COMPARISON METHODS OF MODELING CONTINUOUS RANDOM VARIABLES ON EMPIRICAL DISTRIBUTIONS

Farhadzadeh E.M., Farzaliyev Y.Z., Muradaliyev A.Z.

•

Azerbaijan Scientific-Research and Design-Prospecting
Institute of Energetic AZ1012, Ave. H.Zardabi-94,
e-mail:fem1939@rambler.ru

ABSTRACT

The new method of modeling of continuous random variables on empirical distributions offered. It shown, that discrepancy of accuracy of methods to shown requirements is shown at small number of realizations of random variables, reduced to not casual divergence of estimations of averages and average quadratic values empirical given and modeled samples.

STATEMENT OF THE PROBLEM.

One of the basic stages of imitating modeling is formation of random variables and casual events with the set law of distribution. In conditions of electro power systems (EPS) examples of random variables are: duration of emergency repair of the equipment and devices, intervals of time between non-working conditions of power units, the maintenance of soluble gases in transformer oil, etc. Casual events: short circuits on transmission lines, refusal in switching-off of the switch, false work of relay protection or automatics, etc. the Analytical form of laws of distribution here is in most cases unknown. Laws of change of a continuous random variable set by statistical (empirical) function of distribution (s.f.d.), and a discrete random variable – proceeding from those or other assumptions of probability of occurrence of casual event. This feature brings the certain interrelation between number of intervals s.f.d. $F^*(X)$, and number of intervals m at discrete representation of continuous empirical function of distribution $F(X)$. If for $F(X)$ the number of intervals m gets out equal $(10\div20)$, for $F^*(X)$ $m=n$.

Objectivity of imitating modeling in many respects depends on that, how much realizations of modeled random variables (events) will appear casual and will reflect the set laws of distribution. It is necessary to note also, that in practice often aspire to present set of statistical data one of known laws of distribution. Actually, the law of distribution of the statistical data concerning a class multivariate represents an uncertain composition of many distributions. In other words, difficulties of representation observable s.f.d. objective analytical law in many respects increase.

Methods of statistical modeling. By development of these methods, the greatest attention given a condition when the type of function of distribution of continuous random variable X known. Statistical modeling on empirical distribution is carried out by two methods. According [1] s.f.d. represented the following equations:

$$F_1^*(X)=\begin{cases}0 & \text{если } X=X_0=0\\ \dfrac{1}{n}+\dfrac{(X-X_i)}{n(X_{i+1}-X_i)} & \text{если } X_0<X<X_n\\ 1 & \text{если } X\geq X_т\end{cases} \qquad (1)$$

where i=0,(n-1)

If to designate realization of a random variable with uniform distribution in an interval [0,1] through ξ, that according to (1) calculation corresponding ξ realizations of random variable X it is carried out under the formula:

$$X=X_i+(X_{i+1}-X_i)\cdot(\xi\cdot n-i) \qquad (2)$$

where i=0,(n-1)

Intuitively clearly, that if the divergence (X_n-X_1) is commensurable with X_1, modeling s.f.d. $F_1^{**}(X)$ under the formula (2) leads to regular distinction $F_1^*(X)$ and $F_1^{**}(X)$. This distinction shown in following parities of averages (accordingly $M_1^*(X)$ and $M_1^{**}(X)$) and average quadratic (accordingly $G_1^*(X)$ and $G_1^{**}(X)$) values of random variable X:

$$\begin{aligned}&M_1^*(X)>M_1^{**}(X)\\&G_1^*(X)<G_1^{**}(X)\end{aligned} \qquad (3)$$

Graphic illustration of this method is resulted on fig.1a.

In the second method [2] s.f.d. represented the following equation:

$$F_2^*(X)=\begin{cases}0 & \text{если } X<X_1\\ \dfrac{i-1}{n-1}+\dfrac{(X-X_i)}{(n-1)\cdot(X_{i+1}-X_i)} & \text{если } X_1\leq X<X_n\\ 1 & \text{если } X\geq X_n\end{cases} \qquad (4)$$

Calculation of realization of random variable X spent under the formula:

$$X=X_i+(X_{i+1}-X_i)\cdot[\xi\cdot(n-1)-(i-1)] \qquad (5)$$

where i=1,(n-1)

In [2] it is marked, that obvious lack of this method is modeling random variable X in interval $X_1<X<X_n$, in other words, size X never can be less X_1 and more X_n, that the brings the certain error of an estimation $M_2^{**}(X)$. Graphic illustration $F_2^*(X)$ and components of the formula (5) is resulted on fig.1в.

Features of calculation under formulas (2) and (5) have caused expediency of specification of these methods of modeling. S.f.d. recommends presenting the following the equation [4]:

$$F_3^*(X)=\begin{cases}0 & \text{если } X\leq X_1\\ \dfrac{i-1}{n+1}+\dfrac{(X-X_i)}{(n+1)\cdot(X_{i+1}-X_i)} & \text{если } X_1<X<X_{n+1}\\ 1 & \text{если } X\geq X_{n+2}\end{cases} \qquad (6)$$

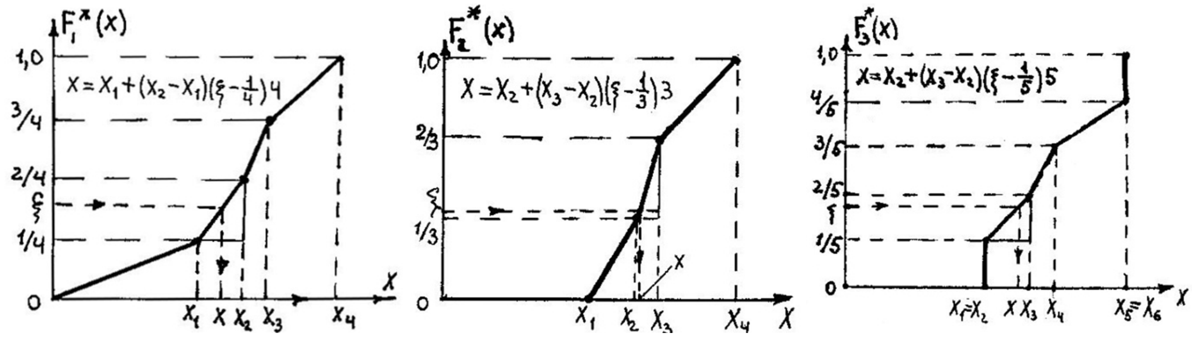
where i=1,(n+1)

Thus, calculation of realization of random variable X carried out under the formula:

$$X=X_i+(X_{i+1}-X_i)\cdot[\xi\cdot(n+1)-(i-1)] \qquad (7)$$

where i=1,(n+1)

The graphic illustration of components $F_3^*(X)$ is resulted on fig.1c



a) b) c)

Fig.1. Illustration methods of modeling continuous random variables on empirical distribution. a - method [1]; b – method [2]; c – method of authors

Algorithm of comparison of methods of statistical modeling. The basic requirement shown to methods of statistical modeling, accuracy of conformity of distribution $F_j^{**}(X)$ to initial distribution F*(X), where j=1,3. Most simple way of the control of a degree of such conformity at small values n is comparison of estimations of average values $M_E^*(X)$ and $M_j^{**}(X)$, and also average quadratic values $G_E^*(X)$ and $G_j^{**}(X)$.

The block scheme of modeling algorithm is resulted on fug.2.

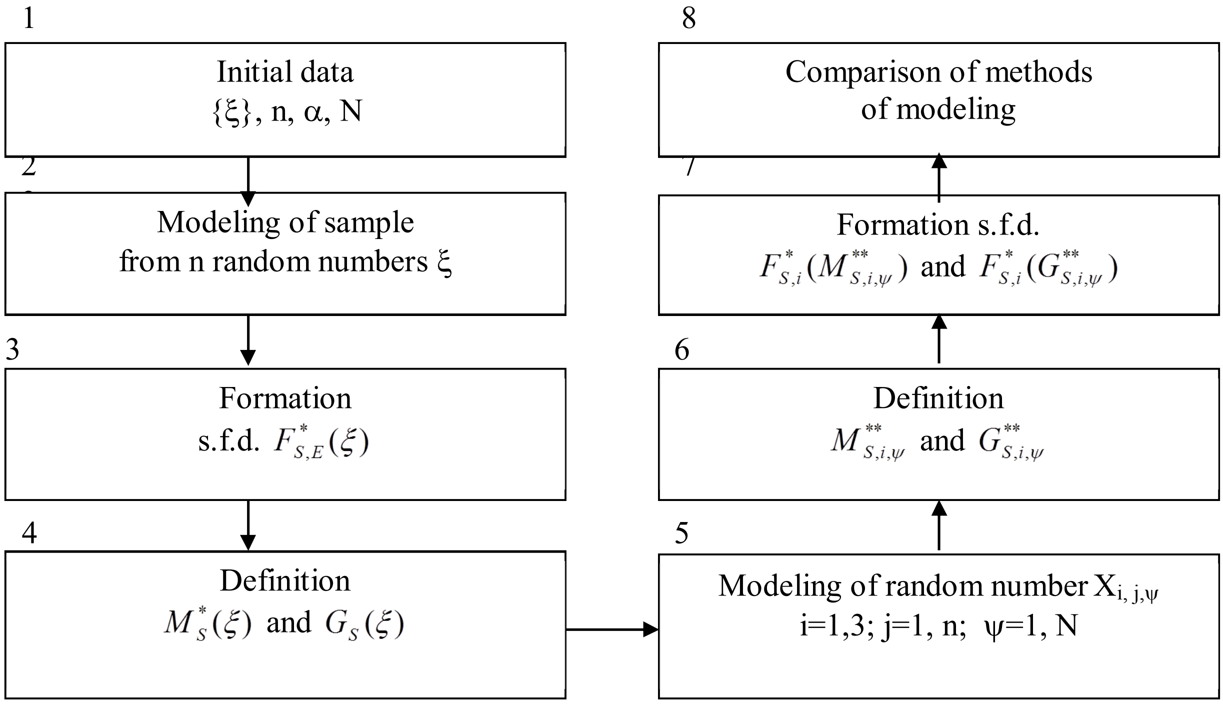


Fig 2. The integrated block diagram of algorithm of comparison of methods of modeling of continuous random numbers

Let's consider features of this algorithm by way of numbering blocks of its block diagram (fig.2.)

1. Initial data are:

− set of pseudo-random numbers {ξ} with uniform distribution in an interval [0,1];

- n – number of random numbers ξ in sample from $\{\xi\}$. Change n allows to establish its influence on result of comparison of methods of modeling of random variables X;
- α - significance value. Allows to estimate influence of a degree of conformity s.f.d. F*(ξ) to the uniform law on result of comparison of methods of modeling of random variables X;
- N – number of imitations of modeled sample $\{X\}_n$

2. Under program RAND, (ξ) is formed n pseudo-random numbers ξ, corresponding the uniform law of distribution in an interval [0,1];
3. Considering, that X=ξ, average M*(X) and an average quadratic G*(X) values on sample is calculated ξ.
4. Under formulas (1), (4) both of (6) and sample $\{\xi\}_n$ are formed s.f.d. $F_1^*(X)$, $F_2^*(X)$ and $F_3^*(X)$;
5. Sample from n pseudo-random numbers with an opportunity of the control of conformity of distribution F*(ξ) is formed to the uniform law with the set significance value α. The method of the general random numbers forms three samples from n random numbers X on distributions $F_1^*(X)$, $F_2^*(X)$ and $F_3^*(X)$. Calculations are spent for significance values (a errors I type) α, and number of realizations of sample N=1000;
6. Estimations of an average $M_{i,\psi}^*(X)$ and average quadratic $G_{i,\psi}^*(X)$ values of modeled random variables on i- th to a method are calculated for ψ- th samples with i=1,3 and ψ=1,N;
7. Are formed s.f.d. $F_i^*[M^*(X)]$ and $F_i^*[G^*(X)]$ for each of three methods i=1,3;
8. Comparison of methods is carried out by comparison M*(X) and G*(X) with similar parameters of distributions $F_i^*[M^*(X)]$ and $F_i^*[G^*(X)]$, i.e. with $M_i^{**}(X) = M_i^*[M_{i,\Psi}^*(X)]$ and $G_i^{**}(X) = M_i^*[G_{i,\Psi}^*(X)]$ i=1,3. Advantage is given a method for which the deviation from $M_E^*(X)$ and $G_E^*(X)$ is minimal

RESULTS OF CALCULATIONS

It is established:

1. Influence of a method of modeling on accuracy of reproduction of distribution F*(ξ) it is shown only for small n. Already at n≥20 divergence between M*(X) and $M_i^{**}(X)$, as well as G*(X) and $G_i^{**}(X)$ with i=1,3 does not exceed 1%. Notice, that at n=4 the divergence between M*(X) and $M_i^{**}(X)$ makes 12%, and between G*(X) and $G_i^{**}(X)$ makes 28.5%;
2. The size of a divergence including the greatest, between F*(X_j) and $F_i^{**}(X_j)$ (designate this size as St_j) does not depend on law of change F*(X_j) and $F_i^{**}(X_j)$, and depends on random variables of sample $\{\xi\}_n$, their numbers n and a way of modeling i=1,3. As an example on fig.3 the graphic illustration of independence St_j with j=1,n from type F*(X) is resulted. In it finds reflection known nonparametric character of criterion of the greatest divergence [3]
3. Comparison of methods of modeling shows, that

$$M^*(X) = M_2^{**}(X) = M_3^{**}(X) >> M_1^{**}(X)$$

$$M^*[G_2^*(X)] << G^*(X) = M^*[G_3^*(X)] << M^*[G_1^*(X)]$$

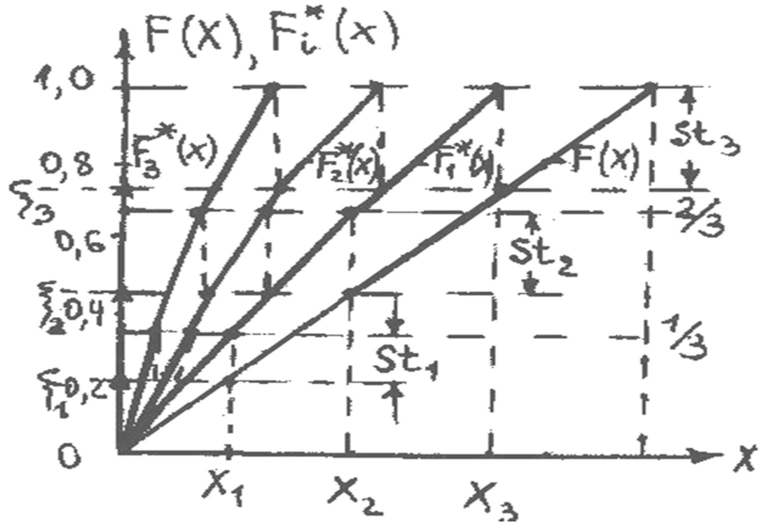
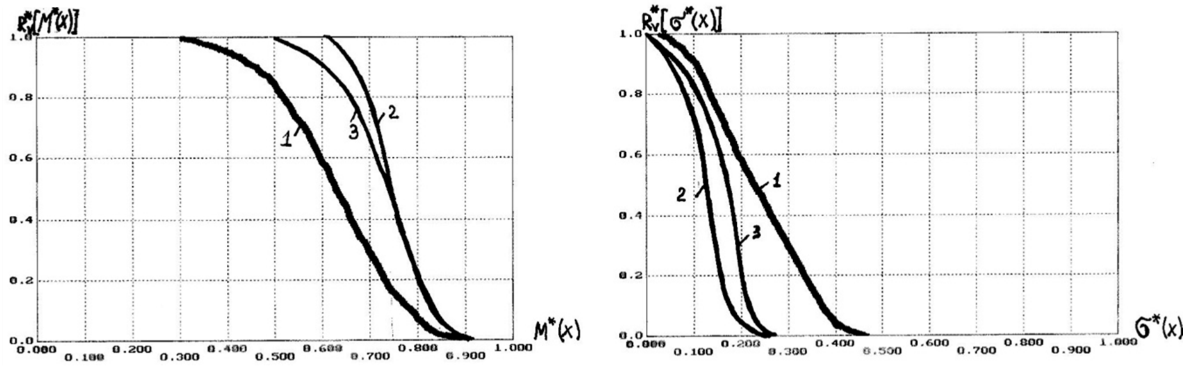


Fig. 3. Graphic illustration of independence St_j with j=1, n from type F*(X)

In other words, the first method does not meet shown requirements to accuracy of calculation both on size $M^*_i(X)$, and on value $G^*_i(X)$. Average values $M^{**}_2(X)$ and $M^{**}_3(X)$, calculated at modeling samples random variables, accordingly, the second (i=2) and the third (i=3) methods, are practically indiscernible and equal to M*(X). However, average quadratic values modeled samples for the second method of modeling $\{X\}_n$ essentially differ from a reference value $G^*(X)$ while the size $M^*[G^*_3(X)]$ practically does not differ from $G^*(X)$.

Graphic illustration of distinction s.f.d. $R^*[M^*_i(X)] = 1 - F^*\left[M^*_i(X)\right]$ and $R^*_i[(G^*_i(X)] = 1 - F^*\left[G^*_i(X)\right]$ for various methods (i=1÷3) and α=0 it is resulted on fig.4.



a) b)

Fig.4. Illustration of distinction s.f.d. averages (a) and averages quadratic (b) values of realizations выборок modeled i=1,2 и3 methods

4. With increase α:

- Average value $M^*_i[M^*_i(X)]$ with i=1,3 i.e. for each method of modeling aspires to the true value and allows to compare with methods more full. At α=0.8 following values are received: $M^*_1[M^*_1(X)] = 0.629$, $M^*_2[M^*_2(X)] = 0.751$ and $M^*_3[M^*_3(X)] = 0.738$ at $M^*(X) = 0.739$;
- The disorder of realizations $M^*_i(X)$ with i=1,3 decreases. If for α=0 for $M^*_1(X)$ it made $G^*_1[M^*_1(X)] = 0.143$, at α=0.8 size $G^*_1[M^*_1(X)] = 0.066$, i.e. disorder of realization $M^*_1(X)$ decreases in 2,2 times. The same reduction of disorder observed for the second and third methods;
- Distinction value of realizations $G^*_1(X)$, on the average, practically invariable also does not exceed 10% for n=4 and 3% for n=16. In the illustrative purposes on fig. 5 distributions $F^*[M^*_i(X)]$

and $F^*[G_i^*(X)]$ are resulted at α=0.8, parities of considered methods confirming independence from α

- Consequences from increase α are similar to consequences of artificial increase in number of modeled random variables n on size $(1-\alpha)^{-1}$.

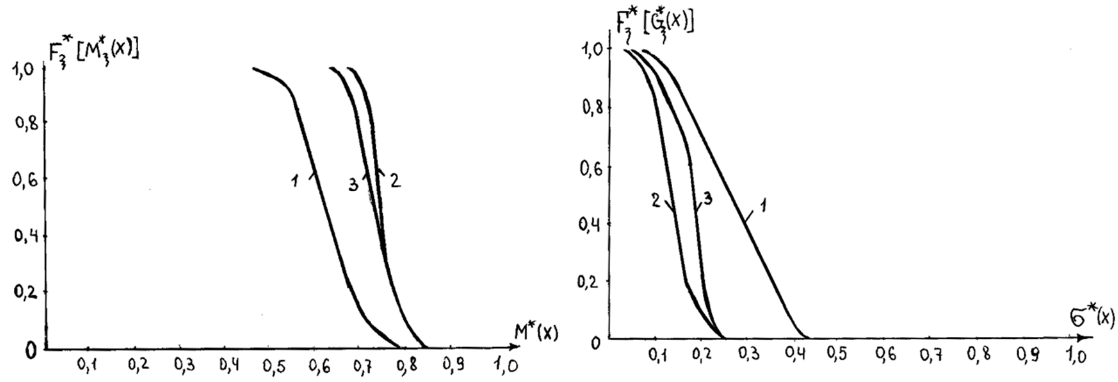


Fig.5. Graphic illustration s.f.d. $F^*[M_i^*(X)]$ and $F^*[G_i^*(X)]$ at α=0.8

CONCLUSION

The lead complex analysis has allowed establishing:

1. Discrepancy of accuracy of methods of modeling of continuous random variables on empirical distributions to shown requirements is shown only at small number of realizations of sample of random variables (n<20)
2. Comparison of methods of modeling can be lead by comparison of modeled estimations of averages and average quadratic values of random variables to empirical values of estimations of these parameters
3. Modeling of continuous random variables on the empirical distribution calculated under the formula (1), at small n leads to essential distinction of averages and average quadratic values of random variables of sample from empirical values, and under the formula (4) – average quadratic values of sample
4. Increase in a significance value α conformity of sample from n pseudo-random numbers to the uniform law on the consequences to similarly artificial increase in number n on size $(1-\alpha)^{-1}$
5. Statistical modeling of random variables on empirical distributions is expedient for spending under the formula (7).

REFERENCE

1. Pole Y.G. Likelihood modeling on electronic computers.–M.: Sov. radio, 1971, 400c.
2. Averill Law, V. Devid's, Kelton M. Imitating modeling. Classics CS,. – SPb.: Peter; Kiev: Publishing group BHV, 2004 847 p.
3. Gnedenko B.V., Beljaev J.K.,Solovyov A.D. Mathematical nightingales methods in the theory of reliability. "Science", 1965, 524 p.
4. Farhadzadeh E.M., Muradaliev A.Z., Rafiyeva N.K., Nazirova U.K. Method of statistical modeling of random variables on empirical distributions. Kazan: News of High schools. Problems of Power №9-10, 2008, 112-120 p