UDC 519.216.1

PSEUDORANDOM PROCESSES OF THE NUMBER SEQUENCE GENERATION

G. Vostrov, A. Khrinenko

Odessa National Polytechnic University

Abstract. *This work investigates problems that occur in modeling of the nonlinear processes and number sequence generation. Dependences of iterative fixed points of nonlinear maps on the function properties and number properties from the functions domain are investigated. This work also analyze prime and sequences obtained using these numbers.*

Key words: *chaos, pseudorandom sequences, nonlinear maps, prime numbers.*

Introduction

Modeling of random and pseudorandom processes is of fundamental importance in mathematics and can be applied in various ranges of mathematical branches, such as dynamic system modeling, functional analysis, theory of functions, cryptography and others. In this regard one of the most important and fundamental problem is number sequence generation. The obtained sequences can be applied for solving different spectrum of tasks, for instance Quasi-Monte Carlo method uses pseudorandom sequences for numerical integration [1], in machine learning theory for obtaining training and test samples [2] and others. Various methods for number generation are based on chaotic processes that raise a question: "What processes can be called chaotic?" Up to now, there is no answer to the question – how to define the concept of randomness accurately and constructively, i.e. axiomatically. Different generators simulate randomness with certain degree of approximation to a given distribution law. Construction of random number generator enable to construct the concept of a formal and constructive definition of randomness, which is essential and necessary for modern probability theory, mathematical statistics, random processes and others. Since an truly random sequence is a mathematical model that is absolutely unpredictable and, therefore, nonperiodical infinite sequence the problem of pseudorandom sequence generation that approximates randomness is formulated [3].

An independent research direction, which is commonly called as deterministic dynamic systems, has been formed apart from previously mentioned research areas. It is based on the study of the dynamics of iterative fixed points via recursive functions. . Iterative cycles or orbits are considered as deterministic chaos [7], since they depend on initial conditions and demonstrate no regularity [4]. Dynamic systems that belong to deterministic class of systems appear as a consequence of an approximation of complex processes in the physical and mental world and can be used as generators of pseudorandom sequences.

It should be noted that in case of PRS generators and in case of dynamic systems it is necessary to take into account properties of iterated function. Main characteristic of these generators is the length of the iteration process period. Meanwhile, properties of the set of numbers, on which given generators are determined, are not taken into account. In this regard direct connection with number theory is observed. Prime numbers are of considerable interest when iterative processes are examined, because they are indecomposable into simple factors and usage of big compound numbers cannot guarantee required length of period. However, some prime numbers also do not provide the longest cycle length; this circumstance refers to those numbers that belong to certain number classes. Such numbers include Fermat, Mersenne, Wagstaff prime numbers and their various generalizations [1]. For an arbitrary choice of prime number it is transpired, that there is a set of large prime numbers on which generators provide sequences where the cycle length is insignificant and clearly not chaotic. Although on adjacent prime numbers the cycle length is commensurable with dimension of the prime number. If these properties of prime numbers are not taken into account, the choice of such an exceptional prime number can lead to very significant errors and incorrect conclusions.

1. Computer analysis of iterative processes in nonlinear maps

In this paper we investigate processes occurring in maps that are presented as examples of simple nonlinear dynamic systems for analyzing previously stated problem. As process we consider a function $F,$ that maps finite sequence (word) into sequences so that if for word x value of $F(x)$ is determined and $y \subset x,$ then $F(y)$ is also determined and

$F(y) \subset F(x)$. Let’s ω - some sequence. Process F will be applied as long as it’s possible. As a result we obtain parts of some new sequence S – result of application F to ω, that is $S = F(\omega)$. More precisely deterministic maps of the following four classes are analyzed: "Tent", "Asymmetric Tent", "Saw-tooth" and the map that generates intermediate steps of multiplicative order of a number modulo p calculation. It is important to remark that maps (1, 2, 3) are continuous functions, whereas map (4) is determined only on set of integers. Analyzed in this paper maps allow to investigate chaotic processes that emerge in complex dynamic systems.

$$t_1(x_n) = x_{n+1} = \begin{cases} 2x_n, & x_n < 1/4 \\ 1 - 2x_n, & x_n \geq 1/4 \end{cases} \qquad (1)$$

$$t_2(x_n) = x_{n+1} = \begin{cases} 2x_n, & x_n < 1/2 \\ 1 - x_n, & x_n \geq 1/2 \end{cases} \qquad (2)$$

$$t_3(x_n) = x_{n+1} = \begin{cases} 2x_n, & x_n < 1/2 \\ 2x_n - 1, & x_n \geq 1/2 \end{cases} \qquad (3)$$

It is worth to note that computer systems use numbers in binary form and of limited length in its calculations, while mathematics operates with infinite length numbers and this circumstance leads to errors in calculations with fractional numbers that ultimately leads to problem of validity and incorrect conclusions about dynamic processes. That circumstance manifests itself due to the fact that any dynamic system has sensitive dependence on initial conditions [4]. To minimize the errors in rounding, a transition to the family of integer maps has performed. The maps are represented as follows:

$$x_{n+1} = \begin{cases} 2x_n, & 4x_n < p \\ p - 2x_n, & 4x_n \geq p \end{cases} \qquad (1)$$

$$x_{n+1} = \begin{cases} 2x_n, & 2x_n < p \\ p - x_n, & 2x_n \geq p \end{cases} \qquad (2)$$

$$x_{n+1} = \begin{cases} 2x_n, & 2x_n < p \\ 2x_n - p, & 2x_n \geq p \end{cases} \qquad (3)$$

$$x_{n+1} = 4x_n (\operatorname{mod} p), \qquad (4)$$

where p – prime number. Graphs, that demonstrate maps (1, 2) and its second iterates on [0,1] and map (3) on [0,1) are shown on Fig. 1, 2, 3 respectively and Fig. 4 shows map (4) on the set of integers. On the graphs projection *FPN* show fixed points of the maps. Diagonal $y = x$ crosses graphs of the maps at some points in their intervals and so there is a periodic points in any such interval. Since the lengths of these intervals are $1/2^n$, where n is a number of iteration, it follows that periodic points are dense on [0,1]. It is worth to mention, that map (1) algebraically congruent to map (4) on set of integers, this means that their cycle length are the same for any prime numbers. But, the map (2) does not satisfy the Fermat’s little theorem since for any prime number p the value of Euler function is not divisible by the cycle length. Despite the simplicity of these maps their iterative cycles, based on prime numbers, have properties that support the above stated hypothesis. According to them, not only properties of maps determine the structure of iteration cycles, but also the properties of numbers from their domain of definition can have a decisive influence on the structure and radically change it. Presented nonlinear maps allow dividing set of prime numbers p into the class system, which is based on the length of the iteration cycles as a function of given prime numbers [5].

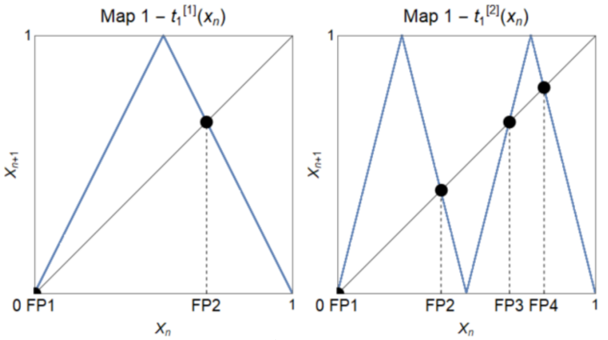

Fig. 1 – 1st and 2nd iteration of the map 1

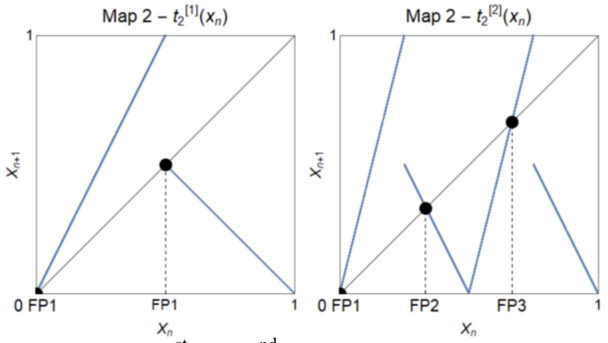

Fig. 2 – 1st and 2nd iteration of the map 2

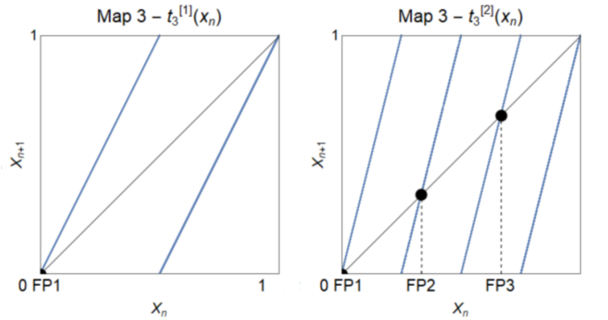

Fig. 3 – 1st and 2nd iteration of the map 3

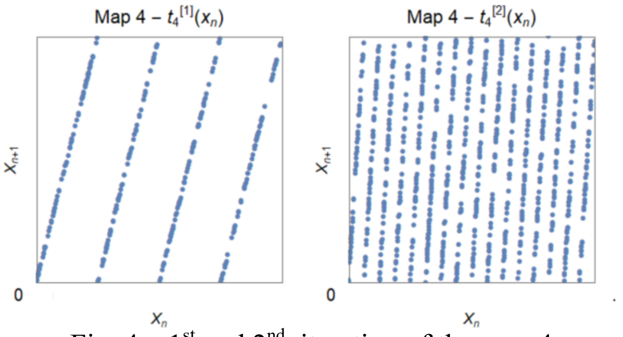

Fig. 4 – 1st and 2nd iteration of the map 4

It should be noted that there is an infinite set of prime numbers for which the length of the period is significantly smaller than the dimension of the number. Table 1 shows the behavior of the triples of some consecutive prime numbers, where the first and third numbers belong to the class that provide large cycle length of sequence, while for the second number the length of the period is incommensurably smaller than the number dimension itself, and the PRS obtained for a given number forms a simple structure.

Table 1

Length of the period for generated sequences

position	prime number	*m(1)*	*m(2)*	*m(3)*	*m(4)*
1	148587941	74293970	111440955	148587940	74293970
2	148587949	142	193	284	142
3	148587953	37146988	55717993	74293976	37146988
1	164511349	82255674	123383511	164511348	82255674
2	164511353	41	49	41	41
3	164511371	82255685	123383528	164511370	82255685
1	168410987	84205493	126308240	168410986	84205493
2	168410989	162	227	324	162
3	168411029	84205514	126308271	168411028	84205514

In this table columns *m(n)* show the period length for the corresponding map *n*, where *n* – number of the corresponding map.

Considering the behavior of individual prime numbers, Fig. 5 represents the internal structure of the iterative process in the maps for the number 160465489, which has a short period length. Filled sectors show repetitive regions in initial sequences

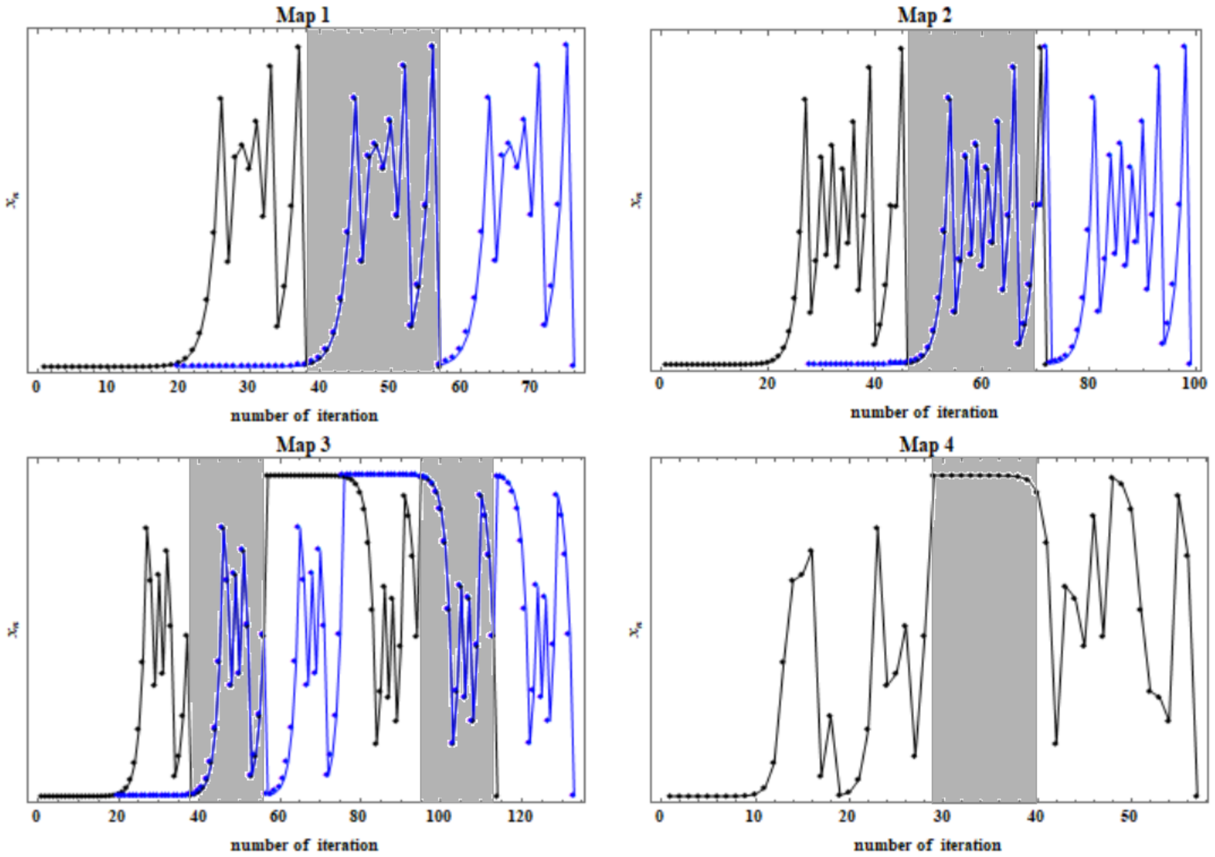

Fig. 5 – The PRS structure for the maps

The thing to note here is that the choice of a large prime number does not allow escaping of the sequences with simple structure and repeated exponential components that are highlighted by filled region on the Fig. 6. In concurrence with this, the preceding prime number and the next prime number form sequences with the length of the period that commensurate with the value of these numbers. This may indicate randomness of the iterative process.

However, these sequences also have periodic components that are observed and highlighted with filled region at compression region on the Fig. 6 for map 1, but it is constant for other maps as well [6].

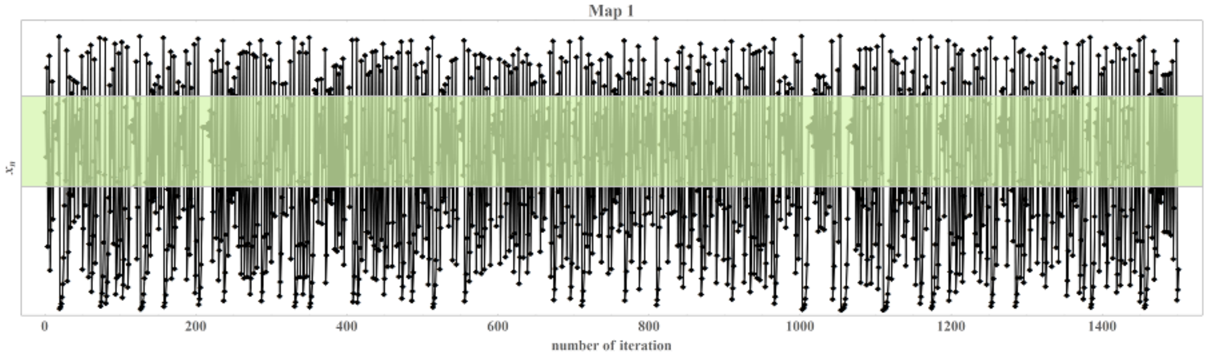

Fig. 6 – Dynamics of number 160465519 for map (1)

Examining prime numbers, particular attention is attracted to prime numbers of a special kind, such as generalized Gaussian-Mersenne prime numbers of the form $p^* = 2^p - 1$ for some rational prime p, where $p' = ((1+i)^p - 1)((1-i)^p - 1)$ is also prime number. There are sequences that have a simple structure with exponential components that repeat with a certain frequency and are slightly different in amplitude. These numbers completely violate the randomness conditions imposed on PRS. It is proved that such an internal regular structure is characteristic of Mersenne, Wagstaff numbers and their various generalizations. Fig. 7 shows the structure of PRS for the generalized Gaussian-Mersenne prime number.

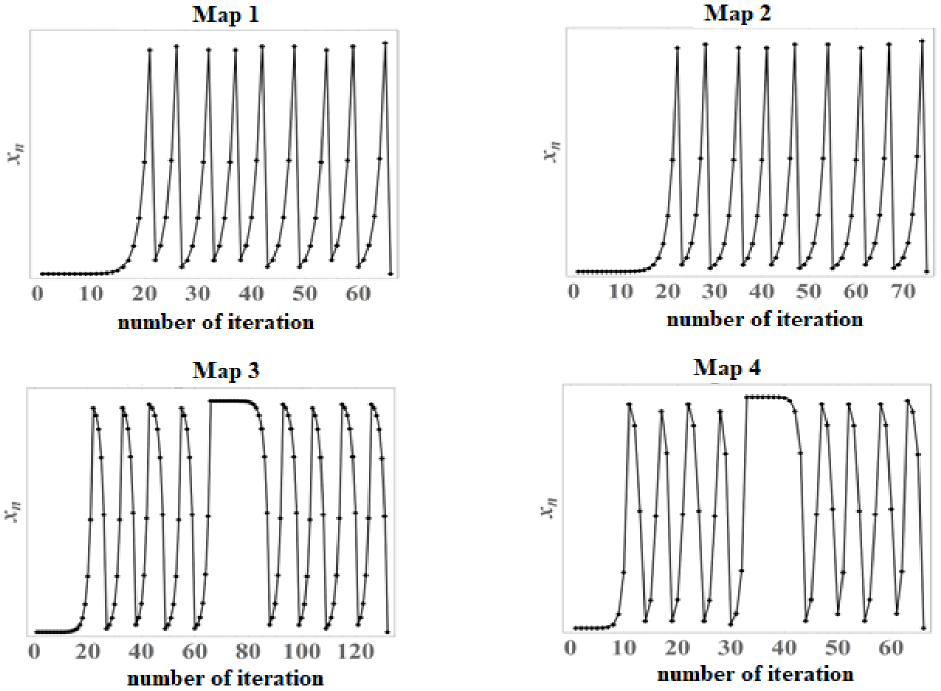

Fig. 7 – PRS, based on Gaussian-Mersenne prime

2. Randomness estimation methods for the iterative processes

There are several approaches and methods for analyzing the “quality” of the generated PRS by estimating some property of a sequence that helps to determine sequences which satisfy requirements that are set for pseudorandomness. Since the maps considered in this paper generate integer sequences for further analysis all the elements of the sequences are converted to binary form. According to Von Mises definition of randomness all sequences are devided into two groups: random and non-random. From the mathematical point of view random sequences forms set of full measure and satisfy all laws of probability theory without exception. According to this approach a sequence is considered as stochastic i.e. random if frequency stability of zeros and ones in the sequence is observed and not only in whole sequence but also in any arbitrary and “right” chosen part of it. Respectively to Von Mises the admissible selection rule is that include a member x_n in the subsequence cannot depend on the value. It is worth noting that the class of "admissible" frequency-stable sequences for which the basic laws of probability theory have been fulfilled is still not defined. It has also been proved that there are sequences that satisfy the Von Mises requirements, but do not satisfy the law of repeat logarithm.

Based on Kolmogorov's ideas, a second approach to the definition of random sequences was formed, based on the fact that the description of a random sequence can not be less than the sequence itself, that is, the algorithmic complexity is approximate to the length of the sequence. The complexity of a sequence y for a given map f is the number $R_f(y) = \min\{|x| : x \epsilon f\text{-}describes\ y\}$, where | x | - the length of the sequence. With this approach, "non-random" are those sequences in which there are quite a lot of regularities. By law, any property of the checked sequence is inherent in a narrow class only. If we measure the "number of regularities" according to the traditional Shannon scale, the last phrase is specified as follows: the measure of the set of those sequences in which more than m bits of regularities can be detected must not be surpassed 2^{-m}. In this case, the tests for checking the sequences can be very diverse.

The third approach is called “quantitative” and is proposed by Martin-Lof. The approach is based on the fact that the sequence is considered to be random if it passes a certain set of statistical tests. The essence of testing is to verify the "zero hypothesis" in relation to the sequence being studied. The statistical test T for binary sequences of length l can be considered as a boolean function $T : V_l \rightarrow \{1,0\}$, that divides the set of sequences V_l into a set of "non-random" sequences $V_{l,0}$ (usually small) and a set of random sequences $V_{l,1}$. The probability pr that a randomly chosen sequence of length l is rejected by the test is equal to $pr = |V_{l,0}| \cdot 2^{-l}$. As a rule, in tests pr is small. Since the maps considered in this paper generate integer sequences, to further testing them on a statistical test, all the elements of the sequences are converted to binary form. Testing methods for obtained sequences are divided into 2 groups.

The first group is connected with the search for regularities, which make it possible to reproduce the sequence over its segment. In this case, the basic requirements for the sequence are reduced to the absence of relatively simple interelement dependencies in it. For example, it is necessary to check the correlation between the elements of the sequence. This task is to conduct an autocorrelation analysis i.e. to construct a correlogram, which shows the value of the correlation coefficient at different shifts of the initial sequence. The data obtained for the number with small cycle length and shown in Fig. 8, 9, 10, 11 show that the correlation for a given number fluctuates at the level of 0.5 and, consequently, this number absolutely does not satisfy the requirements imposed to the PRS.

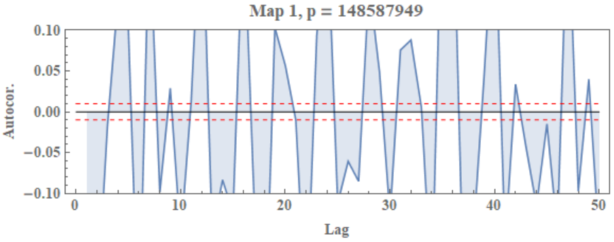

Fig. 8 – Correlation values for map (1)

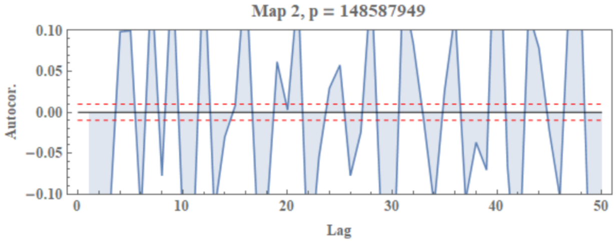

Fig. 9 – Correlation values for map (2)

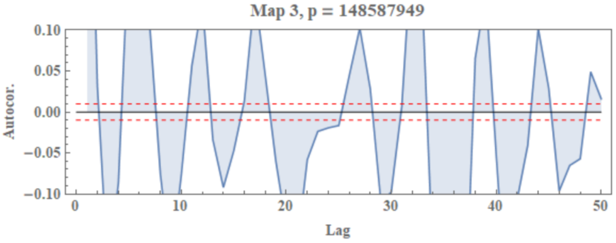

Fig. 10 – Correlation values for map (3)

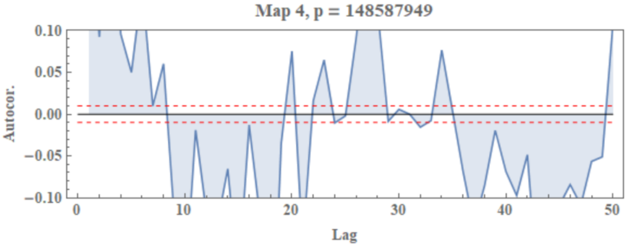

Fig. 11 – Correlation values for map (4)

Whereas, for prime number with commensurate cycle length the autocorrelation values calculated for a part of the sequence containing 10,000 numbers and are shown in Fig. 12, 13, 14, 15 and closer to 0. Thus, this sequence shows a large degree of randomness, but still does not satisfy the requirements of randomness.

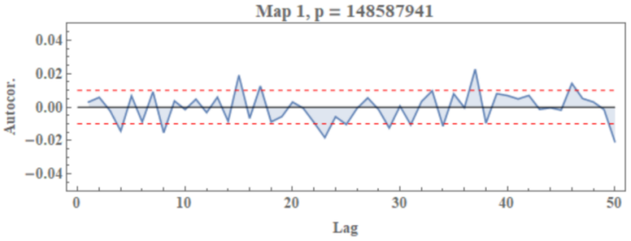

Fig. 12 – Correlation values for map (1)

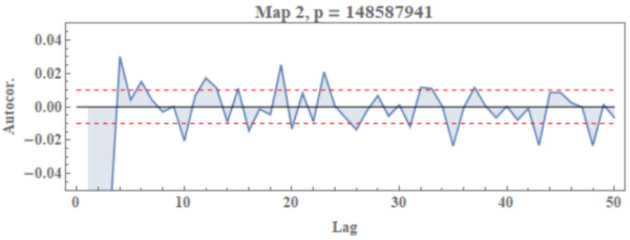

Fig. 13 – Correlation values for map (2)

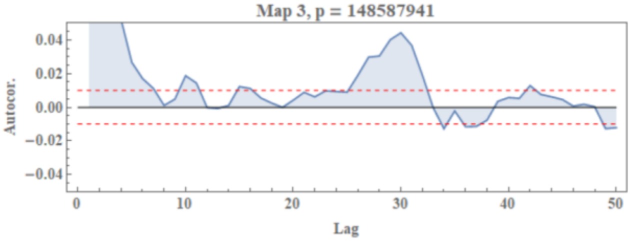

Fig. 14 – Correlation values for map (3)

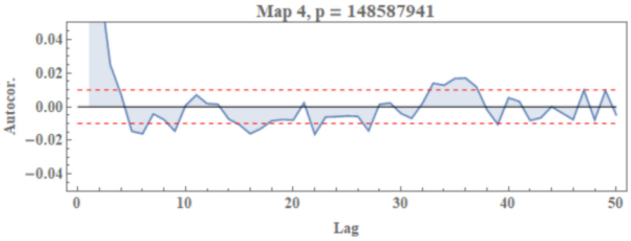

Fig. 15 – Correlation values for map (4)

The obtained data for correlation make it possible to assume that a more acceptable result is achieved by using numbers which period length is commensurable with the dimension of the number itself. Since, analytical proof of some of the necessary properties is possible only for certain classes of sequences to justify the properties of sequences there are a wide range of different statistical tests that allow us to reveal regularities.

As a method for evaluating generated binary sequences, the following tests are considered:

- Frequency Test. This test is aimed to estimate the ratio of zeros and ones in the studied sequence. To accept the sequence as random ratio should be equal, as it could be assumed in the case of a sequence that satisfy the conditions of randomness.
- Non-overlapping Template Matching Test. This test is aimed to determine the number of predefined templates in the initial sequence.
- Approximate Entropy Test. The test is aimed to estimate the frequency of possible overlapping blocks in the entire initial sequence. The test compares the frequencies of blocks of successive length *n* and *n+1* overlapping with the frequency of similar blocks of random sequence.
- Spectral Test. The test is aimed to find the peak heights in the Discrete Fourier Transform of the sequence. This test is trying to detect periodic features i.e. repetitive subsequences that are located close to each other in the tested sequence that would indicate a deviation from the assumption of randomness. The intention is to detect whether the number of peaks exceeding the 95 % threshold is significantly different than 5 %. If the computed *pr*-value is > 0.01, then conclude that the sequence is random.

Table 1 shows the results of statistical tests for sequence fragments containing 10,000 elements and obtained on the basis of *p* = 148587941 (length of period equals 74293970).

Table 2

Results of the tests

Test	**Map**	**p = 148587941**
Frequency Test	1	$1.1373200 \times 10^{-237}$
	2	$2.1654170 \times 10^{-239}$
	3	0.4353908
	4	0.8571525
Non-overlapping Template Matching Test	1	0.0105489
	2	0.0105489
	3	0.9198479
	4	0.0336110
Approximate Entropy Test	1	$2.7360300 \times 10^{-468}$
	2	$2.0647959 \times 10^{-465}$
	3	0.0058075
	4	$2.0665877 \times 10^{-17}$
Spectral Test	1	7.4135928×10^{-9}
	2	$1.4693925 \times 10^{-16}$
	3	0.7830866
	4	0.1988873

The results of these statistical tests show that maps (3) and (4) generate sequences that have better fit to randomness conditions on some test, however, this is not satisfied for the whole group of tests, and it does not allow considering the maps as generators of pseudorandom numbers, since the necessary condition for randomness is not provided.

In practice, the acceptance or rejection of the null hypothesis is based on the results of the application of battery of tests. The generator is considered as appropriate if the admission of the sequences that are recognized as non-random is comparable to the similar characteristic of the truly random generator. Also, the design of the generator should ensure that deliberately weak sequences do not appear at the output, which also requires analyzing the properties of the numbers on which the generator is built. Since the PRS generating methods used in this paper are simple iterative processes, the sequences obtained on the basis of any arbitrary numbers will not satisfy the conditions of randomness. However, the value of *pr* will lead to a conclusion about which numbers possess better statistical properties

Passing from the consideration of separate primes to the analysis of the set of numbers and presenting the number as $p^* = kn + 1$, where n is the length of the period for this number, we can distinguish classes of k primes. Then during the process of generating pseudorandom sequences it is worth to consider numbers with the smallest value of the class. The set of all primes p on the basis of nonlinear maps is divided into the same system of classes that is based on the length of the iterative cycle as a function of the value of a prime number.

To improve the sequence properties and to achieve a longer period, it is possible to use multiple generators for which the output sequences are "blended" with each other to obtain a new sequence. The simplest method for combining sequences is to use the bitwise exclusive disjunction operation that for sequences S and S^* generates a new sequence $R = S \oplus S^*$. Thus, the use of simple maps can provide the generation of sequences with a given degree of approximation to randomness.

Conclusion

The results obtained in this paper show that best possible fit to randomness conditions for generating PRS by nonlinear dynamic maps require to take into account the properties of the set of numbers on which the pseudorandom number generator is built. The best possible fit to randomness conditions for generating PRS can be obtained using prime numbers for which the length of the period is commensurable with the value of the number itself. And to improve the statistical properties of sequences that are generated with simple maps, it is necessary to combine and mix those maps.

References

1. Crandall R., Pomerance K. (2005). Prime numbers: cryptographic and computational aspects. 2nd edition. Springer. p. 597.
2. Balasubramanian, V., Ho, S. and Vovk, V. (2014). Conformal Prediction for Reliable Machine Learning. Morgan Kaufmann. p. 299.
3. Probability and the concept of randomness: to the 75th anniversary of the publication of Kolmogorov's monograph "Basic concepts of probability theory" [Verojatnost' i koncepcija sluchajnosti: k 75-letiju vyhoda v svet monografii A.N. Kolmogorova «Osnovnye ponjatija teorii verojatnostej»]. (2009). p. 92.
4. Hirsch, M., Smale, S. and Devaney, R. (2013). Differential equations, dynamical systems, and an introduction to chaos. Amsterdam: Academic Press. p. 423.
5. Vostrov, G., Opiata, R. (2017). Effective computability of dynamic system structure of prime number formation [Effektivnaya vychislimost' struktury dinamicheskikh sistem formirovaniya prostykh chisel]. ELTECS, 244
6. Sharkovsky, A. (1988). Attractors of trajectories and their pools [Attraktory traektorij i ih bassejny]. Kyiv: «Scientific book». p. 322.
7. Schuster, H. (1995). Deterministic chaos. Weinheim: VCH. p. 320.

ПСЕВДОВИПАДКОВІ ПРОЦЕСИ ПРИ ГЕНЕРАЦІЇ ЧИСЛОВИХ ПОСЛІДОВНОСТЕЙ

Г. М. Востров, А. О. Хріненко

Одеський національний політехнічний університет

Анотація. *В даній роботі проводиться дослідження проблеми, що виникає при моделюванні нелінійних процесів та генерації послідовностей чисел, оскільки створення генератору випадкових чисел дозволяє побудувати концепцію формального та конструктивного визначення випадковості, яке є наріжним у сучасній теорії ймовірності та інших. Досліджена залежність ітеративних нерухомих точок нелінійних відображень від властивостей функцій та властивостей чисел з області визначення цих функцій. Відповідно до довжин розглянутих ітераційних процесів, як однієї з основних* *характеристик псевдовипадкових послідовностей розглянуто прості числа, які відносяться до певних класів, таких як числа Мерсенна, Вагстафа та Ферма. При довільному виборі простого числа існує множина великих чисел для яких використані генератори формують послідовності, що не відповідають умовам випадковості. Для дослідження даної проблеми використано групу відображень, що представляють собою прості ітераційні процеси, однак дозволяють зробити висновки, щодо розглянутих питань. Відповідно до них, внутрішня структура послідовностей, а також їх відповідність умовам випадковості залежать не тільки від властивостей використаних відображень, а також від властивостей чисел з їхньої області визначення. Для аналізу і оцінювання отриманих послідовностей розглядається декілька підходів та виповнено перехід до двійкового представлення. Перший з підходів висуває до послідовностей та їхніх, довільно вибраних, підпослідовностей умову частотної стабільності на основі якої послідовності розділяються на класи випадкових та невипадкових послідовностей. Другий підхід базується на тому, що опис випадкової послідовності не повинен бути меншим за саму послідовність та визначає поняття алгоритмічної складності. Останній підхід заснований на тому, що послідовність вважається випадковою, якщо вона проходить певний набір статистичних тестів, приклади та результати використання яких наведено в роботі. Таким чином, показано, що найкраще наближення до умов випадковості демонструють послідовності для яких довжина ітераційного процесу співмірна з розмірністю використаного простого числа.*

Ключові слова: *хаос, псевдовипадкові послідовності, нелінійні відображення, прості числа.*

ПСЕВДОСЛУЧАЙНЫЕ ПРОЦЕССЫ ПРИ ГЕНЕРАЦИИ ЧИСЛОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Г. Н. Востров, А. О. Хриненко
Одесский национальный политехнический университет

Аннотация. *Данная работа рассматривает проблемы, которые возникают при моделировании нелинейных процессов, а также при генерации псевдослучайных числовых последовательностей. Рассмотрена зависимость итерационного процесса от свойств используемых функций, а также от свойств чисел из их области определения.*

Ключевые слова: *хаос, псевдослучайные последовательности, нелинейные отображения, простые числа.*

Received on 01.04.2018

Востров Георгій Миколайович, кандидат технічних наук, доцент кафедри прикладної математики та інформаційних технологій Одеського національного політехнічного університету. Проспект Шевченко, 1, Одеса, Україна.
E-mail: vostrov@gmail.com, тел. +380503168776

George Vostrov, Ph. D. of Technical Sciences, Associate Professor of the Department of Applied Mathematics and Information Technologies, Odessa National Polytechnic University. Shevchenko ave., 1, Odessa, Ukraine.

ORCID ID: 0000-0003-3856-5392

Хріненко Андрій Олегович, кафедри прикладної математики та інформаційних технологій Одеського національного політехнічного університету. Проспект Шевченко, 1, Одеса, Україна.
E-mail: khrinenko.andrew@gmail.com, тел. +380637515228

Khrinenko Andrii, student of the Department of Applied Mathematics and Information Technologies, Odessa National Polytechnic University. Shevchenko ave., 1, Odessa, Ukraine.

ORCID ID: 0000-0001-6000-2102