Parabola Volume 54, Issue 3 (2018)

From binomial coefficients to primes – Chebyshev revisited

Liangyi Zhao[1]

Abstract

In this paper, we give upper and lower bounds for the number of primes not exceeding x via elementary means using binomial coefficients.

1 Introduction

An integer $p \geq 2$ is said to be a prime number (or prime) if it is only divisible by 1 and itself. By the Fundamental Theorem of Arithmetic, every positive integer has a unique (up to re-arrangements of the factors) factorisation into primes. It has been known since at least Euclid's time that there are infinitely many primes. Indeed, $n! + 1$ must have a prime divisor greater than n. So, there are arbitrarily large prime numbers and hence an infinitude of them.

After Euclid, a natural question to ask is, how infinite are the primes? That is, given $x \in \mathbb{R}$, how many primes are there not exceeding x? It took more than twenty centuries to make any significant progress on this question. Let $\pi(x)$ be the number of primes less than or equal to x. For instance,

$$\pi(1) = 0\,,\ \pi(3) = 2\,,\ \pi(10) = 4\,,\ \pi(1,000,000,000) = 50,847,534\,.$$

Obviously, $\pi(x) < x$ for all x and one can reverse engineer Euclid's proof to show that

$$\pi(x) \geq \ln \ln x\,. \tag{1}$$

This is not a very impressive minorant as $\ln \ln 1,000,000,000 = 3.03\ldots$ [2]. We shall not give a proof of (1) here as it will be greatly superseded by the bounds presented later.

It turns out that $\pi(x)$ is well-approximated by the function $x/\ln x$ for large values of x. A.-M. Legendre was the first to prove that

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1\,.$$

This is known as the Prime Number Theorem. Gauss wrote in 1849 that he had discovered it in a slightly different form during his boyhood in 1792. A major break-through

[1]Liangyi Zhao is a Senior Lecturer at the School of Mathematics and Statistics, UNSW Sydney.

[2]It has been variously said that $\ln \ln x$ is a function that tends to infinity as $x \to \infty$ but has never been observed to do so and that "$\log \log \ldots$" is what an analytic number theorist intones while drowning.

came in 1860. Georg Friedrich Bernhard Riemann, a newly elected member of the Berlin Academy of Sciences who had to report on his most recent research, sent an article titled *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*[3] to the academy. An English translation of this paper can be found in [1]. Considering that it was his only paper in number theory and changed the direction of mathematical research in very significant ways, it is now appropriately and better known as Riemann's Memoir.

The key contribution that Riemann made was the injection of complex analysis into the study of primes. A number of conjectures were made in his memoir. Among other things, he proposed that for $x \geq 2$, the values of $\pi(x)$ resemble those of

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\ln t}\mathrm{d}t\,.$$

We leave it to the reader to check, using integration by parts, that $\mathrm{Li}(x)$ and $x/\ln x$ are good approximations of each other. In 1896, Hadamard and de la Vallée Poussin proved independently this form of the Prime Number Theorem. In fact, $\pi(x)$ coincides with $\mathrm{Li}(x)$ significantly more than with $x/\ln x$. The following table evinces this difference in the quality of the approximations. This contrast can also be made mathematically precise.

x	$\pi(x)$	$\pi(x) - x/\ln x$	$\pi(x) - \mathrm{Li}(x)$
10	4	-0.3	2.2
10^2	25	3.3	5.1
10^3	168	23	10
10^4	1229	143	17
10^5	9592	906	38
10^6	78498	6116	130
10^7	664579	44158	339
10^8	5761455	332774	754
10^9	50847534	2592592	1701
10^{10}	455052511	20758029	3104

It is certainly remarkable that $\pi(10^{10})$ is about half a billion and $\mathrm{Li}(10^{10})$ is off by only three thousand or so. Note also that $\pi(x) - \mathrm{Li}(x)$ generally has half the number of digits as $\pi(x)$. This phenomenon persists for larger values of x and has to do with the Riemann Hypothesis, the only yet unproven conjecture in Riemann's Memoir.

In this paper, we present an elementary proof of upper and lower bounds for $\pi(x)$. Clearly, a lower bound for $\pi(x)$ is more important in this context, since it enables us to conclude that there must at least be a certain number of primes less than x. The bounds we prove here are of the correct order of magnitude (that is, best possible up to a constant multiple). The ideas of the proof were first developed by P.L. Chebyshev

[3] *On the number of primes less than a given magnitude*

in 1852 and are rather simple. We only need to examine a special binomial coefficient and its prime divisors.

From now on, we shall use the convention that p always denotes a prime number. Using this convention, we can express $\pi(x)$ as follows.

$$\pi(x) = \sum_{p \le x} 1\,.$$

In fact, as we shall presently see, it will often be more convenient to count primes with a different weight. We shall work with

$$\theta(x) = \sum_{p \le x} \ln p$$

and extract the bounds for $\pi(x)$ from those for $\theta(x)$.
We will prove the following theorem.

Theorem 1. *If $x \ge 854$, then*

$$\frac{\ln 2}{7} x < \theta(x) \le (4 \ln 2) x\,.$$

From Theorem 1, we readily deduce the following result.

Corollary 2. *For all $x \ge 2$, we have*

$$\frac{\ln 2}{7} \frac{x}{\ln x} \le \pi(x) \le (8 \ln 2 + 2) \frac{x}{\ln x}\,. \tag{2}$$

Proof. One can easily find the list of primes less than 854[4]. From this, $\pi(x)$ can be evaluated for all $x \le 854$ and the bounds in (2) hold for all such values of x. Thus, we may assume that $x \ge 854$ in the sequel.

We first note that, from Theorem 1,

$$\pi(x) \ln x = \sum_{p \le x} \ln x \ge \sum_{p \le x} \ln p = \theta(x) > \frac{\ln 2}{7} x\,.$$

Now, the lower bound follows by dividing $\ln x$ on both sides. The upper bound for $\pi(x)$ is a bit more involved. Observe that

$$\theta(x) \ge \sum_{\sqrt{x} < p \le x} \ln p \ge \ln \sqrt{x} \sum_{\sqrt{x} < p \le x} 1 = \frac{1}{2} \ln x \sum_{\sqrt{x} < p \le x} 1 = \frac{1}{2} \ln x \left(\pi(x) - \pi(\sqrt{x}) \right)\,.$$

So, we have

$$\pi(x) \le \frac{2\theta(x)}{\ln x} + \pi(\sqrt{x}) \le 8 \ln 2 \frac{x}{\ln x} + \sqrt{x}\,.$$

[4]See `https://primes.utm.edu/lists/small/1000.txt`, for example.

One can easily check on a graphing calculator that $\sqrt{x} < 2x/\ln x$ for $x \geq 2$. Therefore, we get that, if $x \geq 2$, then

$$\pi(x) \leq (8\ln 2 + 2)\frac{x}{\ln x}\,.$$

This completes the proof. □

Now, we can also prove the following important fact.

Corollary 3. *As x tends to ∞,*

$$\sum_{p\leq x}\frac{1}{p} \to \infty\,.$$

Proof. We note that $\pi(n) - \pi(n-1)$ is 1 if and only if n is prime and zero otherwise. So,

$$\begin{aligned}\sum_{p\leq x}\frac{1}{p} &= \sum_{n\leq x}\frac{\pi(n)-\pi(n-1)}{n}\\ &= \frac{\pi(x)}{[x]} + \sum_{n\leq x-1}\pi(n)\left(\frac{1}{n}-\frac{1}{n+1}\right)\\ &\geq \sum_{n\leq x-1}\pi(n)\left(\frac{1}{n}-\frac{1}{n+1}\right)\,,\end{aligned}$$

where $[x]$ denotes the integer part of x. Using the lower bound in (2) and the observation that $2n \geq n+1$ for all $n \in \mathbb{N}$, the last sum above is

$$\sum_{n\leq x-1}\frac{\pi(n)}{n(n+1)} \geq \frac{\ln 2}{7}\sum_{2\leq n\leq x-1}\frac{1}{(n+1)\ln n} \geq \frac{\ln 2}{14}\sum_{2\leq n\leq x-1}\frac{1}{n\ln n}\,.$$

Now, the integral test tells us that this last sum tends to ∞ as x tends to ∞, and the proof is complete[5]. □

From the corollaries, we see that there are a lot of primes. More specifically, there are only about $\sqrt{x}$ squares in $[1, x]$ and the sum of the reciprocals of the squares converges to $\pi^2/6$. However, primes are a lot more numerous in the same interval and the sum of their reciprocals diverges to infinity.

Finally, upon a careful inspection of the proofs on these pages, a conscientious reader will realize that she can improve the constants in all of the results. We have not strived for the best constants as those are clearly given by the Prime Number Theorem whose proof will be much more involved and require a long preface. It is our hope that by revisiting these simple and elegant ideas of Chebyshev, we can kindle some interests in number theory in a general audience. For the enthused reader who would like to learn more, we refer her to [1].

[5] In fact, the proof here actually gives that the sum of the reciprocals of primes not exceeding x is at least $c\ln\ln x$ for some positive absolute constant c.

2 Preliminaries

Our proof of Theorem 1 uses the properties of a special binomial coefficient. Let

$$a_n = \binom{2n}{n}.$$

We know that

$$a_n = \frac{(2n)!}{(n!)^2} = \frac{(n+1)\cdots(2n)}{1\cdot 2\cdots n}.$$

Let $\mathrm{ord}_p n$ be the highest power of p that divides n; that is, $p^{\mathrm{ord}_p n}$ divides n but $p^{\mathrm{ord}_p n+1}$ does not. For example,

$$\mathrm{ord}_2 24 = 3\,,\ \mathrm{ord}_3 24 = 1\,,\ \mathrm{ord}_5 24 = 0\,.$$

One can easily check that, for all $n, j \in \mathbb{N}$,

$$n = \prod_p p^{\mathrm{ord}_p n} \quad \text{and} \quad \mathrm{ord}_p n^j = j \cdot \mathrm{ord}_p n\,. \tag{3}$$

Furthermore, if $a, b, a/b \in \mathbb{N}$, then

$$\mathrm{ord}_p \frac{a}{b} = \mathrm{ord}_p a - \mathrm{ord}_p b\,. \tag{4}$$

We shall use the notation $[x]$ to denote the integer part of $x \in \mathbb{R}$ and $\{x\}$ the fractional part of x, so that

$$[x] = \max_{n\in\mathbb{Z}}\{n \le x\}$$

and $\{x\} = x - [x]$. Clearly, $[x] \le x$ and $0 \le \{x\} < 1$ for all x.

Lemma 4.

$$\mathrm{ord}_p n! = [n/p] + [n/p^2] + [n/p^3] + \cdots\,.$$

Proof. Note that the sum is finite as $[n/p^k] = 0$ if $p^k > n$. Now, $[n/p^k]$ gives precisely the number of multiples of p^k not exceeding n. Let $m \in \mathbb{N}$ and $m \le n$. If $\mathrm{ord}_p m = k$, then m gives k factors of p in $n!$, and m is a multiple of p^j for all $1 \le j \le k$. The contribution of m to $\mathrm{ord}_p n!$ is accounted for exactly k times in $[n/p^j]$ for $1 \le j \le k$. □

Lemma 5. *Let $x > 0$. Then $[2x] - 2[x]$ is either 0 or 1.*

Proof. If $0 < \{x\} < 1/2$, then $2\{x\} < 1$ and

$$[2x] - 2[x] = [2[x] + 2\{x\}] - 2[x] = 2[x] - 2[x] = 0\,.$$

Now, if $1/2 \le \{x\} < 1$, then $1 \le 2\{x\} < 2$ and

$$[2x] - 2[x] = [2[x] + 2\{x\}] - 2[x] = 2[x] + 1 - 2[x] = 1\,.$$

This completes the proof. □

3 The Upper Bound

First, we prove the upper bound in Theorem 1.
Proof. We certainly have

$$2^{2n} = (1+1)^{2n} = \sum_{j=1}^{2n} \binom{2n}{j} > a_n = \binom{2n}{n} = \frac{(n+1)\cdots(2n)}{1\cdot 2\cdots n}\,. \tag{5}$$

Recall that a_n is definitely an integer. Moreover, if p is a prime number with $n < p < 2n$, then p appears in the numerator of

$$\frac{(n+1)\cdots(2n)}{1\cdot 2\cdots n} \tag{6}$$

but p does not divide the denominator, which is just $n!$, since $n < p$ and p is prime. So, p is a divisor of a_n if $n < p < 2n$. Therefore, a_n must be a multiple of the product of all such values of p, and, hence,

$$a_n \geq \prod_{n<p<2n} p\,.$$

Together with (5), the above inequality gives

$$2^{2n} > \prod_{n<p<2n} p\,.$$

Now, taking the natural logarithm of both sides, we obtain

$$2n\ln 2 > \sum_{n<p<2n} \ln p = \theta(2n) - \theta(n)\,. \tag{7}$$

Given any natural number m,

$$\theta(2^m) = \big(\theta(2^m) - \theta(2^{m-1})\big) + \big(\theta(2^{m-1}) - \theta(2^{m-2})\big) + \cdots + (\theta(4) - \theta(2)) + (\theta(2) - \theta(1))\ .$$

Note that $\theta(1) = 0$. By applying (7) to the right-hand side of the above m times, we get

$$\theta(2^m) < 2\ln 2 \sum_{j=0}^{m-1} 2^j = \ln 2 \frac{2^{m+1}-2}{2-1} < \ln 2 \cdot 2^{m+1}\,.$$

Now, given any $x > 2$, let 2^m be the smallest power of 2 greater than or equal to x so that $2^{m-1} < x \leq 2^m$. We get

$$\theta(x) < \theta(2^m) < \ln 2 \cdot 2^{m+1} = (4\ln 2)2^{m-1} < (4\ln 2)x\,,$$

and thus the desired majorant is established. □

4 The Lower Bound

Let us now get the lower bound in Theorem 1.
Proof. Note that

$$a_n = \binom{2n}{n} = \left(\frac{n+1}{1}\right)\left(\frac{n+2}{2}\right)\cdots\left(\frac{n+n}{n}\right) \geq 2^n ,$$

as each one of the factors in the product is at least 2. Moreover, for a prime p, let us consider

$$\mathrm{ord}_p a_n = \mathrm{ord}_p \frac{(2n)!}{(n!)^2} = \mathrm{ord}_p (2n)! - \mathrm{ord}_p (n!)^2 = \mathrm{ord}_p (2n)! - 2 \cdot \mathrm{ord}_p (n!) ,$$

upon using (3) and (4). Now, Lemma 4 gives

$$\mathrm{ord}_p a_n = \sum_{j=1}^{\infty} \left[\frac{2n}{p^j}\right] - 2\sum_{j=1}^{\infty} \left[\frac{n}{p^j}\right] = \sum_{j=1}^{\infty} \left(\left[\frac{2n}{p^j}\right] - 2\left[\frac{n}{p^j}\right]\right) . \tag{8}$$

Let t_p be the largest integer such that $p^{t_p} \leq 2n$. So,

$$t_p = [\ln_p 2n] = \left[\frac{\ln 2n}{\ln p}\right] \leq \frac{\ln 2n}{\ln p} . \tag{9}$$

The summands in (8) are zero when $j > t_p$. So, we have

$$\mathrm{ord}_p a_n = \sum_{j=1}^{t_p} \left(\left[\frac{2n}{p^j}\right] - 2\left[\frac{n}{p^j}\right]\right) .$$

By Lemma 5, the summands above are either 0 or 1. By using this fact together with (9), we conclude that

$$\mathrm{ord}_p a_n \leq t_p \leq \frac{\ln 2n}{\ln p} .$$

Observe that only primes less than $2n$ divide a_n.

$$2^n \leq a_n = \prod_{p<2n} p^{\mathrm{ord}_p a_n} \leq \prod_{p<2n} p^{t_p} .$$

Taking the logarithm of both sides of the above yields

$$n \ln 2 \leq \sum_{p<2n} t_p \ln p .$$

From (9), we get

$$\ln 2 \cdot n \leq \sum_{p<2n} \left[\frac{\ln 2n}{\ln p}\right] \ln p \leq S_1 + S_2 ,$$

where

$$S_1 := \sum_{p \le \sqrt{2n}} \left[\frac{\ln 2n}{\ln p}\right] \ln p \quad \text{and} \quad S_2 := \sum_{\sqrt{2n} < p < 2n} \left[\frac{\ln 2n}{\ln p}\right] \ln p \,,$$

say. First,

$$S_1 \le \sum_{p \le \sqrt{2n}} \frac{\ln 2n}{\ln p} \ln p = \sum_{p \le \sqrt{2n}} \ln 2n \le \sqrt{2n} \ln 2n \,.$$

Now, the summation condition of S_2, $\sqrt{2n} < p < 2n$, implies that

$$\frac{1}{2} \ln 2n < \ln p < \ln 2n \,, \text{ so that } 1 < \frac{\ln 2n}{\ln p} < 2 \,.$$

Hence, if $\sqrt{2n} < p < 2n$, then

$$\left[\frac{\ln 2n}{\ln p}\right] = 1 \,.$$

Consequently,

$$S_2 = \sum_{\sqrt{2n} < p < 2n} \ln p \le \theta(2n)$$

and we arrive at

$$\ln 2 \cdot n \le \sqrt{2n} \ln 2n + \sum_{\sqrt{2n} < p < 2n} \ln p \le \sqrt{2n} \ln 2n + \theta(2n)$$

or, equivalently,

$$\theta(2n) \ge \ln 2 \cdot n - \sqrt{2n} \ln 2n \,.$$

Now, once again, one can check on a graphing calculator that, for $n \ge 427$,

$$\frac{2}{3} \ln 2 \cdot n > \sqrt{2n} \ln 2n \,.$$

So, if $n \ge 427$, then

$$\theta(2n) \ge \frac{\ln 2}{3} n \,.$$

Now, if $x \ge 854$ and $[x]$ is even, then $[x]/2 = n \ge 427$ and $2n \le x < 2n + 1$. Therefore,

$$\theta(x) \ge \theta(2n) \ge \frac{\ln 2}{3} n > \frac{\ln 2}{3} \frac{x-1}{2} = \frac{\ln 2}{6}(x-1) > \frac{\ln 2}{7} x \,.$$

The last inequality is true for all $x \ge 8$ (this is easily verifiable by a graphing calculator). If $[x]$ is odd, then $[x+1]$ is even and $\theta(x) = \theta(x+1)$. The above bound gives

$$\theta(x) = \theta(x+1) > \frac{\ln 2}{7}(x+1) > \frac{\ln 2}{7} x \,.$$

Thus, we have completed the proof of Theorem 1. □

References

[1] H.M. Edwards, *Riemann's Zeta Function*, Dover Publications, Inc., New York, 2001.

[2] A. Weil, *Number Theory: An Approach Through History from Hammurapi to Legendre*, Birkhäuser, Basel, 1983.