УДК 519.683+519.24+511.333

Распределение разностей между соседними простыми числами

Ю. К. Чернышев
Национальный аэрокосмический университет имени Н.Е. Жуковского «ХАИ»

Применением ЭВМ для построения гистограммы расстояний между соседними простыми числами (т. н. близнецами различных порядков) обнаружено существование периода величины 6 между локальными максимумами в статистическом ряде распределения этих величин. Для чисел, меньших 2850000000, показано, что наиболее часто встречается расстояние, равное 6. Путём нормирования получена приближённая универсальная модель распределения пробелов. Полученные результаты целесообразно применять для ускорения поиска простых сомножителей, что используется в современных методах шифрования данных.

Ключевые слова: *простые числа, числа-близнецы различных порядков, ряды распределения*

Застосуванням ЕОМ для побудови гістограми відстаней між сусідніми простими числами (т. з. близнюками різних порядків) виявлено існування періоду величини 6 між локальними максимумами в статистичному ряді розподілу цих величин. Для чисел, менших 2850000000, показано, що найбільш часто зустрічається відстань, яка дорівнює 6. Шляхом нормування отримана наближена універсальна статистична модель розподілу відстаней. Отримані результати доцільно застосовувати для прискорення пошуку простих співмножників, що використовується в сучасних методах шифрування даних.

Ключові слова: *прості числа, числа-близнюки різних порядків, ряди розподілення*

Using a computer for constructing a histogram distances between adjacent prime numbers (so-called twins of different orders) has revealed the existence of a period of the magnitude 6 between local maximum in the statistical row. An approximate universal statistical model of the gap distribution has been obtained by the normalization. The obtained results can be applied for acceleration in the search for prime factors, which is used in modern methods of data encryption.

Keywords: *prime numbers, twin numbers of different orders, distribution series.*

1. Общая постановка задачи и её актуальность

В основе современных методов шифрования текстовых сообщений лежат принципы, заложенные ещё Эратосфеном, Диофантом, П. Ферма, Л .Эйлером и многими другими математиками. Для использования, например, метода несимметричного шифрования RSA [1, 2] требуется знание достаточно больших двух простых чисел [3, 4]. Методам поиска чисел, с высокой вероятностью являющихся простыми, посвящено большое количество трудов специалистов теории чисел. Вычислительные сложности требуют применения мощных ЭВМ и изощрённых алгоритмов, поскольку вручную такие задачи неразрешимы. В данной работе предложено осуществить построение статистического ряда распределения для интервалов между соседними простыми числами. В качестве исходной рассматривается отрезок длины m возрастающей последовательности простых чисел $p_i, p_{i+1} > p_i, i = 1,..\infty$, не превышающих некоторого заданного

предельного числа n : $p_m \le n < p_{m+1}$. Пары соседние чисел p_i, p_{i+1} будем называть «близнецами порядка $gap \equiv 2k$ », если разность (gap, пробел [5])равна этому порядку: $p_{i+1} - p_i = 2k$. Порядок – заведомо чётное число ввиду нечётности простых чисел, больших $p_1 = 2$. Вопрос о структуре множества порядков близнецов издавна рассматривался вместе с проблемами, связанными с распределением простых чисел и их генерированием [6, 7]. К настоящему времени известно следующее.

- Для сколь угодно большого числа существуют близнецы порядка, превышающего это число.
- Отношение количества близнецов любого конечного порядка к количеству простых чисел m, не превышающих числа n и содержащих эти близнецы, стремится к нулю при $n \to \infty$.
- Вопрос о конечности числа близнецов остаётся открытым, невзирая на некоторые успехи, достигнутые в 2013 – 2017 гг.
- «…промежутки преимущественно делятся на 6; промежутки между промежутками также проявляют своеобразный характер» [5].

В данной работе предложен статистический подход к изучению и уточнению структуры множества порядков близнецов.

2. Выбор метода построения последовательности простых чисел

Прежде всего, с точки зрения разрешимости поставленной задачи следует выбрать способ построения последовательности простых чисел наибольшей доступной для практических вычислений длины. Базовым решетом *sieve* является булев массив длины $n_0 = 0{,}95 \cdot 10^9$, изначально заполненный булевыми единицами. Исключением составных чисел согласно алгоритму Эратосфена может быть получен массив той же длины $n = n_0$, номера позиций с булевыми единицами которого и образуют отрезок простых чисел на исходном интервале. Цикл обнаружения первичных простых чисел ограничивается пределом $lim = \left\lfloor \sqrt{n_0 + 1} \right\rfloor$.

Простейшая модификация заключается в исключении из рассмотрения позиций с чётными номерами, что даёт возможность получить последовательность простых чисел вплоть до $n = 2n_0 = 1{,}9 \cdot 10^9$ без увеличения требуемой памяти ЭВМ; при этом достаточно ограничиться $lim = \left\lfloor \sqrt{n_0 / 2 + 1} \right\rfloor$. Проверка чётности сводится к выяснению значения младшего бита, т.е. требует очень малого машинного времени.

В данной работе использована модификация, сводящаяся к исключению из рассмотрения тех номеров позиций в модифицированном решете, которые делятся либо на два, либо на три. Это позволило довести предельное значение последовательности простых чисел до $n = 3n_0 = 2{,}85 \cdot 10^9$, причём $lim = \left\lfloor \sqrt{n_0 / 3 + 1} \right\rfloor$. Общее количество простых чисел на достигнутом интервале: $m = 137568155$. Понятный для дальнейшего сжатия путь, основанный на

исключении чисел, делящихся на 5 (или большие простые), сопряжён с потерей времени для определения делимости на числа, большие 2, а потому и не рассматривался. Основной фрагмент алгоритма приведен в Табл. 1. Для иллюстрации использован язык Pascal.

Табл.1. Основная часть построения модифицированного решета

`lim := trunc(sqrt(n /3.0 + 1));`	Предел для внутреннего цикла
`i := 1; repeat inc(i);` `if sieve[i] then begin`	sieve[0]= sieve[1]=false

`If odd(i) then begin`
`p := 3*i-2; d1 := i+p -1; end`
`else begin`
`p := 3*i-1; d1 := i+p; end;`
`d2 := 2*p -d1;`
	Выяснение нечётности номера i
	
	Вычисление «периодов» d1 и d2

`ii:=i; while (ii <= n) do begin`
`ii := ii + d1; if ii <=n then`
`sieve[ii] := false;`
`ii := ii + d2; if ii <=n then`
`sieve[ii] := false;`
`end;`
`until i>=lim;`
	Собственно построение решета. Последовательное вычёркивание элементов, делящихся на найденное простое p

Составление рассмотренного модифицированного решета занимает 2-3 минуты работы персонального компьютера средней мощности. Наибольшее простое число в данных условиях равно 2 849 999 963. Отметим, что те варианты построения решета, которые основываются на активном использовании проверки делимости, требуют гораздо большего машинного времени. Например, алгоритм, описанный в работе [8] и использованный в ней для исследования ряда распределения расстояний между соседними простыми числами, «для получения 1 млн простых чисел требует около 1.5 часа счета».

3. Построение ряда распределения для порядков близнецов

По полученному модифицированному решету *sieve* можно построить массив простых чисел $p_i < n$. Для этого последовательно просматривается массив *sieve*. Предположим, некоторой позиции j отвечает значение «истина»: $sieve[\,j\,] = true$. Тогда в случае, если j чётно, то соответствующее простое число вычисляется как $\widehat{p} = 3j - 1$; в противном случае $\widehat{p} = 3j - 2$. Текущий номер i увеличивается на единицу, и массив простых чисел пополняется числом $p_i = \widehat{p}$. Для построения ряда распределения следует предусмотреть массив *dist*, длина которого заведомо превосходит наибольшее из чисел $(p_{i+1} - p_i)/2$, и обнулить его элементы. В рассмотренных выше условиях максимальное расстояние между простыми числами равно 320. В процессе

просмотра массива простых чисел последовательно вычисляются $k = (p_{i+1} - p_i)/2$, а элемент $dist[k]$ увеличивается на единицу.

Однако для получения искомого ряда распределения нет необходимости в явном построении последовательности простых чисел. Достаточно просмотреть решётку *sieve*, последовательно определить расстояния между соседними булевыми единицами, непосредственно вычислить число k и увеличить на единицу элемент $dist[k]$.

Одним из важнейших результатов является то, что близнецы порядка 6 встречаются наиболее часто. Частота их появления примерно в два раза превышает частоту для обычных близнецов второго порядка и близнецов порядка 4. Результат описанных действий графически представлен на Рис. 1. Возле максимумов приведены соответствующие количества элементов выборки. Дополнительно произведена нормировка путём деления всех элементов ряда на его максимальное значение для $gap = 6$. Очевидна периодичность положения максимумов; период равен 6. Несколько выбиваются из общего правила порядки $gap = 72$ и $gap = 204$. При $n < 5 \cdot 10^7$, например, $dist[35] > dist[36]$. Однако с увеличением предельного числа n неправильности в периодичности устраняются. Этот факт в данной статье не обосновывается; можно считать, что это одно из проявлений «тайной жизни чисел» [9].

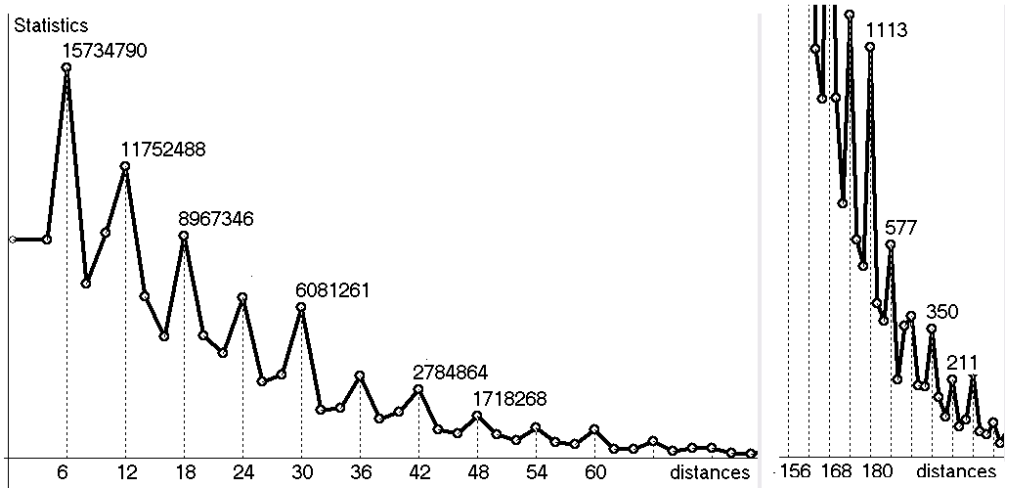
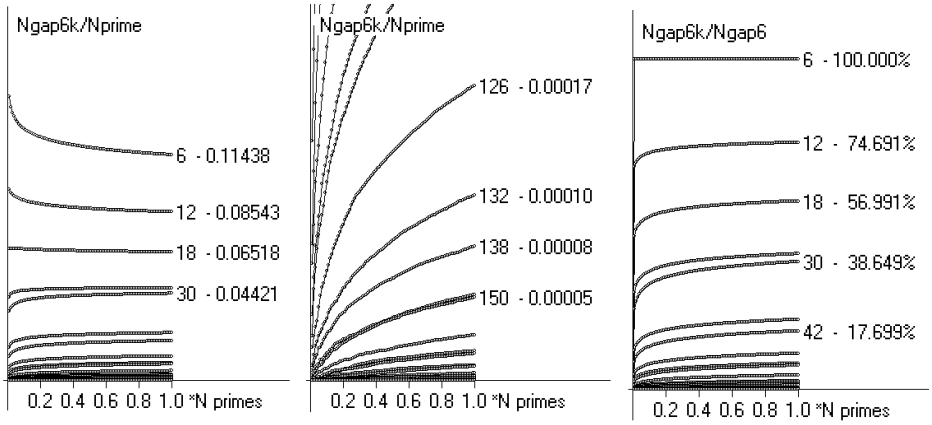


Рис.1. Ряд распределения расстояний между соседними простыми числами, меньшими $n = 2{,}85 \cdot 10^9$

4. Зависимость ряда распределения от предельного числа n

Наличие решета позволяет построить ряды распределения для любых заданных предельных чисел n. Разобьём последовательность простых чисел длины $m = 13756815$ на 100 интервалов длины $\delta = 1375681$ и построим массивы рядов распределения для последовательности количеств простых чисел

$m_i = i \cdot \delta, i = 1..100$. При графическом отображении ограничимся частотами для локальных максимумов соответствующих рядов распределения. На Рис. 2a приведена полученная зависимость частот максимумов от m_i. На Рис. 2b отображена аналогичная зависимость приведенных частот, т.е. отношений частот локальных максимумов к частоте появления пробела $gap = 6$, считая, что соответствующему элементу массива частот $dist[3]$ сопоставляется сто процентов. В рассматриваемых условиях (т.е. $p_i < n = 2{,}85 \cdot 10^9$) прослеживается сохранение порядка подчинённости частот локальных максимумов. Сохраняется ли эта особенность при неограниченном возрастании предельного числа n, – утверждать невозможно. Можно попытаться определить асимптотические значения для отдельных кривых на Рис. 2a, например, по трём равноудалённым точкам на графике, исходя из гипотезы об экспоненциальной зависимости от количества простых чисел. Однако наличие предельного значения, превышающего ноль, противоречит строго доказанной теореме о стремлении этих величин к нулю [6].



a b

Рис.2. Зависимость частот локальных максимумов рядов распределения от длины интервала простых чисел при $Nprimes = 1375680$

a – частоты $dist[3i]/n_k, k = 1..100;$

b – приведенные частоты $dist[3i]/dist[3]$ *в процентном отношении*

5. Гипотезы о предельных значениях частот

Рассмотрим результаты, приведенные на рис. 2.b. Количество точек на каждом графике $\nu_g(m_k), k = 1..100, g = 6,12,18,...$ равно 100. Пусть величина $delta$ имеет некоторое значение порядка 1...25. По результатам статистической обработки, описанной в п.4 , для каждого порядка g определим величины $y_j = \nu_g(m_{100-j \cdot delta}), j = 0,1,2$. Примем следующую гипотезу: приведенные частоты приближённо описываются экспоненциальной зависимостью:

$$\nu_g(x) \approx A_g + B_g \, exp(-\lambda_g x). \quad (1)$$

При интерполировании оказывается, что множитель λ_g отрицателен. Это значит, что для каждой из кривых Рис. 2b существует горизонтальная асимптота (вычисляемая по формуле Эйткена [10]):

$$\nu_g(x) \to A_g = \frac{(y_0 \cdot y_2) - (y_1 \cdot y_1)}{(y_0 + y_2) - (y_1 + y_1)}. \quad (2)$$

Выясняется, что в этом случае относительное расположение частот порядков g сохраняется для сколь угодно больших предельных чисел n.

Аналогичным образом проанализируем графики частот на Рис. 2a. Однако в качестве характерной кривой рассмотрим S_6 - сумму частот для порядков, кратных 6. Согласно гипотезе об экспоненциальной зависимости, приходим к выводу, что S_6 стремится к некоторой константе, превышающей одну треть, т.е. превалирование порядков, кратных шести, сохраняется при стремлении $n \to \infty$. Соответствующее графическое представления характера стремления к асимптотическому положению приведено на рис. 3.

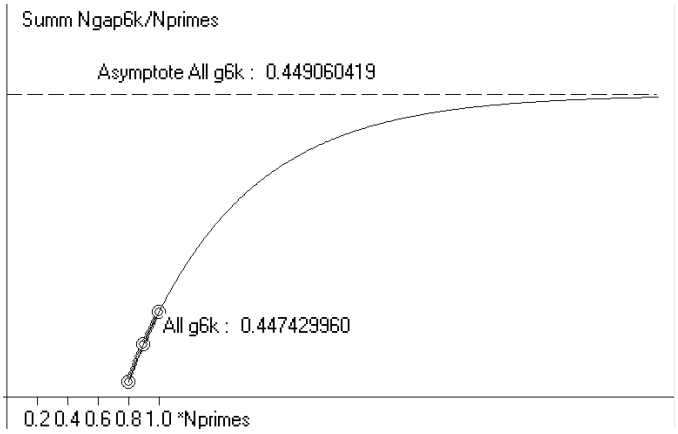


Рис.3. Стремление суммы частот S_6 к асимптотическому положению при $delta = 10$

Предположим, каким-то образом установлена простота некоторого числа p, и требуется найти ещё одно простое число q. Этот вопрос является стандартным для задач несимметричного шифрования. Если рассмотренная гипотеза справедлива, то число q целесообразно искать среди таких, для которых разность $|p - q|$ делится на 6.

6. Одна задача о распределении шаров в ячейках

В теории вероятностей имеется ряд задач, никак не связанных с общей теорий чисел, для которых графическое представление результатов оказывается сходным с теми, которые описаны в п. 3. [11]. Рассмотрим задачу о

произвольном размещении m шаров в $n = 3n_0$ ячейках по одному, $n_0 > m$, причём заполняются лишь те ячейки, номера которых нечётны и не делятся на 3. Для каждого такого размещения составим ряд распределения расстояний между номерами соседних заполненных ячеек. На Рис. 4 представлен графически результат статистической обработки при $n_0 = 40 \cdot 10^6$, $n = 120 \cdot 10^6$, $m_1 = 4 \cdot 10^6$ (тонкая линия), $m_2 = n/6 = 20 \cdot 10^6$ (утолщённая линия). При реализации подобных размещений использован приём модификации, рассмотренный в п. 2. Возле максимумов в ряде распределения проставлены количества пар номеров ячеек, отличающихся на указанные величины пробелов для случая $m_1 = 4 \cdot 10^6$. Анализ полученных графиков приводит к мысли о правдоподобности нескольких гипотез, доказательство справедливости которых достаточно сложно, но несравненно проще, чем доказательство рассмотренных выше гипотез относительно распределения простых чисел, расстояний между соседними простыми и экстраполяции на большие значения предельных чисел.

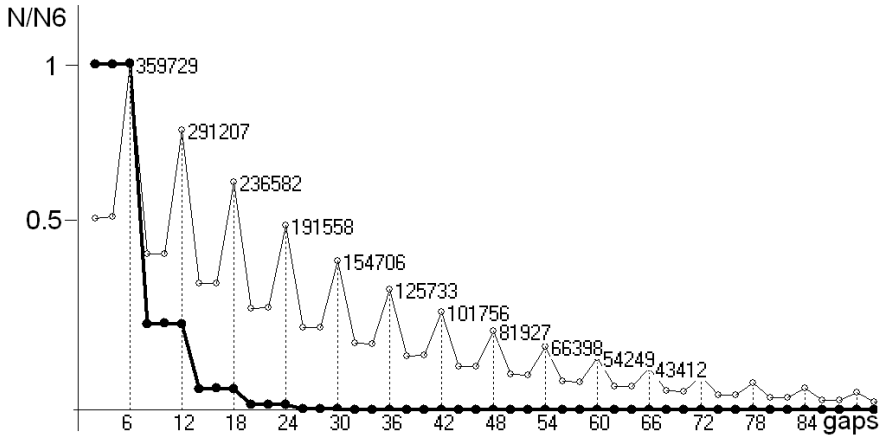


Рис.4. Ряды распределения расстояний между соседними заполненными ячейками при случайном заполнении шарами (не более одного в ячейке) при условии, что номера ячеек взаимно просты с числами 2 и 3

Например, имеются основания доказывать следующие утверждения:

- вероятности $p_{6k-2} = P(gap = 6k-2)$ и $p_{6k-4} = P(gap = 6k-4)$ равны между собой при любых допустимых соотношениях между n и m;
- если $m << n$, то $p_{6k-2} = p_{6k-4} = 0.5\, p_{6k}$;
- при $m = n/6$ вероятности появления следующих троек расстояний совпадают: $p_{6k-4} = p_{6k-2} = p_{6k}$;
- если $m < n/6$, то $p_{6k-4} = p_{6k-2} < p_{6k}$;
- если $m > n/6$, то $p_{6k-4} = p_{6k-2} > p_{6k}$;
- при $n \to \infty$ и $m \approx n_0$ распределение близко к экспоненциальному.

7. Выводы по результатам и направления дальнейших исследований

Статистический подход при изучении структуры множества простых чисел обладает некоторой ограниченностью, поскольку не предполагает никаких априорных теоретических сведений и не позволяет в принципе строго обоснованную экстраполяцию. Но исследование статистических фактов позволяет определить возможные направления в теоретических исследованиях. Простейшим примером является круг вопросов, связанных с количеством простых чисел $\pi(x)$, меньших чем x. Прежде чем П. Л. Чебышевым была строго доказана основная теорема о простых числах: $(x \to \infty) \Rightarrow \left(\frac{\pi(x) \ln x}{x} \to 1 \right)$, – её справедливость была предсказана по результатам статистики.

Некоторая уверенность в обоснованности результатов данной работы может быть получена путём изучения последовательностей простых чисел для значительно больших значений предельных чисел n.

ЛИТЕРАТУРА

1. Введение в криптографию / ред. В. В. Ященко. – СПб, Питер, 2001. –285 с.
2. Коблиц Н. Курс теории чисел и криптография, М.: ТВП, 2001. – 513 с.
3. Певнев В. Я. Методика построения псевдопростых чисел // Системи обробки інформації. Зб. наук. пр./ Х.: Харків. універ. Повітр. Сил, 2016. – С. 30-32.
4. Певнев В. Я. Генератор простых чисел // Каф. сист. інф. НАКУ ім.. М. Є. Жуковського. Зб. наук. пр. – Х.: Тов. «Щ. садиба плюс», 2014. – С. 140-146.
5. G. G. Szpiro. Peaks and gaps: Spectral analysis of the intervals between prime numbers, Physica A, v. 384 (2), 2007, pp. 291–296.
6. Трост Э. Простые числа, – М.: Физ. Мат. Лит., 1959. – 136 с.
7. Прахар К. Распределение простых чисел, – М.: МИР, 1967. – 512 с.
8. Тарунин Е. Л. Возможности вычислительных методов в проблемах теории чисел // Вестник Перм. универ, Сб.науч. тр./Пермь: Изд. Перм.универ., 2010, вып. 2(2). – С. 15-28.
9. G. G. Szpiro. The Secrete Life of Numbers, – J. Henry Pr., Wash., 2006. – 210 p.
10. Чернышев Ю. К. Методы вычисления статистических параметров в событийном моделировании, – Х.: «Фактор», 2014. –244 с.
11. Карлин С. Основы теории случайных процессов, – М.: «Мир», 1971. – 537 с.

Надійшла – 23.06.2017.