

Научная статья

УДК 511-37:004.4

DOI: 10.17853/2587-6910-2024-11-38-48

ЛОКАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ПРОСТЫХ ЧИСЕЛ

Александр Викторович Рожков

доктор физико-математических наук, профессор great.ros.marine2@gmail.com
ФГБОУ ВО «Кубанский государственный университет», Россия, Краснодар

Ирина Алексеевна Черешнюк

магистр 2 курса 89002454755@mail.ru ФГБОУ ВО «Кубанский государственный
университет», Россия, Краснодар

Аннотация. Исследование локального расположения простых чисел, проводимое в КубГУ с 2015 г. Поддержана в 2021 грантом Благотворительного фонда Владимира Потанина. В данной статье исследуются простые числа, и их локальные сгущения.

Ключевые слова: Теория чисел, язык программирования Julia, функция Эйлера, локальное распределение простых чисел

Для цитирования: Рожков А. В., Черешнюк И. А. Локальное распределение простых чисел // Новые информационные технологии в образовании и науке. 2024. № 11 С. 38–48 <https://doi.org/10.17853/2587-6910-2024-11-38-48>

THE LOCAL DISTRIBUTION OF PRIMES

Alexander Viktorovich Rozhkov

Doctor of Physical and Mathematical Sciences, Professor Kuban State University,
Russia, Krasnodar

Irina Alekseevna Chereschnyuk

Kuban State University, Russia, Krasnodar

Abstract. A study of the local location of prime numbers conducted at KubGU since 2015. Supported in 2021 by a grant from the Vladimir Potanin Charitable Foundation. This article examines prime numbers and their local thickenings.

Keywords: Number theory, Julia programming language, Euler's function, local distribution of prime numbers

For citation: Rozhkov AV, Cheresnyuk IA. The local distribution of primes *New information technologies in education and science*. 2024;11: 38-48. (In Russ.). doi:10.17853/2587-6910-2024-11-38-48

Цель проекта – проведение разведочных вычислений в области нерешенных проблем алгебры и теории чисел, с использованием нового перспективного языка программирования Julia [5]. Официальный сайт <https://julialang.org/> текущая версия 1.10.1. В 2021 г. проект был поддержан Благотворительным фондом Владимира Потанина. Промежуточные итоги проделанной работы представлены в [1–3]. В данной статье речь идет о локальном распределении простых чисел. Введены некоторые новые понятия.

Глобальное распределение простых чисел

Формула $\pi(n) \approx \frac{n}{\ln(n)-1}$ задает распределение простых чисел на всей числовой

прямой, но не на конкретном интервале. Это поведение функции $\pi(n)$ на бесконечности.

Однако именно локальное распределение важно для практики, в особенности для нужд криптографии. Сейчас в криптографии часто используются 1024 битные простые числа. Поскольку, $2^{1024} \approx 10^{300}$ то это 300-значные числа и простыми из них являются, в среднем, каждое 700-е число, т.е. плотность расположения простых чисел равна 1/700.

Простые числа расположены на прямой очень неравномерно. Как люди живут, в основном, в городах и поселках, так и простые числа где-то сгущаются, а где-то их встречается относительно мало.

Актуальным для практики является нахождение таких «сгущений» и «пустот». Нахождение, описание, исследование этих сгущений и есть **локальное распределение простых чисел**.

Следует отметить, что локальное распределение простых чисел не является общепринятым термином. Поскольку трудно формализовать само это понятие и предложить инструменты для его изучения.

Экстремальные сгущения простых чисел

Определение. Множество n подряд идущих простых чисел называется *плотной n -кой*, если они расположены на отрезке минимально возможной длины [6].

Очевидно, n -ки — это максимально возможные сгущения простых чисел. Как не уплотняемая застройка в городах. Понятие это не новое, поскольку обобщает многие хорошо известные частные случаи:

1. Пары простых чисел вида $(p, p+2)$ - называются **близнецами**. В ней плотность расположения простых чисел равна 2/3.

Поскольку символ p не несет никакой информации будем записывать плотные n -ки без его участия. Например, близнецы – это $(0,2)$.

2. Тройки простых чисел $(0, 2, 6)$ и $(0, 4, 6)$ называются левыми и правыми **триплетами**. У триплетов плотность 3/7.

4. Четверки простых чисел вида $(0, 2, 6, 8)$ называются *сдвоенными близнецами*. Получаем плотность $4/9$.

5. Пятерки имеют вид $(0, 2, 6, 8, 12)$ и $(0, 4, 6, 10, 12)$. Плотность $5/13$.

6. Шестерка единственная и имеет вид $(0, 4, 6, 10, 12, 16)$. Плотность $6/17$.

Одна из плотных 21-к имеет вид

$(0, 4, 6, 10, 12, 16, 24, 30, 34, 40, 42, 46, 52, 54, 60, 66, 70, 72, 76, 82, 84)$

(<https://pzktupel.de/ktuplets.php>) и ее плотность равна $21/85$.

Следует отметить, что для данного n может быть несколько разных структур, задающих плотную n -ку. Выше мы видели, что разных 3-к и 5-к по две штуки. А, например, при $n = 105$ разных структур 486.

T.J. Engelsma вычислил структуру плотных n -к до $n = 4507$ включительно <http://www.opertech.com/primes/k-tuples.html>.

Наряду с плотными n -ми можно рассмотреть более общие плотные (n, k) -ки, где $k = 0, 1, 2, \dots$. В этих новых n -ках плотность расположения простых чисел меньше, это образно говоря «пригороды сгущений простых чисел».

Определение. Множество n подряд идущих простых чисел называется плотной (n, k) -кой, если они расположены на отрезке длины $k + k_0$, где k_0 — длина отрезка плотной n -ки. Очевидно, $(n, 0)$ -ка — это просто n -ка.

Необходимо найти полные n -ки и (n, k) -ки. Выяснить насколько часто они встречаются, как расположены друг относительно друга. Не образуют ли они некоторую алгебраическую структуру и т. д. Это очень много скажет о локальном распределении простых чисел.

В настоящее время мы умеем (в частных случаях) решать три базовые задачи.

1. Проблема Шаблона. Нахождение структуры максимальной n -ки. Прежде всего нахождение длины минимального отрезка, на котором должна располагаться плотная n -ка. И после этого вычисление того, как на этом отрезке расположены простые числа.

2. Проблема существования n -ок. Шаблон — это необходимое условие существования n -ки из простых чисел. Числа, не входящие в шаблон, заведомо не все являются простыми. А чтобы реализовать теоретическую возможность нужно предъявить конкретный пример простого числа p , такого, что при помещении его в шаблон получится реальный набор из n простых чисел.

В настоящее время (февраль 2024 г.) реально найдены только 21-ки. Минимальные примеры 21-к имеют 30 цифр в десятичной записи <https://pzktupel.de/ktuplets.php>

Нами выдвинута гипотеза, обобщающая имеющиеся данные, что минимальный пример n -ки имеет разрядность $2^n - 11$, при $n > 15$. То есть при переходе от n -ки к $(n+1)$ -ке минимальный пример возрастает на 2 порядка, примерно в 100 раз.

Эта формула (если она верна) важна для теории чисел. Например, опровержением второй гипотезы Харди — Литтлвуда является плотная 447-ка. Потому, что отрезок, на котором она расположена, меньше, чем отрезок, на котором расположены первые 447 простых чисел.

Проблема в том, что такая 447-ка, если она существует, согласно нашей формуле, имеет примерно 900 знаков в десятичной записи.

3. Граф вложений n -к друг в друга. Практика показывает, что часто m -ка при $m < n$ вложена в некоторую n -ку. Поэтому, если мы нашли все m -ки, меньшие числа N , то n -ки, содержащие данную m -ку и меньшие N , нужно искать среди уже найденных m -к, что существенно облегчает вычисления.

Аналогичный граф, но гораздо более сложно устроенный, полезно изучить и для (n, k) -к.

ГРАФ ПЛОТНЫХ N -К

Поскольку четное число не может быть простым, то все четные числа внутри отрезка длины N , внутри которого заключена плотная n -ка, мы будем опускать. Если некоторое нечетное место занято простым числом, мы это пометим цифрой 1, а 0 будет означать отсутствие простого числа.

В этих обозначениях упомянутые выше близнецы, триплеты и сдвоенные близнецы примут вид (шаблон, слово из 0 и 1):

2-ка: (11). 3-ки: (1101); (1011). 4-ка: (11011).

5-ки: (1101101); (1011011). 6-ка: (101101101).

Определение. (графа Γ вложений плотных n -к) Вершинами графа Γ являются кортежи (слова) из 0 и 1, являющиеся шаблонами, задающими формат плотных n -к. Пусть v и w — вершины графа Γ длины $m < n$, они соединены ребром, если слово v является подсловом слова w , причем w имеет минимально возможную длину.

Нетрудно заметить, что это граф частично упорядоченного множества (ч.у.м.), где $v < w$, если слово v является подсловом слова w . На графе ч.у.м. выделяется только связи ближайших вершин, а остальные сравнения доопределяются в силу транзитивности частичного порядка. Это делается, чтобы не перегружать рисунок лишними деталями.

Например, упорядочение $1 < 2 < 3 < 4$ изображается тремя ребрами

$1 \leftarrow 2 \leftarrow 3 \leftarrow 4$ без ребер $1 \leftarrow 3$, $1 \leftarrow 4$, $2 \leftarrow 4$.

Экспериментально подтверждено, что, если есть некоторая плотная n -ка с шаблоном v , то есть и n -ка, ей центрально симметричная, записанная в обратном порядке. Поэтому плотных n -к почти всегда четное число.

Пока известно только 3 исключения. Это 2-ка, 4-ка и 6-ка — их шаблоны центрально симметричны 11, 11011, 101101101 и естественным образом меньшие вложены в большие.

Ниже приведен граф вложений n -к до $n=20$ включительно.

Он, очевидно, симметричен. Не является деревом, поскольку имеет 4 нетривиальных цикла.

Группа его автоморфизмов является элементарной абелевой 2-группой ранга 4.

Возможно, граф вложений симметричен для любого n и для любого n его группа автоморфизмов элементарная абелева 2-группа.

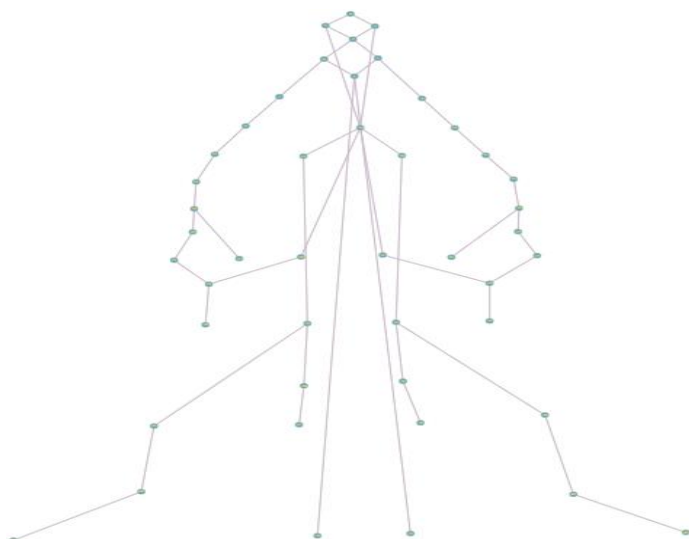


Рис. 1. Граф вложений плотных n -к до $n=20$ включительно

Поиск заготовок шаблонов

Шаблон плотной n -ки — это такие числа из отрезка $[0, N]$, что они не образуют полной системы вычетов ни по какому простому модулю (очевидно, можно ограничиться модулями, меньшими N).

Это нам подсказывает идею как можно описать локальное скопление простых чисел (не обязательно «сгущение», по сравнению с глобальной теоремой

$\pi(n) \approx \frac{n}{\ln(n)-1}$, а может быть даже и «разрежение»).

Нижеследующая гипотеза впервые была опубликована на международной конференции 2018 г. в г. Геленджик [4].

Геленджикская гипотеза. Пусть n - натуральное число, зафиксируем его. Пусть A — множество четных чисел, из отрезка $[0, 2n]$, содержащее 0, но не содержащее полной системы вычетов ни по какому нечетному простому модулю $q < 2n$.

Тогда существует бесконечно много простых чисел p , таких, что:

- а) все числа $\{p+a \mid a \in A\}$, являются простыми (слабая гипотеза);
- б) все остальные числа отрезка $[p, p+2n]$ составные (сильная гипотеза).

Гипотеза обобщает много предположений на тему простых чисел. Она не тривиальна даже в случае, когда множество A одноэлементно.

Например, если $n = 0$ и $A = \{0\}$, то слабая гипотеза означает, что простых чисел бесконечно много, что, конечно, верно. Если $n > 0$ и $A = \{0\}$, то сильная гипотеза означает, что существует бесконечно много простых чисел правее которых расположено не менее $2n$ подряд идущих составных чисел.

Если A — это структура плотной n -ки, то сильная и слабая гипотеза совпадают и означают, что число плотных n -к любой структуры бесконечно, в частности, не верна вторая гипотеза Харди-Литлвуда, поскольку существует плотная 447-ка.

Таким образом, если некий шаблон, не содержит полной системы вычетов ни по какому простому модулю, мы можем его использовать для поиска скоплений простых чисел.

Поскольку структура шаблона фиксирована, то поиск можно сильно облегчить, если перебирать не все нечетные числа, а только соответствующие требованию шаблона.

Программа **Source(M,m)** находит, по шаблону M и по модулю произведения первых m+1 простых чисел, числа, удовлетворяющие шаблону M. При этом используется два пакета языка Julia, а именно Mods — для использования китайской теоремы об остатках и пакет Nemo, содержащий нужные нам сведения по алгебре. Текущие версии Mods v2.2.4 и Nemo v0.42.1.

```
using Mods
using Nemo
function Source(M,m)
    l=1; K=[1];D=[];S=[];
    Pprime
= [2,3,5,7,11,13,17,19,23,29,31,37,41,47,53,59,61,67,71,73];
    for i=1:m
        l= l*Pprime[i]
        L= M.% Pprime[i+1]
        L=sort(unique(L))
        L= setdiff(0:Pprime[i+1]-1,L)
        for j in L
            for k in K
                d=crt(ZZ(k),ZZ(l),ZZ(Pprime[i+1]-j),ZZ(Pprime[i+1]))
                D=push!(D,d)
            end
        end
        K=sort(unique(D))
        D=[]
    end
    S= [K,length(K),l*Pprime[m+1]]
    return(S)
end
```

Например, возьмем шаблон плотной 6-ки — это слово 101101101 и его расшифровка M=[0,4,6,10,12,16]. И вычислим S=Source(M,6).

В этом случае по модулю S[3] = 510510 мы получаем S[2] = 385 претендентов. Т.е. нужно будет проверять всего 1 число из 663 нечетных, что убыстряет вычисления в 663 раза.

Вычисление Source(M,8) дает 85085 претендентов по модулю 223092870. В этом случае мы проверяем только 1 нечетное число из 1311.

Дальнейшее наращивание модуля вычислений ограничено оперативной памятью компьютера.

Не улучшаемое уточнение теоремы Мертенса

Локальное распределение простых чисел влияет на значения многих теоретико-числовых функций, в частности на значения функции Эйлера. Для вычисления этой функции натуральное число нужно разложить на простые множители, что является очень трудоемкой операцией.

Важно отношение $0 < \frac{\varphi(n)}{n} < 1$. Эта дробь достигает максимума, когда число n простое. И дробь, тем меньше, чем больше разных простых множителей имеет число n .

В известной теореме Мертенса [7] о сумме значений функции Эйлера присутствует логарифм, т.е. отклонение может расти неограниченно:

$$\sum_{i=1}^n \varphi(i) = \frac{3}{\pi^2} n^2 + O(n \cdot \ln(n)).$$

Более того, как и глобальное распределение простых чисел, это функция о поведении на бесконечности, ничего не говорящая о значении функции в конкретной точке, что важно для приложений.

Практические вычисления, ведущиеся уже 5 лет, позволили, в пределах первых 9 трлн. натуральных чисел, дать не улучшаемое уточнение формулы, в которой отклонение является знакопеременной функцией, меняющей знак буквально в каждой точке.

Введем среднее значение функции Эйлера $\bar{\varphi}(n) = \frac{1}{n} \sum_{i=1}^n \varphi(i)$ и перепишем формула Мертенса в удобной для нас форме:

$$\bar{\varphi}(n) = t \cdot (n+1) + \frac{O(n \cdot \ln(n))}{n}, \quad t = \frac{3}{\pi^2}.$$

Через $s(n)$ мы обозначим разность $s(n) = \bar{\varphi}(n) - t(n+1)$ — отклонение реального среднего значения функции Эйлера от значения, которое задает формула. Нелегко вычислить рекуррентную формулу для этого отклонения.

$$s(n+1) = \frac{n}{n+1} s(n) + \frac{\varphi(n+1)}{n+1} - 2t, \quad s(1) = 1.$$

Для практических вычислений она нехороша тем, что накапливает ошибки, поскольку следующий член зависит от предыдущего. Если же использовать сумму всех значений функции Эйлера, то поскольку эти значения целые числа, то вычисления производятся абсолютно точно и ошибка не накапливается.

В языке Julia мы использовали для чисел формат Big и поэтому вычисления велись до 78 знака после запятой.

Мы вычисляли 3 параметра. Сумма всех значений функции Эйлера — параметр S. Само отклонение $s(n)$ — параметр t. И сумма всех отклонений T.

Результаты оказались неожиданными и обнадеживающими.

1. Отклонение не выходило за пределы интервала $(-0.46; 0.46)$
2. Знак отклонения менялся в среднем через 1.4 шага.
3. Сумма отклонений качественно ведет себя как синусоида, с увеличивающимся периодом и увеличивающейся амплитудой.

Это говорит о том, что чем дальше от начала координат, тем больше сгущения простых чисел и больше «пустоши», где очень много чисел, имеющих много разных простых делителей, а простых чисел мало.

Вычисления производились по следующей программе.

```

using Nemo
function ros(m, n)
p =
BigFloat(«0.6079271018540266286632767792583658334261526480
3347929307365419136503872577341265»)
S = BigInt(0)
t= BigFloat(0.0)
T::BigFloat = 0.0
for i = m:n
  S += euler_phi(ZZ(i))
  t = BigInt(S)/i-p*(i+1)
  T += t
println(f,»i=«, i,»->«, «S=«, S, «->«, «T=«,T)
end
end
end

```

В программу вводилась сумма S предыдущих значений функции Эйлера и интервал (m, n), на котором нужно было производить новые вычисления.

Теорема. (Уточнение формулы Мертенса).

$$\sum_{i=1}^n \varphi(i) = \frac{3}{\pi^2} n(n+1) + n \cdot s(n),$$

где для всех $n < 9 \cdot 10^{12}$ выполняется неравенство $|s(n)| < 0,46$.

Эта формула очень естественна. Известно [7], что среднее значение функции Эйлера равно $\varphi(n) = \frac{6n}{\pi^2}$. Просуммировав средние значения функции Эйлера от 1 до n, получим арифметическую прогрессию:

$$\varphi(1) + \varphi(2) + \dots + \varphi(n) = \frac{6 \cdot 1}{\pi^2} + \frac{6 \cdot 2}{\pi^2} + \dots + \frac{6 \cdot n}{\pi^2} = \frac{6}{\pi^2} (1 + 2 + \dots + n) = \frac{3}{\pi^2} n(n+1).$$

Дальнейшее уточнение формулы Мертенса невозможно в силу знакопеременности остатка $s(n)$.

Более того, наша формула описывает не только поведение суммы функций Эйлера на бесконечности, но и ограничивает значения этой суммы в жестких пределах.

1. Функция $s(n)$ знакопеременная, средний период знакопостоянства равен примерно 1,4136 (близко к $\sqrt{2} \approx 1.4152$).

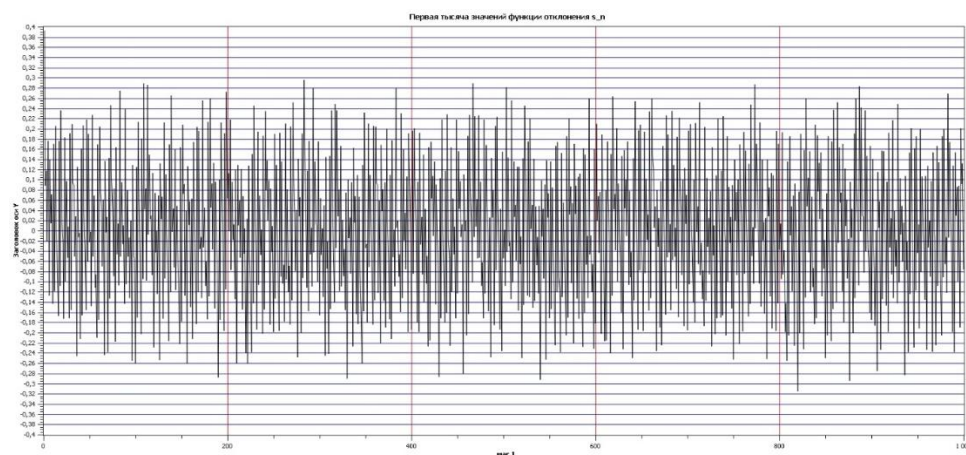


Рис. 2. Первая тысяча значений функции $s(n)$. По горизонтали шаг 1, по вертикали интервал $(-0,4; 0,4)$

2. Среднее положительное значение функции $s(n)$ равно 0.1189, отрицательное - 0.1189.

3. До 9 трлн. в интервал $(-0,45; 0,45)$ не попало только 2 значения:

$n=7016086529723$, $s(n)=0.4553$;

$n=8772508813703$, $s(n)=0.4514$

Возможно, на бесконечности функция отклонения $s(n)$ содержит логарифм.

В настоящее время установлено, что:

- значение отклонения из примерно 16 млрд. удовлетворяет неравенству $|s(n)| > 0.43$;
- значение из 250 млрд. удовлетворяет неравенству $|s(n)| > 0.44$;
- значение из 4 трлн. удовлетворяет неравенству $|s(n)| > 0.45$.

Для увеличения значения отклонения $s(n)$ на 0.01 нужно увеличить аргумент n в 16 раз. Функция отклонения $s(n)$ может выглядеть примерно так

$$s(n) \approx \frac{1}{100} \log_{16} n = \frac{1}{400} \log_2 n.$$

Поэтому переход границы 0.46 нужно ожидать в районе 60-70 трлн., что на домашнем компьютере займет примерно 30 лет.

4. Дисперсия с точностью до 6 знака равна 0,01986,

5. Плотность распределения вероятностей случайной величины $s(n)$ приведена на Рис. 3, с шагом 0.001 до 100 млрд.

Тут же приведен график нормального распределения для матожидания 0 и дисперсии 0,01986.

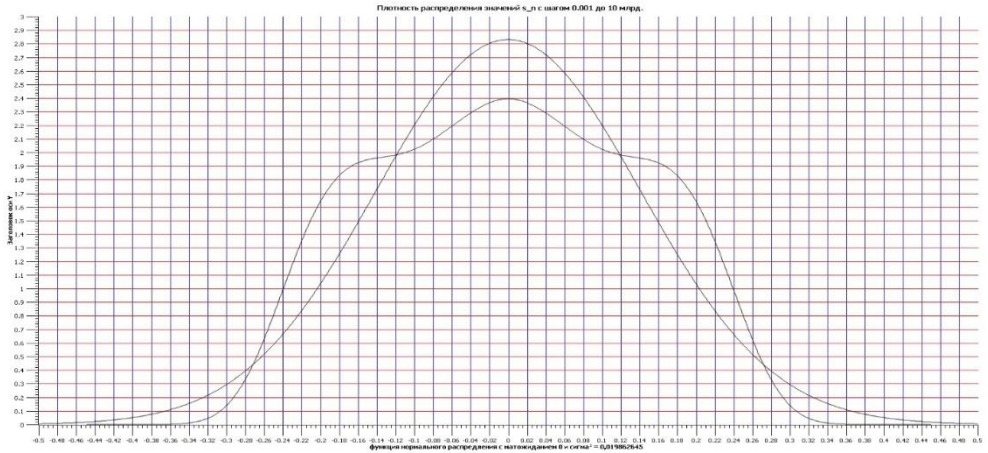


Рис. 3. Плотность распределения вероятностей и нормальное распределение до 100 млрд.

6. Вычислена функция суммы $S(n) = \sum_{i=1}^n s(i)$ Рис.4,

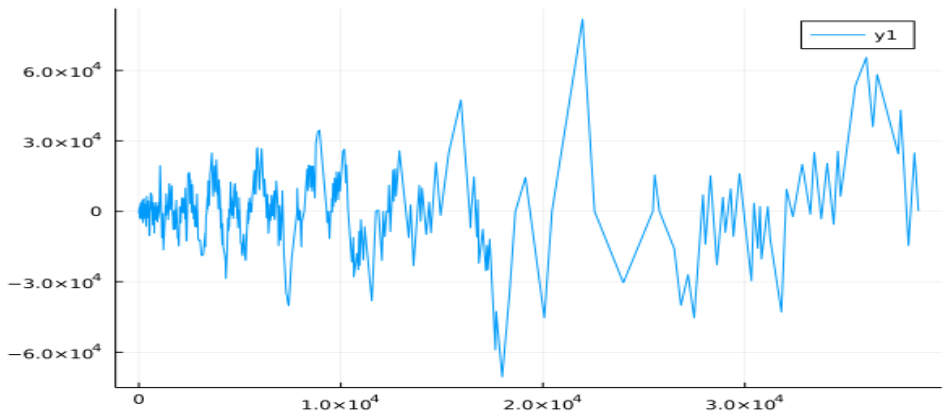


Рисунок 4. Суммы $S(n)$ на интервале до 3860 млрд., шаг 100 млн.

Обобщенный многочлен Эйлера и простые числа

Многочлен Эйлера x^2+x+41 знаменит тем, что задает простые числа для всех $x = 0, 1, 2, \dots, 39$. Интересно было бы найти многочлен вида $f(x) = x^2 + x + p$ такой, что простые числа получаются при всех $x = 0, 1, 2, \dots, p-2$.

Очевидно, такими p являются $\{3, 5, 11, 17, 41\}$. Других простых чисел с таким свойством, пока, не найдено. Мы провели численный эксперимент и проверили для первых 10 тыс. простых чисел p какая часть чисел $f(0), f(1), \dots, f(p-2)$ является простыми.

Результаты вычислений. Выяснилось, что максимальные значения достигаются при $p=101 - 0.68$ и при $p=107 - 0.67$. То есть, соответственно 68 из 100 и 71 из 106 чисел являются простыми. Это очень много.

Минимальные значения достигаются при $p=59053 - 0.0417$ и при $p=90403 - 0.0419$. В обоих случаях простыми являются только одно число из 25.

В среднем же простыми являются примерно 15 % чисел, т. е. каждое 6–7 число. Аналогичные вычисления можно проводить для многочленов любой степени.

Выводы

Поиск формулы простого числа не прекращается уже сотни лет. Тема вечная. Неожиданно, что, даже работая со значениями многочлена 2-й степени мы можем получить множество, где простые числа встречаются в 4 раза чаще, чем простые числа среди нечетных простых чисел натурального ряда.

Мы планируем проводить дальнейшие вычислительные эксперименты в трех направлениях. Искать простые числа среди значений:

1) многочленов вида $f(X)=x^2+x+p$, при p больше 8 млн.

2) других квадратных многочленов.

3) многочленов произвольной степени; произвольных функций $f: N \rightarrow N$, где N — множество натуральных чисел. Используя I и II модели вычисления плотности распределения простых чисел.

Список литературы

1. Рожков А. В., Рожкова М. В. Экспериментальная теория чисел: среднее значение функции Эйлера // Осенние математические чтения в Адигее: материалы II Международной научной конференции, Майкоп, 20–24 октября 2017 г. Барнаул: Алтайский гос. ун-т, 2017. Т. 2. С. 198–203.

2. Рожков А. В., Потапова Н. В., Оконешникова Е. А. Экспериментальная теория чисел – вычисления в программной среде Julia – I // Информационные технологии в математике и математическом образовании: материалы VIII Всероссийской с международным участием научно-методической конференции, посвященной 80-летию профессора Ларина С. В., Красноярск, 13–14 ноября 2019 г. Красноярск: Краснояр. гос. пед. ун-т им. В. П. Астафьева, 2019. Ч. 1. С. 10–15.

3. Рожков А. В., Барсукова А. С. Экспериментальная математика и язык Julia – локальное распределение простых чисел // Новые информационные технологии в образовании и науке. 2022. № 2 (6). С. 82–88. <https://doi.org/10.17853/2587-6910-2022-06-82-88>.

4. Рожков А. В., Потапова Н. В. Автоморфизмы графа вложений сгущений простых чисел // Теория групп и ее приложения: материалы XII международной школы-конференции по теории групп, посвященной 65-летию А. А. Махнева, Геленджик, 13–20 мая 2018 г. Краснодар: Кубан. гос. ун-т, 2018. С. 132–136.

5. Шеррингтон М. Осваиваем язык Julia. М.: ДМК Пресс, 2017. 416 с.

6. Forbes T. Prime clusters and Cunningham chains // Mathematics of Computation. 1999. Vol. 68, no. 228. P. 1739–1747. <https://doi.org/10.1090/S0025-5718-99-01117-5>

7. Чандрасекхаран К. Введение в аналитическую теорию чисел. М.: Мир, 1974. 178 с.