UDC 519.25

V.E. Bakhrushin, I.O. Dudko

DETERMINATION OF CRITICAL VALUES OF THE KOLMOGOROV – SMIRNOV TYPE CRITERION FOR SOME TYPES OF DISTRIBUTION BY THE MONTE-CARLO METHOD

Annotation. In the article it is considered the problem of the critical values estimating of the Kolmogorov – Smirnov type criterion by Monte-Carlo method in the case when the parameters of distribution models are obtained by minimizing the calculated values of this criterion.

Key words: identification, distribution function, Kolmogorov – Smirnov test, Monte-Carlo method.

Statement of the problem. In practice, it is often need to identify models of distribution. It is very often aiming to choose a model from a given class that could best describe the available empirical evidence. This procedure consists of three steps [1, 2]:

1) formation of hypothesis about the law (model) distribution of the audited;

2) estimation of parameters chosen model;

3) check the adequacy of the model with the help of certain statistical criteria.

At the last stage criteria such as the Kolmogorov–Smirnov, omega-square, chi-square and so on are most often used. In [2] it was shown that the distribution model can be identified by minimizing the estimated value goodness-of-fit test. But there is the problem of choosing critical values for acceptance or rejection of the hypothesis of the adequacy of the model to existing data.

Analysis of recent research and publications. H. Lilliefors [3] estimated the critical value of the criterion of Kolmogorov – Smirnov type for the case when the model parameters of a normal distribution estimated as the mean and standard deviation of the studied sample. In his work was studied in 1,000 or more sample from 4 to 30 items. Later, these results were refined by P. Molin and H. Abdi [4] using modern computer technology, allowing us to increase the number of samples studied to 100,000, and their maximum size – up to 50 items. In both cases, the method of statistical tests of Monte Carlo was used.

In [5] we have shown a correlation and defined equation relation that allows you to set a rough correspondence between the critical values for the case when the distribution parameter identification is performed by minimizing the expected value of the Kolmogorov–Smirnov type criterion and the corresponding values of the Lilliefors test. In this paper we consider another approach to determine the critical values of the Kolmogorov – Smirnov type criterion.

Problem. Using Monte-Carlo method to determine the critical value of the Kolmogorov – Smirnov type criterion for normal, gamma, exponential distribution and Weibull distribution.

The main material of research. Among the non-parametric goodness-of-fit test of the empirical distributions by any theoretical model $\{F(x;\Theta), \theta \in \Theta\}$ distinguish the Kolmogorov – Smirnov test criterion [1]. They determine the degree of compliance with the maximum absolute value deviation of models from empirical distribution function. That is, to test the null hypothesis of bilateral $H_0 : F_n(x) = F(x, \theta_0)$ calculate the maximum distance between the empirical $F_n(x)$ and theoretical $F(x)$ distribution functions: $D_n = \sup\limits_{|x|<\infty} |F_n(x) - F(x, \theta_0)|$. Estimated value of the Kolmogorov – Smirnov type criterion determined by the formula:

$$\lambda = \sqrt{n} D_n(\theta_0), \qquad (1)$$

where n is sample size.

Equation (1) is used if the parameters $\theta = \theta_0$ of theoretical distribution functions are known. But in practice usually happens that these parameters are unknown to the researcher. That’s why it is necessary to use a vector of sample estimates θ^*. Estimated value of the criterion in this case is determined by the formula:

$$\lambda^* = \sqrt{n} D_n\left(\theta^*\right) \qquad (2)$$

In this case, the critical values of the criterion are changed. As a result, it is necessary to solve one of these tasks:

- to assess the suitability of the model $F_n(x) = F(x, \theta)$ with known parameters available empirical data (Kolmogorov – Smirnov test);

– to assess compliance with existing empirical data model $F_n(x) = F(x, \theta^*)$ with the parameters that were defined by selective moments (Lilliefors criterion);

– to assess compliance with existing empirical data model $F_n(x) = F_{\min}(x, \theta^*)$ with the parameters that determine from the condition of a minimum indicator of quality models, such as the estimated value of the Kolmogorov – Smirnov type criterion. Problem of this type is solved in this paper.

Method study is similar to the used in [3, 4] and was based on the Monte-Carlo method. It consists in the following. A large number of samples from a given distribution law are generating. For each sample determined sample parameters (for example, for normal distribution this sample mean and standard deviation) and the calculated value D_n^*. Next, using the obtained sample parameters as initial approximation, refine the model parameters by minimizing the calculated values D_n^* using a multidimensional nonlinear optimization method. Then we build an empirical distribution function obtained values $D_n^*(\min)$. Quantile of the respective levels of the distribution function up table values desired statistics such as Kolmogorov – Smirnov type.

At the beginning of conducted research we generated on 10,000 samples quantity N of 6 to 45 items under the standard normal $N(0, 1)$, gamma $\Gamma(1, 0.5)$, exponential $Exp(0.8)$ and Weibull distribution $W(1, 2)$.

The table was made from quintiles of distribution functions corresponding statistics $D_n^*(\min)$. In particular, Table 1 shows the values of 0.05-quantile $D_n^*(\min)$ for the studied models of distribution and different amounts of samples. Also, similar values for the Lilliefors $N_L(\mu, \sigma^2)$ [3] and Kolmogorov – Smirnov test $N_{K-S}(\mu, \sigma^2)$ [4] are shown for comparing.

Table 1

0.05-quintiles of distribution $D_n^*(\min)$

The sample size, N	Quintiles $D_n^*(\min)$				Quintiles D_n	
	$N(\mu, \sigma^2)$	$\Gamma(k, \theta)$	$Exp(\lambda)$	$W(k, \lambda)$	$N_L(\mu, \sigma^2)$	$N_{K-S}(\mu, \sigma^2)$
6	0,226	0,245	0,291	0,313	0,319	0,324
7	0,212	0,230	0,275	0,290	0,300	0,304
8	0,202	0,218	0,265	0,272	0,285	0,288
9	0,192	0,203	0,253	0,263	0,271	0,274
10	0,183	0,193	0,242	0,248	0,258	0,262
11	0,175	0,183	0,234	0,236	0,249	0,251
12	0,169	0,175	0,227	0,228	0,242	0,243
13	0,163	0,169	0,217	0,219	0,234	0,234
14	0,157	0,162	0,212	0,214	0,227	0,226
15	0,152	0,157	0,205	0,205	0,220	0,220
16	0,149	0,153	0,200	0,200	0,213	0,213
17	0,144	0,148	0,193	0,194	0,206	0,207
18	0,140	0,145	0,189	0,187	0,200	0,202
19	0,137	0,140	0,182	0,183	0,195	0,197
20	0,134	0,137	0,180	0,179	0,190	0,192
25	0,120	0,122	0,161	0,161	0,180	0,173
30	0,111	0,112	0,147	0,148	0,161	0,159
35	0,103	0,105	0,139	0,136	0,150	0,155
40	0,096	0,097	0,129	0,129	0,140	0,148
41	0,096	0,096	0,128	0,126	0,138	0,139
42	0,094	0,095	0,126	0,126	0,137	0,135
43	0,093	0,095	0,125	0,123	0,135	0,134

The sample size, N	Quintiles $D_n^*(\min)$				Quintiles D_n	
	$N(\mu,\sigma^2)$	$\Gamma(k,\theta)$	$Exp(\lambda)$	$W(k,\lambda)$	$N_L(\mu,\sigma^2)$	$N_{K-S}(\mu,\sigma^2)$
44	0,092	0,093	0,123	0,123	0,134	0,132
45	0,091	0,093	0,121	0,122	0,132	0,131

In Fig. 1 plots of the dependence of values specified quintiles from sample size for the analyzed distributions are shown.

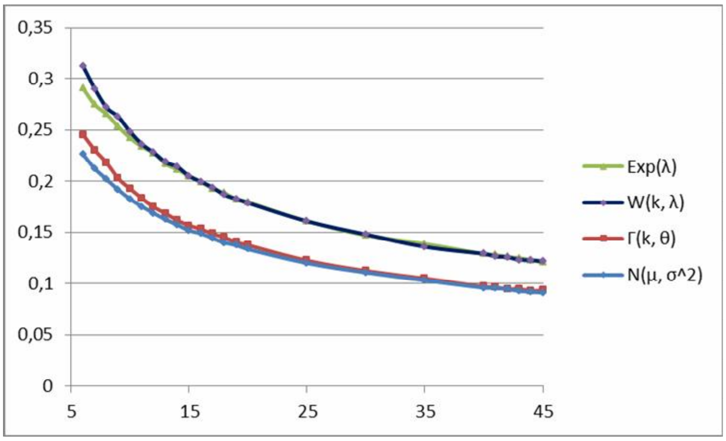

Figure 1 - Dependence of the 0.05 quintiles of the distribution from sample size

At first 10,000 sample size of 1000 elements were generated for determine the boundary estimates of critical values of the Kolmogorov – Smirnov type criterion and then the above-mentioned procedure was repeated. The results are shown in Table 2.

Table 2

Quintiles of distribution $D_n^*(min)$ for n = 1000

	Quintiles $D_n^*(\min)$				
Level of significance, p	0,80	0,85	0,90	0,95	0,99
$N(\mu,\sigma^2)$	0,0171	0,0177	0,0184	0,0195	0,0216
$\Gamma(k,\theta)$	0,0170	0,0178	0,0187	0,0199	0,0227
$Exp(\lambda)$	0,0219	0,0227	0,0242	0,0259	0,0304
$W(k,\lambda)$	0,0222	0,0233	0,0245	0,0263	0,0297

Using formula (2), we get the critical value $\lambda_{\min}^*$ of the Kolmogorov – Smirnov type criterion (Table 3).

Table 3

The critical value $\lambda_{\min}^*$ of the Kolmogorov – Smirnov type criterion

	The critical value $\lambda_{\min}^*$				
Level of significance, p	0,80	0,85	0,90	0,95	0,99

	The critical value $\lambda^*_{\min}$				
Level of significance, p	0,80	0,85	0,90	0,95	0,99
$N(\mu, \sigma^2)$	0,5418	0,5605	0,5845	0,6171	0,6845
$\Gamma(k, \theta)$	0,5407	0,5630	0,5930	0,6296	0,7180
$Exp(\lambda)$	0,6925	0,7178	0,7652	0,8190	0,9613
$W(k, \lambda)$	0,7022	0,7382	0,7735	0,8311	0,9379
$N_L(\mu, \sigma^2)$	0,736	0,768	0,805	0,886	1.031
$N_{K-S}(\mu, \sigma^2)$	0,775	0,819	0,895	0,955	1,035

Results of the research. Critical values of value goodness-of-fit Kolmogorov - Smirnov test for some types of distribution when the parameters of the model distribution were improved by minimizing the estimated value of this criterion were estimated by Monte-Carlo method.

Subsequently we'll plan to specify confidence intervals for these estimates and verify their possible dependence on other factors.

REFERENCES

1. Bakhrushin V.E. Methods of data analysis / V.E. Bakhrushins. – Zaporozhye: CPU, 2011.- 268 p.

2. Bakhrushin V.E. Problems of identification of random distribution models of using modern software / V.E. Bakhrushin // Successes of modern science. – 2011. – № 11. – P. 50-54.

3. Lilliefors H. On the Kolmogorov – Smirnov test for normality with mean and variance unknown. Journal of the ASA, Vol. 62, No. 318 (Jun., 1967), 399-402.

4. Molin P., Abdi H. (1998). New Tables and numerical approximation for the Kolmogorov - Smirnov/Lillierfors/Van Soest test of normality. Technical Report, University of Bourgogne (France) / University of Texas at Dallas (USA).

5. Bakhrushin V.E. Refinement of normal distribution models by minimizing the Kolmogorov – Smirnov criterion / V.E. Bakhrushin, I.O. Dudko // System technologies. Regional Interuniversity collection of scientific papers. – Release 5 (82). – Dnipropetrovsk, 2012. – P. 95-103.