Measurement Techniques, Vol. 49, No. 10, 2006

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

THE ABBÉ INDEPENDENCE TEST WITH DEVIATIONS FROM NORMALITY

S. B. Lemeshko

UDC 519.24

Distributions are examined for the Abbé test statistics for various probability laws. It is shown that the statistic distributions are stable when the normality assumptions are violated. The power of the criterion with respect to various alternatives is examined.

Key words: *Abbé test, independence criterion, criterion power.*

The Abbé test is intended for checking hypotheses of the type H_0: $E[\xi_1] = E[\xi_2] = ... = E[\xi_n]$, i.e., in checking that all the observed quantities $\xi_1, \xi_2, ..., \xi_n$ in a sample of volume n have identical mathematical expectations. A competing hypothesis (alternative) consists in $|E[\xi_{i+1}] - E[\xi_i]| > 0$ for all values or certain of them $i = 1, 2, ..., n - 1$. The test is often used for checking for the absence of systematic changes in a series of measurements.

The Abbé test statistic in its current form [1] is represented by

$$S_A = \frac{1}{2}\left(\sum_{i=1}^{n-1}(\xi_{i+1} - \xi_i)^2 \Bigg/ \sum_{i=1}^{n}(\xi_i - \bar{\xi})^2\right), \tag{1}$$

where $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n}\xi_i$.

It is assumed that $\xi_1, ..., \xi_n$ are mutually independent normally distributed random quantities with identical but unknown dispersions. If a certain alternative is true, then the denominator in the statistic S_A is larger than the numerator and the values of the statistic will as a rule be less than those that are observed when the basic hypothesis of equality for the means is obeyed.

The conditional distribution $G(S_{A_n}|H_0)$ of the (1) statistic when H_0 is true is dependent on the sample volume n, the symmetry with respect to 1, and is determined in the interval $1 \pm \cos(\pi/n)$ [2]; as the sample volume increases, the distribution is closely fitted by a normal one with shift parameter 1 and with standard deviation for volumes $n > 20$ equal to [1]:

$$[(n - 2)/(n^2 - 1)]^{1/2}, \tag{2}$$

and for $n > 60$ [3], by

$$\{(n - 2)/[(n - 1)(n + 2)]\}^{1/2}. \tag{3}$$

Novosibirsk State Technical University; e-mail: skyer@phenix.bulat.info. Translated from Izmeritel'naya Tekhnika, No. 10, pp. 9–14, October, 2006. Original article submitted April 17, 2006.

0543-1972/06/4910-0962

TABLE 1. Significance Levels Attained on Checking the Conformity of a Distribution for the Statistic with a Normal Distribution with the Scale Parameters of (2) and (3)

Conformity test	$n = 10$		$n = 25$		$n = 100$		$n = 200$	
	(2)	(3)	(2)	(3)	(2)	(3)	(2)	(3)
Pearson χ^2	0.0000	0.0000	0.0543	0.0656	0.5849	0.2988	0.2219	0.1725
Kolmogorov	0.0009	0.0000	0.1480	0.0171	0.6073	0.4203	0.2857	0.2701
Mises ω^2	0.0022	0.0000	0.2379	0.0277	0.6134	0.4520	0.2076	0.1800
Mises Ω^2	0.0003	0.0000	0.1517	0.0116	0.5189	0.2935	0.1948	0.1549

The decision on rejecting the hypothesis on equality of the means is taken if the value of the statistic S_A is less than the critical $S_{A_n}(\alpha)$, where α is a given significance level, and the critical value is defined from $P\{S_A < S_{A_n}(\alpha)\} = \alpha$ in accordance with the conditional distribution for the statistic $G(S_{An} | H_0)$.

A basic assumption for using the Abbé test is that $\xi_1, ..., \xi_n$ is a normal distribution; in [4], it was found during research on the numerical characteristics of the (1) statistic distribution when hypothesis H_0 applies and the readings have various symmetrical distributions that the stability of these characteristics and the distribution of the Abbé statistic are unaffected by the assumption of normality for $\xi_1, ..., \xi_n$.

For many statistical tests, deviation from normality leads to substantial changes in the distribution of the test statistic. However, one can say that some parametric tests related to hypotheses on the means can be said to be stable under violation of that assumption (on deviation from a normal distribution within fairly wide limits). Published sources give analytical estimates on the changes in distribution for certain statistics in connection with certain deviations of the measurement errors from a normal law. Numerical studies that confirm this conclusion are to be found in [5]. There is high stability in tests for the homogeneity of the means of two samples (for example, Student's t tests for known and unknown variances) or for a series of samples (for example, the F test); as a rule, the distribution of the statistic deviates significantly from classical with unsymmetrical distributions for the observed random quantities or when there are heavy tails (for laws similar to the Cauchy distribution). The situation is analogous with the Abbé test.

Here I examine how various degrees of deviation from a normal distribution affect the test statistic distribution and give an evaluation of the power of a test in relation to certain alternatives. In particular, I examine what occurs with this statistic distribution if the distribution of the observed quantities is skewed or with heavy tails, or constitutes a symmetrical distribution differing to some extent from normal, or else a symmetrical multimode mixture. I examined the dependence of the statistic distribution on the sample volume. Monte Carlo simulation of the statistic distribution is used.

Closeness of Abbé Statistic Distribution to Normal. I consider how justified one is in using an approximation to the distribution of the (1) statistic in the form of a normal distribution with scale parameter of (2) or (3), and whether one can use such approximation for $n < 20$.

Table 1 gives conformity checks on simulated empirical distributions for Abbé test statistics for normal observed quantities $\xi_1, ..., \xi_n$ and the truth of hypothesis H_0 with sample volumes of $n = 10, 25, 100, 200$ with normal distributions as defined by the scale parameters of (2) and (3). The volume of the sample for simulated values of the statistic in all cases was $N = 10000$. Pearson's χ^2 fit test with asymptotically optimal grouping [6] has been used with nonparametric tests: Kolmogorov's, Mises' ω^2 (Cramer–Mises–Smirnov), and Mises' Ω^2 (Anderson–Darling) [7].

Table 1 shows that for all values of n, (3) gives a result worse than that from (2) for approximating the distribution of the Abbé statistic. Then one is not justified in using (3) for $n > 60$, and it is preferable to use the normal approximation with the scale parameter of (2).

For $n < 20$, in spite of information on the symmetry [2] the distribution of the test statistic becomes skewed. For example, for $n = 10$, the distribution of the Abbé statistic is appreciably skewed and differs substantially from the approximating normal distribution with the scale parameter given by (2).

TABLE 2. Error ε in Simulating the Statistic Distribution in Relation to Number of Monte Carlo Tests

N	Values of ε for $\hat{G}(S)$						
	0.5	0.4	0.3	0.2	0.1	0.05	0.01
1000	0.0260	0.0255	0.0238	0.0208	0.0156	0.0113	0.0052
10000	0.0082	0.0081	0.0075	0.0066	0.0049	0.0036	0.0016
100,000	0.0026	0.00255	0.00238	0.00208	0.00156	0.00113	0.00052
1,000,000	0.00082	0.00081	0.00075	0.00066	0.00049	0.00036	0.00016

A few comments may be made on the simulation accuracy. Large sample volumes are required to construct the empirical distributions for the statistics with high guaranteed accuracy. Table 2 shows the dependence of the estimation error ε for the distribution $G(S)$ of the statistic S as against the number of experiments N in the Monte Carlo method. In the present case, ε defines half the length of the 90% confidence range. Here $G(S)$ and $1 - G(S)$ are simulated with identical accuracy. For example, for the empirical distribution for the statistic $\hat{G}(S) = 0.5$, the confidence range covering the true value is 0.5 ± 0.026 for $N = 1000$. Raising the accuracy by an order of magnitude requires the sample volumes to be increased by two orders of magnitude.

On the other hand, experience shows that on constructing approximate to parametric models for distributions of these kinds, increasing N from 10^4 to 10^6 has little practical value.

Dependence of the Abbé Statistic Distribution on the Observed Law. During research on statistic distributions with an accepted hypothesis H_0, we considered the assignment of an observed normal distribution with density

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

and also the set of various symmetrical and unsymmetrical distributions. In particular, we examined the distribution of the Abbé statistic and the power of the test when the readings corresponded to a family with density

$$f(x) = \frac{\theta_3}{2\theta_1 \Gamma(1/\lambda)} \exp\left\{-\left(\frac{|x-\theta_1|}{\theta_2}\right)^{\theta_3}\right\} \tag{4}$$

and shape parameters $\theta_3 = 0.2$; 0.5; 1; 1.5; 2; 4; 8. In particular, if $\theta_3 = 2$, (4) gives the density of a normal distribution. Figure 1 shows the distributions for the (1) statistic obtained by simulation when the readings belong to the distributions in the family of (4) with various shape parameters: the laws varied from close to a Cauchy distribution to close to a uniform one.

Figure 1 shows that when $\xi_1, ..., \xi_n$ belong to a fairly wide range of distributions for the Abbé test statistic, they do not differ substantially from the distributions when the observations fit a normal law. If the law fitted by the observed quantities is symmetrical and has tails that are not too heavy, then the distribution for the statistic does not differ greatly from the classical one, and when the law is more flat-topped than normal, and also when it is more sharp-peaked (with large lordosis).

For example, it was found when we checked the uniformity of the distributions for the (1) statistic in the cases of a normal law and a law with parameter $\theta_3 = 8$ by the use of the Smirnov and Lehmann–Rosenblatt tests [1, 8] that there is no basis for declining the hypothesis of homogeneity. The test was performed on samples of the statistic values with volume $N = 10000$. The significance levels were 0.823 for the Smirnov test and 0.896 for the Lehmann–Rosenblatt one. As those tests are of high power with such volumes of samples, these significance levels confirm almost complete coincidence between the distributions for the corresponding samples.

An analogous conclusion applies for $\xi_1, ..., \xi_n$ belonging to a uniform distribution.

In (4) with $\theta_3 = 0.2$ and 0.5, the distributions for the (1) statistic differ substantially from a classical one corresponding to the readings fitting a normal distribution.

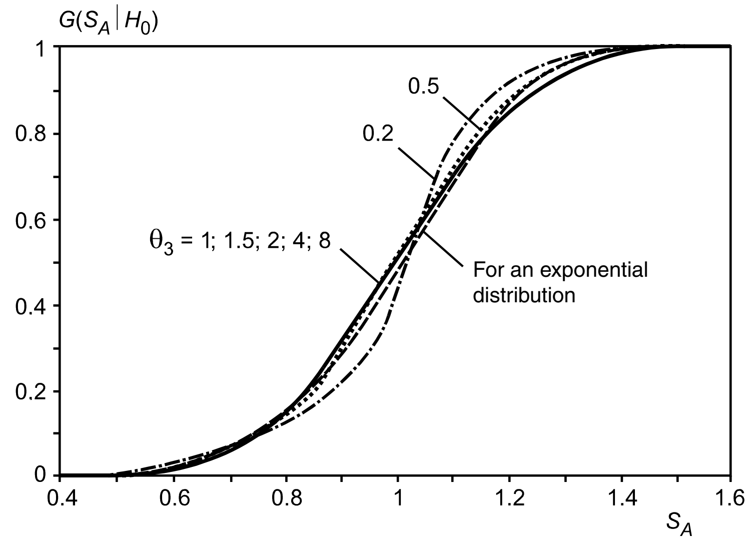

Fig. 1. Distributions for the Abbé test statistic in relation to the shape parameter in the (4) family for $n = 25$.

If there is marked skewness in the distribution of the observed random quantities, the distribution for the (1) statistic becomes different from the classical one. Figure 1 shows the distribution of the (1) statistic for the case of $\xi_1, ..., \xi_n$ belonging to an exponential distribution. It is clear that the effects of the skewness on the statistic distribution are less significant than when the tails are heavy. When $\xi_1, ..., \xi_n$ fits an unsymmetrical distribution, the extreme values (minimal or maximal) in the distribution for the (1) statistic hardly differ from the classical one.

The behavior of the distribution for the Abbé test statistic was examined for cases where the observed distribution is the symmetrical mixture of distributions, e.g., two normal distributions of the form

$$f(x) = \frac{0.5}{\theta_2\sqrt{2\pi}}\exp\left\{-\frac{(x-\theta_1-k\theta_2)^2}{2\theta_2^2}\right\} + \frac{0.5}{\theta_2\sqrt{2\pi}}\exp\left\{-\frac{(x-\theta_1+k\theta_2)^2}{2\theta_2^2}\right\}. \tag{5}$$

The simulation results for distributions for the Abbé test statistic corresponding to the $\xi_1, ..., \xi_n$ representing the (5) mixtures have been examined with shifts in the mixture components of $\pm 1\theta_2$, $\pm 2\theta_2$, $\pm 3\theta_2$ relative to θ_1, which literally are superimposed on one another and on the distribution of the statistic for a normal law. The Smirnov and the Lehmann–Rosenblatt homogeneity tests [1, 8] confirm that there is essential coincidence between the distributions.

From this we conclude that the Abbé test is not sensitive (is robust) for a bimodal distribution of the observed quantities subject to the condition that it is symmetrical and does not have heavy tails.

Abbé Test Power Study. In what follows, the hypothesis H_0 corresponds to obedience to the assumption that the observed quantities $\xi_1, ..., \xi_n$ are independent and belong to a normal distribution with identical but unknown dispersions. Without loss of generality, we can consider the assignment of $\xi_1, ..., \xi_n$ to a standard normal law. As competing hypothesis we have considered various situations in the presence of a trend.

In the case of a linear trend, we simulated the random quantities

$$x_i = at + \xi_i, \tag{6}$$

on which we tested the hypothesis H_0; in (6), the ξ_i are independent random quantities distributed in accordance with the standard normal law $t \in [0, 1]$.

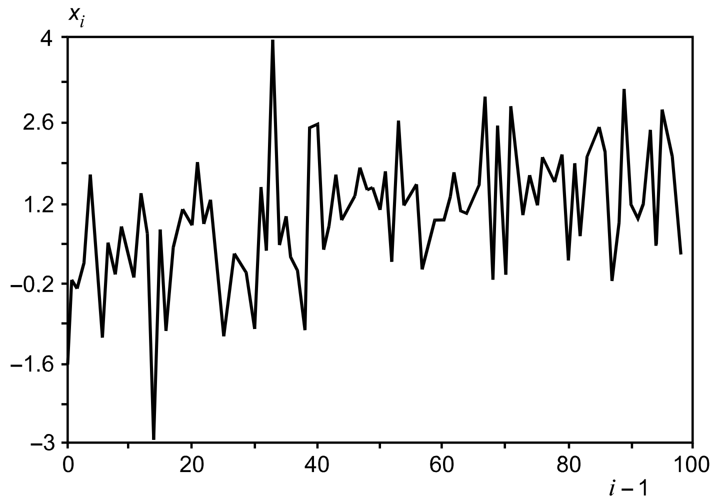

Fig. 2. Time series corresponding to model of (6) with $a = 2$ and $n = 100$.

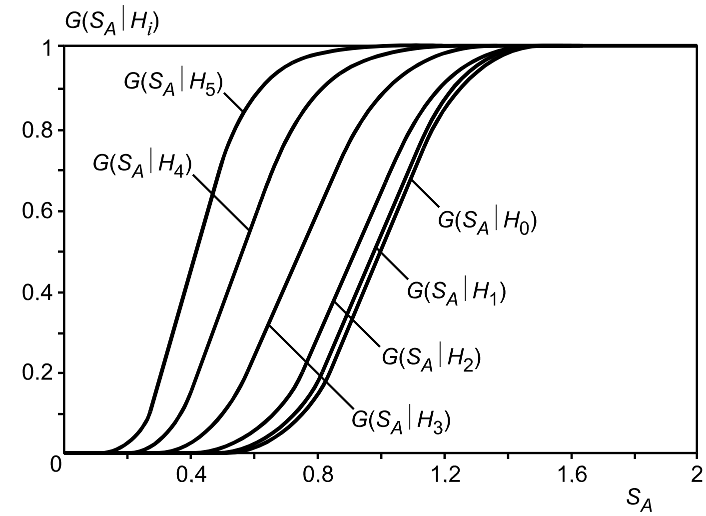

Fig. 3. Distributions of Abbé test statistic in the case of a normal law and linear trend of the form of (6) for $n = 25$.

The x_i of (6) are calculated in accordance with $x_i = a(i-1)\Delta t + \xi_i$, where the step Δt is defined as $\Delta t = 1/n$ in relation to the sample volume n. The random quantities ξ_i were generated in accordance with a standard normal law. We examined the power of the test relative to an alternative with linear trend defined by the parameters $a = 0.5$; 1; 2; 3; 4; the corresponding alternatives are denoted subsequently as $H_1, ..., H_5$. Figure 2 shows an example of a sample corresponding to the alternative H_3 for $n = 100$.

Figure 3 shows the empirical distributions $G(S_{A_n}|H_i)$ for the Abbé statistic constructed by simulation in the presence of a linear trend in the case of a sample volume $n = 25$. The values of the power $1 - \beta$ of the Abbé test, where β is the probability of error of the second kind, is shown relative to the alternatives $H_1, ..., H_5$ with linear trend of (6) in relation to the sample volume n with significance levels $\alpha = 0.025$; 0.05; 0.1 in Table 3.

Similar studies were performed on the distributions of the Abbé test statistics and the power with respect to alternatives with periodic trend:

$$x_i = a\sin(2\pi t) + \xi_i, \tag{7}$$

TABLE 3. Abbé Test Power for a Linear Trend as in (6)

n	α	H_1	H_2	H_3	H_4	H_5
10	0.025	0.034	0.054	0.159	0.349	0.570
	0.05	0.059	0.091	0.232	0.455	0.678
	0.1	0.114	0.159	0.344	0.579	0.799
25	0.025	0.035	0.074	0.298	0.654	0.913
	0.05	0.067	0.128	0.394	0.752	0.953
	0.1	0.125	0.206	0.514	0.840	0.978
50	0.025	0.036	0.087	0.441	0.877	0.995
	0.05	0.071	0.147	0.564	0.935	0.999
	0.1	0.134	0.248	0.691	0.968	1
100	0.025	0.043	0.123	0.699	0.993	1
	0.05	0.079	0.199	0.806	0.997	1
	0.1	0.144	0.306	0.889	0.999	1

TABLE 4. Abbé Test Power with Nonlinear Trend of the (7) Form

n	α	H_1	H_2	H_3	H_4	H_5
10	0.025	0.026	0.033	0.057	0.176	0.391
	0.05	0.051	0.060	0.099	0.267	0.523
	0.1	0.010	0.116	0.179	0.397	0.679
25	0.025	0.026	0.035	0.104	0.431	0.851
	0.05	0.052	0.075	0.156	0.546	0.911
	0.1	0.103	0.135	0.249	0.676	0.953
50	0.025	0.026	0.037	0.126	0.660	0.980
	0.05	0.053	0.078	0.207	0.777	0.993
	0.1	0.105	0.148	0.333	0.866	0.998
100	0.025	0.028	0.051	0.207	0.908	1
	0.05	0.056	0.095	0.312	0.950	1
	0.1	0.110	0.171	0.451	0.977	1

where as in the previous case, ξ_i are independent random quantities distributed in accordance with a standard normal distribution, $t \in [0, 1]$, $\Delta t = 1/n$, while the x_i are derived in accordance with (7) as $x_i = a\sin(2\pi(i-1)\Delta t) + \xi_i$.

We examine the power relative to the alternatives with the nonlinear trend of (7) specified by the parameter $a = 0.1$; 0.25; 0.5; 1; 1.5; the corresponding alternatives are denoted by H_6, ..., H_{10}. Figure 4 shows a sample for the alternative H_8 with $n = 100$.

Table 4 shows the empirical distributions $G(S_{A_n} | H_i)$ produced by simulation for the Abbé statistic in the presence of a (7) nonlinear trend and gives values of the power of the Abbé test relative to the alternatives H_6, ..., H_{10} in dependence on the sample volume n at significance levels $\alpha = 0.025$; 0.05; 0.1.

We then examined the power of the test relative to the alternatives with trend of the form

$$x_i = at + a\sin(2\pi t) + \xi_i. \tag{8}$$

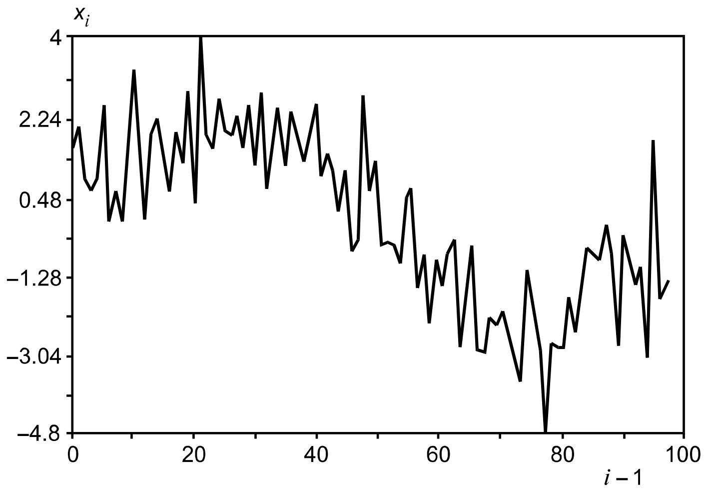

Fig. 4. Time series corresponding to the (7) model with $a = 2$ and $n = 100$.

TABLE 5. Abbé Test Power Relative to Alternatives with Trend of (8)

n	α	H_1	H_2	H_3	H_4	H_5
10	0.025	0.028	0.036	0.067	0.138	0.236
	0.05	0.053	0.067	0.125	0.227	0.375
	0.1	0.104	0.129	0.227	0.380	0.562
25	0.025	0.029	0.059	0.206	0.502	0.809
	0.05	0.062	0.104	0.294	0.623	0.882
	0.1	0.117	0.172	0.415	0.734	0.937
50	0.025	0.030	0.064	0.324	0.759	0.972
	0.05	0.063	0.119	0.447	0.851	0.989
	0.1	0.124	0.215	0.580	0.919	0.996
100	0.025	0.038	0.095	0.555	0.962	0.999
	0.05	0.071	0.162	0.686	0.980	1
	0.1	0.133	0.266	0.790	0.989	1

We considered the alternatives H_{11}, ..., H_{15} defined by the parameters $a = 0.25$; 0.5; 1; 1.5; and 2, respectively.

The simulation gave the power relative to the alternatives H_{11}, ..., H_{15} with a trend of the (8) form in relation to the sample volume for significance levels $\alpha = 0.025$; 0.05; 0.1 (Table 5).

We examined the power of the test for alternatives with a linear trend of the (6) form in cases where ξ_i fits laws in the (4) family with form parameters $\theta_3 = 1$ and $\theta_3 = 4$ and scale parameter $\theta_2 = 1$. In the case of a Laplace distribution ($\theta_3 = 1$), the power relative to H_1, ..., H_5 was lower than that for a normal law, while for $\theta_3 = 4$ it was higher. In that situation, in the case of a Laplace distribution, the dispersion is greater than that for a standard normal law, while in the case of the (4) law with shape parameter $\theta_3 = 4$, it is less.

We examined the power against that of the alternatives for various laws of the (4) form but with identical dispersions for the ξ_i, which gave power estimates virtually coincident with the values given in Table 2. This means that the distribution for the test statistic when the competing hypothesis applies, and consequently the power of the test is almost unaffected by the form of the law, but the power of the test is dependent on the dispersion of the observed quantities (measurement errors), i.e., the more accurate the measurements, the higher the power of the test, which is entirely logical.

From these studies, we consider that the Abbé test is correctly applied when one is dealing with a law substantially different from normal, but the law should not have heavy tails and should be symmetrical. The law may be bimodal and described by a symmetrical mixture of laws. At the same time, moderate skewness in the observed distribution has hardly any effect on the distribution of the test statistic.

These test power estimates enable one to judge the capacity to observe linear and nonlinear trends.

As distribution for the test statistic with $n > 20$, one can use the normal approximation of (2), which is also preferable to (3) for large n.

The research was performed with support from the Russian Federation Ministry of Education and Science (Projects Nos. 15378 and 02.442.11.7103) and the Russian Fund for Basic Research (Project No. 06-01-00059).

REFERENCES

1. L. N. Bol'shev and N. V. Smirnov, *Tables for Mathematical Statistics* [in Russian], Nauka, Moscow (1983).
2. J. Neumann, *AMS*, **12**, 367 (1941).
3. S. A. Aivazyan and V. S. Mkhitaryan, *Applied Statistics and the Principles of Econometrics: A College Textbook* [in Russian], YuNITI, Moscow (1998).
4. V. I. Strunov, *Izmer. Tekh.*, No. 8, 13 (2006).
5. B. Yu. Lemeshko and S. S. Pomadin, *Metrology: An Appendix to the Journal Izmer. Tekh.* [in Russian], No. 3, 3 (2004).
6. R 50.1.033-2001: *Recommendations on Standardization. Applied Statistics. Rules for Checking Conformity of an Experimental Distribution with a Theoretical One. Part I. Tests of* χ^2 *Type* [in Russian], Izd. Standartov, Moscow (2002).
7. R 50.1.037-2002: *Recommendations on Standardization. Applied Statistics. Rules for Checking Conformity of an Experimental Distribution with a Theoretical One. Part II. Nonparametric Tests* [in Russian], Izd. Standartov, Moscow (2002).
8. B. Yu. Lemeshko and S. B. Lemeshko, *Izmer. Tekh.*, No. 12, 9 (2005).