32nd European Modeling & Simulation Symposium
17th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0029 ISBN 978-88-85741-44-7 © 2020 The Authors.
DOI: 10.46354/i3m.2020.emss.003

The high-speed random number generator with the specified two-dimensional probability distribution

Oleg Chernoyarov[1,2,3], Vladimir Litvinenko[4], Boris Matveev[4], Serguei Dachian[5] and Kirill Melnikov[1,2,*]

[1]National Research University "Moscow Power Engineering Institute", Krasnokazarmennaya st. 14, Moscow, 111250, Russia
[2]National Research Tomsk State University, Lenin Avenue 36, Tomsk, 634050, Russia
[3]Maikop State Technological University, Pervomayskaya st. 191, Maikop, 385000, Russia
[4]Voronezh State Technical University, Moscow Avenue 14, Voronezh, 394026, Russia
[5]University of Lille, 42 rue Paul Duez, Lille, 59000, France

[*]Corresponding author. Email address: kirill.a.melnikov@mail.ru

Abstract

In the paper, a pseudorandom number sequence sensor is considered, its design is based on the Markov model of the simulated process. Such a model is derived from either the theoretical two-dimensional probability density or from the random process samples obtained experimentally. There has been developed a simple high-speed algorithm for operating the sensor using a primary source of pseudorandom numbers with a uniform probability distribution, and statistical simulation of such algorithm has been carried out. It is shown that the obtained sequence of numbers possesses probabilistic and correlation properties that are in good agreement with the specified properties of the simulated random processes. When substituting a hardware random number generator for the source of equiprobable pseudorandom numbers, the sensor generates truly random numbers. The possibilities of the hardware implementation of the introduced algorithm in the form of a pseudorandom (random) number generator are demonstrated.

Keywords: Random-number generator; Markov model; matrix of transition probabilities; probability density; histogram; statistical simulation

1. Introduction

For their solution, various problems of statistical simulation (Knuth, 1997; Law and Kelton, 2000; Bardis et al, 2009), cryptography (Schneier, 1996; Ferguson and Schneier, 2003), statistical radio engineering (Bykov, 1971) need generating a sequence of random or pseudorandom numbers with specified properties. Random (truly random (Ferguson and Schneier, 2003)) numbers are generated using physical sources of random processes with subsequent transformation into the required form (Bardis et al, 2009). A produced random realization cannot be repeated again, and that may be either their advantage, or disadvantage, depending on the problem to be solved. Besides, the physical sensor is subject to environmental influences making the statistical properties of the generated numbers dependent on temperature, humidity, supply

voltage, etc.

Pseudorandom number sensors can be implemented in both in software and in hardware. As a rule, they are built in the form of linear congruent generators (Schneier, 1996) or sensors based on shift registers with linear and nonlinear feedbacks (Bykov, 1971; Knuth, 1997). The most common are the generators producing the equiprobable (in a set range) pseudorandom numbers. They are included in both programming systems (C, Pyton, R) and application packages (MathCAD, MathLab). However, in many practical applications, the sources of pseudorandom numbers with specified statistical properties are also required (Bykov, 1971). That applies particularly to the Gaussian random numbers and the numbers obtained from them by means of nonlinear operations.

The structure of the paper includes the following parts. In Section II, there are presented a brief description of the literature on the topic of the paper as well as the recent studies and their conclusions. Then the research problem that this paper addresses is stated. In Section III, the Markov model of discrete random process is introduced, it can simulate sampling of such process at the specific instants in time. The calculation of the model parameters is demonstrated, for this purpose the set theoretical two-dimensional probability density together with the obtained experimental data are used. In Section IV, the algorithm for simulating the sampling of the random process with the required statistical characteristic is introduced. Its block diagram is produced and the sequence of operations is described. In Section V, there are provided the results of the algorithm operation that both represent and confirm the theoretical results. The realizations of the Gaussian and non-Gaussian random processes with the assumed two-dimensional probability densities are simulated. Finally, in Section VI, the conclusions are drawn about the efficiency of the proposed algorithm and the usefulness of its practical application.

2. State of the art

A review of the sensor development was made by Pierre L'Ecuyer (2017), while more detailed reference information was provided by Kroese, Taimre, and Botev (2011).

In tasks of statistical simulation, it is necessary to generate the sequences of pseudorandom numbers (signals) with the specified one-dimensional and two-dimensional probabilistic properties. For example, in radio physics and radio engineering, the development of the algorithms for simulating the random realizations obeying the Nakagami distribution is relevant (Recommendation ITU-R, 2019). In a number of cases, the random process should be simulated for an available experimental realization of the samples with the unknown probabilistic properties. Today, great opportunities are opened for the development of the methods for simulating random radio signals and their propagation medium.

A high-speed pseudo-random number generation algorithm should not include complex computational transformations of samples. It would also be useful to minimize the number of operations with numbers in binary code.

In order to provide the high sensor speed and generality, the random numbers with the specified distribution law can be generated in accordance with the Markov model of the simulated process (Ventsel and Ovcharov, 2000). It is shown below that such a model can be effectively built in terms of either the theoretical two-dimensional probability density or the experimentally obtained sample of the random process.

3. Markov Model of Discrete Random Process

The discrete random process x_n with a finite number of values can be represented by the process $z_n = i$ with integer values. Here $n = \overline{1,N}$ is the sample number, N is the sample size, $i = \overline{1,M}$ is the number of x_n value, M is the number of the possible values of x_n.

In the Markov model of a random process (simple Markov chain), the probability of the value $z_{n+1} = j$ at the time t_{n+1} depends only on the previous value $z_n = i$ at the time t_n and does not depend on earlier values (Ventsel and Ovcharov, 2000). The Markov chain is described by the transition probability matrix of the form

$$[P_{ij}] = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1M} \\ P_{21} & P_{22} & \dots & P_{2M} \\ \dots & \dots & \dots & \dots \\ P_{M1} & P_{M2} & \dots & P_{MM} \end{bmatrix} \quad (1)$$

where P_{ij}, $i,j = \overline{1,M}$ is the probability of the process transition from the value $z_n = i$ to the value $z_{n+1} = j$. The probabilities q_i of the initial values $z_1 = i$ are determined by the matrix

$$[q_i] = [q_1 \quad q_2 \quad \dots \quad q_M]^T \quad (2)$$

where the symbol "T" denotes the matrix transposition.

The time-discrete random process with continuous values x_n is uniformly quantized by an analog-to-digital converter in terms of the amplitude thresholds

$$g_m = \begin{cases} -\infty, & \text{if } m = 0, \\ (m - M/2)d + \bar{x}, & \text{if } m = \overline{1,(M-1)}, \\ \infty, & \text{if } m = M, \end{cases} \quad (3)$$

and then the digital values $z_n = i$, $i = \overline{1,M}$ are generated. Here i is the quantization interval number, $\bar{x} = \langle x(t) \rangle$ is the mathematical expectation (mean value) of the process $x(t)$, and d is the quantization step selected according to the relation $d = (6 \div 10)\sigma/M$ (Rabiner and

Gold, 1975).

For the specified two-dimensional probability density $w(y_1, y_2)$ of the random process values that are x_n and x_{n+1}, the transition probabilities of the Markov model are determined as

$$P_{ij} = \frac{\int_{g_{i-1}}^{g_i}\int_{g_{j-1}}^{g_j} w(y_1, y_2)dy_2dy_1}{\int_{g_{i-1}}^{g_i}\int_{-\infty}^{\infty} w(y_1, y_2)dy_2dy_1} \quad (4)$$

and the probabilities of initial values – as

$$q_i = \int_{g_{i-1}}^{g_i}\int_{-\infty}^{\infty} w(y_1, y_2)\, \mathrm{d}y_2dy_1 \quad (5)$$

In a specific case, when the process $x(t)$ is stationary Gaussian one, one gets

$$w(y_1, y_2) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} \times \\ \times exp\left[-\frac{(y_1-\bar{x})^2 + 2r(y_1-\bar{x})(y_2-\bar{x}) + (y_2-\bar{x})^2}{2\sigma^2(1-r^2)}\right] \quad (6)$$

where σ^2 is the dispersion of the process $x(t)$, r is the correlation coefficient between the samples x_n and x_{n+1}.

A similar Markov model can be built on the experimental signal realization, if the sample size N is great enough. In this case, with obtained realization of the discrete random process z_n, $n = \overline{1,N}$, the estimates of the transition probabilities $\tilde{P}_{ij}$ can be found in terms of the transition numbers of the process values from $z_{n-1} = i$ to $z_n = j$, $n = \overline{2,N}$ as follows

$$\tilde{P}_{ij} = l_{ij} \Big/ \sum_{k=1}^{M} l_{ik} \quad (7)$$

For the estimates $\tilde{q}_i$ of the probabilities q_i one gets

$$\tilde{q}_i = \frac{1}{N-1}\sum_{k=1}^{M} l_{ik} \quad (8)$$

4. Random Number Generation Algorithm

By the transition probabilities, theoretical P_{ij} or experimental $\tilde{P}_{ij}$, the matrix of probability distributions $[F_{ij}]$ ($[\tilde{F}_{ij}]$) is determined as

$$F_{ij} = \sum_{m=1}^{j} P_{im} \text{ or } \tilde{F}_{ij} = \sum_{m=1}^{j} \tilde{P}_{im} \quad (9)$$

A sensor of independent pseudorandom (or random) numbers v_n, that are uniformly distributed within the interval [0,1], produces the n-th value, then selecting the next sample z_n. In cases when it occurs that $z_{n-1} = i$, then the following value $z_n = j$ is selected because j is the minimum value for which the inequality

$$v_n < F_{ij} \text{ or } v_n < \tilde{F}_{ij} \quad (10)$$

is satisfied. Based on the numbers v_n, the integers

$$\mu_n = \{v_n K\} \quad (11)$$

are determined obeying a uniform probability distribution and varying between 0 and $K-1$. Here $\{\cdot\}$ is the integer part, $K = 2^k$ and k is the specified integer number equal to the number of bits of the binary code of the number μ_n used to index the elements of the array.

As a pseudorandom number sensor, the programmable random number generator can use the standard procedure Random or similar. Pseudorandom numbers μ_n can be generated according to (11) or by a binary sequence (for example, M-sequence) generator implemented through shifters (Varakin, 1985). Random numbers μ_n can be also generated using a hardware noise source and comparator.

The procedure (10) for each $i = \overline{1,M}$ is calculated for all the possible numbers μ_n and the obtained values $j = \overline{1,M}$ are stored in the one-dimensional array $R_d = j$ with an index

$$d = (i-1)K + \mu_n. \quad (12)$$

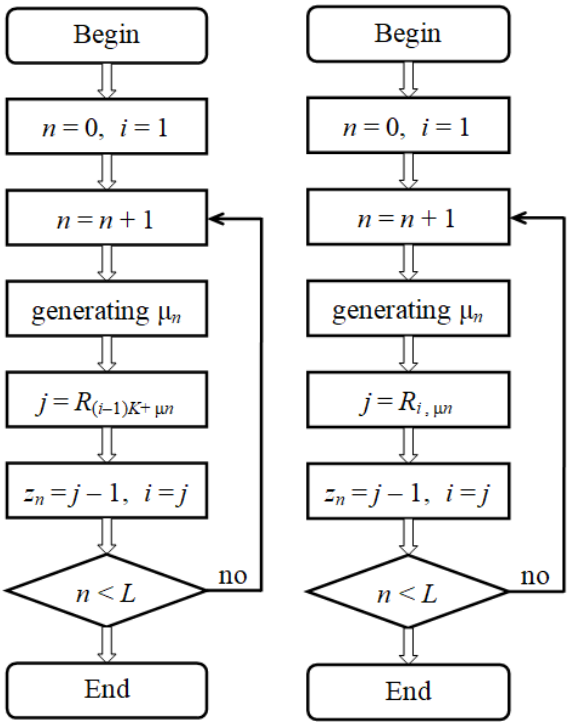
The block diagram of the described algorithm for generating L pseudorandom numbers is shown in Figure 1a. At the start, the first value $i = 1$ is selected (it may be any one from 1 to M). Then the integer μ_1 is generated for $n = 1$, so that the index d of the array is determined and the number j is read at the address (array index) R_d. The resulting value is assigned to the variable $z_n = j - 1$ (so that the numbers are obtained out of the range from 0 to $M-1$) that is at the sensor output.

In Figure 1b, there is presented the possible block diagram of the equivalent algorithm for generating random numbers through the two-dimensional array $j = R_{i\mu_n}$.

As it can be seen, generating the next sample requires a minimum number of operations involving the generation of numbers μ_n and reading data from the storage device. Thus, a high sensor speed is provided for any two-dimensional probabilistic properties of the generated samples.

5. Simulation Results

Let us consider the pseudorandom number sensor based on the two-dimensional Gaussian probability distribution (6) with the correlation coefficient $r = 0.4$ and the quantization step $d = 10\sigma/M$ (12). The transient probabilities are determined according to (4). The matrices P_{ij} in the form of a three-dimensional diagram under $M = 64$ are presented in Figure 2a, while the corresponding probability distribution F_{ij} calculated according to (9) can be seen in Figure 2b.



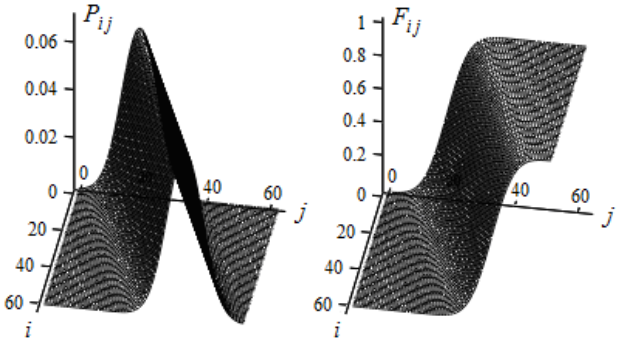
a) b)

Figure 1. The block diagram of the random number generating algorithm

In Figure 3, the part of the array R_d obtained according to (10) is graphically shown. Its indexes d are calculated by means of (12) and normalized to 2^K, while $K = 14$. The shift of the curves for different values of i is the result of the influence of the correlation coefficient.

In Figure 4, there is shown a part of the realization of the digital samples generated by the pseudorandom number sensor, while in Fig. 5a their histogram is presented.

In Figure 5a, by points, the theoretical probabilities are drawn that obviously well coincide with the corresponding experimental data.



a) b)

Figure 2. The probabilistic characteristics of the model

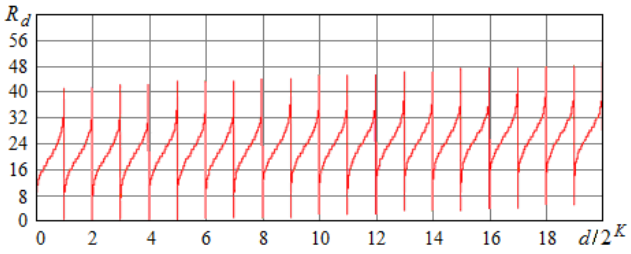


Figure 3. The values of R_d array elements

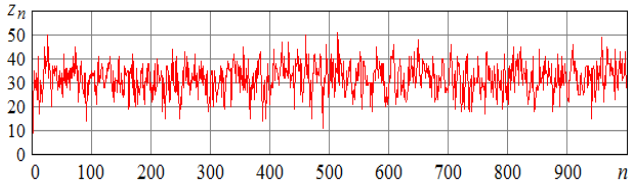
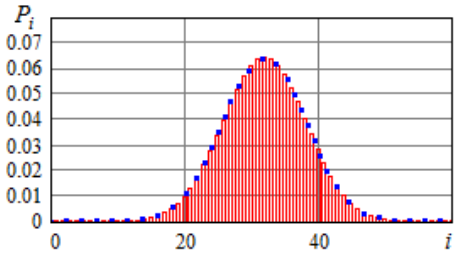
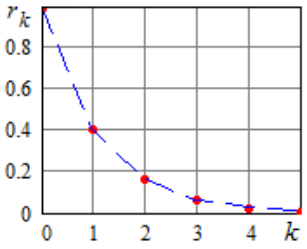


Figure 4. The realization of Gaussian pseudo random numbers with the specified probabilistic characteristics



a)



b)

Figure 5. The histogram (a) and the correlation coefficient (b) of the generated samples

In order to quantify to what degree the obtained and the theoretical statistical models fit together, the testing of χ^2 (Pirson) criterion (Law and Kelton, 2000; Ventsel and Ovcharov, 2000) is run to make sure that the generated realization obeys the Gaussian probability distribution. With the sample size $N = 2^{20} \approx 10^6$, the calculated value of χ^2 is 60.5. On the other hand, with the number of degrees of freedom $M - 1 = 63$ and the alpha level $\alpha = 0.01$, the critical value of criterion is equal to $\chi_0^2 = 90$ (Ventsel and Ovcharov, 2000). As it can be seen, $\chi^2 < \chi_0^2$, and, therefore, the hypothesis that the generated sample belongs to the Gaussian probability distribution is confirmed.

In Figure 5b, by points, the values of the correlation coefficient r_k are shown depending on the shifting of the samples k, while the dashed line represents the

corresponding theoretical values. One can see a good agreement between them.

Now let us consider an example when the pseudorandom number sensor is designed and developed in terms of the probability density significantly different from the Gaussian two-dimensional one. As such a probability density one chooses the function taking the form

$$w(x,y) = sin(x+y)/2,\ 0 \le x,y \le \pi/2. \quad (13)$$

The graph of the function (13) is shown in Figure 6a, while the correlation coefficient r of the simulated process is equal to -0.245.

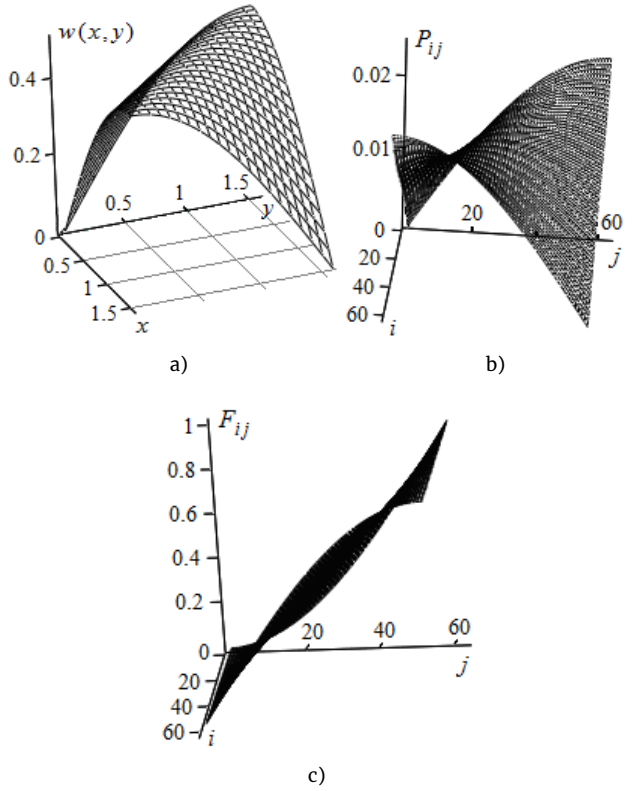


Figure 6. The probabilistic characteristics of the simulated process

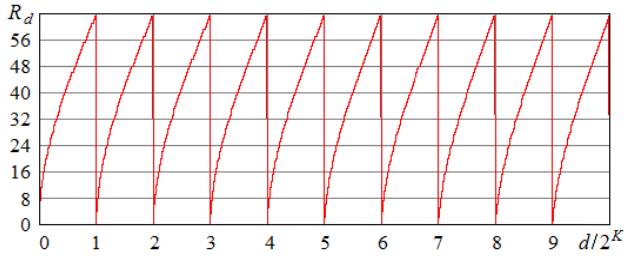


Figure 7. The values of R_d array elements

In Fig. 6b, the three-dimensional diagram of the matrix of transition probabilities P_{ij} (4) is presented; in Fig. 6c, one can see the probability distribution F_{ij} (9); and in Figure 7, the partial graph of the array R_d is demonstrated for the case when $M = 64$ and $K = 14$.

During the simulation, the sequence of pseudorandom numbers is obtained that is shown in Figure 8a. In Figure 8b, by bars, the histogram of the simulated realization is presented when the sample size is 10^6, while by points there are drawn the corresponding values of the theoretical one-dimensional probability density that well fit together with the experimental data.

In Figure 9a, there is plotted the three-dimensional diagram of the matrix formed by joint probabilities of pairs of values $P(i,j)$ calculated using the probability density (13):

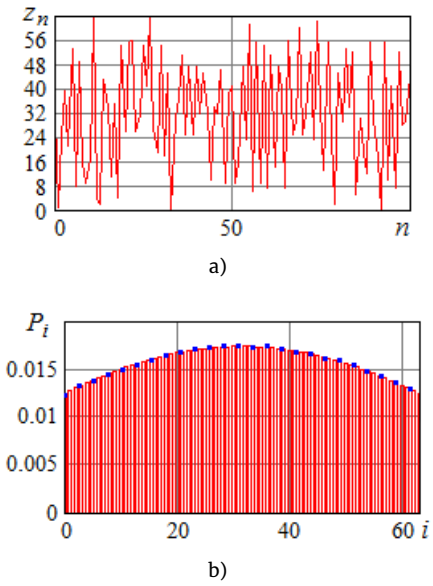


Figure 8. The simulated realization (a) and the histogram (b) of the pseudorandom numbers with the specified non-Gaussian probability distribution

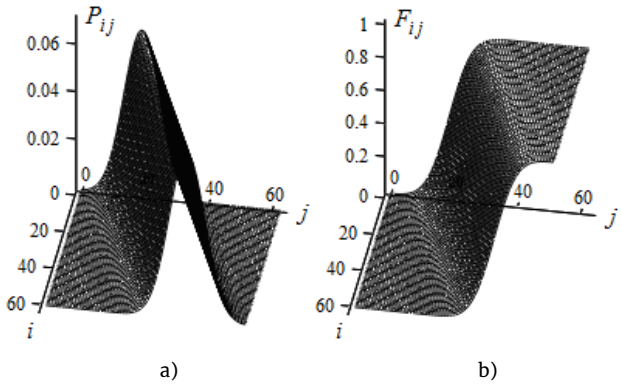


Figure 9. The joint probabilities of values of the simulated non-Gaussian process

$$P(i,j) = \int_{g_{i-1}}^{g_i}\int_{g_{j-1}}^{g_j} w(y_1,y_2)dy_2dy_1 \quad (14)$$

Figure 9b presents the statistical estimation of these probabilities obtained by processing the simulated samples.

The empirical correlation coefficient is equal to –

0.244 that nearly coincides with the theoretical value. Thus, it can be concluded that high simulation accuracy is provided for various probabilistic properties of the simulated processes. During statistical simulation, it is established that small (5-10%) deviations of the ν_n (μ_n) number source from the uniform probability distribution do not lead to the marked performance degradation of the sensor.

6. Conclusions

The introduced pseudorandom number sequence sensor based on a simple Markov model of the simulated process provides the probabilistic and correlation properties of the generated numbers specified by a two-dimensional probability distribution. It allows a simple software implementation with a minimum number of operations independent of the simulated process while high statistical accuracy of the simulation is still achieved. Besides, it is possible to develop a Markov model based on the experimental realization of the sampled random process reproducing its two-dimensional probabilistic properties. The considered algorithm can also be implemented in hardware in the form of a pseudorandom (random) number generator.

Funding

The reported study was funded by RFBR and CNRS, project number 20-51-15001.

References

Bardis N.G., Markovskyi A.P., Doukas N. and Karadimas N.V. (2009). True random number generation based on environmental noise measurements for military applications. *Proceedings of the 8th WSEAS International Conference on Signal Processing, Robotics and Automation*, pp. 68-73, February 21-23, Cambridge (UK).

Bykov, V.V. (1971). *Numerical modeling in statistical radio engineering* [in Russian]. Moscow, USSR: Sovetskoe Radio.

Ferguson, N., and Schneier B. (2003). *Practical cryptography*. New York, USA: Wiley.

Knuth, D.E. (1997). *The art of computer programming, vol. 2. Seminumerical algorithms*. Boston, USA: Addison-Wesley Professional.

Kroese, D.P., Taimre, T., and Botev, Z.I. (2011). *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. New York, USA: Wiley.

Law, A.M., and Kelton, W.D. (2000). *Simulation modeling and analysis*. New York, USA: McGraw-Hill.

L'Ecuyer, P. (2017). History of Uniform Random Number Generation. *Proceedings of 2017 Winter Simulation Conference*, pp. 202-230. December 3-6, Las Vegas (Nevada, USA).

Rabiner, L.R., and Gold, B. (1975). *Theory and application of digital signal processing*. New Jersey, USA: Prentice Hall.

Recommendation ITU-R P.1057-1 (2019, August 14). *Probability distributions relevant to radiowave propagation modeling*. Retrieved from www.itu.int/rec/R-REC-P.1057-6-201908-I/en

Schneier, B. (1996). *Applied cryptography: Protocols, algorithms, and source code in C*. New York, USA: Wiley.

Varakin, L.E. (1985). *Communication systems with noise-like signals* [in Russian]. Moscow, USSR: Radio i Svyaz'.

Ventsel, E.S., and Ovcharov, L.A. (2000). *Theory of stochastic processes and its engineering applications. Textbook for higher technical schools* [in Russian]. Moscow, Russia: Vysshaya shkola.