CHEBYSHEV BIAS

SOUNAK BAGCHI

Abstract. In this paper, we discuss the phenomenon of Chebyshev Bias and find measures for the density of when primes modulo 4, assuming certain parts of the Generalized Riemann Hypothesis (GRH) and Deep Riemann Hypothesis (DRH). We also consider generalized Chebyshev Bias over different modulos and connect this to the Prime Number Theorem (PNT) on Arithmetic Progressions.

1. Introduction

In 1837, Dirichlet proved that there are infinitely many primes p satisfying $p \equiv a \pmod q$ for a, q with $\gcd(a, q) = 1$. 16 years later, Pafnuty Chebyshev, in a letter to P.N. Fuss [1], made the observation that many more primes are congruent to 3 modulo 4 than 1 modulo 4. He phrased it as the assertion

$$\lim_{x\to\infty} \sum_{p>2} (-1)^{(p-1)/2} e^{-p/x} = -\infty$$

where p ranges over primes.

Chebyshev's observation gave rise to the field of **comparative number theory**, which compares how often prime numbers of different residue classes occur. This paper will focus on further advances made in comparative number theory stemming from Chebyshev's observation.

To understand the paper, a surface level knowledge of analytic number theory concepts, such as Dirichlet L-functions, is required.

2. Definitions and Basic Properties

Definition 2.1. Define $\pi(x; q, a)$ to be the number of primes p less than or equal to x that are congruent to $a \pmod q$. We'll make the assumption that x is prime as well throughout the paper, as other cases are trivial.

Chebyshev's assertion is that $\pi(x; 4, 3) > \pi(x; 4, 1)$ more often than not.

Example 2.3. $\pi(100, 3, 4) = 14$ and $\pi(100, 1, 4) = 10$.

It is also useful to define a way of comparing functions:

Definition 2.4. We say that $f(x) \sim g(x)$ for two functions f, g if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

As an example, the Prime Number Theorem [2] states that

$$\pi(x) \sim \frac{x}{\log x}.$$

3. Density of Chebyshev Bias

A natural question that arises with Chebyshev Bias is whether or not $\pi(x;4,3) > \pi(x;4,1)$ is true for all x. It is not; the first few equality cases are at $x = 5, 17, 41, \ldots$, and the first counterexample for the nontsrict inequality is at $x = 26861$. With this information, a second question arises; how much more often does the inequality $\pi(x;4,3) > \pi(x;4,1)$ occur than the inequality $\pi(x;4,3) < \pi(x;4,1)$?

Conjecture 3.1. *Knapowski-Turan Conjecture.* Let $A(x) = \{x_1 < x \mid \pi(x;4,3) > \pi(x;4,1)\}$. The **density** of this set is 1; in other words,

$$\lim_{x\to\infty} \frac{|A(x)|}{x} = 1.$$

This conjecture was disproved by Kaczorowski [3] in 1995 under the assumption of the Generalized Riemann Hypothesis[1] (GRH).

The disproving of the Knapowski-Turan Conjecture shows that the natural density of $A(x)$ is not straightforward. In this section, we will discuss how to better approximate the density of $A(x)$, using another measure of density.

Definition 3.2. Define the **logarithmic density** of a set S to be

$$\delta(S) = \lim_{X\to\infty} \frac{1}{\log X} \int_{\substack{t\in S,\\ 2\le t\le X}} \frac{dt}{t}.$$

As it turns out, $\delta(A) = 0.9959\ldots$ [5], showing a clear bias. As Dirichlet proved in his theorem of arithmetic progressions, the number of primes of the form $4k+1$ and $4k+3$ are both infinite; Chebyshev's Bias claims that primes of the form $4k+3$ occur earlier. Using logarithmic density, earlier terms are weighted more than later terms, hence allowing us to compare the two counts. Logarithmic density works better than natural density in this regard, because natural density weighs earlier and later primes equally.

While we know that $\pi(x;4,3) > \pi(x;4,1)$ much more often than $\pi(x;4,1) > \pi(x;4,3)$, we don't know how to compare $\pi(x;4,3)$ to $\pi(x;4,1)$. Koyama and Kurokawa [6] recently presented. As it turns out, just like we did with logarithmic density, we can adopt another

[1] The Generalized Riemann Hypothesis states that [3], for any Dirichlet character χ modulo k, the Dirichlet L-function $L(\chi, s)$, which in expanded form is

$$L(\chi, s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

has all of it's nontrivial zeroes on the line $\Re(s) = \frac{1}{2}$.

counting function that weighs earlier primes larger, as follows.

Definition 3.3. Define a new counting function, $\pi_s(x; q, a)$, for $s \geq 0$, as

$$\pi_s(x; q, a) = \sum_{\substack{p<x\text{: prime} \\ p\equiv a \pmod q}} \frac{1}{p^s}.$$

Now, to illustrate the use of this counting function, it's also important to understand the Deep Riemann Hypothesis[2] (DRH) for the Dirichlet L-character $L(s, \chi_{-4})$.

Theorem 3.4. *(DRH for $L(s, \chi_{-4})$.)* The limit

$$\lim_{x\to\infty} \prod_{p\leq x} \left(1 - \chi(p)p^{-\frac{1}{2}}\right)^{-1}$$

exists, and is nonzero.

Note that the Dirichlet Character $\chi_{-4}(n)$ is 1 when for $n \equiv 1 \pmod 4$, and -1 for $n \equiv 3 \pmod 4$. Now, we present a proof of the following theorem, from Koyama and Kurokawa [6], which relies on the assumption that DRH is true.

Theorem 3.5.

$$\pi_{\frac{1}{2}}(x; 4, 3) - \pi_{\frac{1}{2}}(x; 4, 1) \sim \frac{1}{2} \log\log x.$$

Proof: Start with DRH for $L(s, \chi_{-4})$ as in Theorem 3.4. Since the limit is finite and nonzero, we can take the logarithm of both sides, which is bounded:

$$\sum_{p\leq x} \log \left(1 - \chi(p)p^{-\frac{1}{2}}\right)^{-1} = O(1).$$

The left hand side can be expanded as

$$\sum_{p\leq x} \sum_{k=1}^{\infty} \frac{\chi(p)^k}{kp^{\frac{k}{2}}}.$$

Note that, for $k \geq 3$, the subseries is absolutely convergent because of the estimation

$$\sum_{p\leq x} \sum_{k=1}^{\infty} \left| \frac{\chi(p)^k}{kp^{\frac{k}{2}}} \right| \leq \frac{1}{3} \sum_{p\leq x} \sum_{k=1}^{\infty} \frac{1}{p^{k/2}} \leq \zeta\left(\frac{3}{2}\right)$$

due to the geometric series calcuation

$$\sum_{k=3}^{\infty} \frac{1}{k^{\frac{3}{2}}} = \frac{\sqrt{p}}{\sqrt{p}-1} \cdot \frac{1}{p^{\frac{3}{2}}} \leq \frac{3}{p^{\frac{3}{2}}}.$$

[2]Goldfeld [7] proved that the DRH imlies the GRH for L functions that are attached to elliptic curves E, while GRH implies the regular Riemann Hypothesis for the Dirichlet Character $\chi(n) = 1$.

Now, considering the subseries for $k = 2$, note from Euler's Theorem [8] that

$$\sum_{p \le x} \frac{\chi(p)^2}{2p} = \sum_{p \le x} \frac{1}{2p} = \frac{1}{2} \log\log x + O(1).$$

What's left is the subseries at $k = 1$, which is simply

$$\sum_{p \le x} \frac{\chi(p)}{\sqrt{p}}.$$

Since the original summation is simply $O(1)$, it follows from subtracting the subseries at $k = 2$ and $k \ge 3$ that

$$\sum_{p \le x} \frac{\chi(p)}{\sqrt{p}} = -\frac{1}{2} \log\log x + O(1).$$

Since $\chi(p) = -1$ for $p \equiv 3 \pmod 4$ and $\chi(p) = 1$ for $p \equiv 1 \pmod 4$, the result follows by taking the negatives of both sides. □

This also gives us another corollary about density:

Corollary 3.6. The natural density of the set

$$A(s) = \{x > 0 \mid \pi_s(x; 4, 3) - \pi_s(x; 4, 1) > 0\}$$

is equal to 1.

4. Generalized Chebyshev Bias

We examined the "original" Chebyshev Bias for modulo 4, but it turns out that the distributions of primes among residue classes in other modulos also show curious differences. For example:

Example 4.1. Consider $\pi(x; 3, 1)$ and $\pi(x; 3, 2)$. More often than not, $\pi(x; 3, 2) \ge \pi(x; 3, 1)$. In fact, the logarithmic density of the set $B = \{x > 0 \mid \pi(x; 3, 2) - \pi(x; 3, 1) \ge 0\}$ is about $0.9990\ldots$. The first x that is not an element of B is 608981813029.

Notice that 2 is a non-quadratic residue (NQR) in modulo 3, as is 3 in modulo 4, while 1 is a quadratic residue (QR) in modulo 3, as it is in modulo 4. Through this observation, the bias can be generalized:

Theorem 4.2. Suppose

$$\left(\frac{m}{a}\right) = -1, \left(\frac{n}{a}\right) = 1.$$

Our only requirements for m and n are that $\gcd(a, m) = \gcd(a, n) = 1$. Then

$$\pi(x, m, a) > \pi(x, n, a)$$

more often than not. This can also be written as

$$\delta(S_{m,n}) > 0.5,$$

where $S_{m,n} = \{x > 0 \mid \pi(x; m, a) - \pi(x; n, a) > 0\}$.

In fact, using this generalization, we arrive at something similar to **Theorem 3.5**, in a more generalized form [9]:

Theorem 4.3. Define the Dirichlet L-Character

$$\chi_q(a) = \left(\frac{a}{q}\right),$$

i.e. the standard **Jacobi Symbol**. It follows that

$$\sum_{\substack{b\in(\mathbb{Z}/q\mathbb{Z})^\times \\ \text{Quadratic Non-Residue}}} \pi_{\frac{1}{2}}(x;q,b) - \sum_{\substack{a\in(\mathbb{Z}/q\mathbb{Z})^\times \\ \text{Quadratic Residue}}} \pi_{\frac{1}{2}}(x;q,a) \sim \left(\frac{1}{2}+m\right)\log\log x$$

where $m = \operatorname{ord}_{s=\frac{1}{2}} L(s,\chi)$. In particular, for $\chi_q = \chi_{-4}$, we get the result from **Theroem 3.5**.

Note that the result in **Theorem 4.2** is not inconsistent with the Prime Number Theorem for Arithmetic Progressions, which states the following:

Theorem 4.4. Let $a_1, \ldots, a_{\phi(m)}$ be the residue classes for modulo m describing numbers relatively prime to m. Then, it follows that

$$\pi(x;m,a_1) \sim \pi(x;m,a_2) \sim \cdots \sim \pi(x;m,a_{\phi(m)}) \sim \frac{1}{\phi(m)}\frac{x}{\log x}.$$

In layman's terms, the Prime Number Theorem for Arithmetic Progression says that the primes are evenly distributed among all residue classes for a certain modulo. Chebyshev Bias does not contradict this. Chebyshev Bias is more of race between residues; for earlier x, the number of primes that are 3 modulo 4 outnumbers the number of primes that are 1 mod 4, less than x. However, in the long run behavior as x goes to infinity, the number of primes that are 1 modulo 4 (catches up).

5. References

[1] P. L. Chebyshev, Lettre de M. le professeur Tch´ebychev ‘a M. Fuss, sur un nouveau th´eoreme r´elatif aux nombres premiers contenus dans la formes 4n + 1 et 4n + 3, Bull. de la Classe phys.-math. de l'Acad. Imp. des Sciences St. Petersburg 11 (1853), 208.

[2] J. Hadamard, Sur la distribution des z´eros de la fonction ζ(s) et ses consequences arithmetiques, Bull. Soc. Math. France 24 (1896), 199–220 (French).

[3] J. Kaczorowski, On the distribution of primes (mod 4), Analysis 15 (1995), no. 2, 159–171.

[4] Léo Agélas. Generalized Riemann Hypothesis. 2019. ffhal-00747680v3ff

[5] Michael Rubinstein. Peter Sarnak. "Chebyshev's bias." Experiment. Math. 3 (3) 173 - 197, 1994.

[6] Shin-ya Koyama. Nobushige Kurokawa. "Chebyshev's bias for Ramanujan's τ-function via the Deep Riemann Hypothesis." Proc. Japan Acad. Ser. A Math. Sci. 98 (6) 35 - 39, June 2022. https://doi.org/10.3792/pjaa.98.007

[7] D. Goldfeld: Sur les produits partiels eul´eriens attach´es aux courbes elliptiques. C. R. Acad. Sci. Paris Ser. I Math. 294 (1982) 471-474.

[8] Euler, Leonhard (1737). "Variae observationes circa series infinitas" [Various observations concerning infinite series]. Commentarii Academiae Scientiarum Petropolitanae. 9: 160–188.

[9] Miho Aoki, Shin-ya Koyama, Chebyshev's bias against splitting and principal primes in global fields, Journal of Number Theory, Volume 245, 2023, Pages 233-262, ISSN 0022-314X, https://doi.org/10.1016/j.jnt.2022.10.005.