DOI 10.1007/s11018-015-0759-2
Measurement Techniques, Vol. 58, No. 6, September, 2015

GENERAL PROBLEMS OF METROLOGY AND MEASUREMENT TECHNIQUE

CHI-SQUARE-TYPE TESTS FOR VERIFICATION OF NORMALITY

B. Yu. Lemeshko

UDC 519.24

The application of the Pearson chi-square test for verification of the normality of a sample is discussed. Tables of percentage points and models for the limiting statistical distributions are constructed. The powers of the Pearson and Nikulin–Rao–Robson chi-square tests are estimated relative to various competing hypotheses. A comparative analysis of the powers of a set of normality tests is given.
Keywords: *Pearson test, Nikulin–Rao–Robson test, test power.*

The application of many classical methods and tests for verification of statistical hypotheses is based on the assumption that the random quantities being analyzed obey a normal law. Only when this assumption is satisfied is it possible to be sure that a correct statistical conclusion has been reached using a given test.

Three groups of tests can be used to verify the hypothesis that a sample obeys a normal law. The use, advantages, and disadvantages of the special Shapiro–Wilk, Epps–Pulley, Frosini, Hegazi–Green, Spigelhalter, Geary, and David–Hartley–Pearson tests are discussed in detail in [1–4]. The use of the nonparametric Kolmogorov, Cramer–Mises–Smirnov, Anderson–Darling, Kuiper, and Watson tests of goodness-of-fit for composite hypotheses is discussed in greatest detail in [3, 5] and their application to tests of normality, in particular, is discussed in [4]. The Kolmogorov test for normality was first used in [6], the Cramer–Mises–Smirnov and Anderson–Darling tests were used for the same purpose in [7], the Kuiper and Watson tests in [8–10], and the Zhang test in [11]. Some disadvantages of the latter have been pointed out in [4].

Chi-square-type tests are traditionally used to test hypotheses regarding the adherence of a given sample to a normal law. The Pearson χ^2 test for composite hypotheses (including tests of normality) assumes that the unknown parameters of the distribution are estimated on the basis of grouped data, since when the estimates are made from an ungrouped sample the distributions of the test statistic differ greatly from χ^2 distributions. For this reason, a number of modified χ^2-type goodness-of-fit tests have been proposed. The best known of these is the Nikulin–Rao–Robson test [12–14].

Here we demonstrate the feasibility of using the Pearson χ^2 test for goodness-of-fit to a normal distribution with estimates of the parameters based on ungrouped data and use statistical modelling techniques to study the power of χ^2-type tests relative to several competing laws. For studying the distributions of the statistics, the number of Monte-Carlo trials was set at 10^6, which ensures an error on the order of $\pm 10^{-3}$ in estimating the probability distribution.

The Pearson χ^2 Test of Goodness-of-Fit. The procedure for hypothesis testing using χ^2-type tests assumes grouping of an original sample $X_1, X_2, ..., X_n$ of volume n. The domain of definition of the random variable is broken into k nonoverlapping intervals bounded by the points:

$$x_0 < x_1 < ... < x_{k-1} < x_k,$$

Novosibirsk State Technical University, Novosibirsk, Russia; e-mail: Lemeshko@ami.nstu.ru. Translated from Izmeritel'naya Tekhnika, No. 6, pp. 3–9, June, 2015. Original article submitted December 17, 2014.

where x_0, x_k are the lower and upper boundaries of the domain of definition of the random variable. The number of observations n_i in the ith interval is counted in accordance with this partition, and the probability of falling in this interval,

$$P_i(\theta) = \int_{x_{i-1}}^{x_i} f(x, \theta)\, dx,$$

corresponding to a theoretical distribution law with a density function $f(x, \theta)$, where

$$n = \sum_{i=1}^{k} n_i, \quad \sum_{i=1}^{k} P_i(\theta) = 1$$

is calculated. Measurements of the deviations n_i/n on $P_i(\theta)$ form the basis of the statistics used in χ^2-type goodness-of-fit tests.

The statistic for the Pearson χ^2 test is calculated using the formula

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

When the simple test hypothesis H_0 is true (i.e., all the parameters of the theoretical law are known) and $n \to \infty$, this statistic obeys a X_r^2 distribution with $r = k - 1$ degrees of freedom. A X_r^2 distribution has the density

$$g(s) = s^{r/2-1} e^{-s/2} / [2^{r/2} \Gamma(r/2)],$$

where $\Gamma(\cdot)$ is the Euler gamma function.

The test hypothesis H_0 is not rejected if the attained level of significance exceeds a specified level of significance α, i.e., if the following inequality holds:

$$P\{X_n^2 > X_n^{2*}\} = \frac{1}{2^{r/2} \Gamma(r/2)} \int_{X_n^{2*}}^{\infty} s^{r/2-1} e^{-s/2}\, ds > \alpha,$$

where X_n^{2*} is the statistic calculated in Eq. (1).

For testing a composite hypothesis and the validity of H_0 under conditions such that an estimate of the parameters is obtained by minimizing the statistic X_n^2 based on the same sample, this statistic obeys a X_r^2 distribution asymptotically with $r = k - m - 1$ degrees of freedom, where m is the number of parameters to be estimated. The statistic X_n^2 has the same distribution if the estimate is obtained by a maximum likelihood method and the estimates are calculated from grouped data by maximizing the likelihood function with respect to θ:

$$L(\theta) = \gamma \prod_{i=1}^{k} P_i^{n_i}(\theta), \tag{2}$$

where γ is a constant and

$$P_i(\theta) = \int_{x_{i-1}}^{x_i} f(x, \theta)\, dx$$

is the probability that an observations falls in the ith interval as a function of θ. This is also true for any estimation techniques based on grouped data leading to asymptotically effective estimates.

For tests of goodness-of-fit to a normal law and for evaluating the parameter vector $\hat{\theta}^T = (\hat{\mu}, \hat{\sigma})$ based on a grouped sample by minimizing the statistic X_n^2 or maximizing with respect to the likelihood function (2), the probabilities of falling into an interval are calculated using

$$P_i(\theta) = \frac{1}{\sqrt{2\pi}} \int_{t_{i-1}}^{t_i} e^{-t^2/2}\, dx,$$

TABLE 1. Optimal Boundary Points of Group Intervals for Testing of Simple and Composite Hypotheses Based on χ^2-Type Tests (for evaluating μ and σ) and the Corresponding Values of the Relative Asymptotic Information A

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
3	–1.1106	1.1106	–	–	–	–	–	–	–	–	–	–	–	–	0.4065
4	–1.3834	0.0	1.3834	–	–	–	–	–	–	–	–	–	–	–	0.5527
5	–1.6961	–0.6894	0.6894	1.6961	–	–	–	–	–	–	–	–	–	–	0.6826
6	–1.8817	–0.9970	0.0	0.9970	1.8817	–	–	–	–	–	–	–	–	–	0.7557
7	–2.0600	–1.2647	–0.4918	0.4918	1.2647	2.0600	–	–	–	–	–	–	–	–	0.8103
8	–2.1954	–1.4552	–0.7863	0.0	0.7863	1.4552	2.1954	–	–	–	–	–	–	–	0.8474
9	–2.3188	–1.6218	–1.0223	–0.3828	0.3828	1.0223	1.6218	2.3188	–	–	–	–	–	–	0.8753
10	–2.4225	–1.7578	–1.2046	–0.6497	0.0	0.6497	1.2046	1.7578	2.4225	–	–	–	–	–	0.8960
11	–2.5167	–1.8784	–1.3602	–0.8621	–0.3143	0.3143	0.8621	1.3602	1.8784	2.5167	–	–	–	–	0.9121
12	–2.5993	–1.9028	–1.4914	–1.0331	–0.5334	0.0	0.5334	1.0331	1.4914	1.9028	2.5993	–	–	–	0.9247
13	–2.6746	–2.0762	–1.6068	–1.1784	–0.7465	–0.2669	0.2669	0.7465	1.1784	1.6068	2.0762	2.6746	–	–	0.9348
14	–2.7436	–2.1609	–1.7092	–1.3042	–0.9065	–0.4818	0.0	0.4818	0.9065	1.3042	1.7092	2.1609	2.7436	–	0.9430
15	–2.8069	–2.2378	–1.8011	–1.4150	–1.0435	–0.6590	–0.2325	0.2325	0.6590	1.0435	1.4150	1.8011	2.2378	2.8069	0.9498

TABLE 2. Optimal Probabilities (frequencies) for Testing of Simple and Composite Hypotheses Based on χ^2-Type Tests (for evaluating μ and σ) and the Corresponding Values of the Relative Asymptotic Information A

k	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	A
3	0.1334	0.7332	0.1334	–	–	–	–	–	–	–	–	–	–	–	–	0.4065
4	0.0833	0.4167	0.4167	0.0833	–	–	–	–	–	–	–	–	–	–	–	0.5527
5	0.0449	0.2004	0.5094	0.2004	0.0449	–	–	–	–	–	–	–	–	–	–	0.6826
6	0.0299	0.1295	0.3406	0.3406	0.1295	0.0299	–	–	–	–	–	–	–	–	–	0.7557
7	0.0197	0.0833	0.2084	0.3772	0.2084	0.0833	0.0197	–	–	–	–	–	–	–	–	0.8103
8	0.0141	0.0587	0.1431	0.2841	0.2841	0.1431	0.0587	0.0141	–	–	–	–	–	–	–	0.8474
9	0.0102	0.0422	0.1009	0.1976	0.2982	0.1976	0.1009	0.0422	0.0102	–	–	–	–	–	–	0.8753
10	0.0077	0.0317	0.0748	0.1438	0.2420	0.2420	0.1438	0.0748	0.0317	0.0077	–	–	–	–	–	0.8960
11	0.0059	0.0243	0.0567	0.1074	0.1823	0.2468	0.1823	0.1074	0.0567	0.0243	0.0059	–	–	–	–	0.9121
12	0.0047	0.0190	0.0442	0.0829	0.1392	0.2100	0.2100	0.1392	0.0829	0.0442	0.0190	0.0047	–	–	–	0.9247
13	0.0037	0.0152	0.0352	0.0652	0.1085	0.1670	0.2104	0.1670	0.1085	0.0652	0.0352	0.0152	0.0037	–	–	0.9348
14	0.0030	0.0124	0.0283	0.0524	0.0862	0.1327	0.1850	0.1850	0.1327	0.0862	0.0524	0.0283	0.0124	0.0030	–	0.9430
15	0.0025	0.0101	0.0232	0.0427	0.0698	0.1066	0.1532	0.1838	0.1532	0.1066	0.0698	0.0427	0.0232	0.0101	0.0025	0.9498

TABLE 3. Percentage Points $\tilde{\chi}^2_{k,\alpha}$ for the Pearson Test Statistic when Evaluating the Parameters μ and σ

k	$p = 1 - \alpha$					Limiting distribution model
	0.85	0.9	0.95	0.975	0.99	
4	2.74	3.37	4.48	5.66	7.26	B_{III} (1.2463; 3.8690; 4.6352; 19.20; 0.005)
5	4.18	5.00	6.39	7.77	9.59	B_{III} (1.7377; 3.8338; 5.5721; 26.00; 0.005)
6	5.61	6.54	8.09	9.61	11.62	B_{III} (2.1007; 4.1518; 4.1369; 26.00; 0.005)
7	6.95	7.98	9.67	11.31	13.43	B_{III} (2.5019; 4.6186; 3.4966; 28.00; 0.005)
8	8.28	9.40	11.21	12.95	15.22	B_{III} (2.9487; 5.8348; 3.1706; 34.50; 0.005)
9	9.56	10.76	12.69	14.53	16.87	B_{III} (3.5145; 6.3582; 3.2450; 39.00; 0.005)
10	10.84	12.11	14.16	16.12	18.58	B_{III} (3.9756; 6.7972; 3.0692; 41.50; 0.005)
11	12.08	13.42	15.55	17.59	20.19	B_{III} (4.4971; 6.9597; 3.0145; 43.00; 0.005)
12	13.34	14.74	16.98	19.10	21.77	B_{III} (5.1055; 7.0049; 3.1130; 45.00; 0.005)
13	14.56	16.01	18.34	20.53	23.30	B_{III} (5.7809; 7.0217; 3.2658; 47.00; 0.005)
14	15.78	17.29	19.68	21.96	24.81	B_{III} (6.6673; 6.9116; 3.5932; 49.00; 0.005)
15	16.98	18.54	21.04	23.40	26.37	B_{III} (7.0919; 7.2961; 3.4314; 51.50; 0.005)

where $t_i = (x_i - \hat{\mu})/\hat{\sigma}$. The test hypothesis H_0 is not rejected if the attained level of significance $P\{X_n^2 > X_n^{2*}\}$ calculated according to the corresponding χ_r^2 distribution exceeds a specified level of significance α or if the value of the statistic X_n^{2*} is smaller than a critical value $\chi^2_{r,\alpha}$ given by

$$\frac{1}{2^{r/2}\Gamma(r/2)}\int_{\chi^2_{r,\alpha}}^{\infty} s^{r/2-1}e^{-s/2}\,ds = \alpha.$$

For maximum likelihood estimates (MLE) based on ungrouped data, this statistic is distributed as the sum of independent terms $\chi^2_{k-m-1} + \sum_{j=1}^{m}\lambda_j\xi_j^2$, where $\xi_1, ..., \xi_m$ are standard normal random quantities that are independent of one another and of χ^2_{k-m-1}; $\lambda_1, ..., \lambda_m$ are numbers between 0 and 1 [15] representing the roots of the equation

$$|(1-\lambda)\mathbf{J}(\theta) - \mathbf{J}_g(\theta)| = 0.$$

Here $\mathbf{J}(\theta)$ is the Fisher information matrix with respect to the ungrouped observations with elements

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

$\mathbf{J}_g(\theta)$ is the information matrix with respect to the grouped observations, with

$$\mathbf{J}_g(\theta) = \sum_{i=1}^{k}\nabla P_i(\theta)\nabla^{T}P_i(\theta)/P_i(\theta).$$

In other words, the distribution of the (1) statistic based on MLE with respect to ungrouped data is unknown and depends, in particular, on the grouping method [16].

The asymptotically optimal groupings (AOG) listed in Tables 1 and 2 can be used for testing of normality based on MLE estimates of the parameters μ and σ for samples with ungrouped data. Here the losses in the Fisher information on the

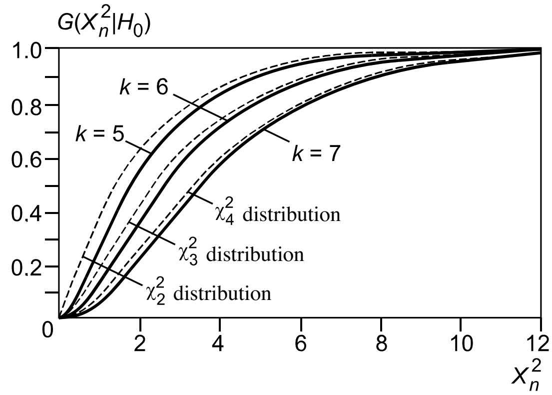

Fig. 1. Distributions of the statistic (1) for maximum likelihood estimates of the parameters of a normal distribution based on ungrouped data together with the corresponding χ^2_{k-m-1} distributions.

parameters of the distribution associated with grouping are minimized [3] and the Pearson χ^2 test has maximal power relative to the very close competing hypotheses [3].

In Table 1, the boundary points for the interval t_i, $i = 1, ..., k-1$ are listed in a form that is invariant with respect to the parameters μ and σ for a normal distribution. For calculating the statistic (1), the boundaries x_i separating the intervals for specified k are found using the values of t_i taken from the corresponding row of the table: $x_i = \hat{\sigma} t_i + \hat{\mu}_i$, where $\hat{\mu}$ and $\hat{\sigma}$ are the MLE of the parameters derived from the given sample. Then the number of observations n_i within each interval are used. The probabilities of falling into a given interval for evaluating the statistic (1) are taken from the corresponding row of Table 2.

When AOG is used in the Pearson χ^2 test, the resulting percentage points $\tilde{\chi}^2_{k,\alpha}$ of the distributions of the statistic (1) and the models of limiting distributions constructed in this paper are shown in Table 3, where B_{III} $(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ is the type III beta distribution with these parameters and the density

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

To make a decision regarding testing the hypothesis H_0, the value of the statistic X_n^{2*} is compared with the critical value $\tilde{\chi}^2_{k,\alpha}$ from the corresponding row of Table 3, or the attained level of significance $P\{X_n^2 > X_n^{2*}\}$, determined using the limiting distribution model in the same row of the table, is compared with a specified level of significance α.

The difference between the real distributions $G(X_n^2 | H_0)$ of the (1) statistic and the corresponding χ^2_{k-m-1} distributions when hypothesis H_0 is true is shown in Fig. 1.

Tables 1 and 2 give Fisher asymptotic information:

$$A = \det \mathbf{J}_g / \det \mathbf{J}.$$

For tests of normality with calculations of an MLE based on the ungrouped sample, only the parameters μ or σ, the required AOG tables, percentage points, and the limiting distribution models can be found in [4].

For AOG relative to the parameter vector and $k = 15$ intervals in the grouped sample, about 95% of the information is preserved. Further increases in the number k of intervals are insignificant; it should be chosen based on the following considerations. For an optimal grouping, the probabilities of falling into an interval are not generally equal (usually these probabilities are minimal for the outermost intervals), so that k should be chosen on the basis of the condition $nP_i(\theta) \geq 5$–10 for any interval. At least, in choosing k the recommendation

$$\min_i \{nP_i(\theta) \mid i = 1, ..., k\} > 1$$

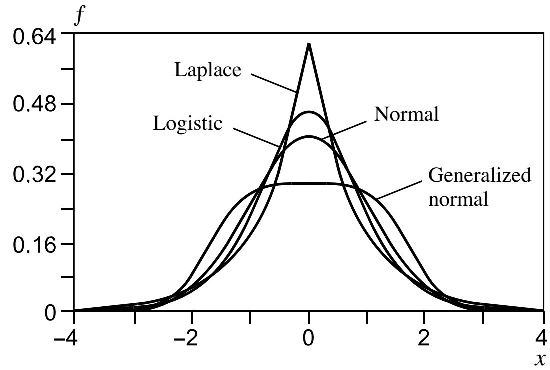

Fig. 2. Probability densities corresponding to the hypotheses H_i examined here.

should be followed. When this condition holds, in the case where the tested hypothesis H_0 is valid, a discrete distribution of the statistic in (1) differs insignificantly from the corresponding asymptotic limiting distribution. If this condition is violated, then the difference between the true distribution of the statistic and the limiting distribution will lead to an increase in the probability of a type I error relative to the specified significance level α. It should also be noted that for small sample sizes, n = 10–20, discrete distributions of the statistics differ substantially from the asymptotic distributions. This condition on the choice of k sets an upper bound estimate on the number of intervals ($k \le k_{max}$). The number of grouping intervals affects the power of the Pearson χ^2 test [17]. It is absolutely unnecessary that its power against a competing distribution (hypothesis) should be maximal for $k = k_{max}$.

In order to compare the power of the Pearson χ^2 test for verification of normality with the power of special and nonparametric goodness-of-fit tests, the power has been estimated relative to the same competing distributions (hypotheses) as in [4].

The test hypothesis H_0 is taken to be that the observed sample obeys the normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

As competing hypotheses for studying the power of the χ^2 distribution, we have considered adherence of the analyzed sample to the following distributions: competing hypothesis H_1 corresponds to a generalized normal distribution (family of distributions) with the density

$$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)} \exp\left\{-\left(\frac{|x-\theta_0|}{\theta_1}\right)^{\theta_2}\right\}$$

and a shape parameter $\theta_2 = 4$; hypothesis H_2 is the Laplace distribution with the density

$$f(x) = \frac{1}{2\theta_1} \exp\left\{-\frac{|x-\theta_0|}{\theta_1}\right\};$$

and hypothesis H_3 is the logistic distribution with the density

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

which is very close to a normal distribution. Figure 2 shows the densities of the distributions corresponding to hypotheses H_1, H_2, and H_3 with scale parameters such that they are closest to a standard normal law. This choice of hypotheses has a certain justification. Hypothesis H_2, corresponding to a Laplace distribution, is the most distant from H_0. Distinguishing them usually

TABLE 4. Power of the Pearson χ^2 Test with Respect to Hypotheses H_1, H_2, and H_3

<table>
<tr><th rowspan="2">n</th><th rowspan="2">k_{max}</th><th rowspan="2">k_{opt}</th><th colspan="5">α</th></tr>
<tr><th>0.15</th><th>0.1</th><th>0.05</th><th>0.025</th><th>0.01</th></tr>
<tr><td colspan="8">H1</td></tr>
<tr><td>10</td><td>4</td><td>4</td><td>0.235</td><td>0.146</td><td>0.043</td><td>0.032</td><td>0.002</td></tr>
<tr><td>20</td><td>4</td><td>5</td><td>0.262</td><td>0.177</td><td>0.100</td><td>0.058</td><td>0.021</td></tr>
<tr><td>30</td><td>5</td><td>5</td><td>0.312</td><td>0.216</td><td>0.136</td><td>0.079</td><td>0.043</td></tr>
<tr><td>40</td><td>6</td><td>5</td><td>0.336</td><td>0.267</td><td>0.168</td><td>0.111</td><td>0.061</td></tr>
<tr><td>50</td><td>6</td><td>5</td><td>0.401</td><td>0.311</td><td>0.204</td><td>0.129</td><td>0.068</td></tr>
<tr><td>100</td><td>9</td><td>5</td><td>0.558</td><td>0.479</td><td>0.352</td><td>0.254</td><td>0.158</td></tr>
<tr><td>150</td><td>10</td><td>7</td><td>0.722</td><td>0.634</td><td>0.486</td><td>0.353</td><td>0.217</td></tr>
<tr><td>200</td><td>11</td><td>9</td><td>0.783</td><td>0.695</td><td>0.548</td><td>0.417</td><td>0.279</td></tr>
<tr><td>300</td><td>13</td><td>11</td><td>0.907</td><td>0.858</td><td>0.756</td><td>0.646</td><td>0.492</td></tr>
<tr><td colspan="8">H2</td></tr>
<tr><td>10</td><td>4</td><td>4</td><td>0.267</td><td>0.206</td><td>0.074</td><td>0.058</td><td>0.01</td></tr>
<tr><td rowspan="2">20</td><td rowspan="2">4</td><td>4</td><td>0.264</td><td>0.177</td><td>0.104</td><td>0.067</td><td>0.037</td></tr>
<tr><td>5</td><td>0.247</td><td>0.189</td><td>0.116</td><td>0.061</td><td>0.024</td></tr>
<tr><td>30</td><td>5</td><td>5</td><td>0.312</td><td>0.261</td><td>0.153</td><td>0.103</td><td>0.044</td></tr>
<tr><td>40</td><td>6</td><td>7</td><td>0.443</td><td>0.358</td><td>0.250</td><td>0.167</td><td>0.101</td></tr>
<tr><td>50</td><td>6</td><td>7</td><td>0.500</td><td>0.423</td><td>0.312</td><td>0.225</td><td>0.138</td></tr>
<tr><td>100</td><td>9</td><td>9</td><td>0.770</td><td>0.708</td><td>0.596</td><td>0.494</td><td>0.379</td></tr>
<tr><td>150</td><td>10</td><td>9</td><td>0.899</td><td>0.860</td><td>0.785</td><td>0.705</td><td>0.596</td></tr>
<tr><td>200</td><td>11</td><td>11</td><td>0.964</td><td>0.946</td><td>0.908</td><td>0.880</td><td>0.786</td></tr>
<tr><td>300</td><td>13</td><td>13</td><td>0.996</td><td>0.993</td><td>0.985</td><td>0.974</td><td>0.950</td></tr>
<tr><td colspan="8">H3</td></tr>
<tr><td>10</td><td>4</td><td>4</td><td>0.221</td><td>0.150</td><td>0.046</td><td>0.034</td><td>0.003</td></tr>
<tr><td>20</td><td>4</td><td>4</td><td>0.194</td><td>0.125</td><td>0.059</td><td>0.038</td><td>0.016</td></tr>
<tr><td>30</td><td>5</td><td>5</td><td>0.169</td><td>0.125</td><td>0.062</td><td>0.034</td><td>0.012</td></tr>
<tr><td>40</td><td>6</td><td>7</td><td>0.204</td><td>0.143</td><td>0.082</td><td>0.045</td><td>0.020</td></tr>
<tr><td>50</td><td>6</td><td>7</td><td>0.214</td><td>0.155</td><td>0.088</td><td>0.050</td><td>0.023</td></tr>
<tr><td>100</td><td>9</td><td>10</td><td>0.303</td><td>0.231</td><td>0.146</td><td>0.090</td><td>0.047</td></tr>
<tr><td>150</td><td>10</td><td>10</td><td>0.359</td><td>0.284</td><td>0.191</td><td>0.124</td><td>0.072</td></tr>
<tr><td>200</td><td>11</td><td>11</td><td>0.432</td><td>0.355</td><td>0.250</td><td>0.175</td><td>0.105</td></tr>
<tr><td>300</td><td>13</td><td>13</td><td>0.566</td><td>0.486</td><td>0.373</td><td>0.280</td><td>0.190</td></tr>
</table>

presents no problem. The logistic distribution (hypothesis H_3) is very close to normal and it is generally difficult to distinguish them by goodness-of-fit tests.

The competing hypothesis H_1, which corresponds to a generalized normal distribution with a shape factor $\theta_2 = 4$, is a "litmus test" for detection of hidden deficiencies in some tests [1, 2, 4]. It turned out that for small sample sizes n and small specified probabilities α of type I error, a number of tests employed for testing goodness-of-fit to normal are not able to distinguish close distributions from normal. In these cases, the power $1 - \beta$ with respect to hypothesis H_1, where β is the probability of a type II error, is smaller than α. This means that the distribution corresponding to H_1 is "more normal than normal" and indicates that the tests are biased.

The power of the Pearson χ^2 test was studied with different number of intervals $k \le k_{max}$ for specified sample sizes n. Table 4 lists the maximum powers of the χ^2 test relative to the competing hypotheses H_1, H_2, and H_3, and corresponding to the optimal number k_{opt} of grouping intervals. To a certain extent, it is possible to orient oneself in choosing k on the basis of the values of k_{opt} as a function of n listed in Table 4.

The Nikulin–Rao–Robson Goodness-of-Fit Test. A variant of the standard statistic X_n^2 was proposed [12–14] in which the limiting distribution of the modified statistic is a X_{k-1}^2 distribution (the number of degrees of freedom is independent of the number of parameters to be estimated). The unknown parameters of the distribution $F(x, \theta)$ have, in this case, to be estimated on the basis of the ungrouped data by a maximum likelihood method. Here the vector $\mathbf{P} = (P_1, ..., P_k)^{\mathrm{T}}$ is assumed to be specified, while the boundary points of the intervals are defined using the relations $x_i(\theta) = F^{-1}(P_1 + ... + P_i)$, $i = 1, ..., k - 1$. The proposed statistic has the form [13]:

$$Y_n^2(\theta) = X_n^2 + n^{-1} a^{\mathrm{T}}(\theta)\Lambda(\theta)a(\theta), \tag{3}$$

where X_n^2 is calculated using Eq. (1). For distribution laws that are determined only by shift and scale parameters,

$$\Lambda(\theta) = [\mathbf{J}(\theta) - \mathbf{J}_{\mathrm{g}}(\theta)]^{-1};$$

in the case of a normal distribution with a parameter vector $\theta^{\mathrm{T}} = (\mu, \sigma)$, the Fisher information matrix has the form:

$$\mathbf{J}(\theta) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{bmatrix},$$

with the elements of the information matrix based on grouped data $\mathbf{J}_{\mathrm{g}}(\theta)$ given by

$$J_{\mathrm{g}}(\mu,\mu) = \sum_{i=1}^{k} \frac{1}{\sigma^2 P_i(\theta)} (f(t_{i-1}) - f(t_i))^2;$$

$$J_{\mathrm{g}}(\sigma,\sigma) = \sum_{i=1}^{k} \frac{1}{\sigma^2 P_i(\theta)} (t_{i-1} f(t_{i-1}) - t_i f(t_i))^2;$$

$$J_{\mathrm{g}}(\mu,\sigma) = J_{\mathrm{g}}(\sigma,\mu) = \sum_{i=1}^{k} \frac{1}{\sigma^2 P_i(\theta)} (f(t_{i-1}) - f(t_i))(t_{i-1} f(t_{i-1}) - t_i f(t_i));$$

where

$$t_i = (x_i - \mu)/\sigma; \quad t_0 = -\infty; \quad t_k = \infty; \quad f(t) = \left[\sqrt{2\pi} e^{t^2/2}\right]^{-1}$$

is the standard normal distribution. The elements of the vector $a^{\mathrm{T}}(\theta) = [a(\mu), a(\sigma)]$ are given by

$$a(\mu) = \sum_{i=1}^{k} \frac{n_i (f(t_{i-1}) - f(t_i))}{\sigma P_i(\theta)},$$

$$a(\sigma) = \sum_{i=1}^{k} \frac{n_i}{\sigma P_i(\theta)} (t_{i-1} f(t_{i-1}) - t_i f(t_i)).$$

As in the case of the Pearson test, when testing for normality with MLE estimation of the parameters μ and σ based on the ungrouped data, Tables 1 and 2 can be used.

TABLE 5. Power of the Nikulin–Rao–Robson Set with Respect to Hypotheses H_1, H_2, and H_3

n	k_{max}	k_{opt}	α				
			0.15	0.1	0.05	0.025	0.01
H_1							
10	4	4	0.348	0.1	0.029	0.009	0.006
20	4	5	0.234	0.143	0.074	0.041	0.016
30	5	5	0.256	0.197	0.102	0.053	0.023
40	6	5	0.293	0.221	0.123	0.079	0.035
50	6	5	0.326	0.240	0.148	0.083	0.040
100	9	5	0.485	0.395	0.271	0.179	0.102
150	10	6	0.619	0.530	0.397	0.284	0.179
		7	0.641	0.539	0.383	0.261	0.148
200	11	9	0.713	0.616	0.464	0.339	0.214
300	13	11	0.872	0.810	0.695	0.573	0.420
H_2							
10	4	4	0.368	0.103	0.055	0.031	0.007
20	4	5	0.250	0.210	0.126	0.065	0.039
30	5	6	0.349	0.265	0.185	0.127	0.078
40	6	7	0.474	0.403	0.297	0.218	0.149
50	6	7	0.548	0.473	0.365	0.281	0.190
100	9	9	0.807	0.755	0.667	0.583	0.482
150	10	9	0.919	0.889	0.834	0.774	0.691
200	11	11	0.973	0.961	0.933	0.900	0.849
300	13	11	0.997	0.995	0.990	0.983	0.968
		13	0.997	0.995	0.990	0.983	0.968
H_3							
10	4	4	0.321	0.083	0.034	0.014	0.005
20	4	5	0.166	0.120	0.065	0.030	0.014
30	5	6	0.198	0.138	0.080	0.047	0.024
40	6	**7**	0.232	0.173	0.104	0.063	0.034
50	6	**7**	0.251	0.188	0.117	0.074	0.040
100	9	10	0.360	0.290	0.202	0.141	0.091
150	10	10	0.432	0.358	0.263	0.195	0.131
200	11	11	0.509	0.436	0.337	0.259	0.183
300	13	13	0.641	0.572	0.469	0.381	0.288

For calculating the (3) statistic, the boundaries x_i separating the intervals for given k are found from the values of t_i in the corresponding row of Table 1 using the formula $x_i = \hat{\sigma} t_i + \hat{\mu}$, where $\hat{\mu}$ and $\hat{\sigma}$ are the MLE parameters found from the sample data. Then the number of observations n_i in each interval is counted. The probabilities $P_i(\theta)$ of falling into an interval when calculating statistic (3) are taken from the corresponding line of Table 2. The elements of the vector $a(\theta)$ and matrix $\Lambda(\theta)$ are calculated using the tabulated data for t_i, P_i and the resulting estimates of $\hat{\sigma}$. To decide on the test results for hypothesis H_0, the value of the statistic Y_n^{2*} is compared with the corresponding critical $\chi^2_{k-1,\alpha}$ or the attained level of significance $P\{Y_n^2 > Y_n^{2*}\}$ is found from the corresponding χ^2_{k-1} distribution. To test for normality with MLE calculation of the parameters μ or σ separately on the basis of ungrouped samples, the required tables of AOG can be found in [4].

Estimates of the power of the Nikulin–Rao–Robson test compared to the competing hypotheses H_1, H_2, and H_3 for k_{opt} are given in Table 5. This test is generally more powerful than the Pearson test (for example, see its powers relative to the competing hypotheses H_2 and H_3). Here we often have $k_{\text{opt}} = k_{\max}$ for

$$\min_i \{nP_i(\theta)\} > 1.$$

However, this is not always so. In terms of its power relative to the "tricky" hypothesis H_1, it is inferior to the Pearson test, and k_{opt} in this case is considerably smaller than $k_{\max}$ with AOG.

If we combine the results of studies of the powers of the tests discussed in this paper with the results of [1, 2, 4], then the entire set of tests used to test for normality can be ordered in terms of their power relative to the competing hypotheses H_1, H_2, and H_3 in the following way:

• relative to competing hypothesis H_1:

D'Agostino z_2 > *David–Hartley–Pearson* > *Geary* > *Shapiro–Wilk** > *Pearson* χ^2 > *Zhang* Z_C^* > *Watson* > *Anderson–Darling* > > *Frosini* > *Royston* > *Kuiper* > *Epps–Pulley** > *Cramer–Mises–Smirnov* > *Nikulin–Rao–Robson* > *Zhang* Z_A^* > *Spiegelhalter** > > *Kolmogorov* > *Zhang* Z_K > *D'Agostino* $z_1^2 + z_2^2$ > *Hegazi–Green* T_1^* > *Hegazi–Green* T_2^*;

• relative to competing hypothesis H_2:

Spiegelhalter > *Hegazi–Green* T_2 > *Geary* > *Hegazi–Green* T_1 > *D'Agostino* $z_1^2 + z_2^2$ > *Anderson–Darling* > *Watson* > > *Epps–Pulley* ~ *Frosini* > *Cramer–Mises–Smirnov* > *Royston* > *Kuiper* > *Zhang* Z_A > *Zhang* Z_K > *Zhang* Z_C > *Kolmogorov* >*Shapiro–Wilk* > *David–Hartley–Pearson* > *D'Agostino* z_2 > *Nikulin–Rao–Robson* > *Pearson* χ^2;

• relative to competing hypothesis H_3:

Hegazi–Green T_2 > *D'Agostino* $z_1^2 + z_2^2$ > *Spiegelhalter* > *Royston* > *Geary* > *Zhang* Z_A > *Zhang* Z_C > *Hegazi–Green* T_1 >*David–Hartley–Pearson* > *Epps–Pulley* ~ *Zhang* Z_K > *D'Agostino* z_2 > *Anderson–Darling* > *Frosini* > *Cramer–Mises–Smirnov* > > *Watson* > *Shapiro–Wilk* > *Kuiper* > *Kolmogorov* > *Nikulin–Rao–Robson* > *Pearson* χ^2.

Here it should be noted that for small n, the series of tests indicated above by an asterisk cannot differ from the normal distribution corresponding to hypothesis H_1 owing to the bias of the distribution of the statistics for these tests [1–4].

This work was supported by the Ministry of Education and Science of the Russian Federation as part of government assignment No. 2.541.2014/K.

REFERENCES

1. B. Yu. Lemeshko and S. B. Lemeshko, "Comparative analysis of tests for verifying deviation of a distribution from normal," *Metrologiya*, No. 2, 3–24 (2005).
2. B. Yu. Lemeshko and A. P. Rogozhnikov, "Features and power of some tests of normality," *Metrologiya*, No. 4, 3–24 (2009).
3. B. Yu. Lemeshko, S. B. Lemeshko, S. N. Postovalov, and E. V. Chimitova, *Statistical Analysis of Data, Modelling, and Study of Probability Behavior. A Computerized Approach: Monograph*, Izd NGTU, Novosibirsk (2011).
4. B. Yu. Lemeshko, *Tests for Verification of Deviation of Distributions from Normal: Application Handbook*, INFRA-M, Moscow (2014).
5. B. Yu. Lemeshko, *Nonparametric Tests of Goodness-of-Fit: Application Handbook*, INFRA-M, Moscow (2015).
6. H. W. Lilliefors, "On the Kolmogorov–Smirnov test for normality with mean and variance unknown," *J. Am. Statist. Assoc.*, **62**, 399–402 (1967).

7. G. V. Martynov, *Omega-Squared Tests*, Nauka, Moscow (1978).
8. B. Yu. Lemeshko and A. A. Gorbunova, "Application and power of the nonparametric Kuiper, Watson, and Zhang tests of goodness-of-fit," *Izmer. Tekhn.*, No. 5, 3–9 (2013).
9. B. Yu. Lemeshko and A. A. Gorbunova, "Application of nonparametric Kuiper and Watson tests of goodness-of-fit for composite hypotheses," *Izmer. Tekhn.*, No. 9, 14–21 (2013).
10. B. Yu. Lemeshko, A. A. Gorbunova, S. B. Lemeshko, and A. P. Rogozhnikov, "Solving problems of using some nonparametric goodness-of-fit tests," *Optoelectr., Instrum., Data Proc.*, **50**, No. 1, 21–35 (2014).
11. J. Zhang, *Powerful Goodness-of-Fit and Multi-Sample Tests: Ph.D. Thesis*, Toronto (2001).
12. M. S. Nikulin, "χ^2 tests for continuous distributions with shift and scale parameters," *Teor. Veroyatn. Primen.*, **XVIII**, No. 3, 583–591 (1973).
13. M. S. Nikulin, "On χ^2 tests for continuous distributions," *Teor. Veroyatn. Primen.*, **XVIII**, No. 3, 675–676 (1973).
14. K. C. Rao and D. S. Robson, "A chi-squared statistic for goodness-of-fit tests within the exponential family," *Commun. Stat.*, **3**, 1139–1153 (1974).
15. H. Chernoff and E. L. Lehmann, "The use of maximum likelihood estimates in χ^2 test for goodness of fit," *Ann. Math. Stat.*, **25**, 579–586 (1954).
16. B. Yu. Lemeshko and S. N. Postovalov, "On the dependence of the limiting distributions of statistics for Pearson χ^2 tests and the likelihood ratio on the means of grouping data," *Zavod. Lab.*, **64**, No. 5, 56–63 (1998).
17. B. Yu. Lemeshko and E. V. Chimitova, "On the choice of the number of intervals in χ^2-type goodness-of-fit tests," *Zavod. Lab. Diagn. Mater.*, **69**, No. 1, 61–67 (2003).