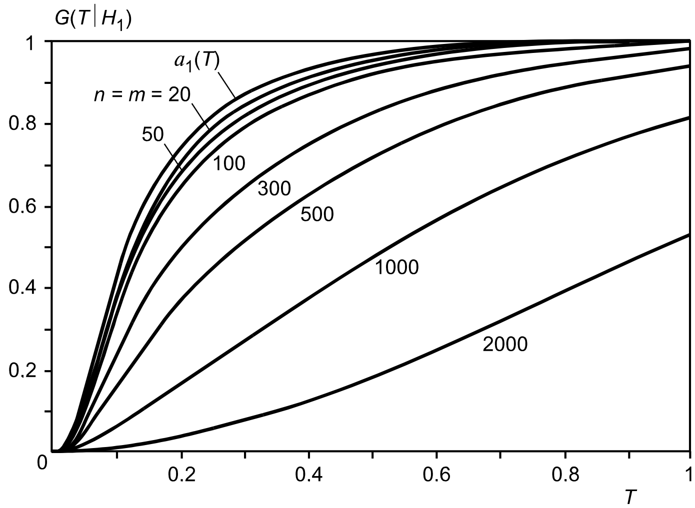

Fig. 5. Distributions of the (4) statistic for H_1 correct.

Similarly, with various sample volumes one can construct the conditional distributions of the (1) statistic with others of these alternatives: $G(S_C | H_2)$, $G(S_C | H_3)$, $G(S_C | H_4)$, $G(S_C | H_5)$. These distributions and the limiting distribution for the statistic $G(S_C | H_0) = K(S_C)$ were used in calculating the power of the test under various alternatives. We derived the values $1 - \beta$ of the power for the Smirnov test, in which β is the probability of an error of the second kind with respect to the competing hypotheses H_1–H_5 with various sample volumes for significance levels (error probabilities of the first kind) $\alpha = \{0.1; 0.05; 0.025\}$ as given in Table 1.

The power figures given in Table 1 are obtained relative to the $1 - \alpha$ quantiles of the limiting Kolmogorov distribution $K(S)$. The $G(S_C | H_0)$ distribution for the (1) statistic differs substantially from $K(S)$, so the actual significance levels differ from the given ones. Table 2 gives the actual significance levels for a Smirnov test corresponding to the power values given in Table 1. As $G(S_C | H_0)$ is stepped, the actual values of α differ quite markedly from the given ones for small sample volumes. For example, with $m = n = 20$ and a given significance level 0.1, there is an actual significance level of 0.0835.

Lehmann–Rosenblatt Homogeneity Test. This was proposed in [10] and was examined in [11]; it is a test of ω^2 type. The statistic takes the form [3]

$$T = \frac{mn}{m+n} \int_{-\infty}^{\infty} [G_m(x) - F_n(x)]^2 \, dH_{m+n}(x),$$

in which $H_{m+n}(x) = \frac{m}{m+n} G_m(x) + \frac{n}{m+n} F_n(x)$ is the empirical distribution constructed from the variation series for the union of the two samples and employed in the form

$$T = \frac{1}{mn(m+n)} \left[n \sum_{i=1}^{n} (r_i - i)^2 + m \sum_{i=1}^{m} (s_i - j)^2 \right] - \frac{mn-1}{6(m+n)}, \qquad (4)$$

in which r_i is the rank of y_i, s_j being the rank of x_j in the combined variational series.

It has been shown [11] that the (4) statistic converges to a limiting distribution for a statistic of the ω^2 Mises fit test $a_1(T)$ [3].

TABLE 3. Power of the Lehmann–Rosenblatt Homogeneity Test with Respect to the Alternatives H_1–H_5 in Relation to the Sample Volumes ($m = n$)

Significance level α	Power for n of						
	20	50	100	300	500	1000	2000
			Relative to alternative H_1				
0.1	0.1241	0.1382	0.1727	0.3125	0.4369	0.6874	0.9114
0.05	0.0615	0.0770	0.0999	0.2078	0.3211	0.5703	0.8469
0.025	0.0324	0.0410	0.0590	0.1333	0.2288	0.4589	0.7681
			Relative to alternative H_2				
0.1	0.4321	0.7628	0.9549	1	1	1	1
0.05	0.3121	0.6473	0.9154	1	1	1	1
0.025	0.2120	0.5355	0.8661	0.9998	1	1	1
			Relative to alternative H_3				
0.1	0.1096	0.1107	0.1147	0.1459	0.1898	0.3265	0.6237
0.05	0.0508	0.0567	0.0563	0.0691	0.0945	0.1675	0.3986
0.025	0.0252	0.0291	0.0283	0.0334	0.0442	0.0805	0.2259
			Relative to alternative H_4				
0.1	0.1655	0.2875	0.5513	0.9875	0.9999	1	1
0.05	0.0801	0.1437	0.3199	0.9470	0.9993	1	1
0.025	0.0361	0.0727	0.1687	0.8587	0.9952	1	1
			Relative to alternative H_5				
0.1	0.1087	0.1069	0.1135	0.1422	0.1826	0.2978	0.5463
0.05	0.0511	0.0549	0.0581	0.0668	0.0910	0.1450	0.3390
0.025	0.0241	0.0276	0.0290	0.0332	0.0431	0.0712	0.1822

A difference from the Smirnov test statistic is that the distribution of statistic T converges rapidly to the limiting one $a_1(T)$ [3]. Simulation of the conditional distributions $G(T|H_0)$ for the (4) statistic with various sample volumes has shown that already for $m = n = 20$ the distribution of the $G(T|H_0)$ statistic is very close to $\alpha_1(T)$, while there is practical coincidence for $m = 100$ and $n = 100$.

Figure 5 shows the simulated conditional distributions for the $G(T|H_1)$ statistic when H_1 is correct. These distributions allow one to estimate the power of the Lehmann–Rosenblatt test with various m and n.

Similarly, distributions were constructed for the statistics $G(T|H_2)$, $G(T|H_3)$, $G(T|H_4)$, $G(T|H_5)$ when the corresponding competing hypotheses are correct. Table 3 gives the calculated values of the power $1 - \beta$ of the Lehmann–Rosenblatt test.

Conclusions. One can compare the power of the test with respect to the alternatives on the basis of the actual significance levels for the Smirnov test (Table 2), which shows that usually the power of the Lehmann–Rosenblatt test is appreciably higher than that of the Smirnov one. However, the power of the Smirnov test is somewhat greater for very closely similar alternatives (see the power for alternative H_5), which is intuitively evident if one remembers that the deviation measure is linear in the Smirnov test, while in the Lehmann–Rosenblatt one it is quadratic.

One is usually dealing with samples of quite restricted or more often small volume in processing measurements concerned with statistical quality control. Homogeneity tests cannot distinguish similar alternatives because of their low power with small sample volumes. Consequently, any hypothesis on the homogeneity of samples, even if incorrect, very often will not be rejected. A shift of 0.1σ or an increase in the spread parameter by 10% with small sample volumes is unlikely to be detected by homogeneity tests, but large deviations in the laws corresponding to the samples will be detected. For example, with the Lehmann–Rosenblatt test, the probabilities of errors of the first kind α and second kind β do not exceed 0.1 with a shift of 0.1σ (alternative H_1), and the sample volume should be about 2000. With a shift of 0.5σ (alternative H_2), the error probabilities do not exceed 0.1 with sample volumes not more than 100.

As the distribution for the (4) statistic converges very rapidly to the $\alpha_1(T)$ distribution, it is correct to use it as the distribution for the statistic for the Lehmann–Rosenblatt test for small m and n.

In the case of the Smirnov test, the stepped distribution for the (1) statistic (especially for $m = n$) means that using the limiting Kolmogorov distribution $K(S)$ will be related to obtaining a very approximate value for the actual significance level (error probability of the first kind), and corresponding critical value. It is therefore recommended as follows for constructing homogeneity tests on the basis of a Smirnov test: 1) to select $m \neq n$ in such a way that they are mutually prime numbers, and their least common multiple k is maximal and equal to mn; 2) to use a statistic of the (2) form in the Smirnov test. Then it will be correct to use the Kolmogorov distribution as the distribution for the (2) statistic for the Smirnov test with relatively small m and n.

It is thus best to use both tests in homogeneity checking.

The research was performed with support from the Russian Federation Ministry of Education and Science (projects No. 15378 and No. 02.442.11.7103).

REFERENCES

1. B. Yu. Lemeshko and S. S. Pomadin, *Metrology: Supplement to the Izmer. Tekh. J.* [in Russian], No. 3, 3 (2004).
2. B. Yu. Lemeshko, S. B. Lemeshko, and E. P. Mirkin, *Proceedings of the 7th International Conference on Current Topics in Electronic Instrumentation APEP-2004*, Vol. 6 [in Russian], Novosibirsk (2004), p. 269.
3. L. N. Bol'shev and N. V. Smirnov, *Tables for Mathematical Statistics* [in Russian], Nauka, Moscow (1983).
4. B. Yu. Lemeshko and E. P. Mirkin, *Izmer. Tekh.*, No. 10, 10 (2004).
5. A. I. Orlov, *Zavod. Labor.*, **69**, No. 1, 55 (2003).
6. N. V. Smirnov, *Byul. MGU, Ser. A*, **2**, No. 2, 3 (1939).
7. A. A. Borovkov, *Izv. AN SSSR, Ser. Matemat.*, **26**, 605 (1962).
8. V. S. Korolyuk, *Probability Theory and Its Applications*, Vol. 4 [in Russian] (1959), p. 369.
9. N. V. Smirnov, *Tr. Matem. Inst. AN SSSR*, **64**, 185 (1961).
10. E. L. Lehmann, *Ann. Math. Statist.*, **22**, No. 1, 165 (1951).
11. M. Rosenblatt, *Ann. Math. Statist.*, **23**, 617 (1952).