ISSN 1990-4789, Journal of Applied and Industrial Mathematics, 2009, Vol. 3, No. 4, pp. 462–475. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © B.Yu. Lemeshko, S.B. Lemeshko, S.N. Postovalov, 2008, published in Sibirskii Zhurnal Industrial'noi Matematiki, 2008, Vol. XI, No. 2, pp. 96–111.

Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses

B. Yu. Lemeshko*, **S. B. Lemeshko****, and **S. N. Postovalov*****

Novosibirsk State Technical University, pr. K. Marksa 20, Novosibirsk, 630092 Russia

Received November 26, 2007; in final form, February 21, 2008

Abstract—Using statistical modeling methods, we analyze the power of a series of goodness-of-fit tests for simple and complex hypotheses. The estimates we give for the power of the tests for simple hypotheses versus some near competing hypotheses enable us to rank the goodness-of-fit tests.

DOI: 10.1134/S199047890904005X

Key words: *goodness-of-fit test, the Kolmogorov test, the Kramer–Mises–Smirnov test, the Anderson–Darling test, the Pearson test, the Nikulin test, power of a test*

INTRODUCTION

The purpose of the goodness-of-fit tests is to test hypotheses on the empirical distributions fitting some theoretical law. Simple and complex hypothesis testings are distinguished. The simple hypothesis H_0 is of the form $F(x) = F(x, \theta)$, where $F(x, \theta)$ is the probability distribution with which an observed sample is being tested to fit, and θ is some known value of a (scalar or vector) parameter. A complex hypothesis H_0 can be expressed as

$$F(x) \in \{F(x, \theta), \theta \in \Theta\},$$

where Θ is the domain of an unknown parameter θ. Differences in the complex hypothesis testing and the corresponding problems arise if the parameter $\hat{\theta}$ of the theoretical distribution is estimated using the same sample as that being fit. Below we usually assume that in the complex hypothesis testing the parameter $\hat{\theta}$ is estimated using the same sample.

Errors of two kinds arise in the statistical hypothesis testing. An *error of the first kind* is that a test rejects a valid hypothesis H_0. An *error of the second kind* is that the test accepts H_0 as a valid hypothesis when actually some competing hypothesis H_1 holds true.

The procedure for testing H_0 assumes known the distribution $G(S|H_0)$ of some test statistic S given the validity of H_0. The critical regions for goodness-of-fit tests are determined by the large values of the statistics. The probability α of an error of the first kind (the significance level) is the probability of the values of the statistics lying in the critical region:

$$\alpha = P\{S > S_\alpha \,|\, H_0\} = 1 - G(S_\alpha \,|\, H_0),$$

where S_α is the critical value. In the hypothesis testing, α is usually given. If the value of the statistic $S^* \leq S_\alpha$ is calculated from a sample or, which is the same, the attained significance level is

$$P\{S > S^*|H_0\} = 1 - G(S^*|H_0) > \alpha$$

then the hypothesis H_0 is not rejected.

*E-mail: lemeshko@fpm.ami.nstu.ru
**E-mail: skyer@mail.ru
***E-mail: Postovalov@ngs.ru

Given a competing hypothesis H_1, the probability of an error of the second kind is defined as

$$\beta = P\{S \le S_\alpha \,|\, H_1\} = G(S_\alpha \,|\, H_1),$$

where $G(S \,|\, H_1)$ is the distribution of the test statistic given the validity of H_1. If the test is completely determined then α uniquely determines β and vice versa. The power $1-\beta$ of a test for H_0 versus H_1 is a function of H_0, H_1, the sample volume n, and possibly some other factors related to the construction of the test.

While making statistical analysis and preferring some test, it is desirable to be sure that the given probability α of an error of the first kind guarantees the minimal probability β of an error of the second kind; i.e., that a test has the greatest power relative to the alternative between H_0 and H_1.

The information available from various sources on the advantages of some goodness-of-fit test in certain situations is ambiguous and often contradictory. The results of studying the asymptotic power of the tests (for instance, see [1–4]) are hard to use as a consequence of a limited volume of samples encountered in practice. The recommendations of various authors are subjective, reflect the stereotypes, and rest on some concrete, particular examples and limited experience with practical applications.

Studying the power of the tests is complicated by the absence of results related to the analytic representations of the distributions $G(S|H_1)$ for concrete goodness-of-fit tests for complex hypotheses, in particular, for the parameter-free and χ^2-type tests, for the parameter estimation from the samples of points (ungrouped observations).

The goal of the study of this article is the comparative analysis of the power of the most popular goodness-of-fit tests on some pairs of sufficiently near competing hypotheses H_0 and H_1. Of interest is precisely the ability of tests to the distinguish some near hypotheses since distinguishing distributions with contrasting laws is usually not a problem.

1. THE TESTS UNDER STUDY

The Kolmogorov Test. In the tests of Kolmogorov type the measured distance between the empirical distribution $F_n(x)$ and the theoretical distribution $F(x,\theta)$ is of the form

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

where n is the sample volume. In the case that a simple hypothesis to be tested is valid, as $n \to \infty$, the statistics $\sqrt{n}\,D_n$ obeys the Kolmogorov distribution $K(s)$ [5]. Most often in the Kolmogorov (Kolmogorov–Smirnov) test, some statistics of the form [6] is used with a correction proposed by L. N. Bol'shev in [7, 8]:

$$S_k = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{1}$$

where

$$D_n = \max\left(D_n^+, D_n^-\right),$$

$$D_n^+ = \max_{1\le i\le n}\left\{\frac{i}{n} - F(x_i,\theta)\right\}, \qquad D_n^- = \max_{1\le i\le n}\left\{F(x_i,\theta) - \frac{i-1}{n}\right\},$$

while n is the sample volume and $x_1, x_2, \ldots, x_n$ are the sample values in increasing order. If a simple hypothesis being tested is valid then the statistics S_k in the limit obeys the Kolmogorov distribution $K(S)$ [6].

Seemingly (and rather unfortunately), the Bol'shev correction escaped the attention of foreign experts. The articles devoted to the Kolmogorov test, as a rule, still use the statistics $\sqrt{n}\,D_n$. As a consequence, for the bounded values of n, they must take into account the essential dependence of the distributions of the statistics on n.

The Kramer–Mises–Smirnov test. The statistics of the ω^2 test of Mises (Kramer–Mises–Smirnov) is of the form [6]

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{F(x_i,\theta) - \frac{2i-1}{2n}\right\}^2. \tag{2}$$

If a simple hypothesis is valid then the statistics in the limit obeys the law with the distribution $a_1(s)$ [6].

The Anderson–Darling test. The statistics of the Mises Ω^2 test (the Anderson–Darling statistics) is defined [6] by

$$S_\Omega = n\Omega_n^2 = -n - 2\sum_{i=1}^{n}\left\{\frac{2i-1}{2n}\ln F(x_i,\theta) + \left(1 - \frac{2i-1}{2n}\right)\ln(1 - F(x_i,\theta))\right\}, \qquad (3)$$

and if a simple hypothesis is valid then in the limit the statistics obeys the law with the distribution $a_2(s)$ [6].

In the expressions for the tests statistics (1)–(3) constructed for the simple hypothesis testing, it is customary to denote the theoretical distribution by $F(x)$, stressing by this that the distribution law and its parameter θ are available. We intentionally denote in (1)–(3) the distribution by $F(x,\theta)$ depending on θ, meaning that, for complex hypotheses, θ will be replaced by an estimate for it. The same remark applies to the Pearson χ^2 test.

For simple hypothesis testing, the limit distributions of the Kolmogorov, Mises ω^2, and Anderson–Darling Ω^2 test statistics are independent of the form of the observed distribution and in particular of its parameters. For this reason they are called *distribution-free*.

For complex hypothesis testing when the parameters of the law are estimated using the same sample, the parameter-free tests lose their distribution-freeness [9]. Moreover, for complex hypothesis testing, the distributions of these test statistics are determined by the character of the complex hypothesis being tested. The distributions $G(S|H_0)$ of the test statistics are affected by the following factors, which determine the "complexity" of the hypotheses [10]:

- the form of the observed distribution $F(x,\theta)$ corresponding to the valid hypothesis H_0;
- the type of the estimated parameter and the number of parameters estimated from the sample;
- in some situations, the actual value of a parameter (for instance, for gamma distributions);
- the method used for estimating the parameters.

The analytic form of the (limit) distributions $G(S|H_0)$ of the parameter-free test statistics for the complex hypotheses is unknown. There are only particular solutions based on various approaches [11–19]. Seemingly, the most promising approach to constructing the distributions of the statistics is a numerical approach resting on the statistical modeling of the empirical distributions of the statistics and the subsequent construction of their approximate analytic models [10, 20–27].

The Pearson χ^2 test. The χ^2-type tests require subdividing the domain of the random variable into k intervals and counting the number of observations n_i lying in each interval and the probability $P_i(\theta)$ of lying in the intervals for the corresponding theoretical law. The Pearson χ^2 goodness-of-fit test statistics is of the form

$$X_n^2 = n\sum_{i=1}^{k}\frac{\left(n_i/n - P_i(\theta)\right)^2}{P_i(\theta)}. \qquad (4)$$

In the case of testing a valid simple hypothesis, in the limit this statistics obeys the χ^2_{k-1}-distribution with $k-1$ degrees of freedom.

If a competing hypothesis H_1 is valid and the sample fits the distribution $F_1(x,\theta_1)$ with a parameter θ_1 then the same statistics in the limit obeys the noncentral χ^2_{k-1}-distribution with the noncentrality parameter

$$\nu = n\sum_{i=1}^{k}\frac{\left(P_i^1(\theta_1) - P_i(\theta)\right)^2}{P_i(\theta)},$$

where $P_i^1(\theta_1)$ is the probability of lying in the interval for the valid hypothesis H_1.

In the case of testing a valid complex hypothesis H_0, provided that the parameters are estimated as a result of minimizing (4) using the same sample, the statistics X_n^2 are asymptotically distributed as χ^2_{k-r-1}, where r is the number of parameters estimated using the sample. The statistics (4) have the same distribution if we use the maximal likelihood method as the estimation method and calculate the estimates from the grouped data [28, 29]. Moreover, it has been shown by statistical modeling methods that this holds if we use other asymptotically effective estimates for the grouped data [30].

In calculating the maximal likelihood estimates (MLE) using ungrouped data, the same statistics is distributed as the sum of independent terms

$$\chi^2_{k-r-1} + \sum_{j=1}^{r} \lambda_j \xi_j^2,$$

where $\xi_1, \dots, \xi_r$ are the standard normal random variables independent of each other and of χ^2_{k-r-1}, and $\lambda_1, \dots, \lambda_r$ are some numbers between 0 and 1 [31–33]. In this case, for complex hypothesis testing with MLE using ungrouped observations, the distributions $G\left(X_n^2 | H_0\right)$ of the test statistic depend essentially on the grouping method [34].

The distributions of the χ^2-type statistics were studied in [34, 35] using statistical modeling methods in the case of simple and various complex hypotheses, the validity of the hypothesis H_0 and the validity of a competing hypothesis H_1 for equiprobable (EPG) and asymptotically optimal (AOG) grouping [36–41]. When AOG is used, the losses in the Fisher information related to grouping are minimized, and the power of the Pearson χ^2 test is maximized relative to the near competing hypotheses.

As a goodness-of-fit test we can use the likelihood-ratio test with the statistics of the form [42]

$$-2\ln l = -2\sum_{i=1}^{k} n_i \ln\left(\frac{P_i(\theta)}{n_i/n}\right),$$

which is asymptotically equivalent to the Pearson χ^2 test [42]. Moreover, the conclusions on the properties of this test in all our studies have coincided with the conclusions on the Pearson test [35, 39, 40]. Thus, in this article we do not consider this test separately.

The Nikulin test. In complex hypothesis testing with χ^2-type tests, the estimation using the ungrouped (point) observations has certain advantages. These estimates have better asymptotic properties in comparison with the grouped estimates. In [43, 44], we proposed a test with MLE using ungrouped data. The articles [45, 46] deal with the same test, and the most complete discussion of it is in [47]. The test has two advantages over the Pearson χ^2 test. Firstly, a remarkable fact distinguishing this test is that when the tested hypothesis is valid the test statistics in the limit obeys the χ^2_{k-1}-distribution independently of the number of parameters of the law estimated by the maximal likelihood method. Secondly, the power of this test is usually higher than that of the Pearson χ^2 test.

K. C. Rao and D. S. Robson in [48] obtain similar results for the exponential family of distributions. The article [49] is devoted to the theory of constructing the tests of this type, which have recently become called the tests with the Rao–Robson–Nikulin statistics [50].

We consider a test with the statistic in the form originally proposed in [43]. The test assumes estimating the unknown parameters of the distribution $F(x, \theta)$ by the maximal likelihood method using the ungrouped data. The vector of probabilities of lying in the intervals $\mathbf{P} = (P_1, \dots, P_k)^T$ is assumed given, and the boundary points of the intervals are determined as

$$x_i(\theta) = F^{-1}(P_1 + \cdots + P_i), \qquad i = \overline{1, (k-1)}.$$

The proposed statistics is of the form [42]

$$Y_n^2 = X_n^2 + n^{-1} a^T(\theta)\Lambda(\theta)a(\theta), \tag{5}$$

where X_n^2 is calculated using (4); the matrix

$$\Lambda(\theta) = \left\| J(\theta_l, \theta_j) - \sum_{i=1}^{k} \frac{w_{\theta_l i} w_{\theta_j i}}{p_i} \right\|^{-1},$$

whose size and entries are determined using the estimated components of the parameter vector θ;

$$J(\theta_l, \theta_j) = \int \left(\frac{\partial f(x,\theta)}{\partial \theta_l} \frac{\partial f(x,\theta)}{\partial \theta_j} \right) f(x,\theta)\, dx$$

are the entries of the information matrix for the ungrouped data; the components of the vector $a(\theta)$ are of the form

$$a_{\theta_l} = w_{\theta_l 1} \frac{n_1}{P_1} + \cdots + w_{\theta_l k} \frac{n_k}{P_k},$$

$$w_{\theta_l i} = -f\left[x_i(\theta), \theta\right] \frac{\partial x_i(\theta)}{\partial \theta_l} + f\left[x_{i-1}(\theta), \theta\right] \frac{\partial x_{i-1}(\theta)}{\partial \theta_l}.$$

2. THE METHODS OF STUDY

In order to evaluate the power of tests we need to know the distributions $G(S|H_0)$ and $G(S|H_1)$ of the statistics whose analytic expressions are usually unknown and depend on n. In order to estimate $G(S|H_0)$ and $G(S|H_1)$ it is most advisable to use statistical modeling methods. In order to construct the empirical distributions of the statistics we model samples of the statistics $S_1, S_2, \ldots, S_N$ of a sufficiently large volume N for a concrete sample volume n of the observed quantities modeled by the laws corresponding to a valid (H_0) or a competing (H_1) hypothesis. Then estimates for the power of the tests can be obtained from the empirical distributions $G_n^N(S|H_0)$ and $G_n^N(S|H_1)$ or by using approximate analytic models constructed from $G_n^N(S|H_0)$ and $G_n^N(S|H_1)$.

In this study, we usually take $N = 10^6$. Henceforth, we omit the index N in the notation for the corresponding empirical functions. In modeling and studying, we use some software for statistical analysis that is under development.

3. THE ALTERNATIVES CONSIDERED

We illustrate the results of our comparative analysis of the power of goodness-of-fit tests on the two pairs of competing hypotheses. The first pair consists of the normal law and the logistic law: the valid hypothesis H_0 corresponds to the normal law with density

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

and the competing hypothesis H_1, to the logistic law with density

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

and parameters $\theta_0 = 1$ and $\theta_1 = 0$. For the simple hypothesis H_0, the parameters of the normal law have the same values. These two laws are similar and hard to distinguish using the goodness-of-fit tests.

The second pair consists of the following by hypotheses: H_0 is the Weibull distribution with density

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

and parameters $\theta_0 = 2$, $\theta_1 = 2$, and $\theta_2 = 0$; while H_1 is the gamma distribution with density

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

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

We study the power for testing the simple and complex hypotheses H_0 versus the simple hypothesis H_1.

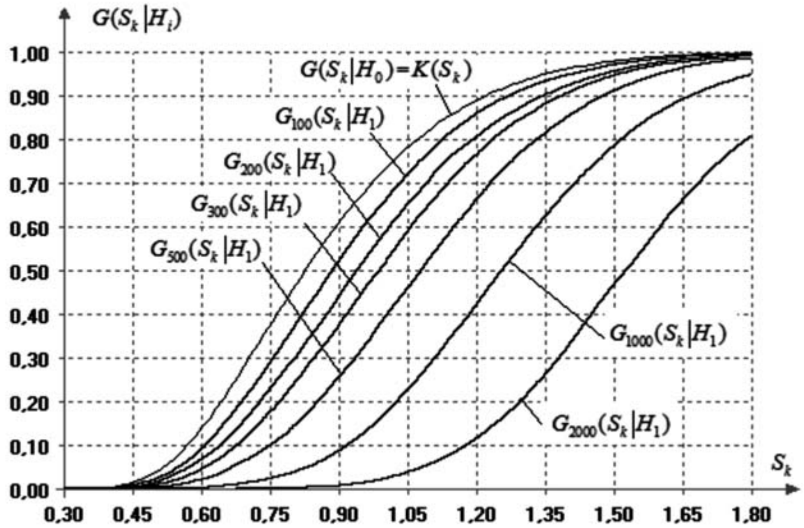

Fig. 1. The distributions $G(S_k|H_0) = K(S_k)$ and $G_n(S_k|H_1)$ of the Kolmogorov type statistic (1) for testing the simple hypothesis H_0 on the fit with the normal law with the competing hypothesis H_1

4. THE POWER OF TESTS FOR SIMPLE HYPOTHESES FOR THE ALTERNATIVE "NORMAL DISTRIBUTION VERSUS LOGISTIC DISTRIBUTION"

The distributions of the normal and logistic laws corresponding to the competing hypotheses H_0 and H_1 are quite similar as we can see by comparing the graphs. If we use samples corresponding to the normal or logistic law to find the MLE parameters of these laws then the estimated values for the translation and scale parameters essentially coincide.

Fig. 1 shows the results of modeling the distribution of the Kolmogorov statistic when the simple hypothesis $G(S_k|H_0) = K(S_k)$ is valid and when the competing hypothesis $G_n(S_k|H_1)$ is valid for the sample volumes of $n = 100, 200, 300, 500, 1000, 2000$ observations. The tested hypothesis H_0 corresponds to the normal law, and the competing hypothesis H_1, to the logistic law.

As we see from the figure, the ability of the Kolmogorov test to distinguish these hypotheses is small. For instance, given the probability $\alpha = 0.1$ of an error of the first kind, the power of the Kolmogorov test for testing H_0 versus H_1 is about 0.127 for $n = 100$; 0.170 for $n = 200$; 0.215 for $n = 300$; 0.309 for $n = 500$; 0.544 for $n = 1000$; and 0.861 for $n = 2000$.

Fig. 2 shows the distributions $G(S_\omega|H_0) = a_1(S_\omega)$ and $G_n(S_\omega|H_1)$ of the Kramer–Mises–Smirnov S_ω statistic for the simple hypothesis testing and the same H_0 and H_1.

Similarly, Fig. 3 shows the distributions $G(S_\Omega|H_i)$ of the Anderson–Darling S_Ω (Mises Ω^2) statistic for testing the simple hypothesis $G(S_\Omega|H_0) = a_2(S_\Omega)$. Fig. 4 shows the distributions of the Pearson X_n^2 statistic for the simple hypothesis in the AOG case with $k = 9$ intervals.

Recall that the Pearson X_n^2 statistic is a discrete random variable. The discreteness of its values shows up particularly strongly for EPG. In this case, the continuous χ^2_{k-1} distribution badly approximates the distribution of the X_n^2 statistic. Fig. 5 illustrates [51] the character of the convergence of the distributions $G\big(X_n^2|H_0\big)$ of the Pearson X_n^2 statistic to the χ^2_8 distribution for 9 equiprobable intervals and simple hypothesis H_0. It is natural that the parameter-free goodness-of-fit tests, as the volume N of the samples of the X_n^2 statistic grows (and as the power of these tests grows), more and more surely reject the hypothesis on the fit of the empirical distributions of the X_n^2 statistic (because they are piecewise constant) to the corresponding continuous χ^2_{k-1} distributions.

Similar distributions of the statistic using AOG have a sufficiently smooth character [51]. For small n, the distributions $G\big(X_n^2|H_0\big)$ in comparison with the χ^2_8 distribution have heavier (right) tail, and in the region of ordinates from 0 to 0.85 they are shifted to the left. However, as n grows, the distributions rapidly converge to the χ^2_{k-1} distributions [51]. For instance, already for $n = 100$, the goodness-of-fit

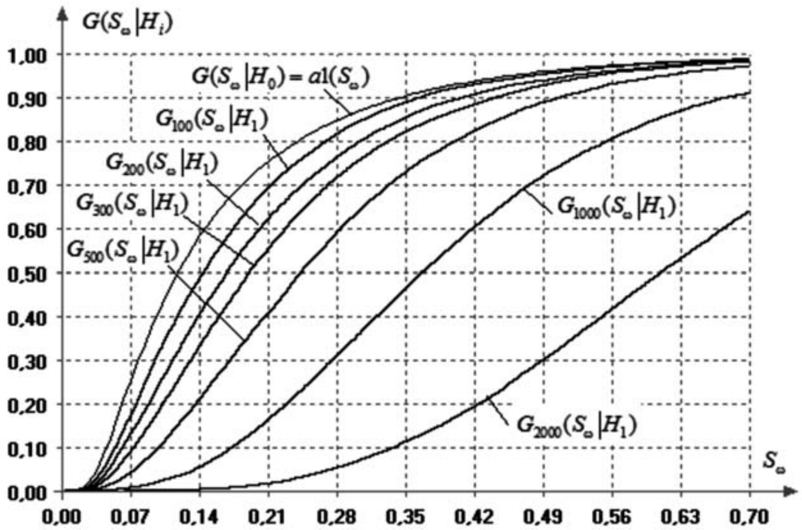

Fig. 2. The distributions $G(S_\omega|H_0) = a_1(S_\omega)$ and $G_n(S_\omega|H_1)$ of the Kramer–Mises–Smirnov ω^2-type statistic (2) for testing the simple hypothesis H_0 on the fit with the normal law with the competing hypothesis H_1

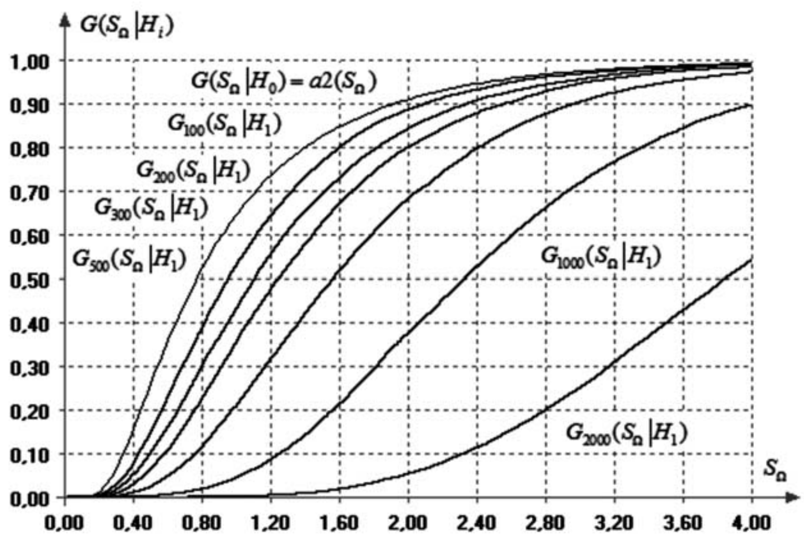

Fig. 3. The distributions $G(S_\Omega|H_0) = a_2(S_\Omega)$ and $G_n(S_\Omega|H_1)$ of the Anderson–Darling Ω^2-type statistic (3) for testing the simple hypothesis H_0 on the fit with the normal law with the competing hypothesis H_1

testing of a model sample of the statistic (4) with the volume of $N = 10000$ for the χ_8^2 distribution attains, for all tests used, the significance level in the interval from $P = 0.0103$ for the Anderson–Darling test to $P = 0.0586$ for the Kramer–Mises–Smirnov test. For $n = 500$, the attained significance levels for all tests we consider already lie in the interval from $P = 0.963$ to $P = 0.992$, which demonstrates the essential coincidence of the empirical distribution of the statistic with the theoretical χ_8^2-distribution.

Table 1 shows the estimates calculated basing on the results of modeling the distributions of the statistics, for the power of the goodness-of-fit tests under consideration for various values of the significance level α in testing the simple hypothesis H_0 corresponding to the normal law with parameters $(0, 1)$ versus the hypothesis H_1 corresponding to the logistic law with the same collection of parameters. The error of these estimates for the testing power for the simple hypothesis and 95% confidence interval is at most $\pm 10^{-3}$. The tests are ordered in decreasing power. The table shows the maximal power of the

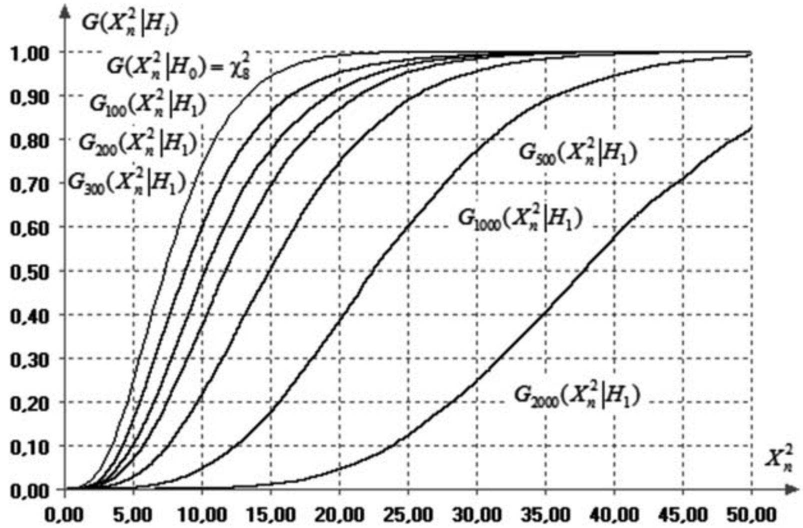

Fig. 4. The distributions $G(X_n^2|H_0) = \chi_8^2$ and $G(X_n^2|H_1)$ of the Pearson χ^2-type statistic (4) for testing the simple hypothesis H_0 on the fit with the normal law with the competing hypothesis H_1 in the AOG case with $k = 9$

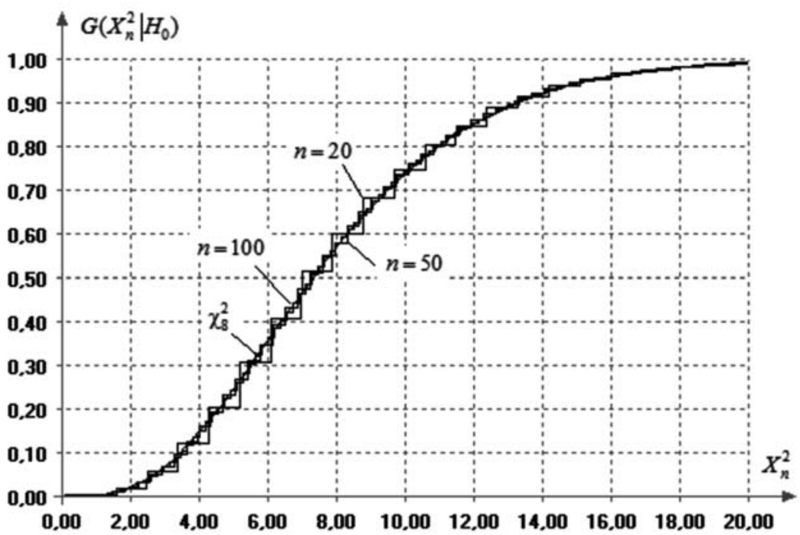

Fig. 5. The character of convergence of the distributions $G(X_n^2|H_0)$ of the Pearson χ^2 statistic (4) to the χ_8^2 distribution for 9 equiprobable intervals and simple hypothesis H_0

Pearson X_n^2 test which, for this pair H_0 and H_1, is for $k = 15$ and AOG. For comparison, Table 2 shows the values of the power for EPG and different number of intervals.

With EPG the Pearson X_n^2 test as regards this pair of hypotheses has the maximal power with $k = 4$ [52], and then with the growth of k the power decreases. However, this maximal level of power is below the power of this test with $k = 9$ and AOG.

With EPG the power of the Pearson χ^2 test as regards this pair of hypotheses is a decreasing function of the number k of the grouping intervals [52], which confirms the results of this study. With AOG the optimal number of intervals, which maximizes the power, shifts into the region of large values of k. The power of this test, as regards near hypotheses with a given k and AOG, is usually higher [40, 41, 52].

Table 1. The power of the goodness-of-fit tests for the simple hypotheses H_0 (normal distribution) versus H_1 (logistic distribution)

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	The power of the Pearson χ^2 test with $k = 15$ and AOG					
0.15	0.349	0.459	0.565	0.737	0.946	0.999
0.10	0.290	0.388	0.490	0.671	0.922	0.998
0.05	0.210	0.292	0.385	0.565	0.871	0.996
0.025	0.154	0.222	0.302	0.472	0.813	0.992
0.01	0.107	0.159	0.221	0.369	0.729	0.983
	The power of the Anderson–Darling Ω^2 test					
0.15	0.194	0.258	0.328	0.472	0.776	0.982
0.1	0.125	0.169	0.222	0.343	0.654	0.957
0.05	0.057	0.079	0.107	0.181	0.439	0.869
0.025	0.026	0.036	0.049	0.088	0.261	0.724
0.01	0.010	0.013	0.017	0.031	0.114	0.491
	The power of the Kolmogorov test					
0.15	0.190	0.246	0.303	0.415	0.662	0.922
0.1	0.127	0.170	0.215	0.309	0.544	0.861
0.05	0.062	0.088	0.116	0.179	0.365	0.721
0.025	0.031	0.044	0.061	0.100	0.231	0.560
0.01	0.012	0.018	0.026	0.044	0.119	0.366
	The power of the Kramer–Mises–Smirnov ω^2 test					
0.15	0.178	0.228	0.283	0.401	0.680	0.947
0.1	0.114	0.147	0.186	0.277	0.542	0.892
0.05	0.052	0.067	0.086	0.136	0.324	0.742
0.025	0.024	0.030	0.039	0.062	0.171	0.548
0.01	0.010	0.011	0.014	0.021	0.065	0.307

5. THE POWER OF TESTS FOR SIMPLE HYPOTHESES IN THE CASE OF A PAIR "THE WEIBULL DISTRIBUTION VERSUS GAMMA DISTRIBUTION"

Let the tested hypothesis H_0 correspond to the Weibull distribution with parameters $\theta_0 = 2$, $\theta_1 = 2$, and $\theta_2 = 0$, while let the competing hypothesis H_1 be the gamma distribution with parameters $\theta_0 = 3.12154$, $\theta_1 = 0.557706$, and $\theta_2 = 0$. The parameters of the gamma distribution are chosen to make it as similar as possible to that Weibull distribution.

Tables 3 and 4 show the calculated estimates for the power of tests for various values of the significance level α in testing the fit with the Weibull distribution (hypothesis H_0) versus the competing

Table 2. The power of the Pearson χ^2 goodness-of-fit test for the simple hypotheses H_0 (normal distribution) versus H_1 (logistic distribution) depending on the grouping method and the number of intervals

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	The power of the Pearson χ^2 test with $k = 9$ and AOG					
0.15	0.269	0.381	0.488	0.670	0.917	0.998
0.10	0.204	0.302	0.403	0.589	0.880	0.995
0.05	0.129	0.203	0.287	0.464	0.806	0.989
0.025	0.084	0.136	0.203	0.359	0.723	0.9797
0.01	0.050	0.081	0.127	0.249	0.608	0.957
	The power of the Pearson χ^2 test with $k = 9$ and EPG					
0.15	0.210	0.282	0.349	0.483	0.747	0.960
0.10	0.152	0.208	0.270	0.392	0.673	0.938
0.05	0.083	0.123	0.170	0.273	0.547	0.890
0.025	0.046	0.072	0.105	0.186	0.435	0.828
0.01	0.020	0.036	0.056	0.109	0.310	0.734
	The power of the Pearson χ^2 test with $k = 15$ and EPG					
0.15	0.192	0.257	0.312	0.432	0.690	0.941
0.10	0.139	0.187	0.237	0.343	0.607	0.911
0.05	0.073	0.106	0.144	0.227	0.477	0.848
0.025	0.040	0.061	0.085	0.149	0.365	0.772
0.01	0.018	0.029	0.043	0.083	0.247	0.662

hypothesis corresponding to the gamma distribution with the indicated parameters (hypothesis H_1) for the simple hypothesis H_0. The tests appear in the tables in the decreasing order of power.

CONCLUSION

Therefore, basing on the results of our analysis of the power of the tests under consideration with respect to a series of pairs of relatively near competing hypotheses for simple hypothesis testing, we can rank the tests by power as follows:

$$\text{Pearson } \chi^2\,(\text{AOG}) \;\succ\; \text{Anderson–Darling } \Omega^2 \;\succ\; \text{Mises } \omega^2 \;\succeq\; \text{Kolmogorov}.$$

This scale holds while using AOG in the Pearson χ^2 test, which minimizes the losses in the Fisher information. For quite near hypotheses, the situation

$$\text{the Kolmogorov test} \;\succ\; \text{the Mises } \omega^2 \text{ test}$$

is possible. This ordering is not rigid. As we see from the tables with the given values of power, sometimes a test has advantages in terms of power for some values of α and n and loses for the others.

Table 3. The power of goodness-of-fit tests for the simple hypothesis H_0 (the Weibull distribution with parameters 2, 2, and 0) versus the hypothesis H_1 (the gamma distribution with parameters 3.12154, 0.557706, and 0)

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	The power of the Pearson χ^2 test with $k = 15$ and AOG					
0.15	0.486	0.621	0.757	0.909	0.996	1.000
0.10	0.418	0.556	0.701	0.876	0.993	1.000
0.05	0.324	0.469	0.611	0.815	0.986	1.000
0.025	0.254	0.403	0.529	0.751	0.974	1.000
0.01	0.191	0.332	0.437	0.668	0.954	1.000
	The power of the Anderson–Darling Ω^2 test					
0.15	0.302	0.446	0.577	0.781	0.976	1.000
0.10	0.223	0.348	0.473	0.689	0.951	1.000
0.05	0.131	0.224	0.326	0.533	0.882	0.998
0.025	0.076	0.141	0.220	0.396	0.785	0.993
0.01	0.037	0.075	0.126	0.257	0.636	0.975
	The power of the Kramer–Mises–Smirnov ω^2 test					
0.15	0.295	0.425	0.539	0.716	0.931	0.998
0.10	0.224	0.343	0.453	0.637	0.894	0.995
0.05	0.138	0.233	0.329	0.508	0.816	0.987
0.025	0.084	0.155	0.233	0.393	0.725	0.970
0.01	0.043	0.088	0.142	0.270	0.597	0.934
	The power of the Kolmogorov test					
0.15	0.294	0.421	0.531	0.700	0.915	0.995
0.10	0.225	0.342	0.450	0.628	0.879	0.992
0.05	0.141	0.237	0.332	0.508	0.806	0.981
0.025	0.087	0.160	0.239	0.401	0.723	0.964
0.01	0.045	0.093	0.150	0.282	0.606	0.930

ACKNOWLEDGMENTS

The authors were supported by the Russian Foundation for Basic Research (project no. 06–01–00059a).

Table 4. The power of the Pearson χ^2 goodness-of-fit test for the simple hypothesis H_0 (the Weibull distribution with parameters 2, 2, and 0) versus the hypothesis H_1 (the gamma distribution with parameters 3.12154, 0.557706, and 0) depending on the grouping method and the number of intervals

α	$n = 100$	$n = 200$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
	The power of the Pearson χ^2 test with $k = 9$ and AOG					
0.15	0.427	0.608	0.748	0.910	0.996	1.000
0.10	0.353	0.534	0.684	0.874	0.993	1.000
0.05	0.261	0.429	0.581	0.807	0.985	1.000
0.025	0.202	0.343	0.488	0.734	0.973	1.000
0.01	0.152	0.255	0.384	0.637	0.950	1.000
	The power of the Pearson χ^2 test with $k = 15$ and EPG					
0.15	0.234	0.347	0.446	0.637	0.908	0.998
0.10	0.174	0.266	0.361	0.549	0.867	0.996
0.05	0.097	0.164	0.245	0.417	0.785	0.9907
0.025	0.056	0.102	0.161	0.311	0.695	0.979
0.01	0.026	0.054	0.092	0.203	0.574	0.958
	The power of the Pearson χ^2 test with $k = 9$ and EPG					
0.15	0.240	0.344	0.440	0.616	0.883	0.995
0.10	0.177	0.262	0.354	0.528	0.835	0.990
0.05	0.100	0.164	0.238	0.399	0.743	0.979
0.025	0.057	0.100	0.157	0.294	0.646	0.960
0.01	0.026	0.053	0.090	0.191	0.520	0.924

REFERENCES

1. D. M. Chibisov, "About Asymptotic Power and Efficiency of the ω_n^2 Test," Dokl. Akad. Nauk SSSR **138** (2), 322–325 (1961).
2. D. M. Chibisov, "Studies of Power of Some Nonparametric Tests," Teor. Veroyatnost. i Primenen. **7** (3), 355–356 (1962).
3. D. M. Chibisov, "About Asymptotic Power of the Goodness-of-Fit Tests for Near Alternatives," Teor. Veroyatnost. i Primenen. **9** (3), 561–566 (1964).
4. D. M. Chibisov, "To Study of the Asymptotic Power of the Goodness-of-Fit Tests," Teor. Veroyatnost. i Primenen. **10** (3), 460–478 (1965).
5. A. N. Kolmogoroff , "Sulla determinazione empirica di una legge di distribuzione," Giornale Ist. Ital. degli Attuari. **4** (1), 83–91 (1933).
6. L. N. Bol'shev and N. V. Smirnov, *Tables of Mathematical Statistics* (Nauka, Moscow, 1983) [in Russian].
7. L. N. Bol'shev, "Asymptotic Pearson Transformations," Teor. Veroyatnost. i Primenen. **8** (2), 129–155 (1963).
8. L. N. Bol'shev, *Probability Theory and Mathematical Statistics: Selected Papers* (Nauka, Moscow, 1987) [in Russian].
9. M. Kas, J. Kiefer, and J. Wolfowitz, "On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods," Ann. Math. Statist. **26**, 189–211 (1955).

10. *Recommendation on Standardization. Applied Statistics. Rules of Check of Experimental and Theoretical Distribution of the Consent*, Vol. 2: *Nonparametric Goodness-of-Fit Test* (Izd. Standartov, Moscow, 2002) [in Russian].
11. G. V. Martynov, *Omega-Square Test* (Nauka, Moscow, 1978) [in Russian].
12. E. S. Pearson and H. O. Hartley , *Biometrica Tables for Statistics*, Vol. 2 (Univ. Press, Cambridge, 1972).
13. M. A. Stephens, "Use of Kolmogorov–Smirnov, Cramer–von Mises and Related Statistics Without Extensive Table," J. Roy. Statist. Soc. Ser. B, **32**, 115–122 (1970).
14. M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," J. Amer. Statist. Assoc. **69**, 730–737 (1974).
15. M. Chandra, N. D. Singpurwalla, and M. A. Stephens, "Kolmogorov Statistics for Test of Fit for the Extreme-Value and Weibull Distribution," J. Amer. Statist. Assoc. **76** (375), 729–731 (1981).
16. Yu. N. Tyurin, "On Limit Distribution of the Kolmogorov–Smirnov Statistics for a Composite Hypothesis," Izv. Akad. Nauk SSSR Ser. Mat. **48** (6), 1314–1343 (1984).
17. N. E. Savvushkina and Yu. N. Tyurin, "An Agreement Criterion for the Weibull–Gnedenko Distribution," Izv. Akad. Nauk SSSR Tekhn. Kibernet., No 3, 109–112 (1984) [Engrg. Cybernet. **22** (4), 113–117 (1984)].
18. Yu. N. Tyurin, *Studies in the Nonparametric Statistics* (*Nonparametric Methods and Linear Model*), Doctoral Dissertation in Mathematical Physics (Moscow, 1985).
19. N. E. Savvushkina, "The Kolmogorov–Smirnov Test for a Logistic and Gamma-Distribution," in *Proceedings of the Institute of System Studies*, No. 8, 50–56 (1990).
20. B. Yu. Lemeshko and S. N. Postovalov, "Some Aspects of Using the Goodness-of-Fit Tests for the Complex Hypothesis Testing," Nadezhnost' i Kontrol' Kachestva, No. 11, 3–17 (1997).
21. B. Yu. Lemeshko and S. N. Postovalov, "Statistical Distributions of Nonparametric Goodness-of-Fit Tests as Estimated by the Sample Parameters of Experimentally Observed Laws," Zavodskaya Lab. **64** (3), 61–72 (1998) [Industrial Lab. **64** (3), 197–208 (1998) (Consultants Bureau, New York)].
22. B. Yu. Lemeshko and S. N. Postovalov, "On the Dependence of the Statistics Distributions and Their Power on the Method for Estimating the Parameters," Zavodskaya Lab. Diagnostika Materialov **67** (7), 62–71 (2001).
23. B. Yu. Lemeshko and S. N. Postovalov, "Using the Nonparametric Goodness-of-Fit Tests for Complex Hypothesis Testing," Avtometriya, No. 2, 88–102 (2001).
24. B. Yu. Lemeshko, S. N. Postovalov, and A. V. Frantsuzov, "Application of the Nonparametric Goodness-of-Fit Tests to Testing Nonparametric Model Adequacy," Avtometriya, No. 2, 3–14 (2002) [Optoelectronics, Instrumentation and Data Processing **38** (2), 3–12 (2002)].
25. B. Yu. Lemeshko and S. N. Postovalov, "Nonparametric Goodness-of-Fit Tests for Complex Hypotheses With the Jonson Distributions," Dokl. Akad. Nauk Vyssh. Shkoly, No. 1(5), 65–74 (2002).
26. B. Yu. Lemeshko and A. A. Maklakov, "Nonparametric Test in Testing Composite Hypotheses on Goodness of Fit Exponential Family Distributions," Avtometriya, No. 3, 3–20 (2004) [Optoelectronics, Instrumentation and Data Processing **40** (3), 3–18 (2004)].
27. S. B. Lemeshko and B. Yu. Lemeshko, "Statistic Distributions of Nonparametric Goodness-of-Fit Tests for Hypotheses Under Beta-Distributions," Dokl. Akad. Nauk Vyssh. Shkoly, No. 2(9), 6–16 (2007).
28. H. Cramer, *Mathematical Methods of Statistics* (Princeton Univ. Press, Princeton, 1946; Mir, Moscow, 1975).
29. Birch M. W. "A New Proof of the Pearson–Fisher Theorem," Ann. Math. Statist. **35** (2), 817–824 (1964).
30. B. Yu. Lemeshko and E. B. Chimitova, "Optimal L-Estimates of the Parameters of Shift and Scale of Distribution Upon the Sample Quantiles," Zavodskaya Lab. Diagnostika Materialov **70** (1), 54–66 (2004).
31. H. Chernoff and E. L. Lehmann , "The Use of Maximum Likelihood Estimates in χ^2 Test for Goodness of Fit," Ann. Math. Statist. **25** (3), 579–586 (1954).
32. D. M. Chibisov, "Some Criteria of the Chi-Square Type for the Continues Distributions," Teor. Veroyatnost. i Primenen. **16** (1), 3–20 (1971).
33. D. S. Moore, "A Chi-Square Statistic With Random Sell Boundaries," Ann. Math. Statist. **42** (1), 147–156 (1971).
34. B. Yu. Lemeshko and S. N. Postovalov, "Application Aspects of Using the Goodness-of-Fit Tests in the Case of Complex Hypotheses," Nadezhnost' i Kontrol' Kachestva, No. 11, 3–17 (1997).
35. B. Yu. Lemeshko and S. N. Postovalov, " Limit Distributions of the Pearson χ^2 and Likelihood Ratio Statistics and Their Dependence on the Mode of Data Grouping," Zavodskaya Lab. **64** (5) 56–63 (1998) [Industrial Lab. **64** (5), 344–351 (1998) (Consultants Bureau, New York)].
36. V. I. Denisov and B. Yu. Lemeshko, "Optimal Grouping for the Experimental Data Treatment," in *Information Measuring Systems* (Novosibirsk, 1979), pp. 5–14.
37. V. I. Denisov, B. Yu. Lemeshko, and E. B. Tsoi, *Optimal Grouping, Estimating Parameters, and Planning Some Regression Experiments*, Vols. 1–2 (Novosibirsk. Gos. Tekhn. Univ., Novosibirsk, 1993) [in Russian].

38. B. Yu. Lemeshko, "Asymptotically Optimal Grouping of Observations Ensures the Maximal Power of the Criteria," Nadezhnost' i Kontrol' Kachestva, No. 8, 3–14 (1997).
39. B. Yu. Lemeshko, "Asymptotically Optimum Grouping of Observations in Goodness-of-Fit Tests," Zavodskaya Lab. **64** (1), 56–64 (1998) [Industrial Lab. **64** (1), 59–67 (1998) (Consultants Bureau, New York)].
40. *Recommendation on Standardization. Applied Statistics. Rules of Check of Experimental and Theoretical Distribution of the Consent*, Vol. 1: *Goodness-of-Fit Tests of the Chi-Square Type* (Izd. Standartov, Moscow, 2002) [in Russian].
41. V. I. Denisov, K.-H. Eger, B. Yu. Lemeshko, and E. B. Tsoi, *Design of Experiments and Statistical Analysis for Grouped Observations* (NSTU Publ. House, Novosibirsk, 2004).
42. M, G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 2: *Inference and Relationship*, 3rd ed. (Charles Griffin, London, 1973; Nauka, Moscow, 1973).
43. M. S. Nikulin, "About Chi-Square Goodness-of-Fit Test for Continuous Distributions," Teor. Veroyatnost. i Primenen. **18** (3), 675–676 (1973).
44. M. S. Nikulin, "Chi-Square Test for Continuous Distributions with Shift and Scale Parameters," Teor. Veroyatnost. i Primenen. **18** (3), 583–592 (1973) [J. Theory Probab. Appl. **18** (3), 559–568 (1973)].
45. M. Mirvaliev and M. S. Nikulin, "Goodness-of-Fit Tests of the Chi-Square Kind," Zavodskaya Lab. **58** (3), 52–58 (1992).
46. N. Aguirre and M. Nikulin, "Chi-Squared Goodness-of-Fit Test for the Family of Logistic Distributions," Kybernetika **30** (3), 214–222 (1994).
47. P. E. Greenwood and M. S. Nikulin, *A Guide to Chi-Squared Testing* (Wiley, New York, 1996).
48. K. C. Rao and D. S. Robson, "A Chi-Squared Statistic for Goodness-of-Fit Tests Within the Exponential Family," Comm. Statist. **3**, 1139–1153 (1974).
49. D. S. Moore and M. C. Spruill, "Unified Large-Sample Theory of General Chi-Squared Statistics for Tests of Fit," Ann. Statist. **3**, 599–616 (1975).
50. A. W. Van der Vaart, *Asymptotic Statistics* (Univ. Press, Cambridge, 1998).
51. S. B. Lemeshko, "Distributions of Chi-Squared Statistics for Goodness-of-Fit Tests in the Case of Finite Sample," in *Proceedings of the* 7^{th} *International Conference "Actual Problems of Electronic Equipment-Making APEP–*2006*"*, Vol. 6 (Novosibirsk, 2006), pp. 78–82.
52. B. Yu. Lemeshko and E. V. Chimitova, "About Choosing the Number of Intervals in Chi-Squared Goodness-of-Fit Tests," Zavodskaya Lab. Diagnostika Materialov **69** (1), 61–67 (2003).