CHEBYSHEV'S BIAS

ROULI FREEMAN

Notice: This paper is refrenced from the "Chebyshev's Bias" paper by Rubinstien and company. [1]

It has been shown that

$$\pi(x,q,a) \sim \frac{\pi(x)}{\phi(q)}$$

Where $\phi(q)$ is the euler totent function, the number of relatively prime to q numbers $\leq q$. $\pi(x,q,a)$ is the number of primes $\leq x$ that are equal to $a \mod q$ and $\pi(x)$ is the number of primes $\leq x$, which is equal to $\pi(x,1,1)$.[1]

The phenomenon called "Chebyshev's Bias" is that for small primes, there are some cases in which $x \mod n$ has more numbers than $y \mod n$. In fact, for most small numbers, $\pi(x,4,3) > \pi(x,4,1)$. The majority of this paper will be talking about similar phenomenons.

Definition: The logarithmatic density is defined as

$$\delta(A) = \lim_{x\to\infty} \frac{1}{\log x} \int_{t\in P\cap[2,x]} \frac{1}{t}$$

for A a set. Higher numbers in the set contribute less, and $\delta(A)+\delta(B)+\delta(C)+\ldots = 1$ if $A, B, C, \ldots$ are disjoint, and $A \cup B \cup C \cup \ldots = [2,\infty]$

In fact, there is a upper logarithmatic density and a lower logarithmatic density, which when equal to each other, converge to $\delta(A)$.

Definition:

$$P_{q;a,b,c,d,e\ldots}$$

is the set of all real numbers $x \geq 2$ such that

$$\pi(x,q,a) > \pi(x,q,b) > \pi(x,q,c) > \pi(x,q,d) > \pi(x,q,e) > \ldots$$

In some cases, the logarithmatic densities are extremely high or low, showing a large bias.

In the case of $P_{3;2,1}$, $\delta(P_{3;2,1}) = 0.9990\ldots$, showing that there are much more numbers in this set than in the other set, $(P_{3;1,2}$. In fact, the first number of $P_{3;1;2}$ is 608981813029. Similarly for $P_{4;1,3}$, the first number in it is 26861.

Definition:

$$E_{q;a,b,c,d,e,\ldots}(x) = \frac{\log x}{\sqrt{x}} \times (\phi(q)\pi(x,q,a) - \pi(x), \phi(q)\pi(x,q,b) - \pi(x), \ldots)$$

[1] The prime number theorem states that

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

which returns a vector.

Definition:

$$\mu_{q;a,b,c,d,e,\ldots}$$

is the limiting distribution of

$$E_{q;a,b,c,d,e,},$$

which exists.

Then $\delta(P_{q;a,b,c,d,e,\ldots}) = \mu_{q;a_1,a_2,\ldots,a_r}(\{x \in \mathbb{R}^r | x_1 > x_2 > x_3 > x_4 > \ldots > x_r\})$ if $\mu_{a,b,c,d,e,\ldots}$ is absolutely continous.

Note: *Assuming the Generalized Riemann Hypothesis* [2] *and the Grand Simplicity Hypothesis* [3]*, then we have the following equation for the Fourier transform of* $\mu_{q;a,b,c,d,e,\ldots}$*:*

$$\hat{\mu}_{q;a_1,a_2,\ldots,a_r}(\xi_1, \xi_2, \ldots, \xi_r) = exp\left(i \sum_{j=1}^{r} c(q, a_j)\xi_j\right)$$
$$\times \sum_{\chi \neq \chi_0, \chi \mod q} \sum_{\gamma_\chi > 0} J_0\left(\frac{2\left|\sum_{j=1}^{r} \chi(a_j)\xi_j\right|}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}\right)$$

where χ_0 *is the principal character, with c being*

$$c(q, a) = -1 + \sum_{b^2 \equiv a(q), 0 \leq b \leq q-1} 1,$$

with J_0 *being the Bessel function.* [4]

Definition:

$$(q; a, b, c, d, e, \ldots)$$

is unbiased if and only if $\mu_{q;a,b,c,d,e,\ldots}$ is same for all permutations of $a, b, c, d, e, \ldots$

In this case,

$$\mu_{q;a_1,a_2,a_3,a_4,a_5,\ldots,a_r} = r!^{-1}$$

Definition:

$$c(q, a) = -1 + \sum_{b^2 = aq, 0 \leq b < q} 1$$

[2] *If s is not a negative real number, than for all* χ *a Dirichlet character, if* $L(\chi, s) = 0$, $\mathbb{R}(s) = \frac{1}{2}$.

[3] *The set of* $\gamma \geq 0$ *such that* $L(\chi, \frac{1}{2} + i\gamma)$*, for all* χ *a Dirichlet character, is linearly independent over* $\mathbb{Q}$.

[4]

$$J_0 = \sum_{m=0}^{\infty} \frac{(-1)^m (\frac{1}{2}z)^{2m}}{m!^2}$$

Theorem: *Assuming the Generalized Riemann Hypothesis* [5] *and the Grand Simplicity Hypothesis* [6], $(q; a_1, a_2, a_3, a_4, a_5, \ldots, r)$ *is unbiased if and only if either* $r = 2$, *and* $c(q, a_1) = c(q, a_2)$ *or* $r = 3$, *and*

$$\exists p \neq 1 : (p^3 \equiv 1 \wedge a_2 \equiv a + 1p \wedge a_3 \equiv a_1 p^2) \mod q$$

Definition:

$$P_{q;N,R} = \{x \geq 2 : \pi_N(x, q) > \pi_R(x, q)\},$$
$$P_{q;R,N} = \{x \geq 2 : \pi_R(x, q) > \pi_N(x, q)\},$$

where

$$\pi_R(x, q) = \sum_{a \leq x} \sum_{p \text{ prime}, a^2 = q \mod p} 1,$$
$$\pi_N(x, q) = \sum_{a \leq x} \sum_{p \text{ prime}, \sim a^2 = q \mod p} 1,$$

Definition:

$$P_1 = \{x \geq 2 : \pi(x) > Li(x)\},$$

where $Li(x)$ is the logarithmatic integral.[7]

It has been proven that P_1 does extend to infinity, but contary to $P_{4;1,3}$ and $P_{3;1,2}$, no numbers are known of this series.

$$\delta(P_1) = 0.00000026$$

This is extremely low, showing that even though a member of this series may exist, the chance of a random number being in P_1 is small.

Definition:

$$E_{q;N,R}(x) = (\pi_N(x, q) - \pi_R(x, q)) \frac{\log x}{\sqrt{x}}$$

Definition:

$$\mu_{q;N,R}$$

is equal to the limiting distribution of $E_{q;N,R}$.

Definition:

$$E_1(x) = (\pi(x) - Li(x)) \frac{\log x}{\sqrt{x}}$$

Definition:

[5] *If s is not a negative real number, than for all χ a Dirichlet character, if $L(\chi, s) = 0$, $\mathbb{R}(s) = \frac{1}{2}$.*

[6] *The set of $\gamma \geq 0$ such that $L(\chi, \frac{1}{2} + i\gamma)$, for all χ a Dirichlet character, is linearly independent over $\mathbb{Q}$.*

[7]

$$Li(x) = \int_2^x \frac{dt}{\log t}$$

$$\mu_1$$

is equal to the limiting distribution of E_1.

We may numerically compute $\delta(P_{q;N,R})$ as:

$$\frac{1}{2\pi}\sum_{-C\leq n\varepsilon\leq C}\varepsilon\frac{\sin n\varepsilon}{n\varepsilon}(1+b_1(n\varepsilon)^2)\times\sum_{0<\gamma\leq M}J_0\left(\frac{2n\varepsilon}{\sqrt{\frac{1}{4}+\gamma^2}}\right)+\frac{1}{2}+e$$

where e is the error,

$$b_1=-T_1(0)+\sum_{0<\gamma\leq M}\frac{1}{\frac{1}{4}+\gamma^2}$$

where e stands for error, C, M stand for cutoffs to approximate the actual value, set to be 25 and 9999,

$$T_1(X)=\sum_{\gamma>X}\frac{1}{\frac{1}{4}+\gamma^2}$$

and J_0 is the Bessel function. [8]

This is found by taking a Fourier transform of the density of $\mu_{q;N,R}$[9], and then taking the integral. This is then transformed into a sum, and as the integral is an improper integral, some values are cut-off, to make it computable to get a close enough answer.

The value $\delta(P_1^{\text{comp}})$ can also be computed by this equation:

$$\frac{1}{2\pi}\sum_{-C\leq n\varepsilon\leq C}\varepsilon\frac{\sin n\varepsilon}{n\varepsilon}(1+b_1(n\varepsilon)^2+b_2(n\varepsilon)^4)\times\sum_{0<\gamma\leq M}J_0\left(\frac{2n\varepsilon}{\sqrt{\frac{1}{4}+\gamma^2}}\right)+\frac{1}{2}+e$$

Where C is 5, M is 88190, and

$$b_2=-\frac{1}{16}\left(2a_1-\gamma^2+\frac{1}{8}\pi^2-1\right)$$

where

$$a_1=\lim_{N\to\infty}\frac{1}{2}\log^2(N+1)-\sum_{m=1}^{N}\frac{\log(m+1)}{m+1}$$

The paper uses a similar enough method to the way to solve for $\delta(P_{q;N,R})$

References

[1] Michael Rubinstein and Peter Sarnak. “Chebyshev’s bias”. In: *Experiment. Math.* 3.3 (1994), pp. 173–197. URL: https://projecteuclid.org:443/euclid.em/1048515870.

[8]

$$J_0=\sum_{m=0}^{\infty}\frac{(-1)^m(\frac{1}{2}z)^{2m}}{m!^2}$$

[9] In the paper, it says that it is $\mu_{q;R,N}$, and then says that $f_{q;N,R}(t)$ is the density function of $\mu_{q;R,N}$, which does not make logical sense.