CALCULATING THE NUMBER OF TWIN PRIMES WITH SPECIFIED DISTANCE BETWEEN THEM BASED ON THE SIMPLEST PROBABILISTIC MODEL

Sarsengali ABDYMANAPOV[1], Serik ALTYNBEK[2],
Alua TURGINBAYEVA[3]

Building modern cryptosystems requires very big primes. This research studies the characterization of primes, the problem of detecting and assessing the number of primes in a numerical series and provides other results regarding the distribution of primes in a natural series. It should be noted that the results were obtained within the framework of the simplest probabilistic model. This research verifies the prospects of obtaining an estimate for the number of r-pairs based on some statistical models of their formation.

Keywords: composite numbers; divisibility criterion, primes; sequence of numbers; twin primes, cryptography.

1. Introduction

Primes are involved in modern cryptography – the science of secret message coding and decoding. Since many cryptographic systems rely on the research of primes and related theoretical and numerical problems, technological and algorithmic advances in this field have become paramount. Our ability to find big primes and prove that they are prime predetermined our ability to influence the situation that makes cryptographers comfortable.

Based on estimate made for the number of twin primes on a particular interval, one can build a pseudorandom number generator or use distribution of primes in cryptography. The latter will be the subject of our research – calculating the number of primes on a specified interval and an attempt to solve the problem of probabilistic distribution of twin primes.

Numerous publications have been devoted to methods for synthesizing the PRBS generators and analyzing their characteristics [1-2]. In [3], there are outlined principles of system approach to certificating PR number generators in information security systems.

Apparently, PRBS generators built based on shifters with linear feedbacks are the most widespread generators. Such methods are widely studied and calculated. Thus, we shall have an alternative method for increasing cryptographic stability.

[1] Kazakh University of Economics, Finance and International Trade, Kazakhstan, email: rector@kuef.kz

[2] Kazakh University of Economics, Finance and International Trade, Kazakhstan, email: serik.altynb@gmail.com, Serik_AA@bk.ru

[3] Eurasian National University by named L.N. Gumileva, Kazakhstan, email: tasheart@mail.ru

Specific instances of the prime number theorem were formulated and proved by Carl Friedrich Gauss and Bernhard Riemann [4]. Existing methods require *n* steps $O(n^e)$ for arbitrary whole numbers, where *e* is a certain fraction, such as 1/3 or 1/4. Thus, experimental results show that each of these methods faces certain untestable numbers at about $n = 10^{65}$. It is easy to show that if the matter at hand is the distribution of primes in arithmetic progressions, there exist similar approaches to specific cases, such as the zeta function in the similar generalized Riemann hypothesis [5].

The local density of primes in a natural series appears to be essentially random. However, the following theorem establishes the asymptotic density of primes. The known prime number theorem gives an asymptotic assessment of the density of primes that are less than or equal to a positive real number.

Let's consider a case of twin primes: two primes that differ by 2. One could find such pairs easily: 11, 13 or 197, 199. Relatively bigger-pairs are not that easy to find, but they can still be found. The biggest ones are as follows:

$835335*2^{39014}$ ± 1, pair $361700055*2^{39020}$ ± 1, pairs of double numbers $2409110779845*2^{60000}$ ± 1, and $665551035*2^{80025}$ ± 1, and $154798125*2^{169690}$ ± 1. Is it possible to predict asymptotically how many such pairs exist within set boundaries? Let us try to think heuristically. The probability of a random number in the neighborhood of x being a prime number was about 1/ln x. Thus, the hypothesis emerged that $\pi_2(x) \sim 2C \int_2^x \frac{dt}{\ln(t)^2}$.

Using the expression for the probability density of primes in the natural series, Davenport (1967), suggested that the number $\pi_2(x)$ of twin primes smaller than *x* is asymptotically equal to

$$\pi_2(x) \approx 2C\int_2^x \frac{dt}{\ln(t)^2}, \quad C \approx 0.66 \tag{1}$$

and, therefore, infinite. Recently received a fundamentally new result in this direction (see Yitang, 2014), which asserts the existence of an infinite number of Primes such that the distance between each pair does not exceed a certain number (70 million). This obviously means that at least one set of pairs with distance *r* less than 70 million is infinite.

Subsequently, Zh. Yitang's method [6] and close methods have shown that infinite, at least one set of pairs with distance *r*, is less than 246 [7].

The proposed work involves verification of the possibility to obtain an estimate of the number of r-pairs on the basis of some statistical models of their formation, the simplest probabilistic models of the empirical formula of the number of r-pairs, including the case of *r=2* is given and numerical comparison with real data provided.

2. Basic definitions

Place the number of positive integers in a matrix *B*. The first *d* series members write down - first column of the matrix B. Second d series members write similarly, as the second column, etc. the result is a matrix of size (d, ∞). In each row of this matrix S_a, there is an arithmetic progression with difference d and with the appropriate initial member of the a-element of the first column of the matrix (1≤a≤d) (close to [8]). If the numbers *a* and *d* are relatively Prime, we will call such string a given string (the given sequence), since this line corresponds to a given deduction. Consider the submatrix B(d,x) of matrix B containing only the first x+1 columns. It fully contains the segment of the natural number from one to *xd*. We denote, as is customary, the number of primes in this interval using π(xd). According to the theorem of Dirichlet, in each given row of the matrix B contains infinitely many primes[1]. Moreover, for any x in each given row of the matrix B(d,x) and the Dirichlet theorem, the number π(d,x) of Primes is the same and proportional to the number of such lines:

$$\pi(d,x) \approx \pi(xd)/\varphi(d) \tag{2}$$

where, $\varphi(d)$ – the Euler function [9].

If the difference between two numbers is equal to $r < d$, they are in different rows of matrix B. Let $a_1 + nd, a_2 + md$ two numbers $(a_1 \neq a_2, n \geq m)$ and

$$\left|(a_1 - a_2) + (n-m)d\right| = r. \tag{3}$$

If $a_1 - a_2 > 0$, then (3) *n-m=0* – numbers with a difference r < d lie in two lines defined by the equation $a_1 - a_2 = r$ and in the same column. If $a_1 - a_2 < 0$, then (3) n-m=1 and $d - (a_1 - a_2) = r$. In this case, the number of the specified distance *r* lie in the lines a_1 and a_2, but in adjacent columns.

Let's place the numbers from the matrix B columns on a circle and define its direction from smaller values to bigger ones. Then any two numbers a_1 and a_2 are the two distances: from a_1 to a_2 and from a_2 to a_1. A necessary condition for the existence of infinitely many twin primes with a distance *r* within a pair is the existence of any matrix B for $r < d$ there are at least two given lines with the distance r between them. Indeed, for any value of d in the matrix B contains the whole row of natural numbers. If there is a matrix B where all rows of c primes are at a distance different from r, then the distance between a pair of Primes is not implemented ever. However, obtained result shows that matrix B does not exist as well [10] any specified matrix B rows are at even distances smaller than *d*. Thus, necessary condition for the existence of twin primes with a distance smaller than *d* for any *d* is met.

3. Probability model

In the framework of presenting natural series as matrix B, we can consider the emergence of twin primes with the given distance as a random event. Namely, let's consider the matrix B=B(d) and let the number of primes n_1 and n_2 be on a certain segment of a sequence from υ to υ'in two specified rows with the distance *r* between them. Let's consider the simplest probabilistic model, which is that the probability of value being a prime number does not depend on the numbers in the sequence and is the same for all positions occupied by odd numbers in this sequence. Based on this assumption (see Appendix), the expected number of cases when the Primes of both sequences (rows of matrix B) are clearly related to each other positions in these sequences is equal to

$$\tilde{\pi}_r((\upsilon' - \upsilon), d) = \frac{n_1 n_2}{\bar{n}} \tag{4}$$

where $\bar{n}$ – an odd number of members of the progression from number υ to number υ'. If d is an even number, then $\bar{n} = (\upsilon' - \upsilon)/d$ if d is odd, then $\bar{n} = (\upsilon' - \upsilon)/2d$. Thus, the number of r-twin primes can be estimated by the ratio

$$\pi_r(\upsilon' - \upsilon) = \frac{1}{\bar{n}} \sum_{\substack{(i,d)=1 \\ (j,d)=1 \\ dist(i,j)=r}} n_i n_j \tag{5}$$

where d - the difference between progression, n_i - the number of Primes in row number *i* in the corresponding segment of the sequence, the summation is for all rows *i, j* of matrix *B*, with numbers co-prime with d ((*i*,*d*)=1, (*j*,*d*)=1) and with distance between them equal to r *(dist(i,j)=r)*. In particular, to estimate the number of pairs with difference not exceeding a given r_0 you can use derived from (5) ratio of

$$\sum_{r \le r_0} \pi_r(\upsilon' - \upsilon) = \frac{1}{\bar{n}} \sum_{r \le r_0} \sum_{\substack{(i,d)=1 \\ (j,d)=1 \\ dist(i,j)=r}} n_i n_j \tag{6}$$

Estimations of the number of r–pairs by the formulas (5, 6) are essentially comparable to a direct calculation of this quantity. In the work of these assessments are used to analyze the adequacy of the accepted model. Calculating the number of Primes $n_1 \le xd$ and $n_2 \le xd$ by the formula (2) and using the known ratio of $\pi(x) \sim \frac{x}{ln(x)}$ we obtain the estimate $\Pi_r(x, d)$ be the number of r-pairs in the interval of the natural numbers from 1 to *xd*:

$$\Pi_r(x,d) = \mu(r)\pi_r(xd) = \frac{d\mu(r)(\pi(xd))^2}{\bar{n}(\varphi(d))^2} \approx C_r \frac{x}{\ln(xd)^2} \tag{7}$$

Where: $C_r = \frac{d\mu(r)}{(\varphi(d))^2}$ at even *d* and $C_r = \frac{2d\mu(r)}{(\varphi(d))^2}$ at odd *d*. (8)

Where: $\mu(r)$ – the number of pairs given strings with a distance *r* in the matrix B.

The relation (7) is essentially Brun's theorem [8] however, the coefficient C_r (7) can be selected in a unique way, depending on the difference d progression. Ratio (8) can be simplified as follows. In computed the quantity $\mu(r)$ equal distances r for all possible pairs of given strings [10].

$$\mu(r) = \varphi(d)\Pi_{\substack{p\backslash r \\ p\backslash d}} \frac{p-2}{p-1} \tag{9}$$

Where: $\varphi(d)$ – Euler's function, *p* is the Prime number "p\r" means that p is not a divisor of r, and "p\d" means that p is a divisor of *d*. If we apply *μ(r)* into (8), we will get: $C_r = \frac{d\Pi_{\substack{p\backslash r \\ p\backslash d}} \frac{p-2}{p-1}}{\varphi(d)}$ at even *d* and $C_r = \frac{d\Pi_{\substack{p\backslash r \\ p\backslash d}} \frac{p-2}{p-1}}{\varphi(d)}$ at odd *d*.

(10)

Thus, formulas (7, 10) give efficiently computable estimate of the number of r-pairs in the interval of the natural numbers, starting with unit[4]. This assessment has a parameter – difference between d, which determines the coefficient C_r. In particular, if *d* is a prime number, then (10) we will be as follows $C_r = 2\frac{d(d-2)}{(d-1)^2}$, and this ratio does not depend on r. If *d* is the product of *m* consecutive primes $p_1, p_2, \dots, p_m$, starting with two, then (10) will be as follows at r=2 (the case of twin primes): $C_2 = \frac{d\Pi_{2\le i\le m} \frac{p_i-2}{p_i-1}}{\varphi(d)} = 2\frac{\Pi_{i=2}^{m}[p_i(p_i-2)]}{\Pi_{i=2}^{m}(p_i-1)^2}$ (11)

This factor asymptotically *m* is close to the value 2C in the formula (1) (Twin Prime Constant). As with any value of *d* the set of simple pairs of twins is the same, there is a possibility to improve the assessment of the choice of *d* parameter.

Another number of pairs with distance smaller than r_0 easily follows from the previous one: $\sum_{r\le r_0} \Pi_r(x,d) = (\sum_{r\le r_0} C_r)\frac{x}{\ln(xd)^2}$ (12)

At *d*=2*p* from (10) we obtain for the case $r_0 < p$:

$$\sum_{r\le r_0} \Pi_r(x,d) = (2pr_0\frac{p-2}{(p-1)^2})\frac{x}{\ln(x)^2} \tag{13}$$

[4] Obviously, a similar formula can be obtained for any segment of the natural numbers, if we use the corresponding estimate of the number of Primes.

4. Numerical experiments

Let’s use the formula (6) to compute the total number of r –pairs for the first 25 even values of distances r = 2,4,...,50 and even values of d ranging from 100 to 400. We have calculated this value twice – (a) for 10000 members in arithmetic progressions, starting with the first ($\upsilon = 1,\ \upsilon' = 10000$) and (b) for 100,000 members of the progression, starting with 10^5 ($\upsilon = 10^5,\ \upsilon' = 10^6$). In addition, let us compute the same value by the formula (12) for case (a). Figure 1 shows the relationship of the real number of r-pairs with the calculated formula for described cases. The average values of relationships computed directly by (6) is close to the unit element. In all cases, the difference is close to 5%. The relationship magnitude depends on the *d* value. Curve calculated according to the formula (12) also deviates from unity by no more than 5%, although it has a different shape.

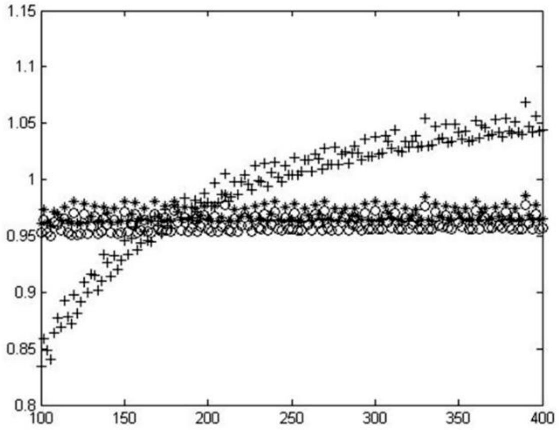

Fig. 1. Relationship between the actual number of r-pairs to the number calculated according to the formulas (6, 12).

In Y-direction, there are relationship values. The points marked with an asterisk and a zero are calculated by (6) for progression fragments in the range from 1 to 10.000 members and from 100.000 to 100.000 members, respectively. Points marked with plus are calculated by (12) for progression fragments in the range from 1 to 10.000 members, as well as the points marked with an asterisk. Calculations with the same parameters, but just for twin primes, are shown in Figure 2. The value of relationships between the actual numbers calculated by (5) for two fragments of the sequence is largely the same. It is seen that for the formula (5) (denote the circle and asterisk), the relations in the ranges 0.65-0.90, sometimes reaching 0.95. There are two groups of values for calculating by formula (7): from 1 to 1.05 and from 0.8 to 0.85.

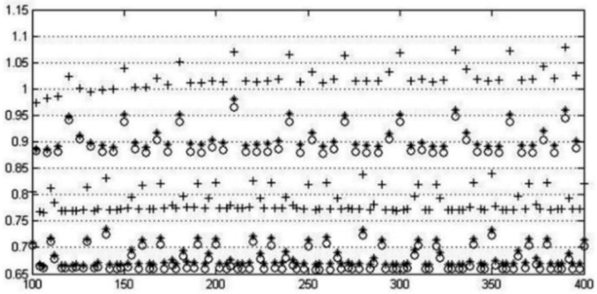

Fig. 2. Relationship between the actual number of 2-pairs (twin primes) and numbers calculated according to equations (5,7,).

In X-direction, value d varies from 100 to 400 with a step of 2. In Y-direction, there are the average values of relations. The points marked with stars and circles calculated by the formula (5) for fragments of the progressions from 1 to 10,000 members and from 100,000 to 100,000 members, respectively. Points marked with a plus sign - the same piece rate, which is used to calculate the stars, but are calculated according to the formula (7). Different values of progression *d* difference provide different estimates for the number of twin primes. We have tested $d = 2 \times 3 \times 5 \times 7 \times i$ series as a candidate for this optimal value, where *i* varies from 1 to 66. Figure 3 shows the results of calculations made by (5, 7). Calculations made by the model formula (5) have 2% -3% error while those made by the formula (7) – up to 10% error.

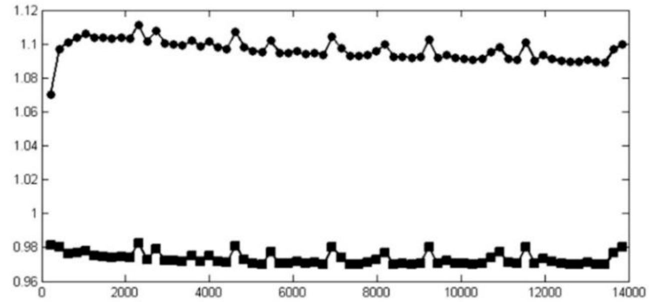

Fig. 3. Relationship between the actual number of twin primes and those calculated according to equations (5,7) when the value is equal to d =2×3×5×7×i, where *i* varies from 1 to 66.

In X-direction, there is the *d* value; in Y-direction – relationship values. Squares fo formula (5), circles for (7). We now compare the numerical standard of a hypothetical estimate of (1) a variety of simple pairs of twins $\pi_2(x)$ with the estimate (7) on the set of matrices B=(d,x) with values of *d* in the sequence of numbers of the form 210i, where i runs through the interval from 1 to 5000, x=10000. The ratio of the number of pairs of twins, calculated by the formula (7) to the number calculated by the formula (1) shown in Figure 4. Although this ratio is not higher than 0.95, it is essential that the schedule clearly has a horizontal asymptote, i.e., misalignment estimates could be corrected by multiplying our estimate of the empirical constant.

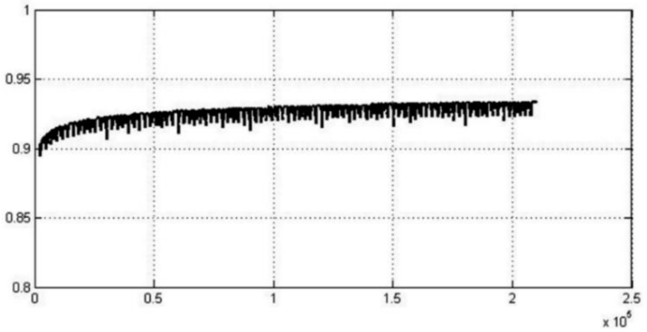

Fig. 4. The ratio of the number of twin primes, calculated by the formula (7) to the number calculated by the formula (1) when the value of d =2×3×5×7×i, where i varies from 1 to 5000. In X-direction, there is the *d* value; in Y-direction – relationship values.

5. Conclusion

The results of numerical experiments show that the estimate of the number of numbers-twins in the simplest probabilistic model correlates with the actual number, however, the absolute error can be quite large. It is shown that for different values of *d*, namely – for different ways to split the natural numbers in arithmetic progression. We have received effectively computable estimate of the number of *r*-pairs on the interval of natural numbers. At certain parameter values, it asymptotically coincides with the known.

Acknowledgment

The present work was carried out on the basis of the scientific and technical program No. BR 05236075 of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan.

REFERENCES

[1] *Y.S. Kharin, V.I. Bernik, G.V. Matveev, S.V. Agievich*, “Mathematical and Computer Foundations of Cryptology”, in Minsk: New Knowledge, 2003, pp. 382.

[2] *B. Schneier*, “Applied cryptology: Protocols, algorithms, source texts in SI language”, in Moscow: Triumph, 2003, pp. 816

[3] *A.V. Potiy, A.K. Pesterev*, “Principles of the system approach to certificating PR number generators in information security systems”, in Radio engineering, **vol. 104**, 1997, pp. 163-172

[4] *C. F. Gauss*, “Disquisitiones arithmeticae”, in Yale University Press, 1966.

[5] *K. Soundararajan*, “The distribution of primes. Equidistribution in number theory, an Introduction”, in Springer Netherlands, 2007, pp. 59-83.

[6] *Zh. Yitang*, “Bounded gaps between primes”, in Annals of Mathematics, **vol. 179**, 2014, pp. 3.

[7] *A. Granville*, “About the cover: a new mathematical celebrity”, in Bulletin of the American Mathematical Society, **vol. 52**, no. 2, 2015, pp. 335–337.

[8] *S. Baibekov, A. Durmagambetov*, “An infinite number Twin prime”, in General Mathematics, **vol. 19**, 2016.

[9] *H. Davenport*, “Multiplicative Number Theory”, in Springer-Verlag, Berlin, 1967.

[10] *M. Orazov*, “The lower bound for a number of representations in additive number theory problems”, in Modern Trends in Engineering Sciences: Proceedings of International Scientific Conference, Ufa, Leto, 2011, pp. 71-73.

[11] *V. Brun, “*On Goldbach's law and the number of prime pairs”, in Archiv for Mathematik og Naturvidenskab, **vol. 34,** no. 8, 1915, pp. 3–19.