Optoelectronics, Instrumentation and Data Processing
Volume 40, no. 3, pp. 3-18, 2004

Avtometriya

UDC 519.2

ANALYSIS AND SYNTHESIS OF SIGNALS AND IMAGES

NONPARAMETRIC TESTS IN TESTING COMPOSITE HYPOTHESES ON GOODNESS OF FIT TO EXPONENTIAL FAMILY DISTRIBUTIONS*

B.Yu.Lemeshko and A.A.Maklakov

Novosibirsk

Statistical modeling methods are used to investigate statistic distributions of nonparametric Kolmogorov, and also ω^2 and Ω^2 Mises tests for composite hypotheses on goodness of fit of empirical data to the exponential family distributions. Models of limiting distributions of statistics of the considered goodness of fit tests are constructed for different values of the parameter of form of exponential family and are recommended for application to problems of statistical analysis.

INTRODUCTION

It is frequently impossible to describe errors of measurement devices and systems based on different physical principles by the normal distribution law [1]. In such situations, in the case of symmetry of laws of observed random variables, a sufficiently good model is an exponential family of distributions with the density

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

Special cases of this law for values 2 and 1 of the form parameter θ_2 are normal and Laplace distributions, respectively. The distribution densities for different values of the form parameter θ_2 are represented in Fig. 1.

The family (1) is recently frequently used as probabilistic models of observation errors in problems of regression and dispersion analysis in the case of violation of classical assumptions when the error distribution law differs substantially from the normal one.

When researchers come up against the necessity of determining the law of distribution of measurement errors of a device or a measurement system as well as observation errors arising in experiments, they should use results of observations and thereby choose a model which is the most similar one to the real law, i.e., identify the error distribution law.

* This research was supported by the Ministry of Education of the Russian Federation, project no. TO2-3.3-3356.

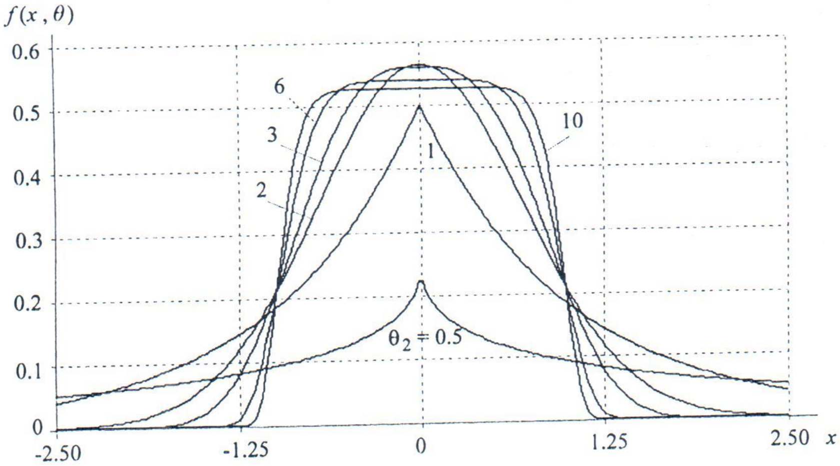

Fig. 1.

The process of identifying the distribution law by experimental observations of measurement errors (or some observed random variable) consists in fact in solving a sequence of problems concerned with estimation of the probabilistic model parameters, testing adequacy of the constructed models by means of goodness of fit tests, and subsequent choice (on the basis of these tests) of the most suitable theoretical law from multiple laws under consideration.

The test for goodness of fit for the obtained experimental distribution and the theoretical one is one of the most common problems of statistical analysis in measurement information processing. We should underline that until now, the actual practice of applying the nonparametric tests for goodness of fit as well as the χ^2-type tests is plagued with examples of their incorrect usage. In the case of using the nonparametric tests for goodness of fit, errors are usually a result of disregarding the factor of complexity of a tested hypothesis.

Applying the tests for goodness of fit, one should distinguish tests of simple and composite hypotheses. A simple tested hypothesis has the form

$$H_0: F(x) = F(x, \theta),$$

where $F(x, \theta)$ is the probability distribution function with which the observed sample is tested for goodness of fit, and θ is the known value of the parameter (vector or scalar parameter). A composite tested hypothesis has the form

$$H_0: F(x) \in \{F(x, \theta), \ \theta \in \Theta\}.$$

In this case, the estimate of the distribution parameter $\hat{\theta}$ is calculated by the sample used to test for goodness of fit.

The nonparametric Kolmogorov, ω^2 Mises (Cramer–Mises–Smirnov), and Ω^2 Anderson–Darling tests [2] refer to the most widely used tests for goodness of fit. In testing simple hypotheses they are "distribution free" tests: the conditional distributions $G(S \mid H_0)$ of statistics S of these tests do not depend on the form of the tested hypothesis H_0 (on the theoretical law tested for goodness of fit).

The Kolmogorov test uses statistic of the form

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{2}$$

where

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

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

Here n is the sample size; $x_1, x_2, \ldots, x_n$ are sampling values in increasing order; $F(x, \theta)$is the function of the distribution tested for goodness of fit. The distribution of S_K under a simple hypothesis in the limit obeys the Kolmogorov law $K(S)$ [2].

In the ω^2 Mises test, one uses a statistic of the form

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \tag{3}$$

which in the case of a simple hypothesis obeys the distribution $a1(S)$ [2].

In the Ω^2 Mises (Anderson–Darling) test, the employed statistic has the form

$$S_\Omega = n\Omega_n^2 = -n - 2\sum_{i=1}^{n} \left\{ \frac{2i-1}{2n} \ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n}\right) \ln(1 - F(x_i, \theta)) \right\}. \tag{4}$$

In the case of a simple hypothesis, in the limit this statistic obeys the distribution $a2(S)$ [2].

In testing composite hypotheses, when parameters of the observed law $F(x, \theta)$ are estimated by the same sample, the nonparametric tests for goodness of fit loose the "distribution-free" property. The conditional distribution law of statistic $G(S \mid H_0)$, while testing composite hypotheses, is affected by a number of factors determining "complexity" of the tested hypothesis H_0: the form of the observed law $F(x, \theta)$corresponding to the true hypothesis H_0; the type of estimated parameter; the number of estimated parameters; sometimes, a concrete value of the parameter (e.g., in the case of gamma and beta distributions); the method used for parameter estimation.

The difference in the limiting distributions of the same statistics while testing simple and composite hypotheses is great, hence, it is absolutely inadmissible to neglect the listed factors when using the nonparametric tests for goodness of fit.

After publication of paper [3] defining the problem of using the nonparametric tests for goodness of fit for composite hypotheses, various approaches were applied to investigation of the limiting distributions of statistics of these tests, such as: analytical [4]; percentage points of distributions were constructed by statistical modeling [5–8]; for approximate calculation of the probabilities of "goodness of fit" of the form of $P\{S > S^*\}$ (an attainable level of significance), where S^* is the statistic value calculated from a sample, authors of [9–13] constructed formulas yielding sufficiently good approximations for small values of the corresponding probabilities. The complexity and labor intensity of obtaining analytical solutions have predetermined the limited number of concrete results.

Authors of papers [14–20] applied computer methods of modeling and investigation of statistical regularities [21] to the problem of investigating the distributions of statistics of the nonparametric tests for goodness of fit for composite hypotheses and to construction of models of these distributions. The obtained results have made it possible to formulate recommendations on standardization [22], which substantially extended the field of correct application of the nonparametric tests for goodness of fit for various composite hypotheses. The aim of this paper is to investigate and construct models of limiting distributions of statistics of the considered tests for composite hypotheses on fitness of empirical distributions to theoretical laws of family (1) for different values of the form parameter θ_2.

Distributions of statistics in tests for goodness of fit to the exponential family. The distributions of statistics $G(S \mid H_0)$ of the considered nonparametric tests for goodness of fit to the exponential family depend on all of the listed factors determining the "complexity" of the tested hypothesis, including the concrete value of the form parameter θ_2. In this case, the uneasy problem of constructing the models of distributions $G(S \mid H_0)$ of statistics of tests is aggravated by the number of parameters determining the law of the form of (1). This means that for a certain method of parameter estimation and a concrete value of the form parameter θ_2, one should construct models of statistic distributions for different combinations of the evaluated parameters of bias θ_0, scale θ_1, and form θ_2 (seven different kinds of composite hypothesis).

In this paper, we consider only one method of parameter estimation, namely, the maximum likelihood method, which allows us to obtain estimates with the best properties. In so doing, the maximum likelihood estimate (MLE) of the parameter θ_0 is determined as a solution of the likelihood equation

$$\frac{\theta_2}{\theta_1}\sum_{j=1}^{n}\frac{|t_j|^{\theta_2}}{t_j}=0, \tag{5}$$

where $t_j = \frac{x_j - \theta_0}{\theta_1}$. MLE of the parameter θ_1 is found as a solution of a likelihood equation of the form

$$\frac{\theta_2}{\theta_1}\sum_{j=1}^{n}\left(|t_j|^{\theta_2}-\frac{1}{\theta_2}\right)=0, \tag{6}$$

and MLE of the parameter θ_2 is found as a solution of the likelihood equation

$$\frac{n}{\theta_2^2}\psi\left(\frac{1}{\theta_2}\right)+\sum_{j=1}^{n}\left(\frac{1}{\theta_2}-|t_j|^{\theta_2}\ln|t_j|\right)=0, \tag{7}$$

where $\psi(\theta) = \Gamma'(\theta)/\Gamma(\theta)$ is a logarithmic derivative of a gamma function (a digamma function).

By simultaneous estimation of several parameters, one maximizes the logarithm of the likelihood function

$$\ln L = \sum_{j=1}^{n}\ln f(x_j, q) = n[\ln\theta_2 - \ln(2\theta_1\Gamma(1/\theta_2))] - \sum_{j=1}^{n}|t_j|^{\theta_2},$$

and the vector of the estimates is the solution of the corresponding system (subsystem) of likelihood equations (5)–(7).

We will emphasize that the estimation method should necessarily be taken into account while testing composite hypotheses [18]. Figure 2 shows dependence of the statistic distributions on the estimation method for this case. It represents distributions of the Kolmogorov statistic while estimating all parameters of distribution (1) for the value of the form parameter $\theta_2 = 2$ and using two estimation methods: a maximum likelihood estimation and an *MD* estimation, for which the estimate of the parameter vector is obtained by minimizing the Kolmogorov statistic (2). To compare, Fig. 2 shows the Kolmogorov distribution which the statistic (2) obeys while testing simple hypotheses.

Statistical modeling and investigation of the obtained empirical distributions of statistics (2)–(4) under validity of hypothesis H_0 corresponding to the law (1) have shown a substantial and unusual dependence of the statistic distributions $G(S \mid H_0)$ on the form parameter θ_2. As a rule, when θ_2 grows from 0 to ≈1.6, the scale parameter of the statistic distribution $G(S \mid H_0)$ decreases, and when θ_2 further grows, the scale parameter increases. For $\theta_2 > 7$, the statistic distributions under the corresponding composite hypotheses do not practically change.

Figure 3 illustrates, in particular, distributions of statistic of the Kolmogorov type versus form parameter θ_2 for the case when all the three parameters of distribution (1) are estimated by the maximum likelihood method. Figure 4 shows a similar situation corresponding to estimation of only two parameters, namely, the parameters of bias θ_0 and scale θ_1, with the known form parameter θ_2.

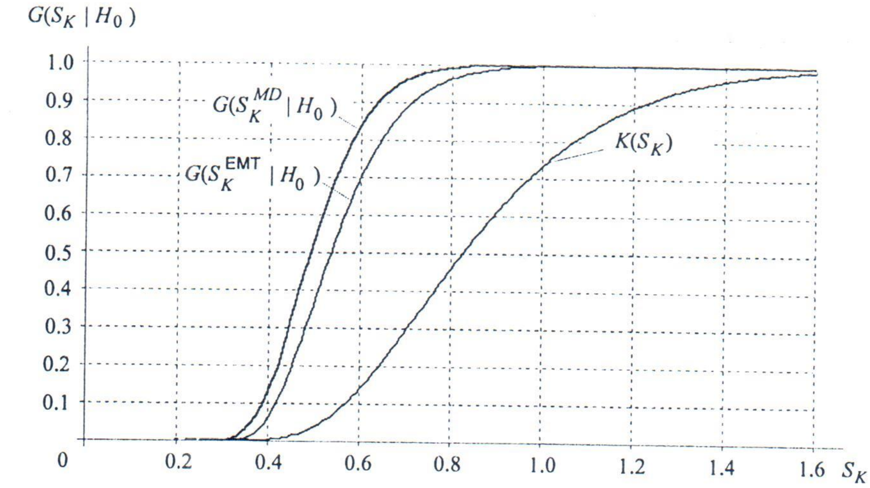

Fig. 2.

Figure 5 represents the character of the dependence of the statistic distribution of the Kolmogorov test on the number and type of parameters estimated by the maximum likelihood method for the form parameter $\theta_2 = 1.6$. This value of the form parameter corresponds to the extreme left statistic distributions of the nonparametric tests for goodness of fit while testing the hypotheses for law (1). (In Figs. 5–8, the corresponding distributions $G(S \mid H_0)$ are denoted by a list of estimated parameters.)

A similar situation for distributions of statistic of the ω^2 Cramer–Mises–Smirnov type is shown in Fig. 6. To compare, we gave there the distribution function $a1(S)$ which the statistic obeys while testing a simple hypothesis.

For an identical situation, Fig. 7 presents distributions of statistic of the Ω^2 Anderson–Darling type while testing composite hypotheses for law (1) with the form parameter $\theta_2 = 1.6$, and also the distribution $a2(S)$ which in the limit obeys this statistic while testing simple hypotheses.

To compare, Fig. 8 shows a situation similar to that in Fig. 5, which represents distribution of statistic of the Kolmogorov type versus the number and type of parameters for the form parameter $\theta_2 = 2$. In this case, the density (1) corresponds to the normal law.

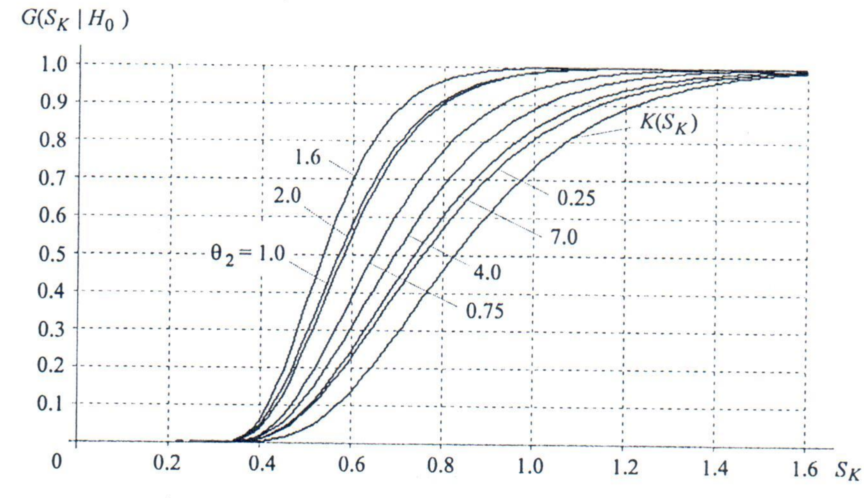

Fig. 3.

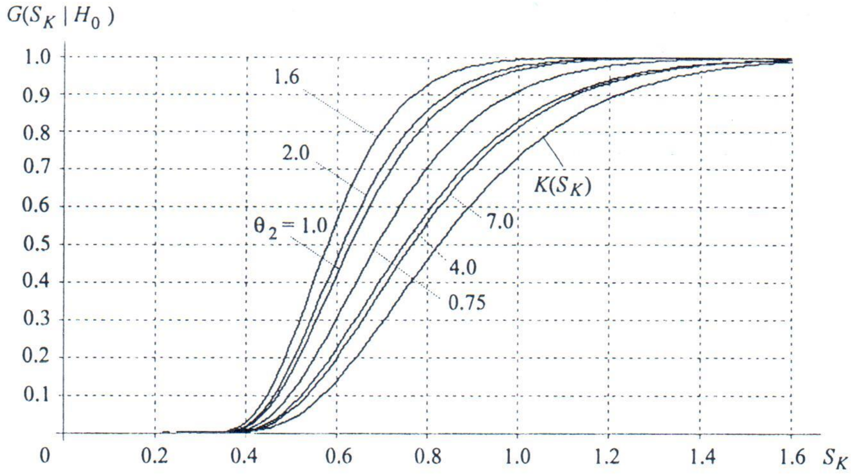

Fig. 4.

The empirical statistic distributions of the tests for goodness of fit, which have been obtained by statistical modeling, were smoothed by different theoretical models of laws included in the system [21]. As a result, a theoretical distribution giving the best description of the empirical one was obtained. As in [14–19], the best analytical models for the distributions $G(S \mid H_0)$ of these statistics most frequently were models corresponding to one of the following three laws: gamma distribution, *Su*-Johnson distribution, or *Sl*-Johnson distribution.

The constructed models for the statistic distributions $G(S \mid H_0)$ of the Kolmogorov, Cramer–Mises–Smirnov, and Anderson–Darling tests for testing composite hypotheses for goodness of fit to exponential family for different values of the form parameter θ_2 are represented in Tables 1–3, respectively. In these Tables containing the distribution $G(S \mid H_0)$ obtained and recommended for usage while testing composite hypotheses with the considered tests, $\gamma(\theta_0, \theta_1, \theta_2)$ denotes the gamma distribution with the density function

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

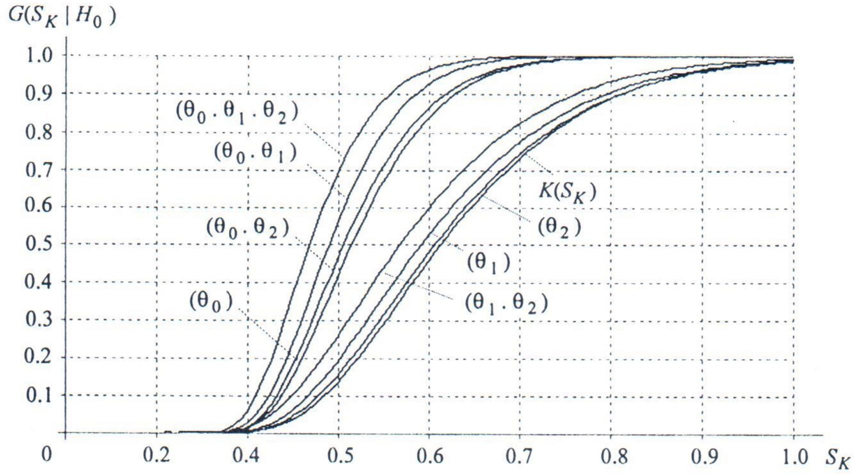

Fig. 5.

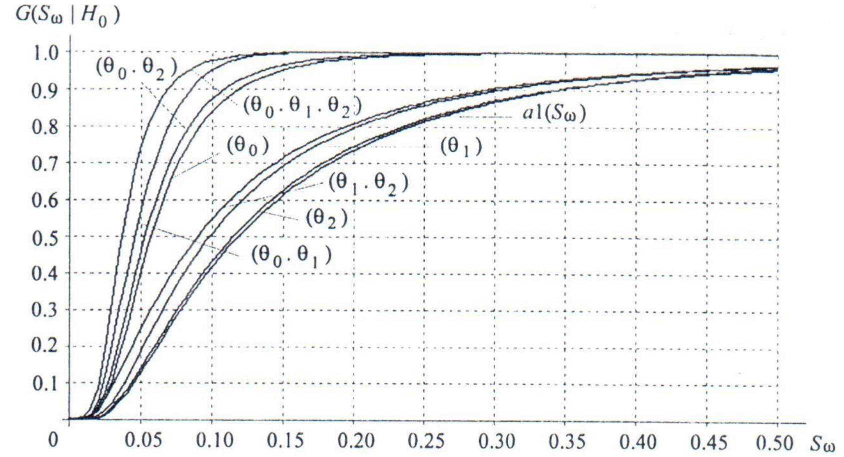

Fig. 6.

$Sl(\theta_0, \theta_1, \theta_2, \theta_3)$ denotes the Sl-Johnson distribution with the density

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}(x-\theta_3)} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln \frac{x-\theta_3}{\theta_2}\right]^2\right\},$$

and $Su(\theta_0, \theta_1, \theta_2, \theta_3)$ denotes the Su-Johnson distribution with the density

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2 + \theta_2^2}} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\left\{\frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2 + 1}\right\}\right]^2\right\}.$$

Application of results. In the course of testing the hypothesis on goodness of fit of the empirical distribution to the theoretical distribution, one calculates from a sample the S^* value of the statistics of the test used.

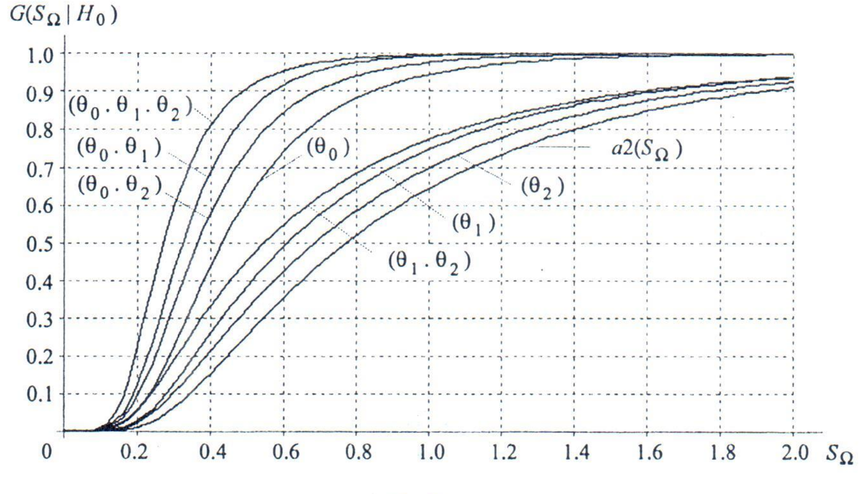

Fig. 7.

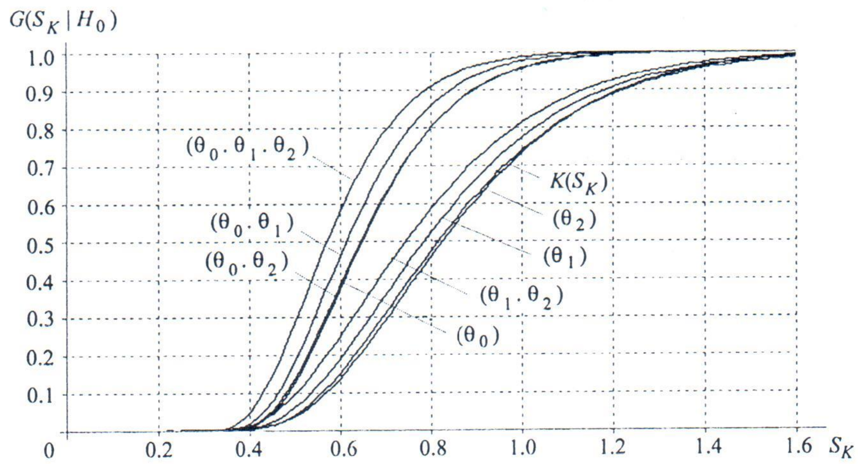

Fig. 8.

Then, in order to conclude whether to accept or reject hypothesis H_0, it is necessary, knowing the conditional distribution $G(S \mid H_0)$ of statistic S for true hypothesis H_0, to calculate the probability

$$P\{S > S^*\} = \int_{S^*}^{+\infty} g(s \mid H_0)\, ds,$$

where $g(s \mid H_0)$ is the conditional density. If the probability is sufficiently large, at least $P\{S > S^*\} > \alpha$, where α is the prescribed level of significance (the probability of an error of the first kind – to reject the true hypothesis H_0), it is generally agreed that there are no grounds for rejecting the hypothesis H_0.

We will exemplify the usage of the obtained results in testing composite hypotheses.

Example. Let we test a hypothesis on membership of a sample of 100 observations:

-1.17	-1.09	-0.91	-0.90	-0.89	-0.88	-0.86	-0.85	-0.85	-0.75
-0.73	-0.66	-0.64	-0.58	-0.58	-0.55	-0.54	-0.53	-0.52	-0.50
-0.50	-0.50	-0.47	-0.45	-0.44	-0.42	-0.41	-0.41	-0.40	-0.40
-0.39	-0.38	-0.37	-0.30	-0.30	-0.29	-0.28	-0.24	-0.23	-0.21
-0.20	-0.18	-0.17	-0.16	-0.15	-0.15	-0.13	-0.09	-0.05	-0.05
-0.03	-0.02	0.00	0.02	0.05	0.08	0.10	0.11	0.11	0.14
0.15	0.16	0.18	0.25	0.25	0.26	0.27	0.28	0.30	0.33
0.33	0.36	0.40	0.41	0.42	0.43	0.44	0.48	0.51	0.52
0.53	0.54	0.55	0.58	0.61	0.63	0.68	0.69	0.71	0.72
0.74	0.76	0.79	0.80	0.88	0.96	0.96	0.97	1.12	1.39

in exponential family with the known form parameter $\theta_2 = 4$. MLEs of the parameters of bias and scale, calculated from this sample, are $\theta_0 = 0.0117$ and $\theta_1 = 0.9625$, respectively. Figure 9 represents the empirical distribution function constructed by the sample (curve 1) and the theoretical distribution function (1) with the obtained parameter vector (curve 2).

We will test the hypothesis on goodness of fit for all tests considered in the paper. In this case, we consider a composite hypothesis with maximum likelihood estimation of the bias and scale parameters. Values cal-

Table 1

Approximation of Limiting Kolmogorov Statistic Distributions while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of only form parameter	Estimation of only scale parameter	Estimation of only bias parameter	Estimation of three parameters (of form, scale, and bias)
1	0.25	(3.3957; 0.1419; 0.3342)	(3.2891; 0.1454; 0.3361)	(4.1678; 0.1293; 0.3334)	(3.5003; 0.1370; 0.3229)
2	0.50	(3.7204; 0.1365; 0.3266)	(3.5436; 0.1387; 0.3232)	(3.7577; 0.1336; 0.3448)	(3.6501; 0.1237; 0.3140)
3	0.75	(3.7924; 0.1367; 0.3213)	(3.4892; 0.1396; 0.3319)	(4.1868; 0.1068; 0.3339)	(3.9498; 0.0936; 0.3106)
4	1..00	(3.4680; 0.1467; 0.3428)	(3.7599; 0.1342; 0.3209)	(4.1175; 0.0935; 0.3408)	(4.5150; 0.0689; 0.2945)
5	1.50	(3.5197; 0.1434; 0.3563)	(3.9228; 0.1296; 0.3155)	(4.5449; 0.0741; 0.3173)	(4.9547; 0.052; 0.2929)
6	1.60	(3.8275; 0.1375; 0.3336)	(3.4907; 0.1424; 0.334)	(5.2114; 0.0676; 0.3011)	(4.7791; 0.0545; 0.2953)
7	2.00	(3.5579; 0.1441; 0.3523)	(3.4864; 0.1424; 0.3371)	(4.4963; 0.078; 0.3250)	(4.1586; 0.0714; 0.2989)
8	3.00	(4.1405; 0.1312; 0.3256)	(3.5201; 0.1401; 0.3411)	(4.2098; 0.0971; 0.3311)	(4.3976; 0.0902; 0.2809)
9	4.00	(3.9274; 0.1356; 0.3394)	(3.6132; 0.1393; 0.3432)	(3.9069; 0.1122; 0.3326)	(3.9126; 0.1095; 0.3033)
10	5.00	(3.7127; 0.1407; 0.3518)	(3.4220; 0.1469; 0.3513)	(3.9228; 0.1183; 0.3308)	(3.7199; 0.1180; 0.3113)
11	6.00	(4.14; 0.1295; 0.3313)	(3.4004; 0.1478; 0.3507)	(4.1123; 0.117; 0.328)	(3.4870; 0.1282; 0.3222)
12	7.00	(3.9512; 0.1351; 0.3385)	(3.8910; 0.1351; 0.3254)	(3.6487; 0.1289; 0.3475)	(3.6734; 0.1262; 0.317)

(continued)

Approximation of Limiting Kolmogorov Statistic Distributions while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of two parameters (of form and scale)	Estimation of two parameters (of form and bias)	Estimation of two parameters (of scale and bias)
1	0.25	(3.2071; 0.1416; 0.348)	(3.6141; 0.1345; 0.329)	(3.2012; 0.1467; 0.3411)
2	0.50	(3.4833; 0.1330; 0.3117)	(3.448; 0.135; 0.3433)	(3.2933; 0.1354; 0.3433)
3	0.75	(3.741; 0.1310; 0.2925)	(4.2278; 0.1025; 0.317)	(4.2809; 0.0956; 0.3132)
4	1.00	(3.2291; 0.1447; 0.3186)	(4.6295; 0.0809; 0.3120)	(4.5961; 0.0760; 0.3084)
5	1.50	(3.3351; 0.1425; 0.3150)	(5.3387; 0.0644; 0.2937)	(4.9614; 0.0588; 0.3066)
6	1.60	(3.5941; 0.1340; 0.3005)	(4.4359; 0.0716; 0.3191)	(6.0113; 0.0529; 0.2807)
7	2.00	(3.3243; 0.1431; 0.3175)	(4.4110; 0.0802; 0.3195)	(4.5532; 0.0722; 0.3106)
8	3.00	(3.2322; 0.1458; 0.3305)	(4.2836; 0.0955; 0.3258)	(3.957; 0.0992; 0.3263)
9	4.00	(3.5176; 0.1428; 0.3195)	(3.9688; 0.1097; 0.3368)	(3.8001; 0.1221; 0.3272)
10	5.00	(3.8539; 0.1361; 0.3016)	(3.4492; 0.1275; 0.3547)	(3.6304; 0.11198; 0.3401)
11	6.00	(3.2838; 0.1493; 0.3337)	(4.0420; 0.1189; 0.3308)	(3.5027; 0.1261; 0.3478)
12	7.00	(3.3896; 0.1464; 0,3339)	(3.7526; 0.1276; 0.3419)	(3.9828; 0.1206; 0.3276)

Table 2

Approximation of Limiting Distributions of ω^2 Mises Statistic while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of only form parameter	Estimation of only scale parameter	Estimation of only bias parameter	Estimation of three parameters (of form, scale, and bias)
1	0.25	*Sl*(0.9970; 1.0278; 0.2233; 0.0157)	*Sl*(1.3009; 1.0402; 0.3113; 0.0104)	*Sl*(0.8362; 0.1012; 0.2234; 0.0125)	*Sl*(1.5828; 1.0581; 0.3775; 0.0097)
2	0.50	*Sl*(1.2174; 1.0729; 0.2863; 0.0103)	*Sl*(0.0880; 1.0126; 0.0926; 0.0121)	*Sl*(1.0095; 1.1215; 0.2377; 0.0117)	*Su*(-2.5153; 0.9719; 0.0105; 0.0157)
3	0.75	*Sl*(1.1095; 1.0685; 0.2635; 0.0114)	*Sl*(1.1818; 1.0525; 0.2757; 0.0111)	*Sl*(0.7165; 1.2655; 0.1404; 0.0091)	*Su*(-2.2482; 1.0158; 0.0097; 0.0169)
4	1.00	*Sl*(0.1324; 1.0347; 0.1083; 0.0130)	*Sl*(0.1218; 1.0473; 0.0990; 0.0122)	*Su*(-2.3848; 1.3757; 0.0188; 0.0173)	*Su*(-2.1532; 1.2270; 0.0104; 0.0154)
5	1.50	*Sl*(1.1294; 1.1011; 0.2796; 0.0117)	*Sl*(0.9558; 1.0559; 0.2173; 0.0125)	*Su*(-2.2563; 1.3959; 0.0159; 0.0170)	*Su*(-2.0280; 1.4821; 0.0113; 0.0152)
6	1.60	*Sl*(1.0768; 1.1009; 0.2770; 0.0114)	*Su*(-2.4611; 1.0097; 0.0141; 0.0187)	*Su*(-2.2391; 1.4169; 0.0171; 0.0176)	*Su*(-1.9635; 1.5327; 0.0132; 0.0151)
7	2.00	*Sl*(0.1533; 1.1027; 0.1170; 0.0121)	*Sl*(0.0713; 1.0480; 0.0970; 0.0118)	*Su*(-2.2789; 1.4329; 0.0190; 0.0164)	*Su*(-2.1446; 1.3616; 0.0110; 0.0152)
8	3.00	*Sl*(1.1683; 1.12; 0.3001; 0.0104)	*Sl*(0.0522; 1.0483; 0.0980; 0.0113)	*Su*(-2.3109; 1.3327; 0.0212; 0.0184)	*Sl*(1.1877; 1.1921; 0.1163; 0.0103)
9	4.00	*Sl*(0.1567; 1.1166; 0.1236; 0.0117)	*Sl*(0.0646; 1.0404; 0.1021; 0.0128)	*Sl*(0.8090; 1.2596; 0.1385; 0.0107)	*Su*(-2.4293; 1.0725; 0.0101; 0.0160)
10	5.00	*Sl*(0.1610; 1.1007; 0.12; 0.0137)	*Sl*(0.1120; 1.0821; 0.1085; 0.0118)	*Su*(-2.3334; 1.2109; 0.0211; 0.0189)	*Sl*(0.9678; 1.1536; 0.1459; 0.0102)

(continued)

Approximation of Limiting Distributions of ω^2 Mises Statistic while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of two parameters (of form and scale)	Estimation of two parameters (of form and bias)	Estimation of two parameters (of scale and bias)
1	0.25	*Sl*(1.4204; 1.0381; 0.3432; 0.0113)	*Sl*(0.9357; 1.0704; 0.2133; 0.0104)	*Sl*(1.2041; 1.0005; 0.2672; 0.0139)
2	0.50	*Sl*(1.3181; 0.9793; 0.2937; 0.0124)	*Su*(-2.4218; 1.0104; 0.0137; 0.0189)	*Su*(-2.4606; 1.0049; 0.0124; 0.0182)
3	0.75	*Sl*(1.2082; 1.0205; 0.2651; 0.0086)	*Sl*(0.8253; 1.2348; 0.1298; 0.0088)	*Sl*(0.7641; 1.1956; 0.1141; 0.0106)
4	1.00	*Sl*(0.1625; 0.9807; 0.0950; 0.0106)	*Su*(-2.1457; 1.4641; 0.0205; 0.0178)	*Su*(-2.2979; 1.3722; 0.0146; 0.0163)
5	1.50	*Sl*(1.4317; 1.0457; 0.3333; 0.0077)	*Su*(-2.2136; 1.4914; 0.0170; 0.0160)	*Su*(-2.0345; 1.5384; 0.0153; 0.0165)
6	1.60	*Sl*(1.2852; 1.0206; 0.2884; 0.0085)	*Su*(-2.2342; 1.4491; 0.0166; 0.0161)	(2.4197; 0.0163; 0.0119)
7	2.0	*Sl*(0.1190; 1.0057; 0.0913; 0.0102)	*Su*(-2.2482; 1.4212; 0.0186; 0.0163)	*Su*(-2.1977; 1.4459; 0.0160; 0.0161)
8	3.00	*Sl*(0.1032; 0.9903; 0.0888; 0.0123)	*Su*(-2.3540; 1.3160; 0.0192; 0.0168)	*Sl*(0.6710; 1.3212; 0.0916; 0.0092)
9	4.00	*Sl*(0.1129; 1.0053; 0.0956; 0.0118)	*Su*(-2.4512; 1.3068; 0.0217; 0.0150)	*Sl*(0.9618; 1.1902; 0.1428; 0.0111)
10	5.00	*Sl*(0.1193; 1.0379; 0.0990; 0.0101)	*Su*(-2.4928; 1.2563; 0.0210; 0.0160)	*Sl*(0.7066; 1.2649; 0.1307; 0.0074)
11	6.00	*Sl*(1.1421; 1.0377; 0.263; 0.0109)	*Su*(-2.4060; 1.2517; 0.0236; 0.0175)	*Sl*(0.8397; 1.1470; 0.1536; 0.0110)
12	7.00	*Sl*(1.2261; 1.0383; 0.2870; 0.0115)	*Sl*(0.7949; 1.1333; 0.1711; 0.0120)	*Sl*(0.8055; 1.1475; 0.1653; 0.0110)

Table 3

Approximation of Limiting Distributions of Ω^2 Mises Statistic while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of only form parameter	Estimation of only scale parameter	Estimation of only bias parameter	Estimation of three parameters (of form, scale, and bias)
1	0.25	*Sl*(0.8745; 1.1415; 1.1035; 0.1002)	*Sl*(1.2310; 1.1339; 1.5755; 0.1026).	*Sl*(1.0109; 1.2891; 1.4503; 0.0940)	*Su*(-2.3812; 1.0889; 0.1015; 0.1485)
2	0.50	*Su*(-2.5811; 1.0889; 0.0908; 0.1502)	*Su*(-2.4437; 1.0649; 0.0919; 0.1550)	*Sl*(0.9215; 1.1284; 1.2930; 0.1159)	*Sl*(1.0190; 1.0398; 1.1220; 0.1135)
3	0.75	*Sl*(1.0242; 1.1631; 1.3162; 0.1026)	*Su*(-2.4034; 1.0931; 0.1025; 0.1527)	*Sl*(0.8710; 1.3996; 0.0112; 0.0804)	*Su*(-2.1784; 1.1030; 0.0802; 0.1439)
4	1.00	*Su*(-2.9616; 1.1617; 0.0850; 0.1204)	*Su*(-2.4983; 1.0832; 0.0911; 0.1548)	*Su*(-2.1865; 1.5178; 0.1810; 0.1669)	*Su*(-2.1179; 1.3890; 0.0939; 0.1309)
5	1.50	*Sl*(1.0533; 1.1878; 1.4052; 0.105)	*Su*(-2.4417; 1.1123; 0.1039; 0.1547)	*Su*(-2.0561; 1.4573; 0.1450; 0.1605)	*Su*(-2.0642; 1.5304; 0.0864; 0.1192)
6	1.60	*Sl*(0.9294; 1.2394; 1.2802; 0.0838)	*Su*(-2.4245; 1.0923; 0.1007; 0.1545)	*Su*(-2.1491; 1.5130; 0.1476; 0.1458)	*Su*(-1.9792; 1.4810; 0.0822; 0.1255)
7	2.00	*Sl*(0.8791; 1.2097; 1.2371; 0.1054)	*Sl*(0.5880; 1.0827; 0.8719; 0.1221)	*Su*(-2.2029; 1.4414; 0.1313; 0.1517)	*Su*(-2.0023; 1.4458; 0.0809; 0.1269)
8	3.00	*Su*(-2.9061; 1.2003; 0.1029; 0.140)	*Su*(-2.6636; 1.1316; 0.0915; 0.1454)	*Su*(-2.1666; 1.4595; 0.160; 0.1518)	*Su*(-2.27; 1.2970; 0.0723; 0.1228)
9	4.00	*Su*(-3.0980; 1.2257; 0.1032; 0.1255)	*Su*(-2.687; 1.1707; 0.1035; 0.14125)	*Su*(-2.2756; 1.4168; 0.1677; 0.1415)	*Su*(-2.3153; 1.1754; 0.0707; 0.1256)
10	5.00	*Su*(-3.1001; 1.2525; 0.1104; 0.1214)	*Su*(-2.7074; 1.1463; 0.1003; 0.141)	*Su*(-2.3443; 1.3488; 0.1511; 0.1536)	*Sl*(0.5525; 1.2411; 0.5018; 0.0831)
11	6.00	*Sl*(0.8671; 1.1959; 1.2956; 0.1158)	*Sl*(1.0228; 1.1518; 1.3609; 0.1083)	*Su*(-2.3719; 1.3537; 0.1623; 0.1462)	*Sl*(0.5166; 1.2025; 0.5222; 0.0882)
12	7.00	*Sl*(1.1286; 1.2642; 1.6235; 0.0929)	*Sl*(0.9102; 1.2147; 1.2198; 0.0939)	*Sl*(0.6061; 1.2797; 0.7946; 0.0980)	*Su*(-2.4112; 1.1873; 0.0905; 0.1194)

(continued)

Approximation of Limiting Distributions of Ω^2 Mises Statistic while Testing for Goodness of Fit to the Class of Exponential Distributions by the Maximum Likelihood Method

No.	Form parameter	Estimation of two parameters (of form and scale)	Estimation of two parameters (of form and bias)	Estimation of two parameters (of scale and bias)
1	0.25	*Sl*(1.0488; 1.0925; 1.2482; 0.0932)	*Su*(-2.3820; 1.1381; 0.1180; 0.1536)	*Su*(-2.4011; 1.0665; 0.0952; 0.1462)
2	0.50	*Su*(-2.4405; 1.0159; 0.0752; 0.1457)	*Sl*(1.2511; 1.14; 1.4661; 0.1067)	*Su*(-2.3619; 1.0352; 0.0814; 0.1621)
3	0.75	*Sl*(1.0707; 1.0907; 1.2632; 0.0941)	*Su*(-2.3367; 1.2714; 0.1190; 0.1369)	*Su*(-2.2040; 1.1573; 0.0965; 0.1602)
4	1.00	*Su*(-2.3917; 0.9957; 0.0772; 0.1414)	*Su*(-2.0747; 1.4923; 0.1427; 0.1533)	*Su*(-1.9642; 1.4128; 0.1289; 0.1568)
5	1.50	*Su*(-2.3473; 1.0625; 0.0970; 0.1335)	*Su*(-2.1035; 1.5579; 0.1265; 0.1432)	*Su*(-1.9488; 1.5350; 0.1071; 0.1406)
6	1.60	*Su*(-2.4953; 1.0236; 0.0736; 0.1305)	*Su*(-2.0957; 1.5279; 0.1241; 0.1367)	*Su*(-1.9403; 1.6314; 0.1232; 0.1409)
7	2.00	*Su*(-2.3736; 0.9870; 0.0734; 0.1472)	*Su*(-2.2201; 1.4813; 0.1249; 0,1282)	*Su*(-2.0349; 1.5403; 0.1173; 0.1372)
8	3.00	*Sl*(0.9651; 1.1157; 1.1347; 0.0897)	*Su*(-2.2047; 1.4; 0.1345; 0.1419)	*Su*(-2.2749; 1.4149; 0.1095; 0.1322)
9	4.00	*Su*(-2.5015; 1.0159; 0.0756; 0.1424)	*Sl*(0.6802; 1.3306; 0.6817; 0.0929)	*Su*(-2.2539; 1.2894; 0.1055; 0.1412)
10	5.00	*Su*(-2.4061; 1.0454; 0.0935; 0.1373)	*Su*(-2.3961; 1.3816; 0.1552; 0.1271)	*Sl*(0.6031; 1.3879; 0.6263; 0.0702)
11	6.00	*Sl*(1.0578; 1.1194; 1.2773; 0.0968)	*Su*(-2.3468; 1.3099; 0.1467; 0.1363)	*Su*(-2.3690; 1.2866; 0.1221; 0.1339)
12	7.00	*Sl*(1.2054; 1.1242; 1.5073; 0.0980)	*Su*(-2.4149; 1.2619; 0.1381; 0.1385)	*Sl*(0.6352; 1.1862; 0.7252; 0.1065)

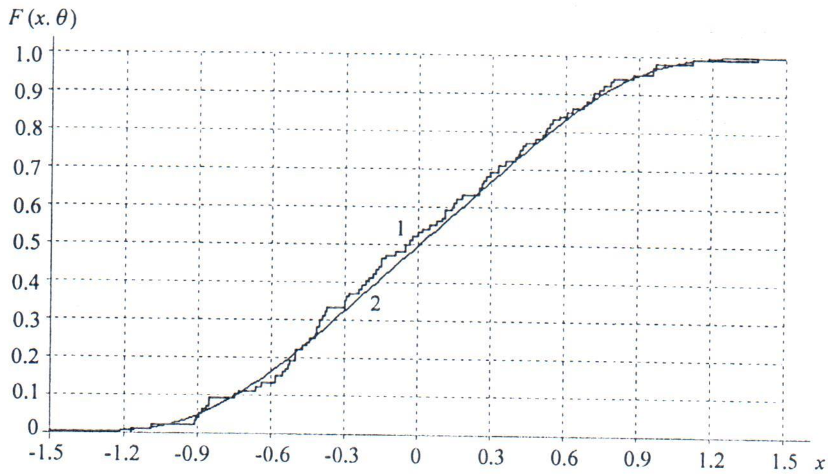

Fig. 9.

culated from the sample: Kolmogorov statistic (2) $S_K^* = 0.5435$, Cramer-Mises-Smirnov statistic (3) $S_\omega^* = 0.0439$, and Anderson-Darling statistic (4) $S_\Omega^* = 0.2647$.

The value of the probability $P\{S_K > 0.5435\} = 0.8727$ calculated according to the gamma distribution γ(3.8001; 0.1221; 0.3272) (see Table 1 for the form parameter $\theta_2 = 4$) for the Kolmogorov test shows that the sample is in good agreement with the theoretical distribution.

Similarly, for the ω^2 Cramer–Mises–Smirnov statistic according to the *Sl*-Johnson distribution *Sl*(0.9618; 1.1902; 0.1428; 0.0111) (see Table 2) we calculate the value of the probability $P\{S_\omega > 0.0439\} = 0.7849$. And for the Anderson–Darling statistic according to the *Su*-Johnson distribution *Su*(-2.2539; 1.2894; 0.1055; 0.1412) (see Table 3) we calculate the value of the probability $P\{S_\Omega > 0.2647\} = 0.8336$.

Thus, we observe good agreement between the analyzed sample and the theoretical model (1) for all of the tests.

CONCLUSION

Using computer methods for investigation of statistical regularities based on statistical modeling of empirical statistic distributions, subsequent analysis of the distributions, and construction of simple analytical models for them we have investigated the distributions of statistics of tests for goodness of fit of the Kolmogorov type, the ω^2 Cramer–Mises–Smirnov type, and the Ω^2 Anderson–Darling type.

Models of the limiting distributions of these statistics while testing composite hypotheses for the case of maximum likelihood estimation of various combinations of parameters of the exponential family of distributions have been obtained. These models were considered for different values of the form parameter θ_2.

The constructed approximations of the limiting distributions of statistics of the nonparametric tests extend the recommendations on standardization [22] and allow one to correctly apply these tests for testing adequacy of models of the form of (1) used in description of measurement system errors and in other problems of statistical analysis of observations.

REFERENCES

1. P.V.Novitskii and I.A.Zograf, *Estimation of Measurement Errors* (in Russian), Energoatomizdat, Leningrad, 1991.
2. L.N.Bolshev and N.V.Smirnov, *Tables for Mathematical Statistics* (in Russian), Nauka, Moscow, 1983.

3. M.Kac, J.Kiefer, and J.Wolfowitz, *Ann. Math. Statist.*, vol. 26, no. 2, p. 189, 1955.
4. G.V.Martynov, *Omega-Square Tests* (in Russian), Nauka, Moscow, 1978.
5. E.S.Pearson and H.O.Hartley, *Biometric Tables for the Statisticians*, University Press, Cambridge, 1972, vol. 2.
6. M.A.Stephens, *Journ. Roy. Statist. Soc.*, vol. B32, p. 115, 1970.
7. M.A.Stephens, *Journ. Amer. Statist. Assoc.*, vol. 69, p. 730, 1974.
8. M.Chandra, N.D.Singpurwalla, and M.A.Stephens, *Journ. Amer. Statist. Assoc.*, vol. 76, p. 375, 1981.
9. Yu.N.Tyurin, *Izv. AN SSSR. Ser. Matem.*, vol. 48, no. 6, p. 1314, 1984.
10. Yu.N.Tyurin and N.E.Savvushkina, *Izv. AN SSSR, Ser. Tekhn. Kibernetika*, no. 3, p. 109, 1984.
11. Yu.N.Tyurin, *Investigation on Nonparametric Statistics (nonparametric methods and a linear model)* (in Russian), MGU, Moscow, 1985, 33 p.
12. N.E.Savvushkina, *Sb. Tr. VNII Sistem. Issled.*, no. 8, 1990.
13. Yu.N.Tyurin and A.A.Makarov, *Computer Data Analysis* (in Russian), INFRA-M, Finansi i Statistika, Moscow, 1995.
14. B.Yu.Lemeshko and S.N.Postovalov, *Nadezhnost i Kontrol Kachestva*, no. 11, p. 3, 1997.
15. B.Yu.Lemeshko and S.N.Postovalov, *Zavodskaya Laboratoriya*, no. 3, p. 61, 1998.
16. B.Yu.Lemeshko and S.N.Postovalov, *Methodical Recommendations*, part II, Nonparametric tests (in Russian), NGTU, Novosibirsk, 1999.
17. B.Yu.Lemeshko and S.N.Postovalov, *Optoelectr., Instrum. and Data Process.*, no. 2, p. 76, 2001.
18. B.Yu.Lemeshko and S.N.Postovalov, *Zavodskaya Laboratoriya. Diagnostika Materialov*, vol. 67, no. 7, p. 62, 2001.
19. B.Yu.Lemeshko and S.N.Postovalov, *Dokl. SO AN VSh.*, no. 1(5), p. 65, 2002.
20. B.Yu.Lemeshko, S.N.Postovalov, and A.V.Frantsuzov, *Optoelectr., Instrum. and Data Process.*, no. 2, p. 3, 2002.
21. B.Yu.Lemeshko and S.N.Postovalov, *Modeling, Automation, and Optimization of Science Technologies* (in Russian), NGTU, Novosibirsk, 2000, p. 44.
22. *R.50.1.037-2002. Recommendations on Standardization. Applied Statistics. Rules for Tests for Goodness of Fit for Experimental and Theoretical Distributions*, part II, Nonparametric tests (in Russian), Izd-vo Standartov, Moscow, 2002.

The Novosibirsk State Technical University

E-mail: headrd@fpm.ami.nstu.ru

13 May 2003