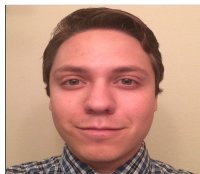# Optimal limits for $\pi(n)$ using Zagier's method

Philip L. Bretz and Connor C. McBryde

Philip L. Bretz is a Mathematics major at the University of Oklahoma. He studies Mathematics to gain a better understanding of the world. After graduating, he plans to continue his studies in graduate school.

Connor C. McBryde is an alumnus of the University of Oklahoma College of Engineering. His research interests include machine vision and signal processing. He plans to attend the University of Michigan for a Masters degree in Electrical and Computer Engineering.

Abstract Zagier [4] showed using elementary methods that $\pi(n)$, the number of primes less than or equal to n, can be bounded by $\frac{2}{3}\frac{n}{\log n} < \pi(n) < 1.7\frac{n}{\log n}$. His proof, which used strong induction and properties of binomial coefficients, yields upper and lower bounds that vary as a function of N, the inductive base step. We improve on Zagier's claims by showing that $0.69314\frac{n}{\log n} < \pi(n) < 1.50746\frac{n}{\log n}$ for all n. We further demonstrate that, as the parameter N tends to infinity, the best possible constants that could be obtained using Zagier's method are $\log 2$ and $\log 4$, respectively.

Introduction

The reader is likely familiar with prime numbers, including their essential role as the building blocks of the natural numbers. The fundamental theorem of arithmetic formalizes this by stating that every natural number n is either prime or a product of primes where the product of primes is unique for every n, meaning that each natural number has only one prime factorization.

The fundamental theorem of arithmetic may cause one to wonder: while there is clearly an infinite amount of natural numbers, can the same be said for the primes? It turns out that there are indeed an infinite number of primes, and this was proven by Euclid over two thousand years ago [1]. Euclid's proof is compellingly simple: if there are a finite number of primes, write them in ascending order as $p_1, p_2, \ldots, p_n$. Then let us consider the number $P = p_1 p_2 \ldots p_n + 1$. Because $P > 1$, it follows that there is some p_i for $1 \le i \le n$ that divides P. Because p_i divides P, it must also divide

$P - p_1 p_2 \ldots p_n$, but $P - p_1 p_2 \ldots p_n = 1$. Since 1 is only divisible by itself, $p_i = 1$, which is impossible. Thus any finite list of primes is not complete.

Euclid's discovery of the infinitude of the prime numbers is accompanied by yet another question: how can we find large primes? In particular, how can we determine if a number is prime or not? Of course, by definition, a prime is an integer greater than 1 that is only divisible by 1 and itself. So, to check if a number n is prime, one may choose every natural number $a < n$, and ensure that none of these numbers divides n. However, this process becomes very time-consuming for large values of n. One clever improvement is to check only the numbers $a \leq \sqrt{n}$, for if some number a divides n, then n/a divides n as well. Further, we may choose to check only prime numbers $p \leq \sqrt{n}$, because if some composite number divides n, then its prime factors do as well. Finally, for numbers $n > 5$, we may safely disregard any number ending in $0, 2, 4, 5, 6,$ or 8, because these endings indicate that the number is either even or divisible by five. There is a famous method for finding primes, called the Sieve of Eratosthenes, a description of which is offered on p. 3 of [3].

Unfortunately, algorithms like the above do not necessarily provide us with insight on the distribution of large primes—they can only tell us whether a number is prime or not. Further, with any algorithm, whether performed by computer or by hand, there are infinitely many primes that are simply too large to verify. This leaves one to wonder if we can say anything conclusive about the distribution of very large primes. This leads to one of the most beautiful relations in mathematics, called the Prime Number Theorem (see [3]). It is first necessary to provide a description of the main component of the theorem, which is called the *prime counting function*. This function, symbolized by $\pi(n)$, denotes the number of primes less than or equal to n.

Figure 1 shows a graph of this function. Note how it increases by 1 every time n is prime and so $\pi(n)$ seems somewhat irregular in its growth. However, if we show the behavior of the function for a larger range of n as in Figure 2, we discover something very counterintuitive: the graph of $\pi(n)$ is remarkably well-behaved. This is in stark contrast to the apparent randomness of primes, and we are led to believe (correctly) that there exists some function to approximate this graph. This function is given in the prime number theorem, which states that

$$\lim_{n \to \infty} \frac{\pi(n)}{\left(\frac{n}{\log n}\right)} = 1.$$

This theorem, first proven in 1896, shows that $\pi(n)$ is well approximated by $n/\log n$ as n tends to infinity (see [4]). It does not state, however, how well the approximation holds for finite values of n, and thus much effort has been put into bounding $\pi(n)$, or providing certain constants k and m such that

$$k \frac{n}{\log n} < \pi(n) < m \frac{n}{\log n}.$$

Unfortunately, many of the proofs for these bounds require extensive mathematical knowledge to understand, and thus are not very accessible. It is therefore noteworthy when one can offer a simpler proof and yield the same (or close) results. Chebyshev accomplished this, proving with elementary methods that $\pi(n)$ could be bounded by $0.92 \frac{n}{\log n} < \pi(n) < 1.11 \frac{n}{\log n}$ for sufficiently large n. More on Chebyshev's proof can be found in [2].

Zagier [4] wrote an interesting synopsis of Chebyshev's proof. Using similar elementary methods, Zagier used a strong inductive approach to obtain an upper bound for $\pi(n)$. The method, which relies primarily on the use of binomial coefficients, used a base step of $n < 1200$ to provide an upper bound of $\pi(n) < 1.7 \frac{n}{\log n}$ for all n. More
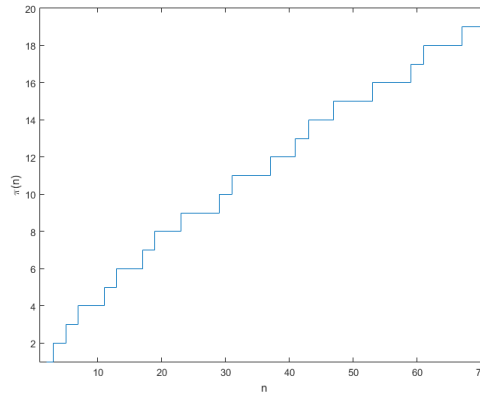
Figure 1: $\pi(n)$ for the first 20 primes
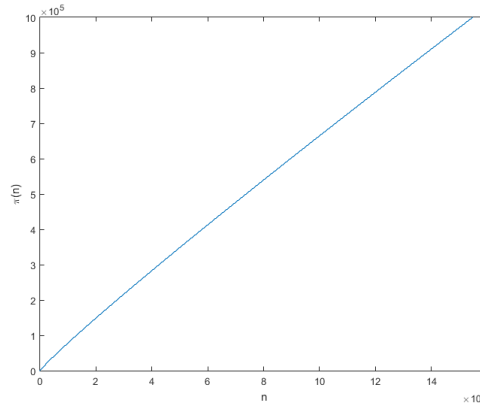
Figure 2: A larger view of $\pi(n)$

precisely, by assuming as a base step that the upper bound holds for all $n < 1200$ an inductive proof of the upper bound for all n is given.

The lower bound $\pi(n) > \frac{2}{3} \frac{n}{\log n}$ is proven for $n > 200$ directly. These bounds are not as good as those obtained by Chebyshev, but this is a consequence of Zagier's method being a simplified case of Chebyshev's. Further, once (i) the lower bound is proven for $n \leq 200$ and (ii) the upper bound is proven for $n \leq 1200$, these bounds hold for all n. This is a convenience not present in Chebyshev's bounds, which require that n be sufficiently large.

With the approach Zagier used, higher values of the inductive base step (i.e., higher values of N for the base step $n < N$) yield better upper bounds. The main results of this paper are as follows. It will be shown that the upper bound constant m, which is a function of the base step N, asymptotically approaches $\log 4$ as $N \to \infty$. The lower bound can likewise be improved and we show that the lower bounding constant k, as a function of N, approaches $\log 2$ as $N \to \infty$.

As an application of these results we will then show that with $N = 1.55 \times 10^7$ (the

approximate value of the one-millionth prime) we can verify that (i) the lower bound holds for $n \leq N$ and (ii) the upper bound holds for $n \leq N$, and thus safely conclude that $\pi(n)$ can be bounded by

$$0.69314 \frac{n}{\log n} < \pi(n) < 1.50746 \frac{n}{\log n},$$

for all n.

We first demonstrate Zagier's proof for comparison purposes. The proof can be found in [4].

Zagier's Proof

Assume as a base for an induction that $\pi(n) < 1.7n/\log n$ for $n < 1200$. (This has been verified for $n < 1200$ and we will discuss more about how to check this later in this paper.) Then, consider the binomial expansion

$$2^{2n} = (1+1)^{2n} = \binom{2n}{0} + \cdots + \binom{2n}{n} + \cdots + \binom{2n}{2n}.$$

It follows that $\binom{2n}{n} \leq 2^{2n}$. Now, note that

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} = \frac{2n(2n-1)(2n-2)\ldots(4)(3)(2)(1)}{n!(n)(n-1)\ldots(2)(1)}.$$

For every factor in $n!$, there is twice that factor in $(2n)!$. Dividing through by $n!$, one obtains

$$\binom{2n}{n} = \frac{2^n(2n-1)(2n-3)\ldots(3)(1)}{n!}.$$

Since every prime less than $2n$ appears in the numerator, but no prime greater than n appears in the denominator, $\binom{2n}{n}$ is divisible by every prime between n and $2n$. That is,

$$\prod_{n<p\leq 2n} p \text{ divides } \binom{2n}{n},$$

where here and elsewhere p always denotes a prime unless otherwise stated. The product above has $\pi(2n) - \pi(n)$ factors, since the product contains all primes greater than n and less than $2n$. Note that the product must be smaller than 2^{2n}, since $\binom{2n}{n} \leq 2^{2n}$. Each prime in the product is greater than n, and this leads to

$$n^{\pi(2n)-\pi(n)} < \prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq 2^{2n}. \tag{1}$$

Taking logarithms of both sides and dividing by $\log n$, one obtains

$$\pi(2n) - \pi(n) < \log 4 \frac{n}{\log n}. \tag{2}$$

Using the inductive assumption that $\pi(n) < 1.7n/\log n$ yields

$$\pi(2n) < (1.7 + \log 4)\frac{n}{\log n} < 3.09\frac{n}{\log n}. \tag{3}$$

For $n \geq 1200$, one has $3.09\frac{n}{\log n} < 1.7\frac{2n}{\log 2n}$, and this leads by (3) to the desired inequality, $\pi(2n) < 1.7\frac{2n}{\log 2n}$, which completes the induction for $\pi(2n)$. For $\pi(2n+1)$, note that

$$\pi(2n+1) \leq \pi(2n) + 1 < 3.09\frac{n}{\log n} + 1. \tag{4}$$

For $n \geq 1200$, the inequality $\pi(2n) + 1 < 3.09\frac{n}{\log n} + 1 < 1.7\frac{(2n+1)}{\log(2n+1)}$ holds. It follows that $\pi(2n + 1) < 1.7(2n + 1)/\log(2n + 1)$. Thus, the inductive step is now verified for both even and odd cases, which allows us to conclude the upper bound $\pi(n) < 1.7n/\log n$ for all $n \geq 1200$.

The lower bound for $\pi(n)$ will now be shown by utilizing a short lemma, a proof of which is provided by Zagier on p. 14 of [4].

Lemma 1. *Let p be a prime and $n \geq k > 0$. If v_p is the largest power of p such that $p^{v_p}|\binom{n}{k}$, then $p^{v_p} \leq n$.*

A consequence of the fact that $p^{v_p} \leq n$ is

$$\prod_{p \leq n} p^{v_p} \leq n^{\pi(n)}, \tag{5}$$

which we now show. The left hand side of (5) is the prime factorization of $\binom{n}{k}$. The inequality holds because there are $\pi(n)$ distinct primes in the product, and by Lemma 1 each p^{v_p} is less than n.

Since $\binom{n}{k} \leq n^{\pi(n)}$, and $2^n = \sum_{k=0}^{n} \binom{n}{k}$, it follows that $2^n \leq (n + 1)n^{\pi(n)}$. Taking logarithms yields the inequality $n \log 2 \leq \pi(n) \log n + \log(n + 1)$. Finally, rearranging terms gives

$$\pi(n) \geq \left(\log 2 - \frac{\log(n + 1)}{n}\right)\frac{n}{\log n}. \tag{6}$$

For $n > 200$, the bound (6) yields $\pi(n) \geq 0.6667\frac{n}{\log n} > \frac{2}{3}\frac{n}{\log n}$. The value 0.6667 is very close to the value of $2/3$ that Zagier chose. Combining the inequalities for the upper and lower bounds yields the final inequality obtained by Zagier,

$$\frac{2}{3}\frac{n}{\log n} < \pi(n) < 1.7\frac{n}{\log n}, \tag{7}$$

which holds for all n. Note that in order to complete the proof that (7) holds for **all** n, one must first independently verify both (i) that the upper bound holds for all $n < 1200$ (as this was used as the base for the inductive proof of the upper bound for all $n > 1200$), and (ii) that the lower bound holds for $n \leq 200$ (though this is not needed to assert the lower bound for all $n > 200$).

Improving on Upper and Lower Bounds

We will now examine, using Zagier's method, the optimal values of k and m for the inequality

$$k\frac{n}{\log n} < \pi(n) < m\frac{n}{\log n}. \tag{8}$$

Here k and m will be referred to as the lower bound constant and upper bound constant, respectively. Zagier gave an upper bound constant of $m = 1.7$ and a lower bound constant of $k = 2/3$. We shall use a generalized version of the approach Zagier used in his proof.

First, assume as a base for our induction that $\pi(n) < m\frac{n}{\log n}$, for $n < N$; this would then need to be verified independently as in the discussion following (7). From (2), we have that $\pi(2n) - \pi(n) < \log 4\frac{n}{\log n}$. Since by induction $\pi(n) < m\frac{n}{\log n}$, it follows that

$$\pi(2n) < (m + \log 4)\frac{n}{\log n}. \tag{9}$$

We seek to find which values of m satisfy

$$(m + \log 4)\frac{n}{\log n} \leq m\frac{2n}{\log 2n}. \tag{10}$$

If m satisfies (10), it will satisfy the desired inequality $\pi(2n) < m\frac{2n}{\log 2n}$.

Lemma 2. *The constant m satisfies* (10) *if and only if* $m \geq \log 4 \frac{\log n + \log 2}{\log n - \log 2}$.

Proof. Dividing both sides of (10) by $\frac{n}{\log n}$ yields $m + \log 4 \leq m\frac{2\log n}{\log 2n}$. Subtracting m from both sides, and noting that $\log 2n = \log n + \log 2$, one obtains

$$\log 4 \leq m\left(\frac{2\log n}{\log n + \log 2} - \frac{\log n + \log 2}{\log n + \log 2}\right) = m\frac{\log n - \log 2}{\log n + \log 2}.$$

Finally, multiplying both sides by $\frac{\log n + \log 2}{\log n - \log 2}$, we have

$$m \geq \log 4 \frac{\log n + \log 2}{\log n - \log 2}. \tag{11}$$

Since the above steps are reversible the lemma follows. $\qquad\square$

Now, note that the best bound that can be obtained is the minimum value that satisfies (11), i.e., $m = (\log 4)(\log n + \log 2)/(\log n - \log 2)$. Recall that the upper bound constant in Zagier's proof was 1.7 with a base step of $n < 1200$. With this base step (i.e., $N = 1200$) we have that the optimal upper bound constant that can be obtained is $m = (\log 4)(\log 1200 + \log 2)/(\log 1200 - \log 2) \approx 1.6867$.

From (4), we know that $\pi(2n + 1) \leq \pi(2n) + 1$. Since $\pi(2n) < (m + \log 4)\frac{n}{\log n}$, it follows that $\pi(2n + 1) < (m + \log 4)\frac{n}{\log n} + 1$. We now seek values of m that satisfy

$$(m + \log 4)\frac{n}{\log n} + 1 \leq m\frac{(2n + 1)}{\log(2n + 1)}. \tag{12}$$

If m satisfies (12), then the inequality $\pi(2n + 1) < m\frac{(2n+1)}{\log(2n+1)}$ holds. Note that the upper bound must hold for the case of the evens and the odds. That is, the same constant m must work for both cases. Thus, we are restricted to the larger of the two, i.e. $m = \max\{m_{\pi(2n)}, m_{\pi(2n+1)}\}$, where the subscripts denote which case each constant belongs to. From the case of the evens, we already have the restriction that $m > \log 4$. Because of this, we assume that $m > \log 4$ for the case of the odds. Now consider the inequality

$$m \geq \left(\log 4 + \frac{\log n}{n}\right)\frac{\log(2n + 1)}{\left(\frac{2n+1}{n}\right)\log n - \log(2n + 1)}. \tag{13}$$

Lemma 3. *The constant m satisfies* (12) *if and only if m satisfies* (13).

Proof. With some algebra (13) is transformed into the following.

$$m\left(\left(\frac{2n + 1}{n}\right)\log n - \log(2n + 1)\right) \;\geq\; \left(\log 4 + \frac{\log n}{n}\right)\log(2n + 1)$$

$$m\left(\left(\frac{2n + 1}{n}\right)\log n\right) \;\geq\; \left(m + \log 4 + \frac{\log n}{n}\right)\log(2n + 1)$$

$$m\frac{(2n + 1)\log n}{\log(2n + 1)} \;\geq\; (m + \log 4)n + \log n$$

$$(m + \log 4)\frac{n}{\log n} + 1 \;\leq\; m\frac{(2n + 1)}{\log(2n + 1)}$$

which is (12). $\qquad\square$

The best upper bound constant that can be obtained for the odds is the minimum value of m that satisfies (13). If m does not satisfy (13), then it does not satisfy (12), by the same reasoning as presented for the evens. We will now show that the optimal value of m approaches $\log 4$ as $n \to \infty$ for both the even and the odd case. To prove it for the evens, we refer to (11), which states that the optimal constant is given by $m = \log 4 \frac{\log n + \log 2}{\log n - \log 2}$. To determine the optimal upper bound constant, we take the limit as n tends to infinity:

$$m_{optimal} = \lim_{n \to \infty} \log 4 \frac{\log n + \log 2}{\log n - \log 2} = \log 4.$$

The optimal upper bound constant for $\pi(2n+1)$ is obtained using a similar procedure. L'Hôspital's rule quickly gives $\lim_{n \to \infty} \frac{\log n}{n} = 0$ and $\lim_{n \to \infty} \frac{\log (2n+1)}{\log n} = 1$ which are now used together with (13) to show that the optimal upper bound constant for the odds is

$$m_{optimal} = \lim_{n \to \infty} \left(\log 4 + \frac{\log n}{n}\right) \frac{\log (2n + 1)}{(2 + \frac{1}{n}) \log n - \log (2n + 1)} = \log 4.$$

Thus, as was the case for $\pi(2n)$, the optimal upper bound constant for $\pi(2n + 1)$ is $m = \log 4$. Since this is the same constant as that for the evens, the optimal upper bound constant for $\pi(n)$ using Zagier's method is $m = \log 4$.

For the lower bound constant, Zagier's proof shows that $\pi(n) > \frac{2}{3} \frac{n}{\log n}$ for $n > 200$. We seek to find the optimal value of k for the inequality $\pi(n) > k \frac{n}{\log n}$. Recall from (6) that $\pi(n) > (\log 2 - \frac{\log (n+1)}{n}) \frac{n}{\log n}$. Hence by L'Hospital's rule the lower bound constant

$$k = \log 2 - \frac{\log (n + 1)}{n} \tag{14}$$

approaches $\log 2$ as $n \to \infty$. Thus, the optimal value for the lower bound constant is $\log 2$.

Finally, combining the optimal lower and upper bound constants, we have proven the optimal bounds obtainable from Zagier's method are

$$\log 2 \frac{n}{\log n} < \pi(n) < \log 4 \frac{n}{\log n}.$$

We emphasize that it does *not* follow that these bounds actually hold for all n, as this would require letting the base step in the induction tend to infinity. It is unrealistic to assume that one may actually verify all values of $\pi(n)$ for $n \to \infty$. Indeed, if it were possible, then all primes would become known and there would no longer be any need to come up with bounds for $\pi(n)$. Thus, we are left with the conclusion that one may not obtain the exact values $\log 2$ and $\log 4$, but may achieve near-optimal constants using very large values of N for the inductive base step. In other words (11), (13), and (14) state that a proof can yield bounds no better than m and k for a given base step N.

To accompany this discussion, we determined how well Zagier's method performs using a larger value of the inductive step. To accomplish this, we obtained a list of the first one million primes, with the maximum prime being 15,485,863. With $N = 15,485,863$ as the inductive base step, we used MATLAB to find the optimal m in (11) and (13), yielding upper bound constants of 1.507450382 for the even case and 1.507451449 for the odd case. Since the constant provided by (13) is the larger of the two, this is the best constant that this inductive base step can provide. We rounded this constant to 1.50746. The lower bound constant was determined through (14)
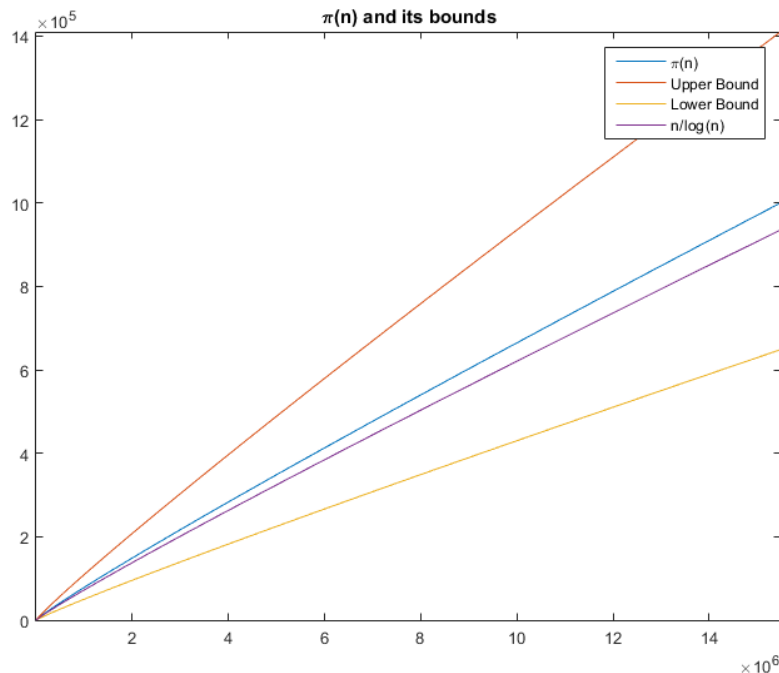
Figure 3: Top to bottom the graphs are of the Upper Bound, $\pi(n)$, $\frac{n}{\log n}$, and the Lower Bound.

to be 0.693146111, which we rounded down to 0.69314. We then used MATLAB's plotting tools to generate the graphs (see Figure 3) of four functions: $\pi(n)$, its upper and lower bounds, and $\frac{n}{\log n}$ for comparison. This graph suggests that the base for the induction is verified for $n < 15, 485, 863$.

Although graphically it is quite certain that the inequality holds for $n < N$, we verified this numerically. We therefore obtain the bounds

$$0.69314 \frac{n}{\log n} < \pi(n) < 1.50746 \frac{n}{\log n}$$

for all n. Note for comparison that the best possible upper and lower bound constants stemming from Zagier's method are $\log 2 \approx 0.693147$ and $\log 4 \approx 1.386294$.

References

[1] Dunham, William. Euclid and the Infinitude of Primes. *The Mathematics Teacher*, 1987, Vol. 80(1), pp.16 - 17.

[2] Goldfeld, Dorian. The elementary proof of the prime number theorem: an historical perspective. (Online). Retrieved from `http://www.math.columbia.edu/~goldfeld/ErdosSelbergDispute.pdf`. Accessed June 2, 2016.

[3] Monks, H. V. The Prime Number Theorem. *ProQuest Dissertations & Theses Global*, 1965. Retrieved from `http://search.proquest.com/docview/302191538`. Accessed June 2, 2016.

[4] Zagier, Don. The first 50 million prime numbers. *The Mathematical Intelligencer*, 1977, Vol. 1(1), pp.7-19.