

Chi-Squared Goodness-of-Fit Tests: The Optimal Choice of Grouping Intervals

Ekaterina V. Chimitova^(✉) and Boris Yu. Lemeszko

Novosibirsk State Technical University, Novosibirsk, Russia
chimitova@corp.nstu.ru

Abstract. When using the chi-squared goodness-of-fit tests, the problem of choosing boundary points and the number of grouping intervals is always urgent, as the power of these tests considerably depends on the grouping method used. In this paper, the investigation of the power of the Pearson and Nikulin-Rao-Robson chi-squared tests has been carried out for various numbers of intervals and grouping methods. The partition of the real line into equiprobable intervals is not an optimal grouping method, as a rule. It has been shown that asymptotically optimal grouping, for which the loss of the Fisher information from grouping is minimized, enables to maximize the power of the Pearson test against close competing hypotheses. In order to find the asymptotically optimal boundary points, it is possible to maximize some functional (the determinant, the trace or the minimum eigenvalue) of the Fisher information matrix for grouped data. The versions of asymptotically optimal grouping method maximize the test power relative to a set of close competing hypotheses, but they do not insure the largest power against some given competing hypothesis. For the given competing hypothesis H_1 , it is possible to construct the chi-squared test, which has the largest power for testing hypothesis H_0 against H_1 . For example, in the case of the Pearson chi-squared test, it is possible to maximize the non-centrality parameter for the given number of intervals. So, the purpose of this paper is to give the methods for the choice of optimal grouping intervals for chi-squared goodness-of-fit tests.

Keywords: Chi-squared goodness-of-fit tests · Optimal grouping · Fisher information · Test power

1 Introduction

The χ^2 Pearson goodness-of-fit test is very popular in various applications, including the investigation of distributions of measurement error in problems of metrological support.

The correct usage of the Pearson χ^2 test for composite hypotheses (including testing normality) provides estimation of unknown parameters by the grouped data, as in the case of calculating parameter estimates by the original non-grouped sample the test statistic distribution differs from the χ^2 – distribution significantly [1, 2]. By this reason, a series of modified χ^2 tests has been offered, the most famous of which is the Nikulin-Rao-Robson test [3–5]. Moreover, it is necessary to take into account, that the power of Pearson test depends on the number of grouping intervals [6] and the grouping method used [7].

2 The Pearson χ^2 Test of Goodness-of-Fit

The procedure for hypothesis testing using χ^2 type tests assumes grouping an original sample X_1, X_2, \dots, X_n of size n . The domain of definition of the random variable is divided into k non-overlapping intervals bounded by the points:

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

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

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

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

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

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 of Pearson χ^2 test is calculated using the formula

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

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

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

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

The test hypothesis H_0 is not rejected if the achieved significance level (p -value) 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 (1).

When testing a composite hypothesis and estimating parameters by minimizing the statistic X_n^2 basing on the same sample, this statistic asymptotically obeys the χ_r^2 distribution with $r = k - m - 1$ degrees of freedom, where m is the number of parameters estimated.

The statistic X_n^2 has the same distribution if parameter estimate is obtained by the maximum likelihood method 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 observation falls into i -th interval. This result remains for any estimation technique based on grouped data leading to asymptotically effective estimates.

If unknown parameters are estimated by the maximum likelihood method basing on non-grouped data, then the Pearson statistic is distributed as the sum of independent terms [1] $\chi_{k-m-1}^2 + \sum_{j=1}^m \lambda_j \xi_j^2$, where ξ_1, \dots, ξ_m are standard normal random quantities that are independent from each other and from χ_{k-m-1}^2 ; $\lambda_1, \dots, \lambda_m$ are numbers between 0 and 1, 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 non-grouped observations with elements

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

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

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

In other words, the distribution of statistic (1), based on maximum likelihood estimates (MLE) calculated by non-grouped data, is unknown and depends, in particular, on the grouping method [2].

3 The Choice of Grouping Intervals

When using chi-squared goodness-of-fit tests, the problem of choosing boundary points and the number of grouping intervals is always important, as the power of these tests considerably depends on the grouping method used. In the case of complete samples (without censored observations), this problem was investigated in [7–10]. In particular, in [11], the investigation of the power of the Pearson and NRR tests for complete samples has been carried out for various numbers of intervals and grouping methods. The partition of the real line into equiprobable intervals (EPG) is not an optimal grouping method, as a rule. In [12], it was shown for the first time that asymptotically optimal grouping, for which the loss of the Fisher information from grouping is minimized, enables us to maximize the power of the Pearson test against close competing hypotheses. For example, it is possible to maximize the determinant of the Fisher information matrix for grouped data $\mathbf{J}_g(\theta)$, i.e. to solve the problem of D-optimal grouping

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \det(\mathbf{J}_g(\theta)). \tag{3}$$

In the case of the A-optimality criterion, the trace of the information matrix $\mathbf{J}_g(\theta)$ is maximized by the boundary points

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \text{Tr}(\mathbf{J}_g(\theta)), \tag{4}$$

and the E-optimality criterion maximizes the minimum eigenvalue of the information matrix:

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} \min_{i=1,2} \lambda_i(\mathbf{J}_g(\theta)). \tag{5}$$

The problem of asymptotically optimal grouping by the A- and E-optimality criteria has been solved for certain distribution families, and the tables of A-optimal grouping are given in [13]. The versions of asymptotically optimal grouping maximize the test power relative to a set of close competing hypotheses, but they do not ensure the highest power against some given competing hypothesis. For the given competing hypothesis H_1 , it is possible to construct the χ^2 test, which has the highest power for testing hypothesis H_0 against H_1 . For example, in the case of χ^2 Pearson test, it is possible to maximize the non-centrality parameter for the given number of intervals k :

$$\max_{x_0 < x_1 < \dots < x_{k-1} < x_k} n \sum_{j=1}^k \frac{(p_j^1(\theta^1) - p_j^0(\theta^0))^2}{p_j^0(\theta^0)}, \tag{6}$$

where $p_j^0(\theta^0) = \int_{x_{j-1}}^{x_j} f_0(u, \theta^0) du$, $p_j^1(\theta^1) = \int_{x_{j-1}}^{x_j} f_1(u, \theta^1) du$ are the probabilities to fall into j -th interval according to the hypotheses H_0 and H_1 , respectively. Let us refer this grouping method to as optimal grouping.

Asymptotically optimal boundary points, corresponding to different optimality criteria, as well as the optimal points, corresponding to (6), are considerably different from each other.

For example, the boundary points maximizing criteria (3)–(6) for the following pair of competing hypotheses are given in Table 1. The null hypothesis H_0 is the normal distribution with density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}, \tag{7}$$

and parameters $\mu = 0, \sigma = 1$ and the competing hypothesis H_1 is the logistic distribution with density function

$$f(x) = \frac{\pi}{\theta_1\sqrt{3}} \exp\left\{-\frac{\pi(x - \theta_0)}{\theta_1\sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x - \theta_0)}{\theta_1\sqrt{3}}\right\}\right]^2, \tag{8}$$

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

Table 1. Optimal boundary points for $k = 9$

Optimality criterion	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
A-optimum	2.3758	1.6915	1.1047	0.4667	0.4667	1.1047	1.6915	2.3758
D-optimum	2.3188	1.6218	1.0223	0.3828	0.3828	1.0223	1.6218	2.3188
E-optimum	1.8638	1.1965	0.6805	0.2216	0.2216	0.6805	1.1965	1.8638
Optimal grouping	3.1616	2.0856	1.2676	0.4601	0.4601	1.2676	2.0856	3.1616

Moreover, in the case of the given competing hypothesis, we can use the so-called Neyman-Pearson classes [14], for which the random variable domain is partitioned into intervals of two types, according to the inequalities $f_0(t) < f_1(t)$ and $f_0(t) > f_1(t)$, where $f_0(t)$ and $f_1(t)$ are the density functions, corresponding to the competing hypotheses. For H_0 and H_1 from our example, we have the first-type intervals

$$(-\infty; 2.3747], (0.6828; 0.6828], (2.3747; \infty),$$

and the second-type intervals

$$(2.3747; 0.6828], (0.6828; 2.3747].$$

Figures 1 and 2 illustrate the power of the Pearson χ^2 test for the hypotheses H_0 and H_1 of our example in the case of different grouping methods, depending on the number of intervals ($\alpha = 0.1, n = 500$). The powers of the well-known nonparametric Kolmogorov, Cramer-von Mises-Smirnov and Anderson-Darling goodness-of-fit tests are given for the comparison.

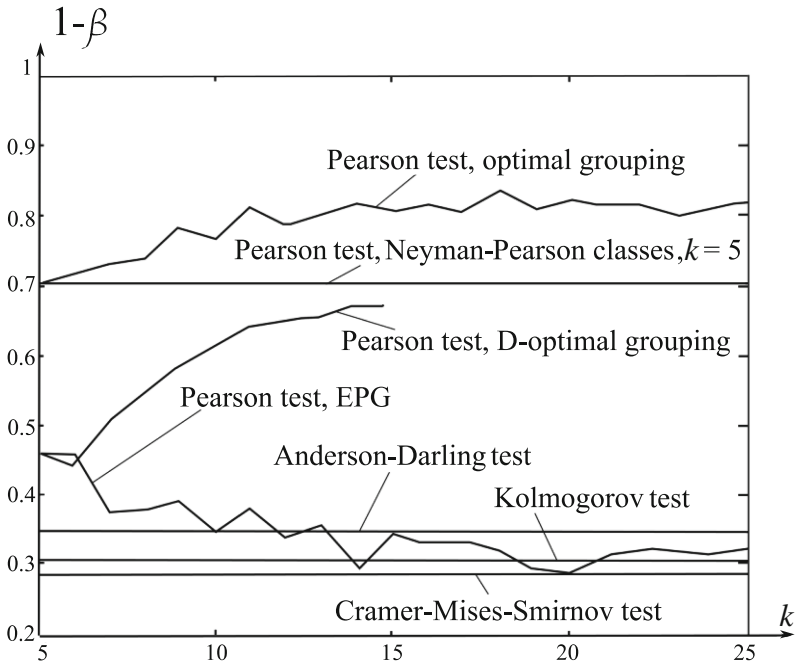


Fig. 1. The power of the χ^2 Pearson test for simple hypothesis

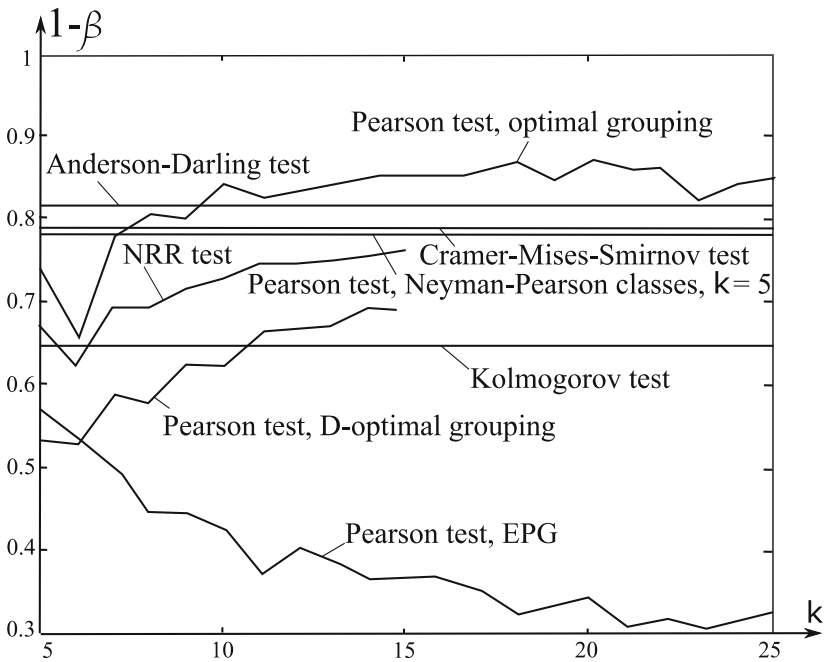


Fig. 2. The power of the χ^2 Pearson test for composite hypothesis

4 The Pearson χ^2 Test When Checking Normality

The asymptotically D-optimal groupings (AOG) given in Tables 2 and 3 can be used for testing normality using MLE estimates of the parameters μ and σ . Here, the losses in the Fisher information associated with grouping are minimized [13] and the Pearson χ^2 test has maximal power relative to the very close competing hypotheses [13].

Table 2. 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	1.11	1.11	–	–	–	–	–	–	–	–	–	–	–	–	0.41
4	1.38	0	1.38	–	–	–	–	–	–	–	–	–	–	–	0.55
5	1.7	0.69	0.69	1.7	–	–	–	–	–	–	–	–	–	–	0.68
6	1.88	1	0	1	1.88	–	–	–	–	–	–	–	–	–	0.76
7	2.06	1.27	0.49	0.49	1.27	2.06	–	–	–	–	–	–	–	–	0.81
8	2.2	1.46	0.79	0	0.79	1.46	2.2	–	–	–	–	–	–	–	0.85
9	2.32	1.62	1.02	0.38	0.38	1.02	1.62	2.32	–	–	–	–	–	–	0.88
10	2.42	1.76	1.21	0.65	0	0.65	1.21	1.76	2.42	–	–	–	–	–	0.90
11	2.52	1.88	1.36	0.86	0.31	0.31	0.86	1.36	1.88	2.52	–	–	–	–	0.91
12	2.6	1.9	1.49	1.03	0.53	0	0.53	1.03	1.49	1.9	2.6	–	–	–	0.93
13	2.68	2.08	1.61	1.18	0.75	0.27	0.27	0.75	1.18	1.61	2.08	2.68	–	–	0.94
14	2.74	2.16	1.71	1.3	0.91	0.48	0	0.48	0.91	1.3	1.71	2.16	2.74	–	0.943
15	2.81	2.24	1.8	1.42	1.04	0.66	0.23	0.23	0.66	1.04	1.42	1.8	2.24	2.81	0.95

Table 3. 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	0.13	0.73	0.13	–	–	–	–	–	–	–	–	–	–	–	–	0.41
4	0.08	0.42	0.42	0.08	–	–	–	–	–	–	–	–	–	–	–	0.55
5	0.05	0.20	0.51	0.20	0.05	–	–	–	–	–	–	–	–	–	–	0.68
6	0.03	0.13	0.34	0.34	0.13	0.03	–	–	–	–	–	–	–	–	–	0.76
7	0.02	0.08	0.21	0.38	0.21	0.08	0.02	–	–	–	–	–	–	–	–	0.81
8	0.01	0.06	0.14	0.28	0.28	0.14	0.06	0.01	–	–	–	–	–	–	–	0.85
9	0.01	0.04	0.10	0.20	0.30	0.20	0.10	0.04	0.01	–	–	–	–	–	–	0.88
10	0.01	0.03	0.08	0.14	0.24	0.24	0.14	0.08	0.03	0.01	–	–	–	–	–	0.90
11	0.01	0.02	0.06	0.11	0.18	0.25	0.18	0.11	0.06	0.02	0.01	–	–	–	–	0.91
12	0.01	0.02	0.04	0.08	0.14	0.21	0.21	0.14	0.08	0.04	0.02	0.01	–	–	–	0.93
13	0.004	0.02	0.04	0.07	0.11	0.17	0.21	0.17	0.11	0.07	0.04	0.02	0.004	–	–	0.94
14	0.003	0.01	0.03	0.05	0.09	0.13	0.19	0.19	0.13	0.09	0.05	0.03	0.01	0.003	–	0.94
15	0.003	0.01	0.02	0.04	0.07	0.11	0.15	0.18	0.15	0.11	0.07	0.04	0.02	0.01	0.003	0.95

In Table 2, the boundary points $t_i, i = 1, \dots, k - 1$ 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 3.

Table 4. Percentage points $\tilde{\chi}_{k,\alpha}^2$ 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	BIII (1:2463; 3:8690; 4:6352; 19:20; 0:005)
5	4.18	5	6.39	7.77	9.59	BIII (1:7377; 3:8338; 5:5721; 26:00; 0:005)
6	5.61	6.54	8.09	9.61	11.62	BIII (2:1007; 4:1518; 4:1369; 26:00; 0:005)
7	6.95	7.98	9.67	11.31	13.43	BIII (2:5019; 4:6186; 3:4966; 28:00; 0:005)
8	8.28	9.4	11.21	12.95	15.22	BIII (2:9487; 5:8348; 3:1706; 34:50; 0:005)
9	9.56	10.76	12.69	14.53	16.87	BIII (3:5145; 6:3582; 3:2450; 39:00; 0:005)
10	10.84	12.11	14.16	16.12	18.58	BIII (3:9756; 6:7972; 3:0692; 41:50; 0:005)
11	12.08	13.42	15.55	17.59	20.19	BIII (4:4971; 6:9597; 3:0145; 43:00; 0:005)
12	13.34	14.74	16.98	19.1	21.77	BIII (5:1055; 7:0049; 3:1130; 45:00; 0:005)
13	14.56	16.01	18.34	20.53	23.3	BIII (5:7809; 7:0217; 3:2658; 47:00; 0:005)
14	15.78	17.29	19.68	21.96	24.81	BIII (6:6673; 6:9116; 3:5932; 49:00; 0:005)
15	16.98	18.54	21.04	23.4	26.37	BIII (7:0919; 7:2961; 3:4314:51:50; 0:005)

When AOG is used in the Pearson χ^2 test, the resulting percentage points $\tilde{\chi}_{k,\alpha}^2$ of the distributions of the statistic (1) and the models of limiting distributions constructed in this paper are shown in Table 4, where $\beta_{\text{III}}(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ is the type III beta distribution with these parameters and the density

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

To make a decision regarding testing the hypothesis H_0 , the value of the statistic X_n^{2*} is compared with the critical value $\tilde{\chi}_{k,\alpha}^2$ from the corresponding row of Table 4, 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 statistic (1) and the corresponding χ_{k-m-1}^2 distributions, when hypothesis H_0 is true, is shown in Fig. 3.

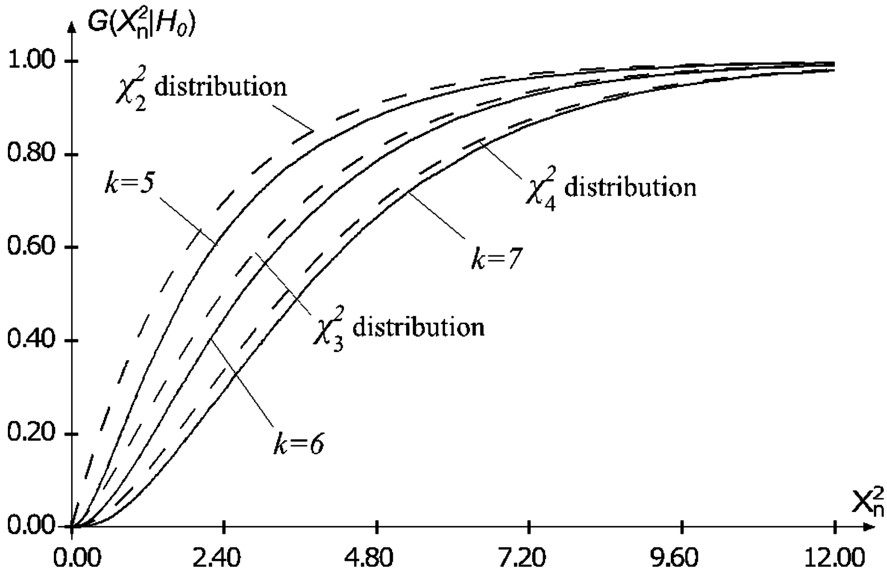


Fig. 3. Distributions of statistic (1) for maximum likelihood estimates of the parameters of a normal distribution based on non-grouped data together with the corresponding χ_{k-m-1}^2 distributions

Tables 2 and 3 give Fisher asymptotic information:

$$A = \det \mathbf{J}_\Gamma / \det \mathbf{J}.$$

For tests of normality with calculations of an MLE based on the non-grouped sample, only the parameters μ or σ , the required AOG tables, percentage points, and the limiting distribution models can be found in [15].

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 $nP_i(\theta) \geq 5 \dots 10$ for any interval. At least, in choosing k the recommendation

$$\min_i \{nP_i(\theta) | i = \overline{1, k}\} > 1$$

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, $n = 10\text{--}20$, 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 ($k \leq k_{\max}$). The number of grouping intervals affects the power of the Pearson χ^2 test [6]. 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 checking normality with the power of special normality tests and nonparametric goodness-of-fit tests, the power has been estimated relative to the same competing distributions (hypotheses) as in [15].

The test hypothesis H_0 is taken to be that the observed sample obeys the normal distribution (7).

As competing hypotheses for studying the power of the χ^2 test, 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 (8), which is very close to a normal distribution. Figure 4 shows the densities of the distributions corresponding to hypotheses H_1 , H_2 and H_3 with scale parameters such that they are the closest to the

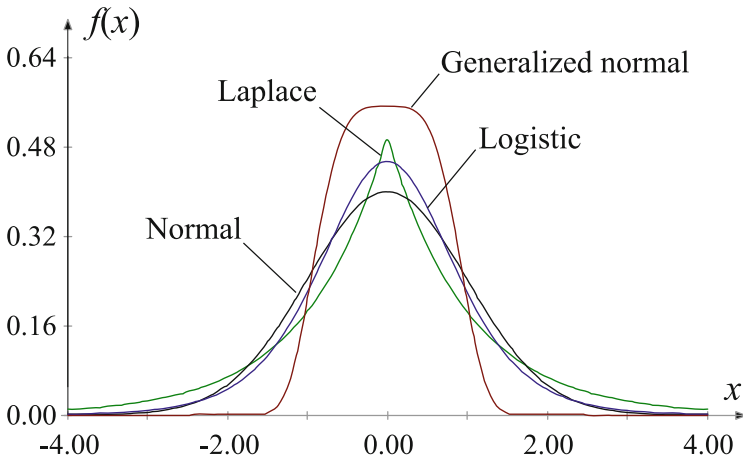


Fig. 4. Probability density functions corresponding to the considered hypotheses H_i

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 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 [15–17]. It turned out that for small sample sizes n and small specified

Table 5. Power of the Pearson χ^2 test 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	0.235	0.146	0.043	0.032	0.002
20	4	5	0.262	0.177	0.1	0.058	0.021
30	5	5	0.312	0.216	0.136	0.079	0.043
40	6	5	0.336	0.267	0.168	0.111	0.061
50	6	5	0.401	0.311	0.204	0.129	0.068
100	9	5	0.558	0.479	0.352	0.254	0.158
150	10	7	0.722	0.634	0.486	0.353	0.217
200	11	9	0.783	0.695	0.548	0.417	0.279
300	13	11	0.907	0.858	0.756	0.646	0.492
H_2							
10	4	4	0,267	0,206	0,074	0,058	0,01
20	4	4	0,264	0,177	0,104	0,067	0,037
		5	0,247	0,189	0,116	0,061	0,024
30	5	5	0,312	0,261	0,153	0,103	0,044
40	6	7	0,443	0,358	0,25	0,167	0,101
50	6	7	0,5	0,423	0,312	0,225	0,138
100	9	9	0,77	0,708	0,596	0,494	0,379
150	10	9	0,899	0,86	0,785	0,705	0,596
200	11	11	0,964	0,946	0,908	0,88	0,786
300	13	13	0,996	0,993	0,985	0,974	0,95
H_3							
10	4	4	0.221	0.15	0.046	0.034	0.003
20	4	5	0.194	0.125	0.059	0.038	0.016
30	5	6	0.169	0.125	0.062	0.034	0.012
40	6	7	0.204	0.143	0.082	0.045	0.02
50	6	7	0.214	0.155	0.088	0.05	0.023
100	9	10	0.303	0.231	0.146	0.09	0.047
150	10	10	0.359	0.284	0.191	0.124	0.072
200	11	11	0.432	0.355	0.25	0.175	0.105
300	13	13	0.566	0.486	0.373	0.28	0.19

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 the normal law and indicates that the tests are biased.

The power of the Pearson χ^2 test was studied with different number of intervals $k \leq k_{\max}$ for specified sample sizes n . Table 5 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 5.

5 The Nikulin-Rao-Robson Goodness-of-Fit Test

A modification of the standard statistic X_n^2 was proposed [3–5] in which the limiting distribution of the modified statistic is a χ_{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, must be estimated on the basis of the non-grouped data by a maximum likelihood method. Here the vector $\mathbf{P} = (P_1, \dots, P_k)^T$ is assumed to be specified, while the boundary points of the intervals are defined using the relations $x_i(\theta) = F^{-1}(P_1 + \dots + P_i), i = 1, \dots, k - 1$. The proposed statistic has the form [4]:

$$Y_n^2(\theta) = X_n^2 + n^{-1} a^T(\theta) \lambda(\theta) a(\theta), \tag{9}$$

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

$$\lambda(\theta) = [\mathbf{J}(\theta) - \mathbf{J}_g(\theta)]^{-1}.$$

In the case of the normal distribution with parameter vector $\theta^T = (\mu, \sigma)$, 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 $\mathbf{J}_g(\theta)$ given by

$$\begin{aligned} J_g(\mu, \mu) &= \sum_{i=1}^k \frac{1}{\sigma^2 P_i(\theta)} (f(t_{i-1}) - f(t_i))^2, \\ J_g(\sigma, \sigma) &= \sum_{i=1}^k \frac{1}{\sigma^2 P_i(\theta)} (t_{i-1} f(t_{i-1}) - t_i f(t_i))^2, \\ J_g(\mu, \sigma) = J_g(\sigma, \mu) &= \sum_{i=1}^k \frac{1}{\sigma^2 P_i(\theta)} (f(t_{i-1}) - f(t_i))(t_{i-1} f(t_{i-1}) - t_i f(t_i)), \end{aligned}$$

where

$$t_i = (x_i - \mu)/\sigma, t_0 = -\infty, t_k = \infty, f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

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

Table 6. 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	5	0.349	0.265	0.185	0.127	0.078
40	6	5	0.474	0.403	0.297	0.218	0.149
50	6	5	0.548	0.473	0.365	0.281	0.190
100	9	5	0.807	0.755	0.667	0.583	0.482
150	10	7	0.919	0.889	0.834	0.774	0.691
200	11	9	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

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

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

As in the case of the Pearson test, when testing for normality with MLE estimation of the parameters μ and σ based on the non-grouped data, Tables 2 and 3 can be used.

For calculating statistic (9), the boundaries separating the intervals for given k are found from the values of t_i in the corresponding row of Table 2 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 (9) are taken from the corresponding line of Table 3. The elements of the vector $\alpha(\theta)$ 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 $\chi_{k-1,\alpha}^2$ or the achieved level of significance (p -value) $P\{Y_n^2 > Y_n^{2*}\}$ is found from the corresponding χ_{k-1}^2 distribution.

To test for normality with MLE calculation of the parameters μ or σ separately on the basis of non-grouped samples, the required tables of AOG can be found in [15].

Estimates of the power of the Nikulin-Rao-Robson test for the competing hypotheses H_1 , H_2 and H_3 for k_{opt} are given in Table 6. 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 $k_{opt} = k_{max}$ 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.

6 Conclusion

The power of the Pearson χ^2 test and Nikulin-Rao-Robson test can be maximized by the optimal selection of the number of intervals and interval boundary points.

Combining the obtained results of the power analysis for the Pearson χ^2 test and Nikulin-Rao-Robson test with the results presented in [15–17], we can see that in regard to the competing hypothesis H_1 , the Pearson χ^2 test shows very good results, yielding in power only to some special normality tests.

At the same time, in regard to competing hypotheses H_2 and H_3 , the Pearson χ^2 test and Nikulin-Rao-Robson test inferior in power to most special normality tests and to nonparametric goodness-of-fit tests (Anderson-Darling, Cramer-Mises-Smirnov, Watson, Kuiper, Zhang, Kolmogorov tests).

This work is supported by the Russian Ministry of Education and Science (project 2.541.2014 K).

References

1. Chernoff, H., Lehmann, E.L.: The use of maximum likelihood estimates in χ^2 test for goodness of fit. *Ann. Math. Stat.* **25**(3), 579–586 (1954)
2. Lemeshko, B.Y., Postovalov, S.N.: Limit distributions of the Pearson chi 2 and likelihood ratio statistics and their dependence on the mode of data grouping. *Ind. Lab.* **64**(5), 344–351 (1998)
3. Nikulin, M.S.: χ^2 tests for continuous distributions with shift and scale parameters. *Teor. Veroyatn. Primen.* **XVIII**(3), 583–591 (1973)
4. Nikulin, M.S.: On χ^2 tests for continuous distributions. *Teor. Veroyatn. Primen.* **XVIII**(3), 675–676 (1973)
5. Rao, K.C., Robson, D.S.: A chi-squared statistic for goodness-of-fit tests within the exponential family. *Commun. Stat.* **3**(1), 1139–1153 (1974)
6. Lemeshko, B.Y., Chimitova, E.V.: On the choice of the number of intervals in χ^2 -type goodness-of-fit tests. *Zavod. Lab. Diagn. Mater.* **69**(1), 61–67 (2003)
7. Lemeshko, B.Y.: Asymptotically optimum grouping of observations in goodness-of-fit tests. *Ind. Lab.* **64**(1), 59–67 (1998)
8. Voinov, V., Pya, N., Alloyarova, R.: A comparative study of some modified chi-squared tests. *Commun. Stat. Simul. Comput.* **38**(2), 355–367 (2009)
9. Denisov, V., Lemeshko, B.: Optimal grouping in estimation and tests of goodness-of-fit hypotheses. *Wissenschaftliche Schriftenreihe der Technischen universitat Karl-Marx-Stadt.* **10**, 63–81 (1989)
10. Lemeshko, B., Postovalov, S., Chimitova, E.: On statistic distributions and the power of the Nikulin χ^2 test. *Ind. Lab.* **67**(3), 52–58 (2001)
11. Lemeshko, B., Chimitova, E.: Maximization of the power of χ^2 tests. *Papers of Siberian Branch of Academy of Science of Higher Education*, vol. 2, pp. 53–61 (2000)
12. Denisov, V., Lemeshko, B.: Optimal grouping in the analysis of experimental data. In: *Measuring Information Systems*, Novosibirsk, pp. 5–14 (1979). [In Russian]
13. Lemeshko, B.Y., Lemeshko, S.B., Postovalov, S.N., Chimitova, E.V.: *Statistical Data Analysis, Simulation and Study of Probability Regularities. Computer Approach*, 888 pp. NSTU Publisher, Novosibirsk (2011). [In Russian]
14. Greenwood, P.E., Nikulin, M.S.: *A Guide to Chi-Squared Testing*. John Wiley & Sons Inc., New York (1996)
15. Lemeshko, B.Y.: Tests for checking the deviation from normal distribution law. In: *Guide on the Application*. INFRA-M, Moskow (2015). doi:10.12737/6086
16. Lemeshko, B.Y., Lemeshko, S.B.: Comparative analysis of tests for verifying deviation of a distribution from normal. *Metrologiya* **2**, 3–24 (2005). [In Russian]
17. Lemeshko, B.Y., Rogozhnikov, A.P.: Features and power of some tests of normality. *Metrologiya* **4**, 3–24 (2009). [In Russian]