Statistical testing of hypotheses about the form of the factor law of influence by the Kolmogorov criterion

A Tumanov\textsuperscript{1,}, A Sabanaev\textsuperscript{1}, A Solovyov\textsuperscript{1} and V Tumanov\textsuperscript{1}

\textsuperscript{1}Peter the Great St.Petersburg Polytechnic University, Polytechnicheskaya, 29, St. Petersburg, 195251, Russia
E-mail: toumanov@mail.ru

Abstract. When affected by the impact of emergency sources on the same type of elementary objects, the depth of the introduction of shrapnel into the barrier changes randomly. It is necessary to determine the law of distribution of random depth of inserting fragments into the barrier. In the course of the work, hypotheses about three distribution laws were tested. Based on the results of experimental studies, it is proposed to assess the suitability of the law of distribution, the statistics of which, according to the Kolmogorov criterion, is intended to check the alignment of empirical distribution with this continuous theoretical distribution. With a small number of tests on the criterion Kolmogorov confirmed the consent of even clearly incompatible samples. However, with a large number of observations (more than 50) the Kolmogorov Criterion is able to detect any deviation from the hypothesis.

1. Introduction
In the case of potentially dangerous sea-based energy projects (hereinafter referred to as the SBEP) is available at the same time complicated system of risk factors (RF) emergency situations (ES), because security concerns SBEP overlap set of security problems in maritime transport and the problems of nuclear and radiation safety, for example in the case of floating nuclear thermal power stations (hereinafter referred to as the FNTPS) and floating energy blocks (FEB).

When affected by the impact of emergency sources, the depth of the shrapnel in the barrier changes randomly. It is necessary to determine the law of distribution of random magnitude of the depth of inserting fragments into the barrier. Testing the hypothesis about the law of distribution. It is proposed to assess the suitability of distribution laws through the results of experimental studies (table 1).

Let's look at three hypotheses about the form of distribution. Hypothesis 1: The depth of the introduction of shrapnel into the barrier is subject to the normal law of distribution. Point estimates of the parameters of the total population were calculated by a sample of the table 1: $X \sim 16.672$ mm; $\Gamma = 408$. Hypothesis 2: The depth of the introduction of shrapnel into the barrier is subject to the lognormal law of distribution. Hypothesis 3: The depth of the introduction of shrapnel into the barrier is subject to the law of Weibull-Gnedenko. The Kolmogorov criterion was used to test the hypotheses.

In statistics the Kolmogorov test is used to determine whether two empirical distributions follow the same law, or to determine whether the resulting distribution follows the proposed model [1,2]. The criterion of the Soviet mathematician A. N. Kolmogorov (1903-1987) is based on determining the maximum discrepancy between the accumulated frequencies of empirical or theoretical distributions. The transition from the parametric statistics of Pearson, Student, and Fisher to the nonparametric
statistics of Kolmogorov and Smirnov corresponds to the transition from the old paradigm to the new one, from outdated and not relevant reality formulations of data analysis problems to modern statistical methods [1,2]. The statistics of the Kolmogorov criterion is intended to verify the agreement of the empirical distribution with a given continuous theoretical distribution.

The purpose of the research is to verify how much the empirical distribution is consistent with a given continuous theoretical distribution.

Table 1. An example of the results of a pilot study of the depths of implementation in the barrier obtained by ballistic tests at the speed of shrapnel \( v = 180 \text{ m/s} \).

| \( n \) | depth, \( x_0 \), mm | \( n \) | depth, \( x_0 \), mm | \( n \) | depth, \( x_0 \), mm | \( n \) | depth, \( x_0 \), mm |
|-------|-----------------|-------|-----------------|-------|-----------------|-------|-----------------|
| 1     | 15.04           | 16    | 19.00           | 31    | 16.23           | 46    | 18.13           |
| 2     | 15.49           | 17    | 17.07           | 32    | 14.96           | 47    | 17.48           |
| 3     | 15.80           | 18    | 15.63           | 33    | 16.66           | 48    | 15.76           |
| 4     | 14.66           | 19    | 17.51           | 34    | 18.34           | 49    | 16.77           |
| 5     | 14.86           | 20    | 17.21           | 35    | 16.64           | 50    | 16.61           |
| 6     | 13.50           | 21    | 16.61           | 36    | 15.76           | 51    | 15.58           |
| 7     | 16.60           | 22    | 16.38           | 37    | 15.73           | 52    |                 |
| 8     | 17.72           | 23    | 16.89           | 38    | 15.81           | 53    |                 |
| 9     | 14.24           | 24    | 17.35           | 39    | 14.65           | 54    |                 |
| 10    | 18.65           | 25    | 17.93           | 40    | 16.74           | 55    |                 |
| 11    | 17.32           | 26    | 17.55           | 41    | 16.89           | 56    |                 |
| 12    | 19.20           | 27    | 16.63           | 42    | 16.57           | 57    |                 |
| 13    | 17.37           | 28    | 16.94           | 43    | 17.33           | 58    |                 |
| 14    | 20.16           | 29    | 15.82           | 44    | 18.36           | 59    |                 |
| 15    | 18.46           | 30    | 13.61           | 45    | 18.06           | 60    |                 |

2. Methods

Methods the Kolmogorov criterion is calculated by the formula in accordance with [1]:

\[
\lambda_n := D_n \cdot \sqrt{n}
\]

(1)

For lognormal distribution:

\[
\lambda_l := D_l \cdot \sqrt{n}
\]

(2)

The critical values for the statistics \( \lambda_{cr} \) are given in table 2.

Table 2. Critical values for the statistics \( \lambda_{cr} \) [1].

| \( \alpha \) | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|-------------|-----|-----|-----|-----|------|-------|------|-------|-------|-------|
| \( \lambda_{cr} \) | 0.85 | 0.94 | 1.08 | 1.22 | 1.36 | 1.48 | 1.63 | 1.73 | 1.95 | 2.03 |

In this case, the Kolmogorov criterion is used to test the hypothesis that the observed sample belongs to the normal, lognormal, or Weibull-Gnedenko law, the parameters of which are estimated from this very sample using the maximum likelihood method. That is, a complex hypothesis is tested and sample estimates of the mean and variance are used as an estimate of the parameters of the normal law.

In this case, in accordance with Hubert W. Lilliefors [3], modified statistics of the form were used:

\[
\lambda_x = D_n (\sqrt{n} \cdot 0.01 + \frac{0.05}{\sqrt{n}})
\]

(3)

It should be noted that the critical values for the statistics \( \lambda_{cr} \) (Table 1) can no longer be used [3]. The critical values for the \( \lambda_x \) statistics are given in table 3.
The calculations. We define the Kolmogorov criterion and the modified Kolmogorov criterion for sample 1.

| Table 3. Critical Values for Statistics $\lambda_2$ [3]. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $\alpha$       | 0.15            | 0.10            | 0.05            | 0.03            | 0.01            |
| $\lambda_{xp}$ | 0.775           | 0.819           | 0.895           | 0.955           | 1.035           |

The initial data are taken from [4].
Expected value $m = 5.58$.
Standard deviation $\sigma = 1.329$.
Below is a description of the calculations performed to obtain Kolmogorov statistics in the environment and in the notation MathCAD.

```plaintext
ORIGIN := 1
n := 27
i := 1..n
x := 2.2.1..10
Pe_i := i/n
Fe := 1 - Pe
m := mean(V)
m = 5.58
\sigma := Stdev(V)
\sigma = 1.329

For normal law
Fn(x) := pnorm(x, m, \sigma)
Fne_i := Fn(V_i)
\Delta Fn_i := Fe_i - Fne_i
Dn := max(|\Delta Fn|)
Dn = 0.239
\gamma := Dn \cdot \sqrt{n}
\gamma = 1.241

For lognormal law
y := ln(V)
lnm := mean(y)
```
\begin{align*}
\ln n &= 1.693 \\
\lambda_0 &:= \text{Stddev}(\gamma) \\
\lambda &:= 0.232 \\
F_l(x_0) &:= \text{pinorm}(x_0, \lambda n, \lambda) \\
F_{lev} &:= F_l(V_i) \\
\Delta F_l &:= F_{lev} - F_{lev} \\
D_l := \max(\left| \Delta F_l \right|) \\
D_l &= 0.176 \\
\lambda &= D_l \cdot \sqrt{n} \\
\lambda &= 0.913
\end{align*}

Modified Kolmogorov criterion for normal law

\begin{align*}
\nu_n := D_n \cdot \left( \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right) \\
\nu_n &= 1.278
\end{align*}

Modified Kolmogorov criterion lognormal law,

\begin{align*}
\nu_n := D_l \cdot \left( \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right) \\
\nu_n &= 0.94
\end{align*}

for Weibull-Gnedenko law

\begin{align*}
f(x, \alpha, q) &:= \alpha \cdot \exp[-\alpha \cdot (x - q)] - \exp[-\alpha \cdot (x - q)] \\
\alpha &:= \frac{1}{\sigma} \\
q &:= m \\
\alpha &= 0.752
\end{align*}

Given

\begin{align*}
m &= \int_{0}^{100} x \cdot f(x, \alpha, q) \, dx
\end{align*}
\[
\sigma^2 = \int_0^{100} (x - m)^2 \cdot f(x, \omega, q_0) \, dx
\]

\[
\text{Find} (\omega, q_0) = \begin{pmatrix} 0.965 \\ 4.982 \end{pmatrix}
\]

\[
\omega \quad := 0.965
\]

\[
q_0 \quad := 4.982
\]

\[
f_{\omega}(x) := \omega \cdot \exp[-\omega \cdot (x - q_0) - \exp[-\omega \cdot (x - q_0)]]
\]

\[
Fv_1(x) := \int_1^x f_\nu (t) \, dt
\]

Fv1(5) = 0.374

\[
Fv_{e_i} := Fv_1 (V_i)
\]

\[
\Delta Fv_i := Fv_{e_i} - Fv_{e_i}
\]

\[
Dv := \max(\left|\Delta Fv \right|)
\]

Dv = 0.181

\[
\omega v := Dv \cdot \left( \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right)
\]

\[
\omega v = 0.971
\]

\[
S2n := \sum_{i=1}^{n} (\Delta F_{n_i})^2
\]

S2n = 0.0571

\[
S2l := \sum_{i=1}^{n} (\Delta F_{l_i})^2
\]

S2l = 0.03085

\[
S2v := \sum_{i=1}^{n} (\Delta F_{v_i})^2
\]

S2v = 0.03291
3. Results and Discussion

Figure 1 shows graphs of the integral theoretical and empirical distribution functions for sample 1.

![Graph of integral theoretical and empirical distribution functions for sample 1](image1)

**Figure 1.** Graphs of the integral theoretical and empirical distribution functions for the sample 1.

We define the Kolmogorov agreement criterion for sample 2. The data are taken from [4]. Figure 2 shows graphs of the integral theoretical and empirical distribution functions for sample 2.

![Graph of integral theoretical and empirical distribution functions for sample 2](image2)

**Figure 2.** Graphs of the integral theoretical and empirical distribution functions for sample 2.

We define the Kolmogorov agreement criterion for sample 5. The initial data are taken from [4] Figure 3 shows graphs of the integral theoretical and empirical distribution functions for sample 5. In the form of hypotheses, it was put forward that the integral function of the distribution of probabilities of the introduction of fragments of industrial equipment into the barrier of various materials in an accident at a dangerous technical object is subject to one of the known laws distribution. Descriptions of the research site, the materials of the barriers, techniques and devices for the tests are given in [5-11].

![Graph of integral theoretical and empirical distribution functions for sample 5](image3)

**Figure 3.** Graphs of the integral theoretical and empirical distribution functions for the sample 5.
4. Conclusion

An analysis of the results shows that the obtained values of the criteria for samples 1 and 2 are less than their tabular critical values with an appropriate significance level. For these samples, the null hypothesis is not rejected. For the lognormal distribution, sample 1 shows better results than normal.

For sample 5, the null hypothesis is not confirmed. With a small number of observations by the Kolmogorov criterion, the agreement of even clearly incompatible distributions is confirmed. However, with a large number of observations (more than 50), the Kolmogorov criterion is able to detect any deviation from the hypothesis. This means that any difference in the distribution of the sample from the theoretical one will be found with its help if there are a lot of observations. The practical significance of this property is significant in our case, since it is possible to obtain a sufficiently large number of observations under constant conditions, although the theoretical concept of the distribution law to which the sample must follow is always approximate, and the accuracy of statistical checks should usually not exceed the accuracy of the selected model.

References

[1] Kolmogoroff A 1933 Sulla determinazione empirica di una legge di distribuzione // Guornale dell’ Instituto Italiano degli Attuari V 4 №1 P 83-91

[2] Orlov A I 2014 Neparametricheskie kriterii coglasiya Kolmogorova, Smirnova, Omega-kvadrat i oshibki pri ih primenenii Nauchnyj zhurnal KubGAU №97(03)

[3] Hubert W 1967 Lilliefors On the Kolmogorov-Smirnov Test for Normality with mean fnd Variance J. of the American statistical association Vol 62, no 318 pp 399-402

[4] Tumanov A Yu 2019 Nauchno-metodicheskie osnovy ocenki opasnosti tekhnogennyh avarij na potencial'nop-opasnyh ob"ektah: monografiya (POLITEKH-PRESS) pp 241

[5] Ahmedov I M, Naumov V A 2017 O zakone raspredeleniya razrynoj nagruzki niti Vestnik nauki i obrazovaniya Severo-Zapada Rossii

[6] Sikan A V 2011 Prakticheskie priemy ocenki parametrov raspredeleniya Vejbulla pri vypolnenii gidrologicheskikh raschetov Uchenye zapiski Rossijskogo gosudarstvennogo gidrometeorologicheskogo universiteta Nauchno-teoreticheskij zhurnal SPb (RGGMU) № 19 pp 195

[7] Tumanov A Yu, Gumenyuk V I 2018 Scientific methodological fundamentals of risk assessment natural and man-caused emergencies at hazardous technical objects monograph. (Spb. Polytechnic), pp 135.

[8] Petrov G and Tarmaev A 2018 Modeling of railway vehicles movement having deviations in the content of running parts Proceedings of the International Conference «Aviamechanical Engineering and Transport» (Series: Advances in Engineering Research, Atlantis Press) vol 158, pp 410-415 DOI:10.2991/avent-18.2018.79

[9] Tumanov A, Gumenyuk V, Tumanov V 2017 Development of advanced mathematical predictive models for assessing damage avoided accidents on potentially-dangerous sea-based energy facility IOP Conference Series: Earth and Environmental Science 19, Cep. "Energy Management of Municipal Transportation Facilities and Transport, EMMFT 2017" C 012027

[10] Kulinkovich A V, Sakova N V and Tumanov A Yu 2019 Development of the Express Method for Controlling Uranium Compounds in Natural Waters in Emergency Situations on Floating Nuclear Thermal Power Plants 2019 IOP Conf. Ser.: Earth Environ Sci. 272 № 022016

[11] Tumanov A 2019 Risk Assessment of Accidents During the Transportation of Liquid Radioactive Waste in Multimodal Transport IOP Conference Series: Earth and Environmental Science, 272 № 032078