

UDC 519.246

Estimation of the length of the process simulation

Valeriy A. Naumov*, Yuliya V. Gaidamaka^{†‡}, Anna A. Platonova[†]

** Service Innovation Research Institute,
Annankatu 8 A, Helsinki, 00120, Finland*

*† Peoples' Friendship University of Russia (RUDN University),
6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*

*‡ Federal Research Center "Computer Science and Control" (FRC CSC RAS),
44-2 Vavilov St, Moscow, 119333, Russian Federation*

Email: valeriy.naumov@pfu.fi, gaydamaka-yuv@rudn.ru, aaplatonova@list.ru

The paper is devoted to the method of determining the simulation duration sufficient for estimating of unknown parameters of the distribution law for given values of the relative error and the confidence level. The method is developed for nonnegative random variables, for nonnegative regenerating sequence and for nonnegative regenerating piecewise-constant process. An algorithm for simulation a piecewise constant random process is proposed, the result of which is the number of experiments or the duration of the simulation as well as the estimates of unknown parameters obtained as a result of simulation.

The results of numerical experiments illustrating the application of the method for determining the simulation duration are given for a predetermined relative error and confidence level. For the examples, two distributions are chosen from the families of one-parameter and two - from two-parameter distributions. The simulation of random variables with the chosen distributions took into account the simulation duration, sufficient for estimating the unknown parameters of the distribution law for given relative error and confidence level. For chosen distributions, series of experiments were performed at different values of the relative error and the confidence probability.

Key words and phrases: confidence, simulation, estimation of an unknown parameter.

1. Introduction

One way to analyze the characteristics of a random process is to simulate it, when the model of the calculated process is fully reproduced. Simulation simulation relies on Monte Carlo methods - numerical methods for solving mathematical problems by simulation random variables. The most important technique in this case is to bring the problem to the calculation of mathematical expectations [1]. If the mathematical expectation is unknown, the problem arises of estimating it with the required accuracy. Methods are known for obtaining estimates for unknown parameters of the distribution law of a random variable on the basis of a limited number of experiments [2, 3]. In this case, it is necessary to determine the duration of the simulation, sufficient to estimate the unknown parameters of the distribution law for given relative error and confidence level. The method for determining the length of the simulation, based on the central limit theorem, is proposed in [4]. The approach is also applicable to estimating the characteristics of nonnegative regenerating process both in discrete or in continuous time. The method makes it possible to obtain estimates of the characteristics of a random process both in discrete and continuous time, i.e. the characteristics that depend on the number of jumps in states and the characteristics that depend on the time of stay in states. The method is applicable for the case of known values of the distribution characteristics corresponding to the experiments, and for the case of unknown values of the mathematical expectation and variance, for which they are obtained from existing experiments.

In Section 2, the method is described for the case of simulation independent identically distributed random variables. Section 3 presents the simulation algorithm. In Section 4, a numerical experiment was carried out for some of the laws of distribution — the exponential distribution, the Weibull distribution, the Rayleigh distribution, and the Pareto distribution, the characteristics of which are known [5]. For these distributions, for a number of values of the relative error, for several values of the confidence probability, a series of experiments were conducted and estimates of the mathematical expectation \tilde{m} and variance \tilde{D} were obtained as well as a sufficient length of the series of experiments (the duration of the simulation). A numerical experiment illustrates the operation of the method for determining the length of simulation for a given relative error and confidence level [4].

2. Method for determining the length of the simulation

In this section the method for determining the length of the simulation is described according to [2].

Let n be the number of independent experiments on a nonnegative random variable X , on the basis of which we estimate \tilde{m} and \tilde{D} for the mathematical expectation $m > 0$ and variance D :

$$\tilde{m} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

$$\tilde{D} = \frac{\sum_{i=1}^n (x_i - \tilde{m})^2}{n - 1}. \quad (2)$$

Suppose that the quantities $m > 0$ and D are known [2]. Let us find the number of experiments n for which

$$P \left\{ \left| \frac{\tilde{m} - m}{m} \right| < \epsilon \right\} \geq \beta, \quad (3)$$

where ϵ is the required relative error, and β is the confidence probability.

We express the probability on the left-hand side of the inequality in terms of the normal distribution function $\Phi^*(x)$:

$$P \left\{ \frac{|\tilde{m} - m|}{m} < \epsilon \right\} = 2\Phi^* \left(\frac{m\epsilon}{\sigma_{\tilde{m}}} \right) - 1, \tag{4}$$

where $\sigma_{\tilde{m}} = \sqrt{\frac{D}{n}}$ is a standard deviation \tilde{m} . From the inequality

$$2\Phi^* \left(\frac{m\epsilon}{\sigma_{\tilde{m}}} \right) - 1 \geq \beta \tag{5}$$

follows

$$\frac{m\epsilon}{\sigma_{\tilde{m}}} \geq t_\beta, \tag{6}$$

where $t_\beta = \arg\Phi^* \left(\frac{1+\beta}{2} \right)$ and $\arg\Phi^*(x)$ is the inverse of $\Phi^*(x)$ meaning the value of the argument at which the normal distribution function is equal to x .

If mathematical expectation m and variance D are not known, we use their estimates $m = \tilde{m}$ and $\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}}{n}}$.

Thus, the required number n of experiments will be found from the inequality [3]:

$$\tilde{m}\epsilon \geq t_\beta \sqrt{\frac{\tilde{D}}{n}}. \tag{7}$$

Inequality (7) is also applicable in the case when the values x_1, x_2, \dots, x_n form a nonnegative regenerating sequence, i.e. discrete time random process [6].

Denote $V_n = x_1 + x_2 + \dots + x_n$ and $W_n = x_1^2 + x_2^2 + \dots + x_n^2$, then

$$\tilde{m} = \frac{V_n}{n}, \quad \tilde{D} = \frac{W_n}{n} - \left(\frac{V_n}{n} \right)^2. \tag{8}$$

Therefore, the required number of experiments can be found from the inequality

$$t_\beta \sqrt{\frac{1}{n} \left(\frac{W_n}{n} - \left(\frac{V_n}{n} \right)^2 \right)} \leq \frac{V_n}{n} \epsilon, \tag{9}$$

those $W_n \leq (V_n)^2 \left(E + \frac{1}{n} \right)$.

In order to avoid overflow when exponentiation of accumulated amounts, it is more convenient to use its record in the following form:

$$\frac{W_n}{V_n} \leq V_n \left(E + \frac{1}{n} \right), \tag{10}$$

where $E = \left(\frac{\epsilon}{t_\beta} \right)^2$.

The same approach can be applied in case of estimating characteristics of a continuous time random process.

We consider a nonnegative regenerating piecewise-constant process $\mathbf{X}(t)$, which assumes a constant value x_i on the interval $[t_{i-1}, t_i)$, $t_0 = 0 < t_1 < t_2 < \dots < t_{n-1} < t_n = T$ [5], and denote

$$V_T = \sum_{j=1}^n x_j (t_j - t_{j-1}), \quad W_T = \sum_{j=1}^n x_j^2 (t_j - t_{j-1}). \tag{11}$$

The mathematical expectation and variance of this process on the interval $[0, T)$ are estimated by formulas

$$\tilde{m} = \frac{V_T}{T}, \quad \tilde{D} = \frac{W_T}{T} - \left(\frac{V_T}{T} \right)^2, \quad (12)$$

and the required simulation duration of the T is found from the inequality

$$\frac{W_T}{V_T} \leq V_T \left(E + \frac{1}{T} \right). \quad (13)$$

3. Simulation Algorithm

Let $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_L(t))$ be a piecewise constant process with L components, either in discrete or in continuous time.

Let nonnegative functions $F_d^{(i)}(\mathbf{X}(t))$ and $F_c^{(i)}(\mathbf{X}(t))$ be characteristics of the process $\mathbf{X}(t)$ in case of discrete or continuous time correspondingly with the mathematical expectations

$$m_d^{(i)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n F_d^{(i)}(\mathbf{X}(t_k)), \quad i = 1, 2, \dots, N_d, \quad (14)$$

$$m_c^{(i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T F_c^{(i)}(\mathbf{X}(t)) dt, \quad i = 1, 2, \dots, N_c, \quad (15)$$

where N_d and N_c are the corresponding numbers of characteristics.

The goal is to estimate $m_d^{(i)}$ and $m_c^{(i)}$ with predetermined values of relative error ϵ and confidence probability β . In addition the following input data should be determined: T_{min} - duration of transit period before beginning of accumulating statistics; T_{max} - the maximum length of the simulation in case of failure to achieve the specified accuracy; eps - infinitesimal increment in time units; δ - small quantity to avoid zero divide in step **d**.

The algorithm for simulation a piecewise constant process to obtain estimates of required characteristics is given below.

a) Set the duration of the transit period $T_{min} > 0$, the maximum simulation duration $T_{max} > T_{min}$ and determine the initial state of the process $\mathbf{X}(t): \mathbf{X} = \mathbf{X}(t_0)$.

For each characteristic $F_d^{(i)}$ of the process $\mathbf{X}(t)$, we zero out the quantities $V_d^{(i)}$ and $W_d^{(i)}$, $i = 1, 2, \dots, N_d$. For each characteristic $F_c^{(i)}$ of the process $\mathbf{X}(t)$, we zero out the quantities $V_c^{(i)}$ and $W_c^{(i)}$, $i = 1, 2, \dots, N_c$. We set $n = 1$ and $\tau_{next} = 0$.

b) Set $\tau_{last} = \tau_{next}$, and then for each component of the process $X_j(t)$ determine the time of the next jump τ_j .

We determine the instant $\tau_{next} = \min\{\tau_j\}$ of the nearest jump of the process $\mathbf{X}(t)$ and the corresponding component $J_{next} = \operatorname{argmin}\{\tau_j\}$, where $1 \leq J_{next} \leq L$.

If $\tau_{next} \geq T_{max}$, we set $\tau_{next} = T_{max} + eps$.

We change the accumulated data:

$$\begin{aligned} V_d^{(i)} &:= V_d^{(i)} + F_d^{(i)}(\mathbf{X}), \\ W_d^{(i)} &:= W_d^{(i)} + \left(F_d^{(i)}(\mathbf{X}) \right)^2, \quad i = 1, 2, \dots, N_d, \\ V_c^{(i)} &:= V_c^{(i)} + F_c^{(i)}(\mathbf{X}) (\tau_{next} - \tau_{last}), \end{aligned} \quad (16)$$

$$W_c^{(i)} := W_c^{(i)} + \left(F_c^{(i)}(\mathbf{X}) \right)^2 (\tau_{next} - \tau_{last}), \quad i = 1, 2, \dots, N_c. \quad (17)$$

c) If $\tau_{next} < T_{max}$, then we set $n := n + 1$ and find the following state \mathbf{X} of the process $\mathbf{X}(t)$: $\mathbf{X} = \mathbf{X}(\tau_{next})$.

If $\tau_{next} \geq T_{max}$, go to step **f** (the specified accuracy is not achieved).

d) If $\tau_{last} > T_{min}$ and for some values $V_c^{(i)} < \delta$ or $V_d^{(i)} < \delta$, then go to step **b**.

e) If $\tau_{last} > T_{min}$ and for some characteristic $F_d^{(i)}$, we have $\frac{W_d^{(i)}}{V_d^{(i)}} > V_d^{(i)} \left(E + \frac{1}{n} \right)$,

then go to step **b**. If for some characteristic $F_c^{(i)}$, we have $\frac{W_c^{(i)}}{V_c^{(i)}} > V_c^{(i)} \left(E + \frac{1}{\tau_{last}} \right)$, then go to step **b**.

f) We calculate the estimates of the average values of the characteristics

$$\tilde{m}_d^{(i)} = \frac{V_d^{(i)}}{n}, \quad i = 1, 2, \dots, N_d, \quad (18)$$

$$\tilde{m}_c^{(i)} = \frac{V_c^{(i)}}{T_{max}}, \quad i = 1, 2, \dots, N_c. \quad (19)$$

g) End of simulation.

At the end of simulation one obtain the estimates $\tilde{m}_d^{(i)}$ for discrete time process or $\tilde{m}_c^{(i)}$ for continuous time process as well as the number of experiments or simulation duration. In case specified accuracy was not achieved the values $\tilde{m}_d^{(i)}$ and $\tilde{m}_c^{(i)}$ correspond to the estimates during determined period T_{max} .

4. Numerical results

For the simulation, we selected the distributions of non-negative continuous random variables — the exponential distribution and the Rayleigh distribution (one-parameter), the Weibull distribution and the Pareto distribution (two-parameter) [5]. For all experiments three values of relative errors $\epsilon \in \{10^{-2}, 10^{-3}, 10^{-4}\}$ and the confidence probability $\beta=0.95$ are chosen.

Table 1 shows the simulation results for the exponential distribution

$$F_{exp}(x) = 1 - e^{-\lambda x} \quad (20)$$

with the parameter $\lambda=3$.

For several values of the relative error ϵ the number n of experiments necessary to achieve the required accuracy with given confidence level $\beta=0.95$ is indicated, together with the corresponding estimates \tilde{m} for mathematical expectation and \tilde{D} for variance.

Figure 1 and Figure 2 illustrate the convergence of estimates \tilde{m} and \tilde{D} to the exact values

$$m_{exp} = \frac{1}{\lambda}, \quad (21)$$

$$D_{exp} = \frac{1}{\lambda^2} \quad (22)$$

depending on the number of experiments.

In Figure 1 and Figure 2 the curves correspond to the estimate of the mathematical expectation \tilde{m} and variance \tilde{D} , the straight line corresponds to the exact value of the mathematical expectation $m_{exp} \approx 0.333$ (21) and variance $D_{exp} \approx 0.111$ (22).

Table 1
The number n of experiments and numerical characteristics for the exponential distribution with the parameter $\lambda=3$

$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
$n = 372$	$n = 38261 \approx 10^4$	$n = 3840310 \approx 10^6$	$n = 384177934 \approx 10^8$
$\tilde{m} = 0.3$	$\tilde{m} = 0.33$	$\tilde{m} = 0.333$	$\tilde{m} = 0.3333$
$\tilde{D} = 0.1$	$\tilde{D} = 0.11$	$\tilde{D} = 0.111$	$\tilde{D} = 0.1111$

Table 2
The number n of experiments and numerical characteristics for the Weibull distribution with the parameters $\lambda=1$ and $k=5$

$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
$n = 25$	$n = 1939 \approx 10^3$	$n = 201621 \approx 10^5$	$n = 20162100 \approx 10^7$
$\tilde{m} = 0.9$	$\tilde{m} = 0.92$	$\tilde{m} = 0.918$	$\tilde{m} = 0.9181$
$\tilde{D} = 0.05$	$\tilde{D} = 0.04$	$\tilde{D} = 0.044$	$\tilde{D} = 0.0442$

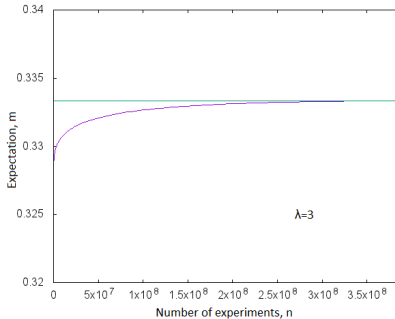


Figure 1. Estimation of the mathematical expectation as a function of the number of experiments for the exponential distribution

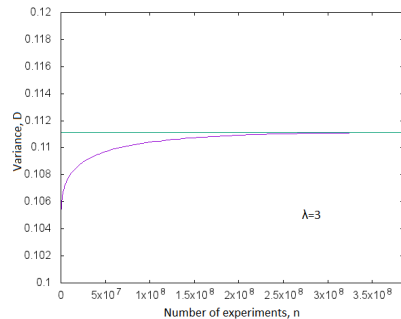


Figure 2. Estimation of the variance as a function of the number of experiments for the exponential distribution

Table 2 shows the simulation results for the Weibull distribution

$$F_{Weibull}(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k} \quad (23)$$

with the parameters $\lambda=1$ and $k=5$.

Figure 3 and Figure 4 illustrate the convergence of estimates \tilde{m} and \tilde{D} to the exact values

$$m_{Weibull} = \lambda \Gamma\left(1 + \frac{1}{k}\right), \quad (24)$$

$$D_{Weibull} = \lambda^2 \Gamma\left(1 + \frac{2}{k}\right) - \mu^2. \quad (25)$$

In Figure 3 and Figure 4 the curves correspond to the simulation estimate of the mathematical expectation \tilde{m} and variance \tilde{D} , the straight line corresponds to the analytical value of the mathematical expectation $m_{Weibull} = 0.9181$ (24) and variance $D_{Weibull} \approx 0.0442$ (25).

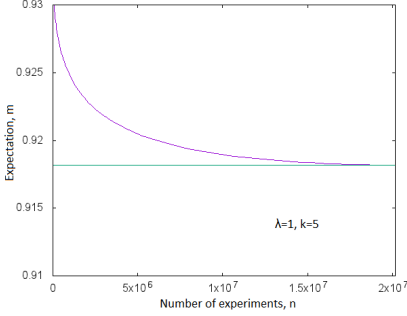


Figure 3. Estimation of the mathematical expectation as a function of the number of experiments for the Weibull distribution

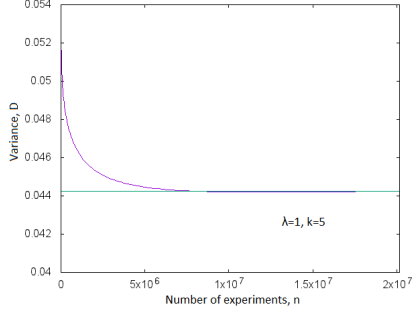


Figure 4. Estimation of the variance as a function of the number of experiments for the Weibull distribution

Figure 5 and Figure 6 for the exponential distribution (20) and for the Weibull distribution (23) show the dependence of the number of experiments on the relative error.

Table 3 shows the simulation results for the Rayleigh distribution

$$F_{Rayleigh}(x) = 1 - e^{-\frac{x}{2s^2}} \tag{26}$$

with the parameter $s=2$.

Figure 7 and Figure 8 illustrate the convergence of estimates \tilde{m} and \tilde{D} to the exact values

$$m_{Rayleigh} = \sqrt{\frac{\pi}{2}}\sigma, \tag{27}$$

$$D_{Rayleigh} = \left(2 - \frac{\pi}{2}\right)\sigma^2. \tag{28}$$

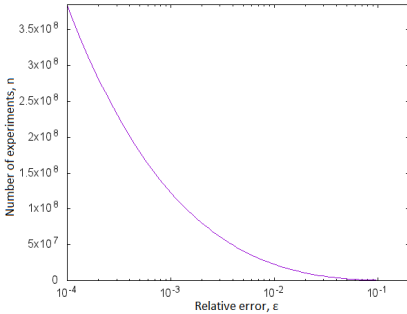


Figure 5. The number of experiments, depending on the relative error for the exponential distribution

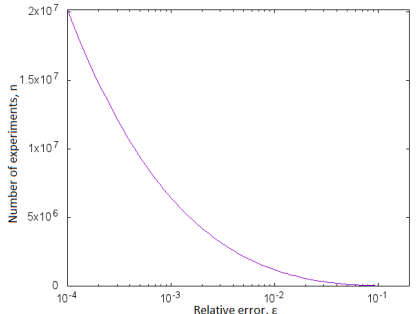


Figure 6. The number of experiments, depending on the relative error for the Weibull distribution

Table 3

The number n of experiments and numerical characteristics for the Rayleigh distribution with the parameter $s=2$

$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
$n = 73$	$n = 10334 \approx 10^4$	$n = 1050831 \approx 10^6$	$n = 110337255 \approx 10^8$
$\tilde{m} = 2.4$	$\tilde{m} = 2.49$	$\tilde{m} = 2.506$	$\tilde{m} = 2.5066$
$\tilde{D} = 1.1$	$\tilde{D} = 1.67$	$\tilde{D} = 1.713$	$\tilde{D} = 1.7168$

Table 4

The number n of experiments and numerical characteristics for the Pareto distribution with the parameters $x_m=2, k=3$

$\epsilon = 10^{-1}$	$\epsilon = 10^{-2}$	$\epsilon = 10^{-3}$	$\epsilon = 10^{-4}$
$n = 68$	$n = 10547 \approx 10^4$	$n = 1222897 \approx 10^6$	$n = 158976610 \approx 10^8$
$\tilde{m} = 1.4$	$\tilde{m} = 1.49$	$\tilde{m} = 1.499$	$\tilde{m} = 1.5$
$\tilde{D} = 0.3$	$\tilde{D} = 0.61$	$\tilde{D} = 0.715$	$\tilde{D} = 0.75$

In Figure 7 and Figure 8 the curves correspond to the estimate of the mathematical expectation \tilde{m} and variance \tilde{D} , the straight line corresponds to the analytical value of the mathematical expectation $m_{Rayleigh} = 2.5066$ (27) and variance $D_{Rayleigh} = 1.7168$ (28).

Table 4 shows the simulation results for the Pareto distribution

$$F_{Pareto}(x) = 1 - \left(\frac{x_m}{x}\right)^k \quad (29)$$

with the parameters $x_m=2, k=3$.

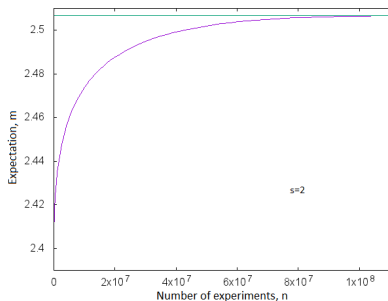


Figure 7. Estimation of the mathematical expectation as a function of the number of experiments for the Rayleigh distribution

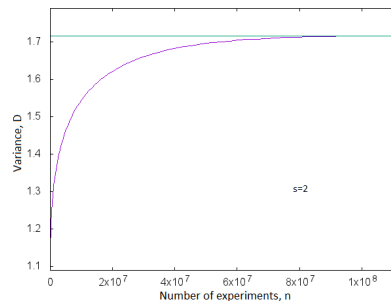


Figure 8. Estimation of the variance as a function of the number of experiments for the Rayleigh distribution

Figure 9 and Figure 10 illustrate the convergence of estimates \tilde{m} and \tilde{D} to the exact values

$$m_{Pareto} = \frac{kx_m}{k-1} \quad (\text{for } k > 1), \quad (30)$$

$$D_{Pareto} = \left(\frac{x_m}{k-1} \right)^2 \frac{k}{k-2} \quad (\text{for } k > 2). \tag{31}$$

In Figure 9 and Figure 10 the curves correspond to the estimate of the mathematical expectation \hat{m} and variance \hat{D} , the straight line corresponds to the exact value of the mathematical expectation $m_{Pareto} = 1.5$ (30) and variance $D_{Pareto} = 0.75$ (31).

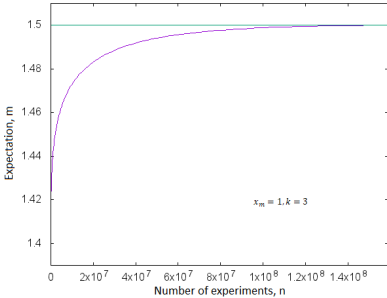


Figure 9. Estimation of the mathematical expectation as a function of the number of experiments for the Pareto distribution

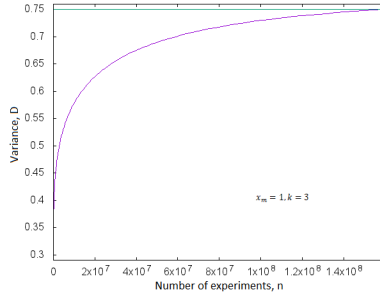


Figure 10. Estimation of variance as a function of the number of experiments for the Pareto distribution

Figure 11 and Figure 12 for the Rayleigh distribution (26) and for the Pareto distribution (29) show the dependence of the number of experiments on the relative error.

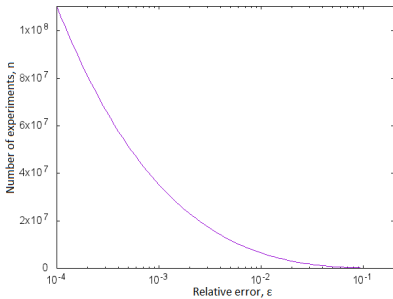


Figure 11. The number of experiments, depending on the relative error for the Rayleigh distribution

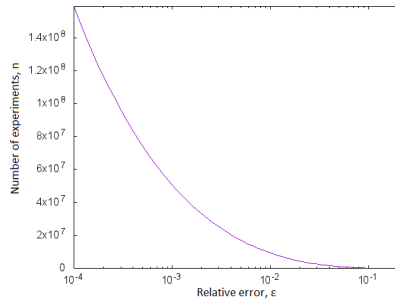


Figure 12. The number of experiments, depending on the relative error for the Pareto distribution

The results obtained in the series of the numerical experiment confirm the conclusion that an increase in the number of experiments increases the accuracy of the estimates of unknown parameters obtained on their basis.

5. Conclusions

In this paper we propose an algorithm for simulation of a regeneration multi-component process. The length of the simulation is specified by the relative errors of the process parameters. The algorithm makes it possible to obtain estimates the characteristics of the process, both in discrete and continuous time, which depend on the number of jumps in the states, and on the characteristics of the stay time in states.

The result of the algorithm is the number of experiments (the number of process jumps) or the duration of the simulation as well as the estimates of unknown parameters obtained as a result of simulation.

The numerical experiment is an illustration of the work of the method for determining the simulation duration, sufficient for estimating the unknown parameters of the distribution law for given relative error and confidence level. The evaluation of the parameters obtained on the basis of a series of experiments, in the case considered - mathematical expectation and variance, was compared with the exact values of these parameters, the analytical form of which for taken distributions is known. We note that the method makes it possible to determine a sufficient length of a series of experiments also for the case when the values of the distribution law parameters are unknown, and estimates for these parameters are obtained on the basis of simulation [2]. This method is planned to be used in simulating random access [7] and adaptive radio access [8] for LTE networks as a development of previous research.

Acknowledgments

The publication has been prepared with the support of the “RUDN University Program 5-100” and funded by RFBR according to the research projects No. 18-07-00576, 19-07-00933. This work has been developed within the framework of the COST Action CA15104, Inclusive Radio Communication Networks for 5G and beyond (IRACON).

References

1. Metropolis N., Ulam S. The Monte Carlo Method // *Journal of the American Statistical Association*, Vol. 44, No. 247 (Sep., 1949).— pp. 335-341.
2. Wentzel E.S., Ovcharov L.A. *Applied problems in probability theory*. — Moscow: Mir.— 1986.
3. Ross S.M. *A first course in probability* / 9th ed. University of Southern California. Upper Saddle River, New Jersey London, Prentice Hall.— 2014.
4. Naumov V.A., Gaidamaka Yu.V. Opredelenie momenta ostanovki modelirovaniya pri zadannyh oshibke i urovne doverija [Determination of the stopping time for simulation with a given error and confidence level]. *Trudy 21 Mezhdunarodnoj konferencii «Raspredelennye komp'yuternye i telekommunikacionnye seti: teorija i prilozhenija» DCCN-2018, 17-21 sentjabrja 2018. Moskva. – 2018. – S.287-292. [Proc. 21st International Conference Distributed Computer and Communication Networks (DCCN 2018), Moscow, Russia, September 17–21, 2018, pp. 287-292].*
5. Johnson N., Kotz S., Balakrishnan N. *Continuous Univariate Distribution, Volume 1*. Wiley, New York.— 1994.
6. Iglehart D. L., Shedler G. S. *Regenerative Simulation of Response Times in Networks of Queues*. Springer-Verlag Berlin Heidelberg, 1980. doi: 10.1007/BFb0044445.
7. Borodakiy V.Y., Samouylov K.E., Gudkova I.A., Markova E.V. Analyzing Mean Bit Rate of Multicast Video Conference in LTE Network with Adaptive Radio Admission Control Scheme // *Journal of Mathematical Sciences (United States)*.— 218 (3).— 2016.— pp. 257-268.
8. Gudkova I., Samouylov K., Buturlin I., Borodakiy V., Gerasimenko M., Galinina O., Andreev S. Analyzing impacts of coexistence between M2M and H2H Communication on 3GPP LTE System // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8458.— 2014.— pp. 162-174.