The simulation system and research of functions of random variables

Pavel Yu. Blinov, Boris Yu. Lemeshko
Novosibirsk State Technical University, Novosibirsk, Russia

Abstract **–The questions of research of the distribution laws of one-dimensional functions of a set of independent random variables are considered. The subsystem that allow for implementing the imitation modeling of different functions of a set of independent random variables and investigating these distribution laws is implemented. System capabilities illustrated the efficiency of this method is demonstrated in set of examples.**

Index terms **– function of random variables, distribution function of random variables, simulation, program system.**

I. Introduction

In practice of statistical analysis the problem statements appear substantially more than the proposed solution in classical mathematic statistic.

The problem of detection the distribution law of function of random variable is one of widely sought. The solution of this problem with application of classical method represents time-consuming process, but it could find a solution in analytic form in exceptional cases. The various distribution laws and different complexity of the function of random variable complicates the problem.

Researchers encountered very often with the need to construct the distribution function of random variables in the problems of indirect measurement error analysis when the measured value Y is a function of random variables $Y = \varphi(X_1, X_2, ..., X_n)$. The determination of the distribution law of Y under known distribution laws of random variables X_i is possible by analytical method in analytical form in exceptional cases. The use of the linearization functions $\varphi(\cdot)$ and the construction of approximate solutions, as usually, suffer from serious inaccuracies.

We can consider the application of statistical modeling as real escape from deadlock under searching the distribution law $F(y)$. The investigation of distribution $F(y)$ in depend on distribution laws $F_i(x_i)$ by the methods of statistic modeling with corresponding program support doesn't generate the fundamental difficulties [1-3].

One of aims of present work is development of software package allowing to simulate the empirical distribution laws for any functions of random variables distributed by different laws in system ISW[4], with further research.

We solve the next tasks consistently to achieve this problem:

- the analysis of methods and approach used for simulation of pseudorandom variables in method of statistic test;
- the modeling of distribution laws of one-dimensional function of random variables;
- debugging syntactic and semantic analyzers of input expressions;
- the research of distributions received in result of simulation.

The methods of statistical modeling, probability theory, mathematic statistic is used under performance of work.

Also the aim of this paper is concluded in demonstration of opportunities of simulation system for investigating of distribution laws of functions of random variables. Considered examples related with known results demonstrate the change of distribution laws in depend on distribution laws of the arguments.

In this paper, examples have been studied by the method of statistical simulations. The number of experiments carried out for statistical modeling is usually assumed equal to $N = 1.66\times10^6$ in the study of the distributions. One the one hand, such number of experiments allows tracing the qualitative picture of distributions in depend on various factors. In the other hand, this number of experiments provides acceptable accuracy of unknown probabilities.

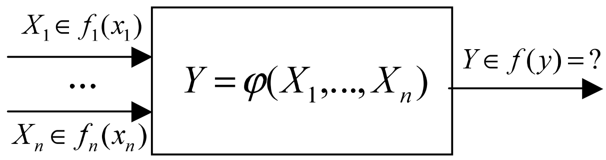
II. Problem Definition

Quite frequently solved problem in metrology is problem of definition of probability characteristics of quantity Y, which isn't available directly for measurement, on base of quantities $X_1, X_2, ..., X_k$ available for multiple measurements. Suppose that

$$Y = \varphi(X_1, X_2, ..., X_k),$$

where $\varphi(\cdot)$ – some known function (or in vector form $Y = \varphi(\overline{X})$). Suppose that distribution law of vector $\overline{X}$ or distribution laws $X_1, X_2, ..., X_k$ (or distribution law of error measurements) in the case of independence of its components are known or can be construct based on results of statistic analysis.

The function $Y = \varphi(\overline{X})$ can be the result of functioning of some information measuring system.



The classical approach of determination the probability distribution law from set of random variables suggests that joint density $f(x_1, x_2, ..., x_k)$ of a set of random variables $X_1, X_2, ..., X_k$ is known.

Let be $X : \Omega \to R^n$ is random variable and $g : R^n \to R^n$ is continuously differentiable function such as $J_g(x) \neq 0,\ \forall x \in R^n$, where $J_g(x)$ is jacobian function g in point x. Then random variable absolutely continuous too and it density has form:

$$f_Y(y) = f_X(g^{-1}(y))\left|J_{g^{-1}}(y)\right|.$$

However analytical decision with using the classical approach can be finding only for some special cases of $Y=\varphi\left(\overline{X}\right)$ and $f\left(x_1,x_2,...,x_k\right)$.

In this paper, the various functions of a set of random variables satisfying to the different distribution laws are considered. The main method of distribution definition of interested random variable Y is statistical method, which based on Monte-Carlo method [2]. The subsystem of simulation of function of random variables is implemented in the program ISW[4].

III. APPLICATIONS

Some examples demonstrated the opportunity and accuracy research of the behavior of distribution function law of random variable are considered below.

We used different goodness-of-fit tests for testing that modeling empirical distribution $F(y)$ is theoretic law. However, tables contain the results of use only for χ^2 Pearson test, Kolmogorov test, ω^2 Cramer–Mises–Smirnov (CMS) test and Ω^2 Anderson-Darling (AD) test. The significance level $\alpha=0.01$ is used in all considered examples.

At the start, we consider a few examples of simulation of functions of one random variable.

Example 1. $Y=X^n$, where $X\in U(0,1)$. In this case theoretic law describing distribution Y is beta distribution of I type with shape parameters equals $1/n$ and 1. Table 1 contains p_{value} under goodness-of-fit testing of empirical distributions $F_N(y)$ obtained in result of simulations with corresponding beta-distribution.

TABLE I
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY $Y=X^n$ WITH BETA-DISTRIBUTION

n	Y	χ^2 Pearson	ω^2 CMS
0.1	$B(10,1)$	0.493	0.880
0.2	$B(5,1)$	0.426	0.089
0.5	$B(2,1)$	0.494	0.291
2	$B(0.5,1)$	0.807	0.803
3	$B(0.333,1)$	0.668	0.789
5	$B(0.2,1)$	0.422	0.614
n	Y	Kolmogorov	Ω^2 AD
0.1	$B(10,1)$	0.837	0.820
0.2	$B(5,1)$	0.072	0.098
0.5	$B(2,1)$	0.395	0.313
2	$B(0.5,1)$	0.892	0.819
3	$B(0.333,1)$	0.899	0.796

Example 2. Now we verify the property of mirror-image symmetry for beta distribution. If $X\in B(k,n)$, then $Y=(1-X)\in B(n,k)$, where n and k - shape parameters. The results of testing that $F_N(y)$ is $B(n,k)$ are demonstrated in table 2.

TABLE II
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY Y WITH CORRESPONDING BETA-DISTRIBUTION

X	Y	χ^2 Pearson	ω^2 CMS
$B(0.8,1)$	$B(1,0.8)$	0.493	0.880
$B(5,3)$	$B(3,5)$	0.886	0.914
$B(10,25)$	$B(25,10)$	0.234	0.458
X	Y	Kolmogorov	Ω^2 AD
$B(0.8,1)$	$B(1,0.8)$	0.837	0.820
$B(5,3)$	$B(3,5)$	0.938	0.887
$B(10,25)$	$B(25,10)$	0.702	0.435

After analyzing the previous two examples, we obtain the following relationship between the uniform distribution and beta-distribution of I type.

Example 3. $Y=1-X^n$, where $X\in U(0,1)$. Theoretic law describing distribution Y is beta-distribution with parameters of form equals 1 and $1/n$. Table 3 includes the quantities of p_{value} under goodness-of-fit testing of modeling distibutions $F_N(y)$ with corresponding beta-distribution.

TABLE III
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY $Y=1-X^n$ WITH BETA-DISTRIBUTION

n	Y	χ^2 Pearson	ω^2 CMS
0.1	$B(1,10)$	0.181	0.535
0.2	$B(1,5)$	0.942	0.998
0.5	$B(1,2)$	0.604	0.873
2	$B(1,0.5)$	0.651	0.661
3	$B(1,0.333)$	0.588	0.885
5	$B(1,0.2)$	0.653	0.652
n	Y	Kolmogorov	Ω^2 AD
0.1	$B(1,10)$	0.590	0.651
0.2	$B(1,5)$	0.999	0.986
0.5	$B(1,2)$	0.869	0.938
2	$B(1,0.5)$	0.495	0.642
3	$B(1,0.333)$	0.758	0.811

Example 4. $Y=\left|X-\mu\right|/\sigma$, where $X\in Laplace(\mu,\sigma)$. Theoretic law describing distribution Y is standard exponentional distribution. The values of p_{value} obtained under goodness-of-fit testing of modeling distribution $F_N(y)$ with exponential distribution law are contained in Table 4.

TABLE IV
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY Y WITH STANDART EXPONENTIAL DISTRIBUTION

Random variable	χ^2 Pearson	ω^2 CMS
$\lvert Laplace(0,1)\rvert$	0.731	0.883
$\lvert Laplace(2,3)-2\rvert/3$	0.749	0.488
$\lvert Laplace(3,5)-3\rvert/5$	0.973	0.845
$\lvert Laplace(10,7)-10\rvert/7$	0.103	0.266
Random variable	Kolmogorov	Ω^2 AD
$\lvert Laplace(0,1)\rvert$	0.925	0.907
$\lvert Laplace(2,3)-2\rvert/3$	0.559	0.639
$\lvert Laplace(3,5)-3\rvert/5$	0.828	0.820
$\lvert Laplace(20,10)-20\rvert/10$	0.219	0.349

Example 5. $Y=\mu-\sigma\log\frac{e^{-X}}{1-e^{-X}}$, где $X\in Exp(0,1)$. Theoretic law describing distribution Y is logistic distribution with shift parameter μ and scale parameter σ. Table 5 shown the values of p_{value} under goodness-of-fit testing $F_N(y)$ with logistic distribution law.

TABLE V
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY Y WITH LOGISTIC DISTRIBUTION

μ	σ	Y	χ^2 Pearson	ω^2 CMS
0	1	$Log(0,1)$	0.786	0.880
-1	0.5	$Log(-1,0.5)$	0.292	0.395
5	3	$Log(5,3)$	0.528	0.521
10	0.5	$Log(10,0.5)$	0.641	0.539
2.5	50	$Log(2.5,50)$	0.973	0.933
μ	λ	Y	Kolmogorov	Ω^2 AD
0	1	$Log(0,1)$	0.837	0.820
-1	0.5	$Log(-1,0.5)$	0.541	0.358
5	3	$Log(5,3)$	0.471	0.657
10	0.5	$Log(10,0.5)$	0.764	0.512
2.5	50	$Log(2.5,50)$	0.956	0.914

Now we consider a few examples with function from two random normal distributing variables.

Example 6. Let be $Y=X_1/X_2$, where $X_1\in N(0,\sigma_1)$ and $X_2\in N(0,\sigma_2)$ are independent. Theoretic law describing distribution Y is Cauchy distribution with null shift parameter. Received results of goodness-of-fit testing of empirical distribution $F_N(y)$ of quantity Y with

TABLE VI
THE RESULTS OF GOODNESS-OF-FIT TESTING THE QUANTITY $Y=X_1/X_2$ WITH CAUCHY DISTRIBUTION

X_1	X_2	Y	χ^2 Pearson	ω^2 CMS
$N(0,1)$	$N(0,1)$	$C(0,1)$	0.884	0.841
$N(0,1)$	$N(0,2)$	$C(0,0.5)$	0.804	0.962
$N(0,2)$	$N(0,1)$	$C(0,2)$	0.878	0.638
$N(0,3)$	$N(0,7)$	$C(0,0.428)$	0.828	0.829
$N(0,1.5)$	$N(0,10)$	$C(0,0.15)$	0.386	0.162
$N(0,10)$	$N(0,2)$	$C(0,5)$	0.659	0.966
$N(0,1)$	$N(0,20)$	$C(0,0.05)$	0.980	0.920

TABLE VI(CONTINUED)

Y	Y	Y	Kolmogo-rov	Ω^2 AD
$N(0,1)$	$N(0,1)$	$C(0,1)$	0.828	0.660
$N(0,1)$	$N(0,2)$	$C(0,0.5)$	0.982	0.142
$N(0,2)$	$N(0,1)$	$C(0,2)$	0.537	0.706
$N(0,3)$	$N(0,7)$	$C(0,0.428)$	0.804	0.902
$N(0,1.5)$	$N(0,10)$	$C(0,0.15)$	0.235	0.202
$N(0,10)$	$N(0,2)$	$C(0,5)$	0.974	0.965
$N(0,1)$	$N(0,20)$	$C(0,0.05)$	0.811	0.950

Values of p_{value} for all tests indicate very well goodness of obtained in result of empirical distribution simulation with Cauchy distribution.

Example 7. $Y=X_1/X_2$, where $X_1, X_2\in N(1,1)$ are independent. In this case the distribution law Y is not Cauchy distribution. The estimation of parameters of Cauchy with density $f(y)=C(\theta_1,\theta_2)=\frac{\theta_1}{\pi\left(\theta_1^2+(y-\theta_2)^2\right)}$ by modeling samples gives the maximum likelihood estimate of scale parameter equals $\theta_1=0.6962$ and shift parameter equals $\theta_2=0.7536$. Empirical distribution $F_N(y)$ received in result of modeling and approximated distribution Cauchy are demonstrated in figure 1.

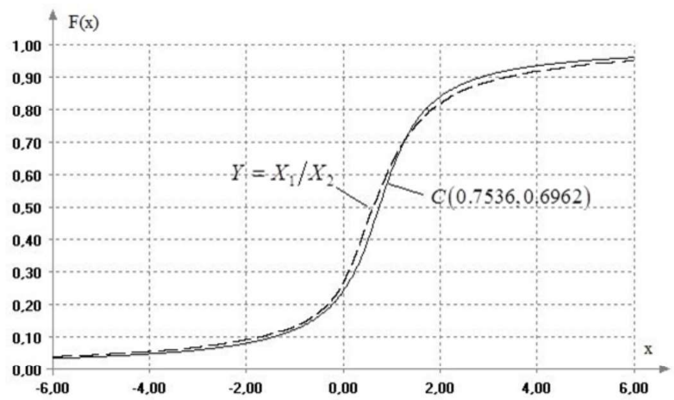


Fig. 1 – The distribution of $Y=X_1/X_2$ where $X_1, X_2\in N(1,1)$

Example 8. It is more general case, when $Y = X_1/X_2$, where $X_1 \in N(\mu_1,\sigma_1)$ and $X_2 \in N(\mu_2,\sigma_2)$ independent. Table 7 demonstrates the results of goodness-of-fit testing of modeling samples with selected distribution laws. The distribution $F_N(y)$ tends to normal distribution law under equality of dispersions and with growth of absolute value μ_2 compared with μ_1. You can see that normal distribution law is becoming good model for random variable Y in case of significant excess of parameter μ_2 over μ_1 ($\mu_1 < \mu_2$)

TABLE VII
THE GOODNESS-OF-FIT TESTING THE QUANTITY $Y = X_1 / X_2$ WITH NORMAL DISTIBUTION

Random variable	Tested distribution law	χ^2 Pearson	ω^2 CMS
$N(1,1)/N(10,1)$	$N(0.1,0.1)$	0	0
$N(1,1)/N(25,1)$	$N(0.04,0.04)$	0.001	0.067
$N(1,10)/N(25,1)$	$N(0.04,0.4)$	0.228	0.881
$N(1,1)/N(25,4)$	$N(0.04,0.04)$	0	0
$N(1,1)/N(37,1)$	$N(0.027,0.027)$	0.050	0.196
$N(1,1)/N(50,1)$	$N(0.02,0.02)$	0.617	0.699
Random variable	Tested distribution law	Kolmogorov	Ω^2 AD
$N(1,1)/N(10,1)$	$N(0.1,0.1)$	0	0
$N(1,1)/N(25,1)$	$N(0.04,0.04)$	0.054	0.004
$N(1,10)/N(25,1)$	$N(0.04,0.4)$	0.876	0.379
$N(1,1)/N(25,4)$	$N(0.04,0.04)$	0	0
$N(1,1)/N(37,1)$	$N(0.027,0.027)$	0.218	0.077
$N(1,1)/N(50,1)$	$N(0.02,0.02)$	0.818	0.683

The empirical distributions has more heavy right tail when there is slight excess of μ_2 over μ_1. Y start to deviates from normal distribution law with increase dispersion of X_2. Under these conditions the distribution Y better approximates by normal distribution law with increase dispersion of X_1 towards dispersion of X_2.

The figure 2 represents $F(y)$ in case of significant excess of absolute value μ_1 over μ_2.

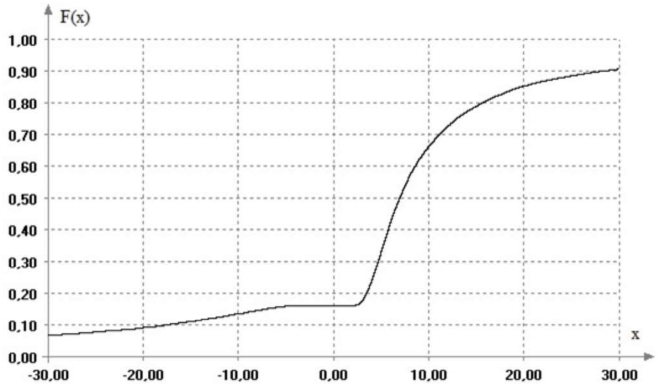


Fig. 2 – The empirical distribution of Y, $X_1 \in N(10,1)$, $X_2 \in N(1,1)$

If standard deviation $\sqrt{D[X_i]}$ much less than mathematic expectation $E[X_i]$ and distributions X_i close to normal law (such situations don't rare), then distribution $Y = X_1/X_2$ is approximated well by normal law. In this case, the function of linearization $\varphi(\cdot)$ often used in practice does not provide to large inaccuracies.

IV. The study of random variables function convergence

Example 9. There is interrelation between the distribution of Fisher and distribution χ^2

$$d_1 * F(d_1,d_2) \to \chi^2(d_1) \text{ under } d_2 \to \infty$$

where d_1 and d_2 are degrees of freedom of distributions. We study the convergence of this form in example of distribution of random variable $4*F(4,d_2)$ under sample size equals 1 660 000. The Kolmogorov differences between theoretical distribution $\chi^2(d_1)$ and empirical distribution $d_1 * F(d_1,d_2)$ are shown in table 8 and figure 3.

Obviously, that distribution of given quantity actually convergences to distribution χ^2, and the trend line confirms that. In this case we can propose that goodness-of-fit hypothesis of $4*F(4,d_2)$ with $\chi^2(4)$ will not deviates under $d_2 > 218$.

TABLE VIII
THE DIFFERENCES BETWEEN DISTRIBUTION OF RANDOM VARIABLE $4*F(4,d_2)$ AND DISTIBUTION $\chi^2(4)$

d_2	difference	d_2	difference
1	0.909795488	20	0.086142562
2	0.735759456	40	0.04559945
3	0.560978724	80	0.022972802
4	0.435409205	120	0.015452631
8	0.223739526	180	0.010450386
10	0.179828623	199	0.009542026

You can see the degree of closing between the theoretical and the empirical distributions in depend on n in figure 4.

Example 10. The interrelation between beta distribution and gamma distribution has the next form:

$$n * B(k,n,1,0) \to \Gamma(k,1,0) \text{ under } n \to \infty,$$

where k and n are shape parameters of distributions. This convergence is studied in example of distribution of random variable $n * B(2,n,1,0)$ under sample size equals 1660000. Table 9 contains the differences between theoretical and empirical distributions.

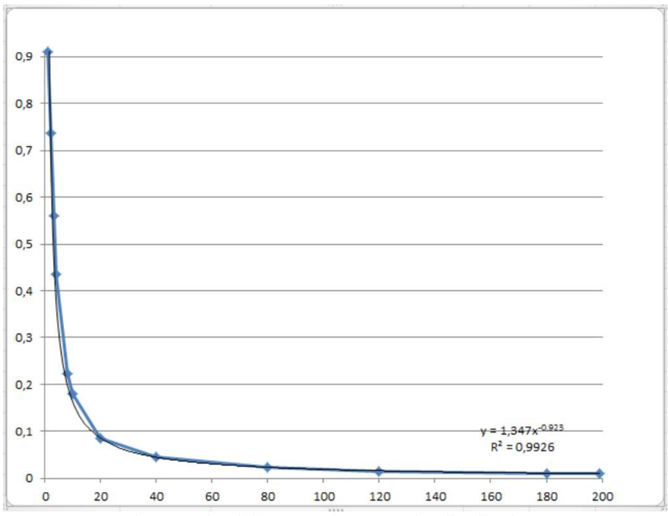


Fig. 3 – The differences between the distributions

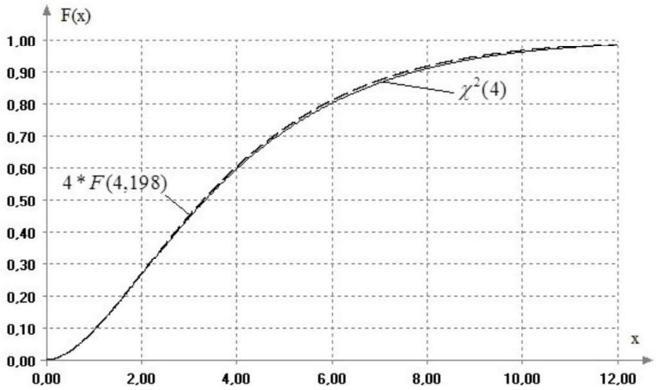


Fig. 4 – The distribution of $4*F(4,198)$

TABLE IX
THE DIFFERENCES BETWEEN DISTRIBUTION OF RANDOM VARIABLE $n*B(2,n,1,0)$ AND DISTIBUTION $\Gamma(2,1,0)$

n	difference	n	difference
1	0,735759456	20	0,036542575
2	0,435409205	36	0,02548311
3	0,296347784	50	0,018458541
4	0,223739526	64	0,014513949
5	0,179828623	75	0,012450055
8	0,112964985	80	0,011699978
10	0,090685139	90	0,010450386
16	0,056916171	99	0,009542026

The trend line for distribution convergence in depend on n is demonstrate in figure 5.

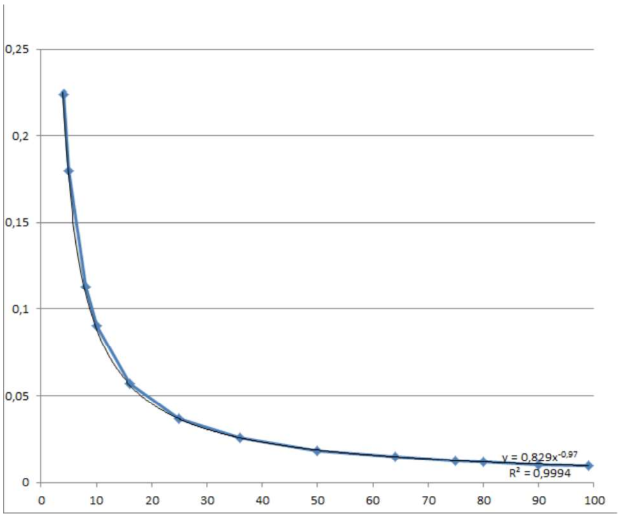


Fig. 5 – The differences between the distributions

In this case the goodness-of-fit hypothesis of $4*F(4,d_2)$ with $\chi^2(4)$ will not deviates under $d_2 > 218$ as predicted.

Also this property performed for special case of given interrelation between gamma-distribution and exponential distribution. In this case, distributions of random variable $n*(1-U(0,1)\wedge(1/n))$, where $U(0,1)$ if uniform on interval (0,1) random variable, must convergences to exponential distribution. The results of testing are demonstrated in table 10. Figure 6 shows distribution of random variable under $n = 500$.

TABLE X
GOODNESS-OF-FIT TESTING OF THE QUANTITY $n*(1-U(0,1)\wedge(1/n))$ WITH EXPONENTIAL DISTRIBUTION

n	χ^2 Pearson	Kolmogorov
100	0	0
250	0.000003	0.032
500	0.014	0.524
750	0.095	0.561
1000	0.749	0.552
1250	0.894	0.971
n	ω^2 CMS	Ω^2 AD
100	0	0
250	0.034	0.004
500	0.362	0.171
750	0.649	0.378
1000	0.878	0.942
1250	0.802	0.822

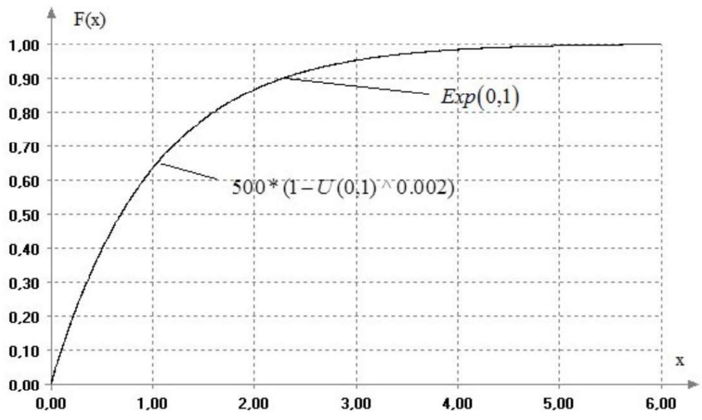


Fig. 6 – The distribution of $n*(1-U(0,1)\,^\wedge(1/n))$ under $n=500$

The results of research are shown advanced features of developed subsystem under investigation of distribution laws of uncertainty of measurements and constructions of its model.

V. The description of user interface

The program ensuring allowed to simulate the samples of random variables functions is elaborated for investigating of distribution laws functions of random variables. The interface allows setting the arbitrary functions of random variables distributed by voluntary different one-dimensional distribution laws. The inlay "Modeling of functions of random variables" is shown in figure 7.

The description of user interface:

1 – is module of choice of distributions from distribution laws loaded in system. The normal distribution with one scale and null shift is chosen in figure 7;

2 – is module of choice of random variables from non-used yet variables;

3 – is module contained the chosen random variables. The relations of these variables with distribution ID are given in module in figure 7 (ID is not shown to users);

4 – are buttons for adding and deleting of random variables in module 3;

5 – is module contained the information about variables, such as its distribution including its parameters (copy from module 1) and the initial value of random-number generator, which you can change ;

6 – is button for loading of distributions type in module 1;

7 – is button for change of distribution parameters. The click by this button creates the form demonstrated in figure 8;

8 – is module for change of modeling parameters, contained the number of random variables created in module 2 and size of modeling samples;

9 – is module for entry of modeling mathematical expression. The operations accessed in this module will present bellow;

10 – is button for start of modeling function of random variables. In case of incorrect entry of expression in module 9 will be shown the error and modeling don't start;

11 – is module for entry of file name for saving of result sample. The file name will be generated by default after the ending of modeling according from entry of modeling expression;

12 – is button for confirmation of saving already simulated sample in file, which name selected in module 11.

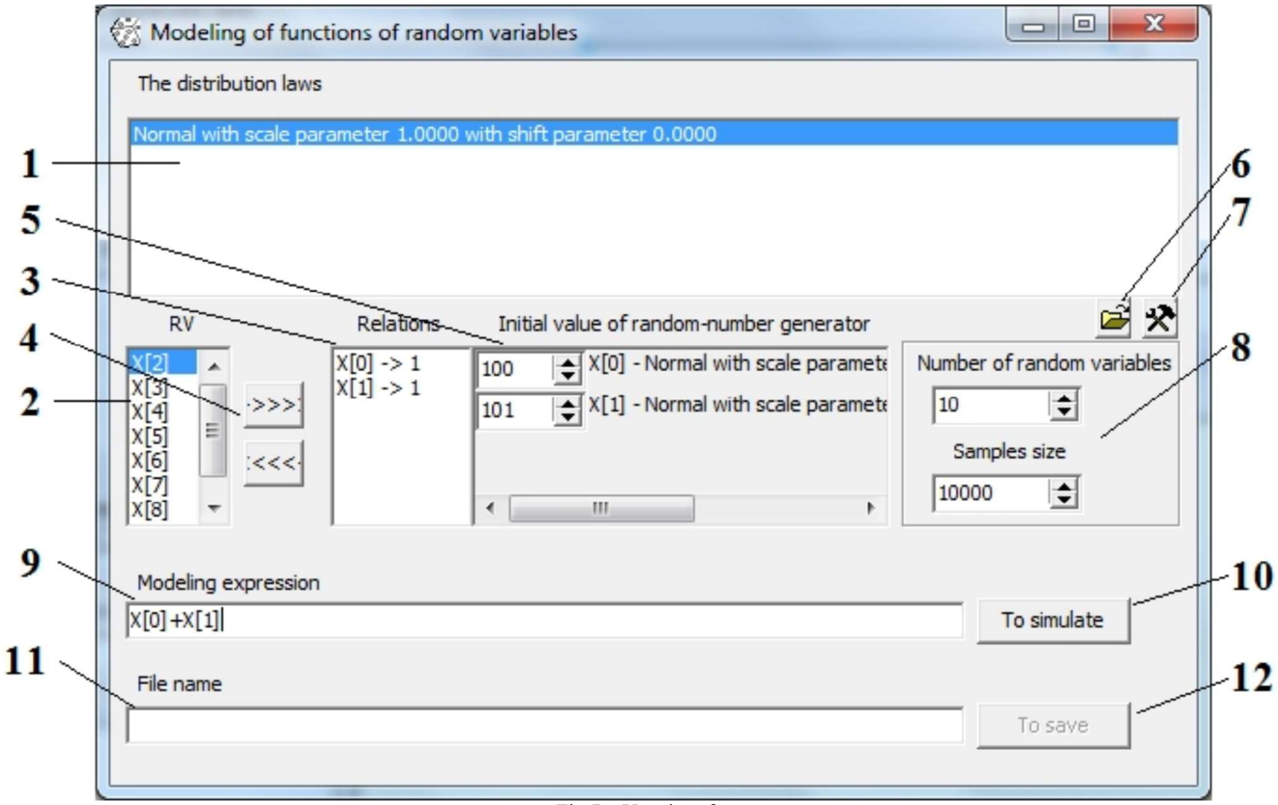


Fig.7 – User interface

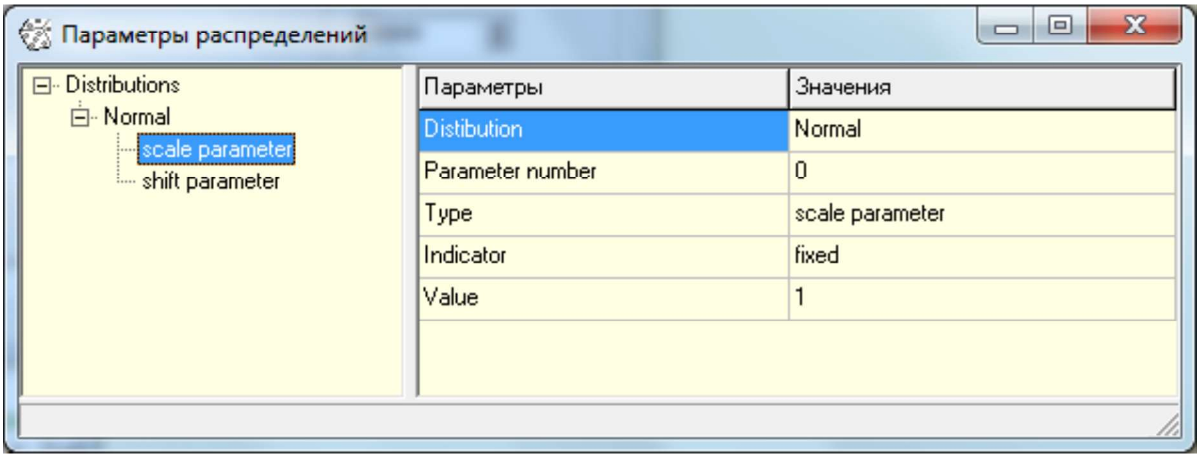


Fig. 8 – The “distribution parameters” form

VI. Conclusions

The software intended for investigating of distribution laws of probability of functions of set of independent random variables had been realized in accordance with the objectives of research.

The developed program allows: to simulate the samples of pseudorandom variables distributed by different distribution laws with preset parameters and initial value of random-number generator; to simulate the samples of values of functions of set of independent random variables.

The form of independent random variables functions can be “arbitrary” and it is set by user in dialog mode.

The following operations are accessible in program system:

- binary operations: addition (+), subtraction (–), multiplication (*), division (/), division modulo (%), involution (^);
- unary operations: unary minus, rootsquaring (sqrt), modulo (abs), trigonometric operations (sin, cos, tg, ctg, arcsin, arccos, arctg, arcctg, sh, ch, th, cth, exp), logarithmic operations (lg, ln);
- n-ary operations: finding the minimum (min), finding the maximum (max), finding the average (avg) and sum (sum) of many random variables.
- Also well-known constants e (e) and π (pi) are accessible in system.

In addition to the simple simulated function, which the entry is shown in Figure 7, we give examples of more complex entries for n-ary operations:

- avg(X[i(0)],i(0)=0:10) – average value of 11 elements of X with indexes 0 to 10 inclusively;
- min(X[i(0)],i(0)=0:49) – minimum among the 50 elements of X with indexes 0 to 49 inclusively;
- syntax for maximum finding is similar;
- Sum(X[(i(0)],i(0)=0:24)+Sum(X[i(1)],i(1);=50:75) – the finding the sum of X with indexes 0 to 24 and от 50 to 75 inclusively.

The application results show that software is effective instrument for investigating the distribution laws of different functions of random variables set. A wide set of instruments allows to simulate the distribution laws of quite complex dependencies interested for applications.

It is shown that methods of statistical modeling together with software allowing to construct the approximate mathematical models for obtained empirical distributions (including in the form of mixture of different parametric laws), represents effective instrument for investigating the distribution laws of functions of random variables and investigating the probabilistic laws.

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

References

[1] Lemeshko B.Y., Lemeshko S.B., Postovalov S.N., and Chimitova E.V. Statistical data analysis, simulation and study of probability regularities. Computer approach: monograph. Novosibirsk: NSTU Publishing House, 2011. 888 pp. (in Russian).

[2] Lemeshko B.Yu., Ogurtsov D.V. Statistical modeling as an effective instrument for investigating the distribution laws of functions of random variables // Metrology. 2007. – № 5. – P. 3-13. (in Russian)

[3] Lemeshko B.Yu., Ogurtsov D.V. Statistical modeling as an effective instrument for investigating the distribution laws of functions of random quantities // Measurement Techniques, 2007. V.50, № 6. – P. 593-600.

[4] ISW – Program system of statistical analysis of one-dimensional random variables. URL: http://ami.nstu.ru/~headrd/ISW. html

Pavel Yurievich Blinov

Undergraduate, Novosibirsk State Technical University, Department of Applied Mathematics

Boris Yurievich Lemeshko

Professor of the Department of Applied Mathematics of NSTU, Doctor of technical sciences