A GENERALIZED PROBABILISTIC MODEL OF COMPUTER PROOF OF THE ARTIN HYPOTHESIS

G. Vostrov, R. Opiata
Odessa national polytechnic university
Odessa, UKRAINE
e-mail: vostrov@gmail.com, roma.opyata@gmail.com

Abstract

The analysis of the probabilistic approach to solving the problem of the distribution of primes in the generalized Artin's hypothesis is given. The foundations of a computer approach to solving problems in the field of pure and applied number theory are formulated. On the basis of the generalized Artin's hypothesis, it is shown how probabilistic methods of nonlinear dynamic systems can be obtained with sufficiently accurate solutions.
Keywords: data science, Artin hypothesis, computer proof

1 Introduction

An important problem in the theory of numbers is the description of the law of the distribution of primes. This problem was solved by Hadamard and Valle-Poussin, independently of each other, in 1896 [1]. They proved that the number of primes $(\pi(x))$ is less than or equal to x is determined by the expression:

$$\pi(x) = \int_2^x \frac{\mathrm{d}t}{lnt} + O\left(xe^{-\frac{c}{2}\sqrt{lnx}}\right) \tag{1}$$

where c is an absolute constant. This analytically proved form of representation of the law of distribution of prime numbers has already become universally recognized in the mathematical world. Yet two things should be noted. First, it was obtained on the basis of the analytic zeta-Riemann function, which, until it is proved, adequately describes the distribution of primes in a complex space. According to the Riemann hypothesis, all the zeros of the zeta function are on the line passing through the point equal to 1/2. This millennium hypothesis has not yet been proved. And this fact is the basis for criticizing all the results obtained on the basis of the zeta-Riemann function.

The second circumstance is that simultaneously with this fact the dynamics of the change of $O\left(xe^{-\frac{c}{2}\sqrt{lnx}}\right)$ [2] is investigated. In [1, 5], an estimate of the entropy of this estimate is obtained and it is proved that it is fractal in nature. These facts are the basis for the formation of proposals on the need to study other models for the distribution of prime numbers. Another problem related to the distribution of prime numbers appeared in 1927, when the well-known mathematician Artin formed a hypothesis about the distribution of prime numbers for which the natural number $a > 1$ is given is their primitive root [1, 5].

According to the Artin conjecture [5], the set of such prime numbers has the distribution law $\pi(x,a)$ in the form of the expression:

$$\pi(x,a) = c(a)\pi(x) \tag{2}$$

where $\pi(x)$ is the distribution of primes, and $c(a)$ is a constant depending on a. So far, despite numerous studies, this hypothesis has not been resolved. At the same time, it is not known whether this is true for any values of a. If the hypothesis is correct, the question remains how to evaluate the constant $c(a)$ for each particular a and what properties of the number a affect its value. Answers to these questions are still lacking. In [1, 5] a detailed analysis of all research results in the field of the Artin hypothesis solution is given.

It should be noted that the proof of Artin's hypothesis is important both from a theoretical point of view in number theory, and from an applied rhenium point, because its positive solution is important in cryptography, coding theory, and the theory of dynamical systems. In [6], a generalized Artin hypothesis was formed for any $a > 1$, i.e. and at the same time a may not be a primitive root. According to Artin's generalized theory, the following equality is true:

$$\pi(x,a,i) = c(a,i)\pi(x) \tag{3}$$

where $a > 1$, i is the index of the subgroup of the group $(Z/pZ)^*$ of primes in the classification of prime numbers generated by the numbers a, $c(a,i)$ is a constant. According to the classification built in [6]:

$$\mathcal{P}(a,i) = \left\{ p \in \mathcal{P} | \frac{(p-1)}{card_a(p)} = i \right\} \tag{4}$$

where $card_a(p)$ is the length of the recursion $x_{n+1} \equiv ax_n(modp)$ at $x_0 = 1$, $\mathcal{P}$ is the set of all primes.

It is not difficult to show that for any $a > 1$ the equality:

$$\sum_{i=1}^{\infty} c(a,i) = 1 \tag{5}$$

This means that primes are evenly distributed in classes $\mathcal{P}(a,i)$ for any a. By uniformity is meant that within each class of primes $\mathcal{P}(a,i)$ a logarithmic law of the distribution of primes is preserved. The constant $c(a,i)$ determines the measure of puncturing prime numbers based on the value a. If $i = 1$ then a is the primitive root of all primes $\mathcal{P}(a,1)$. For an arbitrary natural number x, the equality:

$$\pi(x,a,i) = c(a,i,x)\pi(x) \tag{6}$$

Moreover, if $x \to \infty$, then $c(a,i,x)$ tends to the limit value $c(a,i)$. If we put $i = 1$ then $c(a,1)$ will be Artin's constant for primitive roots. In this case $a \neq \pm 1$, and $a \neq k^2$ for none $k \in N$. This is true according to Fermat's theorem [3,4]. Wherein, a is the primitive root of the group of residues $(Z/pZ)^*$ for any $p \in \mathcal{P}$ such that:

$$\mathcal{P}(a,1) = \left\{ p \in \mathcal{P} | \frac{(p-1)}{card_a(p)} = 1 \right\} \tag{7}$$

It is important to investigate the classes of primes $\mathcal{P}(a,i)$ for $i>1$ since in this case the positive integer a will be the primitive root for the subgroups of the group $(Z/pZ)^*$ with the index defined by the relations:

$$\mathcal{P}(a,i)=\left\{p\in\mathcal{P}|\frac{(p-1)}{card_a(p)}=ind_a(p)\right\} \tag{8}$$

where $ind_a(p)=i$ is the index of the subgroup of $(Z/pZ)^*$. The classes of primes $\mathcal{P}(a,i)$ have not yet been studied and the distribution of primes in these classes is not known. In [1], an assumption was made that $\mathcal{P}(a,i)$ at $i>1$ is proportional to $\mathcal{P}(a,1)$ with a factor of $1/i^2$. Since $i>1$ is considered, in this case it is important to know the distribution of prime numbers for the value $a=k^2$. This is an important generalization of Artin's hypothesis. At the same time, the probability of:

$$P(p\in\mathcal{P}(a,i))=\left\{p\in\mathcal{P}|\frac{|\mathcal{P}(a,i)|}{|\mathcal{P}|}=c(a,i)\right\} \tag{9}$$

membership agrees exactly with the provisions of the theory of probability, and therefore, estimating $c(a,i)$ on the basis of successive statistical tests and the law of large numbers is parity.

The determination of $c(a,i)$ for any a,i using analytical methods is unlikely in the near term. However, the formation and development of experimental mathematics [1, 2] opens up another way to solve this problem by using computer simulation of nonlinear dynamic processes for the formation of classes of prime numbers.

2 Modeling of dynamic processes of distribution of simple numbers in the generalized artin hypothesis

The process of modeling the distribution of primes in classes $\mathcal{P}(a,1),\mathcal{P}(1,2),...,\mathcal{P}(a,k)$ was reduced to choosing a set of consecutive primes from a set of a sufficiently large sample of these classes. The number of primes analyzed at each interval of natural numbers was chosen to be 500,000. This choice was largely due to the fact that it was previously established that reducing this value leads to more significant fluctuations in estimates, although convergence to the limit over the entire set of any intervals, even if they are not placed consistently, has the same character.

The process of statistical testing of $p\to\mathcal{P}$ primes for checking their belonging to class $\mathcal{P}(a,i)$ was reduced to calculating for the selected number p the recursive procedure $x_0=1$, $x_{n+1}=ax_n(modp)$ until the pairs $ax_l\equiv 1(modp)$ were reached at some step i. Then $card_a(p)=i$ and according to Fermat's theory and the cyclic group theorem the number $p-1$ is divisible by i and then $ind_a(p)=(p-1)/card_a(p)=i$, and therefore $p\in\mathcal{P}(a,i)$ and if $i=1$, then a is the primitive root of the cyclic group $(Z/pZ)^*$, and otherwise it is the primitive root of some subgroup. At $i>1$, we obtain the primitive roots of the subgroups of the $(Z/pZ)^*$ residue group with the index $i>1$.

The study of the distribution law of prime numbers p on their belonging to $\mathcal{P}(a,i)$ had the character of consistent statistical tests on the set of natural numbers contain-

ing the first 500,000 primes. At the first stage, primes p were chosen from the set $p_1, ..., p_{500000}$. With this $x = p_{500000}$.

For each $n \in \{2, ..., x\}$, we had to solve two problems: check n for simplicity, and if $n = p \in \mathcal{P}$, then $p - 1$ was decomposed into simple factors, i.e. systematically solved two non-simple problems of checking numbers for simplicity and decomposition into simple factors. An effective algorithm for solving them was created based on probabilistic methods in the theory of elliptic curves.

As a result of analyzing $a \in \{2, ..., x\}$, $\mathcal{P}(a, 1), ..., \mathcal{P}(a, l)$ sets were obtained for some $l < x$ and absolutely exact values of their powers were calculated, i.e. $|\mathcal{P}(a, 1)|, ..., |\mathcal{P}(a, l)|$, and then estimates of $c(a, 1) = \frac{|\mathcal{P}(a,1)|}{\pi(x)}, ..., c(a, l) = \frac{|\mathcal{P}(a,l)|}{\pi(x)}$ were obtained.

At the next stage, work was also carried out for primes from the interval as $\{p_{500001}, ..., p_{1000000}\}$ interval and the values of the $c(a, 1), ..., c(a, l)$ constants were calculated using the same scheme. At the same time l increases. The $\{p_1, ..., p_{500000}\}$ and $\{p_{500001}, ..., p_{1000000}\}$ sequences were combined, and the estimates of the generalized Artin constants were again calculated and the process of their refinement was studied on the basis of the theory of large numbers in probability theory. This procedure continued until $x = p = 179424673$ and this is a ten million prime numbers. It was found that $c(a, 1), ..., c(a, k)$ in probability converges to some values, the exact values of which are irrational and possibly transcendental numbers. In the process of estimating the $c(a, i)$ constants, two important theorems were proved:

Theorem 1. *For any $a \in \{2, 3, ..., k, ..\}$ that is not a square, i.e. $a \neq k^2$ The number of non-empty classes of primes tends to infinity at $x \to \infty$.*

Theorem 2. *For any $a \in \{2, 3, ..., k, ..\}$ that is not a square, i.e. $a \neq k^2$ The number of prime numbers in $\mathcal{P}(a, i)$ tends to infinity at $x \to \infty$.*

These theorems are the basis of the convergence of a sequence of statistical tests to marginal values. Since for any $x \in N$ it is obvious that $\bigcup_{i=1} \mathcal{P}(a, i) = \pi(x)$ and $\mathcal{P}(a, i) \bigcap \mathcal{P}(a, j) = \phi$ at $i \neq j$, it follows from this that $\sum_{i=1}^{k} c(a, i) = 1$ and this is true for all values of $x \to \infty$. The review [5] provides an estimate of $c(2, 1)$, which is identified by $c(2, 1)$ in our sense, but $c(2, 1)$ differs from the estimate of $c(2, 1)$ starting from the fifth decimal place and this is a theoretical error of the survey works.

For different $a \in \{2, 3, 5, 6, 7, 8, 10, 11\}$, the behavior of the $c(a, i)$ constants is complex group-theoretic and number-theoretic. The study of their dynamic properties is beyond the scope of this work. It should be noted that the results of computer simulation of the processes of distribution of primes are calculated with an accuracy of the eleventh decimal place for estimates of $c(2, 1), c(3, 1), c(5, 1), c(6, 1), ...$ values. This cannot be asserted for classes by the $i \geq 2$ index. To achieve the same accuracy with $i \geq 2$, it is necessary to significantly increase the number of prime numbers. With an increase in the i class index $\mathcal{P}(a, i)$ more than three requirements and the volume of the analyzed primes increases in accordance with the unexplored laws.

3 Statistical analysis of the distribution of prime numbers in classes

Probability-theoretic interpretation of the constant $c(a) = \frac{\pi(x,a)}{\pi(x)}$ at $x \to \infty$. Consider the probability space (Ω, F, P) based on $\Omega = \{\omega_1, ..., \omega_n, ...\} = \{p_1, ..., p_n, ...\}$. Obviously at $x \to \infty$ the numbers are $\pi(x) \to \infty$, $\pi(x,a) \to \infty$, but $\pi(x,a) = |\mathcal{P}(a,1,x)|$, $\pi(x) = |\mathcal{P}(x)|$, $c(a,1,x) = \frac{|\mathcal{P}(a,1,x)|}{|\mathcal{P}(x)|}$ and at $x \to \infty$ it is obvious that $\frac{|\mathcal{P}(a,1,x)}{|\mathcal{P}(x)|} \to c(a,1)$ is where $x \in \mathcal{P}$, $\mathcal{P} \to \infty$, $\mathcal{P}(a,i,x) = \left\{p|p \le x \& \frac{(p-1)}{card_a(p)} = i\right\}$ is at $x \to \infty$ $\mathcal{P}(a,i,x) \to \mathcal{P}(a,i)$. Thus $c(a) = lim\frac{\pi(x,a)}{\pi(x)}$.

It follows from Artin's hypothesis that with $c(a,1)$ there is precisely the probability of a random event $\mathcal{P}(a,1)$ consisting of a choice of $\Omega = \{p_1, ..., p_n, ..\}$ of a prime number p for which a is an original root of the cyclic group $(Z/pZ)^*$. To estimate this probability, the law of large numbers and the method of successive statistical tests were used. The essence of the method is that the first test group was reduced and calculated for $\{p_1, p_2, ..., p_{500000}\}$ for each $a \in \{2, 3, ..., 16\}$ evaluation of the values of $c(a,i,x)$ at $x = p_{500000}$ for all possible values of $i \in \{1, 2, ..., k, ..\}$, that is, $\tilde{c_1}(a,1,x), ..., \tilde{c_1}(a,k,x), ...$ was calculated on the next iteration, the same tests were performed for the second iteration on the set $\{p_{500001}, ..., p_{1000000}\}$. $\tilde{c_1}(a,1,x), ..., \tilde{c_k}(a,1,x), ...$ Estimates were obtained at the same time $\tilde{c_1}(a,1,x), ..., \tilde{c_k}(a,k,x), ...$, provided that the first and second samples were combined and computed values and were determined by $|\tilde{c}(a,i,x) - \tilde{c}(a,k,x)| \le \varepsilon$ for all x. The main focus was on $c(a,1,x)$. As a result of some iterations, it was found that for all a the estimates obtained:

$$\mathcal{P}(x) = \{p|p \le x\} \tag{10}$$

$$\mathcal{P}(a,i,x) = \left\{p|p \le x \& \frac{(p-1)}{card_a(p)} = i\right\} \tag{11}$$

the order of the cyclic group of the subgroup $(Z/pZ)^*$. If $l = p-1$, then a is an original root, and if $l < p-1$ is the original form of the $c(a)$ Artin measure, $c(a,i)$ is a measure of classes by $\mathcal{P}(a,i)$ in $\mathcal{P}$. At that $c(a,i) = \frac{|\mathcal{P}(a,i)|}{|\mathcal{P}|}$ and $\sum_{i=1}^{\infty} c(a,i) = 1$.

This applies only to classes with indexes $i = 1$. For $i \ge 2$ it is necessary to increase the number of statistical tests. This is naturally due to the fact that the classes $\mathcal{P}(a,i,x)$ for $i \ge 2$ from numerical theorems contain less than prime numbers. In [1] it is stated that this decrease should be of the order of $1/i^2$, but this is an erroneous assertion. The degree of decline essentially depends on the properties of a and requires a separate study. Case $a \in \{4, 9, 16\}$ requires separate investigations, because these numbers cannot be primitive roots of that number p, in accordance with the Fermat theorem [3]. cannot be generating elements of groups $(Z/pZ)^*$. However, they are generating elements of the subgroups of the group $(Z/pZ)^*$ with even indices. All classes with odd indices are empty sets. Table 1 shows the constants for $c(a,1)$ for all a except $\{4, 9, 16\}$. Analysis of the table. The table contains over a thousand

columns. The analysis of these data is numerically theoretical and group-specific and goes beyond the scope.

The simulation process of the dynamics of the formation of prime numbers was constructed on the following assumptions. Suppose that an ordered set of prime numbers $\mathcal{P} = \{p_1, p_2, ..., p_k, ...\}$ is given, whose elements are ordered in ascending order. All this set was split into a subset of 500,000 primes. The number of 500,000 is due to the limitations of MS Excel, as a statistical analysis tool, on a number of characteristics of the process of generating prime numbers. Only one restriction is important. We always select 500,000 consecutive primes of the set $\mathcal{P}$. In the current version of Excel, this number can be increased to one million. If you use a powerful computer, you can choose a larger number instead of a million.

The implemented version of the study of dynamic processes for the formation of primes includes the following indicators: the number of a simple number in the p in the ordered set of $\mathcal{P}$, the value of a simple number of p, the value of the recursion length of the numbers $card_a(p)$ at the same value of a for all prime numbers $\mathcal{P}$, the index $ind_a(p)$ of the index of the class, $ind_a(p) = \frac{(p-1)}{card_a(p)}$, the value of the residues modulo any natural module $n > 1$, for all classes and any other analytic properties of primes or factors of the decomposition of the number of $p-1$ into simple factors. For each simple multiplier p_i in the $p - 1 = \prod_{i=1}^{n} p_i^{a_i}$ decomposition, one parameter of the dynamic process of generating primes is presented, with separate indicators that can be analyzed for any other indicators, the values for them are deducted by the modulus of the natural number $n > 1$. The only exception is $ind_a(p)$. The number of controlled indicators analyzed in the Excel environment can be expanded.

According to the idea of experimental mathematics on the first iteration, we proceed from hypothetically known data. But it is also the basis for obtaining experimental information on the basis of which the analytical methods of the theory of numbers yield an expanded representation of the hypothesis in the form H_i. It is possible that at the same time the hypothesis can be corrected or even rejected as not true. From the point of view of information technology in mathematics, the hypothesis H_i is used to develop from the point of view of deepening the experimental mathematics of the model of in-depth studies at the level I_1.

The iterations process is continued until an analytically based solution of the generated hypothesis is obtained. Since the Artin generalized hypothesis is considered in the paper, we present the results of the estimation of the constant $c(a, i)$ for the case $a = 4$ and $i = 2$. The number $a = 4$ is a perfect square, and therefore it cannot be a primitive root. In terms of Artin's generalized hypothesis, this is as interesting and important as in the case when a is an original root.

Based on the data presented in [6], we obtained estimates for $c(a, i)$ for $a = 2, ..., 10$ and $i = 1, 2, ...9, ...$ It is shown that their values are stable for class $\mathcal{P}(4, 2)$ ie class with $ind_4(p) = 2$ to within a fourth decimal place. They are presented in the table 1.

An analysis of the data in the tables shows that for these numbers Artin's hypothesis is true on the set of primes $|\mathcal{P}| = 10^9$.

Table 1: The distribution of prime numbers in 1 to 9 classes in the generalized artin conjecture

a	$\mathcal{P}(a,1)$	$\mathcal{P}(a,2)$	$\mathcal{P}(a,3)$	$\mathcal{P}(a,4)$	$\mathcal{P}(a,5)$	$\mathcal{P}(a,6)$	$\mathcal{P}(a,7)$	$\mathcal{P}(a,8)$	$\mathcal{P}(a,9)$
2	0.3740	0.2805	0.0664	0.0467	0.0189	0.0498	0.0089	0.0351	0.0074
3	0.3739	0.2992	0.0666	0.0561	0.0190	0.0332	0.0089	0.0140	0.0074
4	0	0.5609	0	0.0935	0	0.0997	0	0.0701	0
5	0.3937	0.2657	0.0700	0.0664	0	0.0473	0.0094	0.0166	0.0078
6	0.3741	0.2805	0.0665	0.0748	0.0189	0.0498	0.0089	0.0140	0.0074
7	0.3741	0.2827	0.0664	0.0684	0.0188	0.0503	0.0089	0.0170	0.0074
8	0.2243	0.1683	0.1995	0.0281	0.0114	0.1496	0.0054	0.0211	0.0222
9	0	0.5983	0	0.1122	0	0.0666	0	0.0281	0
10	0.3741	0.2804	0.0665	0.0713	0.0189	0.0499	0.0089	0.0166	0.0074

4 CONCLUSIONS

The results of experimental mathematics in table 1 of the first iteration confirm that Artin's hypothesis is correct. The estimates of the constants are obtained with the accuracy of the third decimal place. These tables confirm Artin's generalized hypothesis for $a = 2, ..., 8$ and the assumption that $\sum_{i=1}^{\infty} c(a, 2i) = 1$. The results obtained are the basis for constructing an analytical proof of Artin's hypothesis and its generalization.

References

[1] Ambrose D. (2014). *On Artin's Primitive Root Conjecture*. der Georg-August-Universitat Gottingen.

[2] Artin E. (1982). *Collected papers*. Edited by Serge, Lang and T. John, Springer-Verlag, New York.

[3] Crandall R., Pomerance C. (2005). *Prime Numbers A Computational Perspective*. Springer, Portland.

[4] Manin Yu. I., Panchishkin A. A. (2009). *Introduction to the modern theory of numbers*. MTSNMO, Moscow.

[5] Moree P. (2012). Artin's Primitive root conjecture a survey, *Journal*. Vol. **12**, pp. 1305-1416.

[6] Vostrov G., Opiata R. (2018). Computer modeling of dynamic processes in analytic number theory, *Jornal* Vol, **26**, pp. 240-247.