Are there infinitely many twin primes?

Primes, the Prime Number Theorem, Twin Primes, and the Twin Prime Conjecture.

Kathryn Roper and Chris Conrad
MATH 4513: Math Capstone, Dr. Albert
April 23, 2008.

The Twin Prime Conjecture is an interesting unsolved problem in mathematics. Despite its apparent simplicity, there exists no conclusive answer to the question: "Are there infinitely many twin primes?" Some basic properties of prime numbers and apparent trends suggest that there are infinitely many primes. However, to this point, no one has found a way to prove this.

Prime Numbers:

Let's start with a really basic concept: prime numbers. A prime number is a counting number p such that p has only two factors: 1, and p itself.

Twin Primes:

Twin primes are pretty simple too. They are pairs of prime numbers that are two apart. Looking at a list of the first few prime numbers, it is easy to spot the smallest pairs of twin primes:

2,3,5,7,11,13,17,19,23,29,31,...

(3,5), (5,7), (11,13), (17,19), and (29,31) are the first five pairs of twin primes.

First, it is necessary to know a few things about the quantity and distribution of ordinary prime numbers. These facts will help in attempting to draw conclusions about the twin primes.

Number of Prime Numbers:

There would seem to be an infinite number of prime numbers. In fact, the Greek mathematician Euclid proved this over 2000 years ago, and many other more complicated proofs exist. It's actually not that difficult to prove.

Theorem:

There exist an infinite number of primes.

Proof:

Assume there are only finitely many primes, and proceed to a contradiction. Denote the finite number of primes as $p_1, p_2, \ldots, p_n$. This list then contains every number that is a prime number.

Then consider a new term, z. Let z be 1 plus the product of all the prime numbers.

$z = 1 + p_1 p_2 \ldots p_n$.

Then let a new number x be a prime number that divides z.

However, x obviously cannot be any of the prime numbers $p_1 p_2 \ldots p_n$. If x was one of these numbers, then it would divide $p_1 p_2 \ldots p_n$. It would then divide $z-1 = p_1 p_2 \ldots p_n$. But x cannot divide both z-1 and z and be a prime number.

Therefore x must be an entirely new prime number not contained in a finite list of primes. Thus, there are an infinite number of primes. But, can we say the same thing about twin primes?

Another interesting thing about Prime Numbers:

After 2, 3, and 5, prime numbers take on some characteristic behavior. Finding prime numbers can be made a little bit easier by this simple theorem that's also not too bad:

Theorem:

All primes greater than 3 are of the form $6n + 1$ or $6n - 1$, where n is a counting number.

Proof:

2 and 3 are primes. Then all multiples of 2 and 3 greater than 2 and 3 are not primes. 5 is of the form $6n - 1$, where $n = 1$. 5 is prime. Then all multiples of 5 greater than 5 are not primes. Then consider modular arithmetic in base 6:

There are six subdivisions of the counting numbers mod 6:

0 mod 6: 6, 12, … = 6n, n is a counting number
1 mod 6: 1, 7, 13, … = 6n + 1, n is a counting number
2 mod 6: 2, 8, 14, … = 6n + 2, n is a counting number
3 mod 6: 3, 9, 15, … = 6n + 3, n is a counting number
4 mod 6: 4, 10, 16, … = 6n + 4, n is a counting number
5 mod 6: 5, 11, 17, … = 6n + 5, n is a counting number.

Any counting number that is equivalent to 0 mod 6, 2 mod 6, or 4 mod 6 is even, and thus is divisible by 2. $6n/2 = 3n$, $(6n + 2) / 2 = 3n + 1$, $(6n + 4) / 2 = 3n + 2$. $3n$, $3n + 1$, and $3n + 2$ are all counting numbers. Thus any such number is not prime, unless it is 2 itself.

Any counting number that is equivalent to 3 mod 6 is divisible by 3, since $(6n + 3) / 3 = 2n + 1$, a counting number. Thus any such number is not prime, unless it is 3 itself.

Any counting number that is equivalent to 1 mod 6 or 5 mod 6 then may be prime or may not be prime. Not all numbers of the form $6n + 1$ or $6n + 5 \equiv (6n - 1) \bmod 6$ are prime, but all primes greater than three fall into one of those two categories.

How this relates to Twin Primes:

Since primes can only be written as $6n + 1$ or $6n - 1$ when they're greater than 3, it's blatantly obvious that for twin primes to exist, they have to be a pair of numbers that can be written as $(6n - 1, 6n + 1)$. Otherwise, there's no way for them to be primes and still be two apart.

Fun Fact:

Computers somewhere out there are hard at work trying to find amazingly large twin primes. There is something called the "Twin Internet Prime Search" trying to find enormously large twins. So far, the biggest pair found is: $2{,}003{,}663{,}613 \times 2^{195000} \pm 1$, which has 58711 digits.

Number of Twin Primes:

Here's where it gets tricky. Similar reasoning seems like it ought to work for twin primes, just like it worked for regular primes. On the other hand, nobody has found a way to make a similarly simple argument. Let's consider more facts about regular prime numbers, and how mathematicians have tried to approximate the distribution of prime numbers over the integers, to see if any of this can help in understanding the more complicated case of twin primes.

Distribution of Primes:

One of the easiest ways to visualize the distribution of the prime numbers is to look at a graph comparing the number of prime numbers with the number of counting numbers. That graph is at the end of the paper in the graphs section. (GRAPH 1) This ratio can be denoted by $\pi(x)$. Let n be a natural number, and let $\pi(n)$ be the number of counting numbers less than or equal to n that are prime. For example, $\pi(3) = 2$, since 2 and 3 are prime numbers, but 1 is not. And $\pi(100) = 25$, since there are 25 primes in the first 100 counting numbers, which are all the integers from 1 to 100. Considering only the first 100 counting numbers, there are already 25 primes. The graph is all jagged and stepped. However, the overall rate at which they occur is quite clearly decreasing and leveling out.

Now take a look at the next graph. It shows the behavior of $\pi(x)$ over wider and wider ranges. Looking at the number of primes between 0 and one million, the graph of $\pi(x)$ really resembles a smooth curve that's slowly decreasing in slope, not just a jagged series of steps like the first graph. (GRAPH 2)

Prime Number Theorem:

It would be convenient, and it is extremely likely, that a curve that looks that smooth approaches a normal and simple function as x goes out toward infinity. The Prime Number Theorem makes this simple projection:

$\pi(x) \sim x / \ln(x)$, when $x \to \infty$.

That is, $\pi(x)$ is related to x divided by the natural logarithm of x (when x approaches infinity). The ratio of the two functions is 1 as x goes to infinity. It's not strictly equal, just related, as you can see by the next graph. This also means that the nth prime number is about $n \ln(n)$. The number of prime numbers is growing slightly faster than $x / \ln(x)$. (GRAPH 3)

Probability of a number being Prime:

This is also equivalent to saying that $1 / \ln(x)$ is closely related to the probability that a number between 1 and x is a prime number, since there are x such numbers in this range and about $x / \ln(x)$ of them are prime. We can narrow this down for a particular number n. Its probability of being prime is then about $1 / \ln(n)$.

Using this relation and these facts about probability, along with some calculus, we can get an even closer estimate.

Logarithmic Integral:

Instead of just using a single function, we can take the sum of all the point probabilities. This is directly related to integrating the density function, which is $1/\log(x)$. This is a big complicated function called li(x).

The summation of all the point probabilities less than x can be written as:

$$\sum_{2 \leq n \leq x} 1/\ln(n).$$

This is directly related to the integral of the density function, li(x). li(x) is also known as the logarithmic integral of x.

$$\mathrm{li}(x) = \int_2^x 1/\ln(n)\, dn.$$

If you work this integral out, using integration by parts:

$$li(x) = 0!x / \ln(x) + 1!x / [\ln(x)]^2 + 2!x / [\ln(x)]^3 + \ldots$$

$$= \sum_{a=0}^{\infty} a!x / [\ln(x)]^{a+1}$$

This gives a remarkably accurate approximation of $\pi(x)$. It is many times closer than $x/\ln(x)$ at values just larger than a million, though it overshoots the value of $\pi(x)$ by just a little bit. (GRAPH 4)

Twin Prime Conjecture:

Now that we've seen graphical methods to show the distribution of individual primes, and seen a proof that there are an infinite number of primes, can we use this to definitively answer the question "Are there infinitely many twin primes?" Can we even use these methods to make an intuitive guess? It does appear that these methods will suggest that the Twin Prime Conjecture is true, and that there are infinitely many twin primes. It also appears that there will be no simple or easy way to conclusively prove the conjecture, despite the evidence that suggests it is true.

Twin Primes:

There are obviously less twin primes than primes as a whole, first off since it takes two prime numbers to make a pair of twin primes, and secondly because there's only the possibility that they'll come on either side of a multiple of six. In fact, there are an awful lot less. The rate of increase of the number of twin primes is much slower than that of ordinary primes.

Where there were 25 primes total less than 100, there are only 8 pairs of twin primes between 1 and 100. (3,5), (5,7), (11,13), (17,19), (29, 31), (41, 43), (59, 61), (71,73) are those eight pairs. Let's call the function representing the integers versus the twin primes ß(x). Let it increase by one every time it encounters the first prime in a pair of twin primes. Then ß(3) = 1, since it's encountered the first prime in the twin prime pair (3,5). ß(11) = 3, since it's encountered the first prime in three pairs: (3,5), (5,7), and (11,13). ß(100) = 8, since there are eight pairs of twin primes with the first number in the pair less than 100. ß(1000) = 35, since there are 35 pairs of twin primes with the first number in the pair less than 1000. This can be represented graphically. The graph is even more staggered and has more irregular steps than that of $\pi(x)$, and increases much more slowly than $\pi(x)$.(GRAPH 5)

At large values, it's even easier to see how much slower the twin prime function ß(x) increases when compared to $\pi(x)$. Where there were 80000 primes in the first million integers, there are only 8000 pairs of twin primes. That means one in 5 primes is part of a pair of twin primes. (GRAPH 6)

Approximation of ß(x):

Like with the regular primes, there ought to be a function that does a pretty good job of approximating the distribution of twin primes. One good guess would be $li_2(x)$, since we're dealing with two primes at a time instead of one.

$$li_2(x) = \int_2^x 1/[\ln(n)]^2 \, dn.$$

Another guess would be $x/[\ln(x)]^2$, but since that was already looking inaccurate for regular primes, there's not a good chance it'll do any better for twin primes.

When we graph ß(x) and these two plausible approximations, it's clear neither approximation is correct. (GRAPH 7).

However, if we correct $li_2(x)$ with a constant c, c $li_2(x)$ ends up being a remarkably good approximation of the number of twin primes. This relies on something called a Cramér model, proposed by the Swedish mathematician Harald Cramer in the 1930s. It basically states that the probability of any integer x being prime is $1/\ln(x)$. Then for two integers k and k+2 to be prime, the probability of them both being prime is $1/\ln(k)[1/\ln(k+2)]$. But since k+2 is very closely related to k, we can simplify this to $1/\ln(k)^2$. This allows us to find c to be about 1.32032. This gives the best approximation possible, with c $li_2(x)$ just slightly less than ß(x) at values around a million. (GRAPH 8)

This gives a good approximation that's related to the distribution of regular prime numbers. This leads mathematicians to think that it's likely that there are an infinite number of twin primes, just like there are an infinite number of normal primes.

Warning:

However, it's risky business to just blindly assume that large numbers will always keep following the same trends. There might be a decent chance they will, but it's not always a guaranteed thing. So just because the graph of the twin primes appears to be increasing, and just because it seems intuitively right that there will always be a bigger set of twin primes, there's no absolute guarantee that this will happen. So we can't just say we've proved the twin prime conjecture by looking at graphs. There's always the risk something unexpected will happen. In this case, the gamble we can't take is that there would stop being pairs of twin primes outside the scope of the graphs we've included.

Other, Non-Graphical Ways Someone Might Conclusively Solve the Twin Prime Conjecture:

Mathematicians have tried to use sieve theory to minimize gaps between primes.

Sieve Theory:

Sieve theory involves sorting through a list of numbers to make it possible to count how many numbers in the list fit a certain property.

Sieve of Eratosthenes:

The first and most famous example was the Sieve of Eratosthenes. Eratosthenes was a Greek mathematician who figured out a method for eliminating all composite numbers out of a list of integers from 2 to n, to leave only a list of the prime numbers between 2 and n.

The process is as follows:

1. Write out the list of all integers from 2 to n.
2. Include 2 in the list of prime numbers, since it is obviously prime. Then cross out all multiples of 2 greater than 2: 4, 6, 8, ….
3. Include 3 in the list of prime numbers, since it is obviously prime. Then cross out all multiples of 3 greater than 3 that haven't already been crossed out: 6(already crossed out), 9, 12(already crossed out), 15, ….
4. Continue including the next largest known prime p in the list of prime numbers, and crossing out all multiples of p. Continue doing so until $p \geq \text{sqrt}(n)$.

When p = sqrt(n), every composite number less than n will have been eliminated. Take n to be the product of two factors x and y. If $x \geq \text{sqrt}(n)$, then $y \leq \text{sqrt}(n)$. If y was not, then $y > \text{sqrt}(n)$, and $xy > \text{sqrt}(n)\text{sqrt}(n) = n^2$.

This will leave only the list of all prime numbers between 2 and n.

Sieves and the Twin Prime Conjecture:

Sieves like the Sieve of Eratosthenes have been used to look for twin primes and to try to prove the twin prime conjecture.

Goldston, Pintz, and Yildirim, mathematicians from the US, Hungary, and Turkey, looked for abnormally small gaps between prime numbers. To give some small examples: The gap between the first two prime numbers, 2 and 3, is 1. The gap between the next two, 3 and 5, is 2. These gaps have a tendency to get larger as the numbers you're considering get larger, since primes are less dense at greater values. The average gap between two primes is about ln(p) for any number p.

In the 1980s, Dr. Helmut Maier from Michigan had been able to prove there exist infinitely many gaps between prime numbers that are around ¼ the size of an average gap, or ln(p)/4. However, the average gap between prime numbers gets bigger than 2/(1/4) = 8 at or around e^8, which is about 2981. Let's just say 3000 for simplicity. After that, even though we can say there are infinitely many gaps that are smaller than ln(p) / 4, we can't say those gaps are less than two, so we can't say that there exist infinitely many twin primes at these large values.

Goldston, Pintz, and Yildirim found a way to prove that for sufficiently large numbers, they could get these gaps down to an arbitrarily small number. Basically, they can generate primes that are a certain distance apart, but only by assuming an unproved theorem called the Elliott – Halberstam conjecture, which is well over our heads. Since this depends on an unproved conjecture, they can't definitely prove a way to generate these primes that are a fixed distance apart. If they could, they could choose 2 as the fixed distance and generate twin primes, thus proving the Twin Prime Conjecture. However, no one can yet prove the Elliott – Halberstam conjecture.

Conclusion:

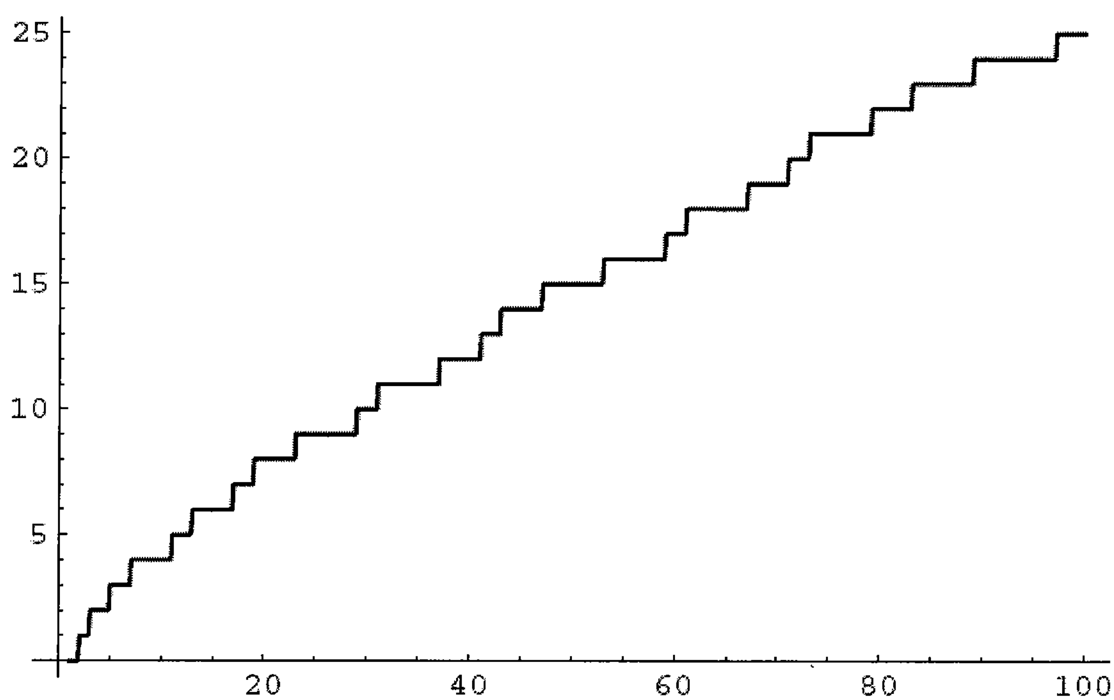
We cannot definitively say whether the Twin Prime Conjecture is true or not. Based on graphical inferences, it appears to be true for reasonably small numbers. However, it cannot be proved by graphical means, and has not yet been proved by other mathematical means, though Goldston, Pintz, and Yildirim have made progress recently. The number of twin primes can be approximated, and the process for doing so follows from the process for approximating the number of regular prime numbers. Properties of ordinary prime numbers help build a foundation for the study of twin primes, and may assist in the eventual proof of the Twin Prime Conjecture.

Graphs:

1. D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 5.

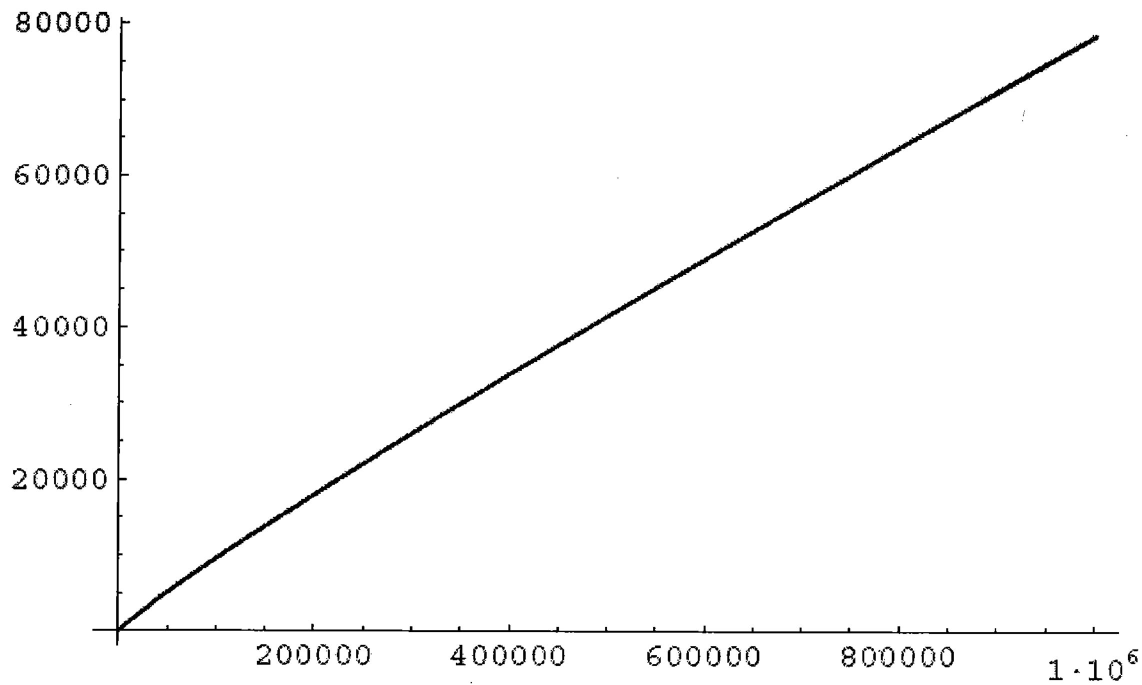
The graph shows non-negative integers on the horizontal axis versus the number of primes less than or equal to an integer on the vertical axis. This function, $\pi(x)$, makes a bunch of little steps and is not continuous. The rate of increase appears to be slowly decreasing.

ARE THERE INFINITELY MANY TWIN PRIMES ? 5



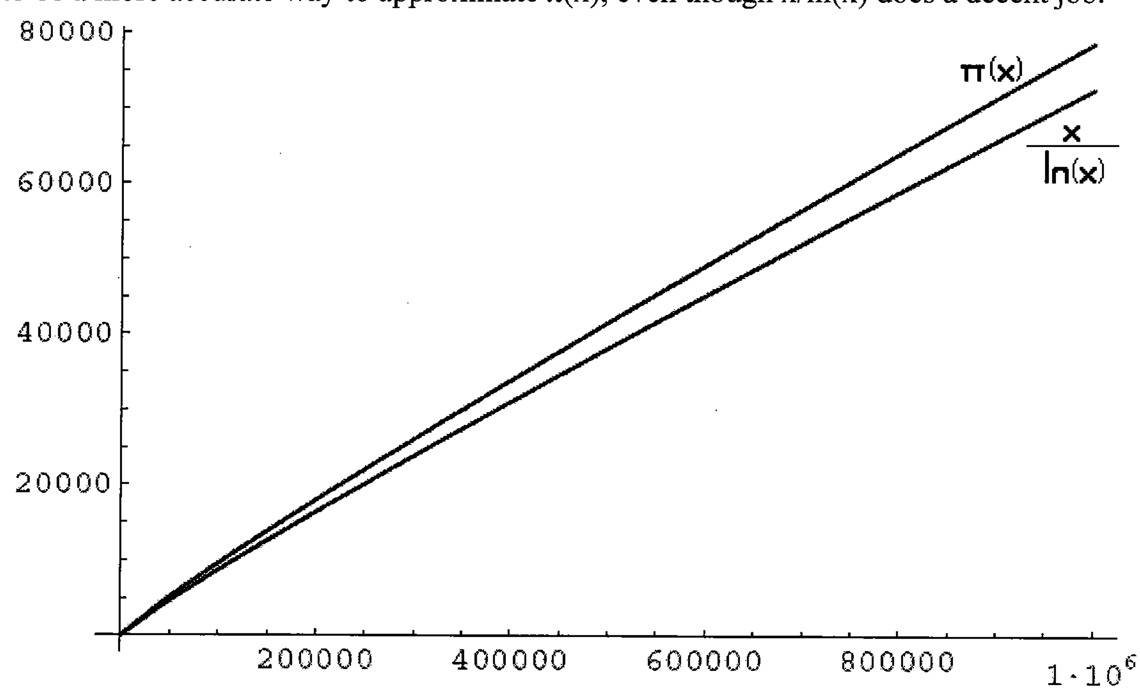
2. D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 6.

This represents $\pi(x)$, same as the first graph, just on a much larger scale. There were 25 primes in the first 100 integers, so the first hundred is 25% prime. But the 10000^{th} prime is 104729, so the first 104729 integers are 10000/104729 = 9.55% prime. In the first million integers, there look to be only about 80000 primes, so the first million is about 8% prime. The rate of increase is definitely decreasing, since there are less primes percentage wise. However, for larger and larger integers, the graph decreases at a much slower rate.



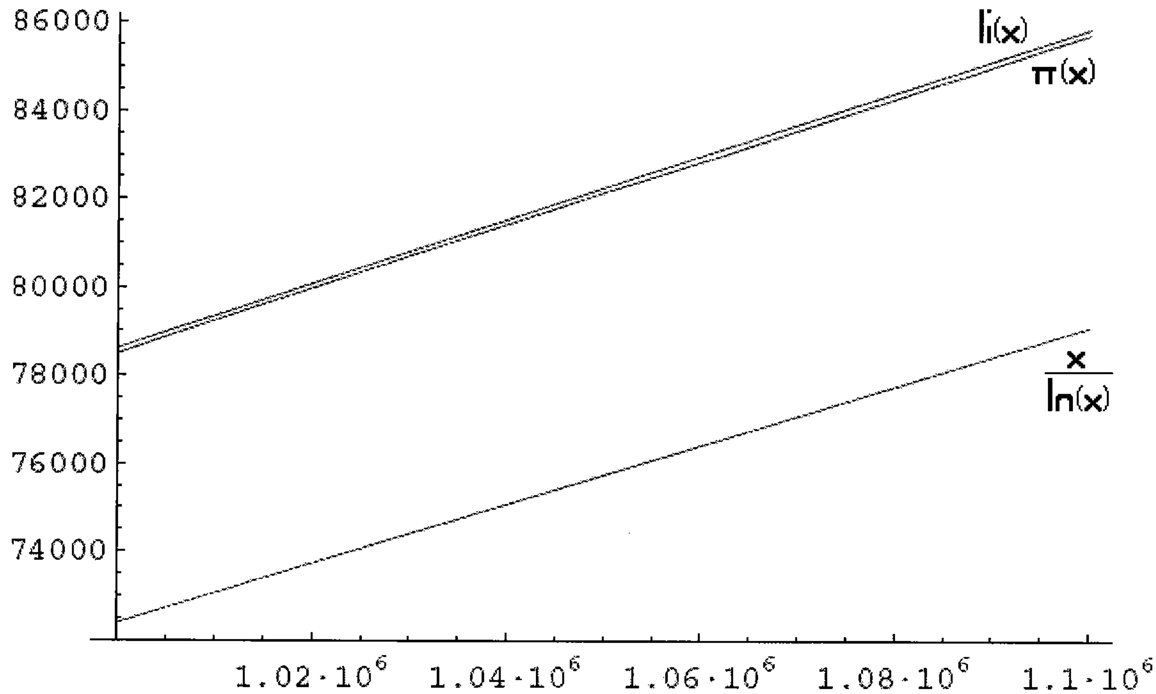
3. Base image from D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 7. Labels added for clarity.

The top function is $\pi(x)$, it's just the number of primes less than or equal to an integer. The bottom function is the approximation given by the prime number theorem, $x/\ln(x)$. It gives a decent approximation at small values, but then the functions diverge. There ought to be a more accurate way to approximate $\pi(x)$, even though $x/\ln(x)$ does a decent job.



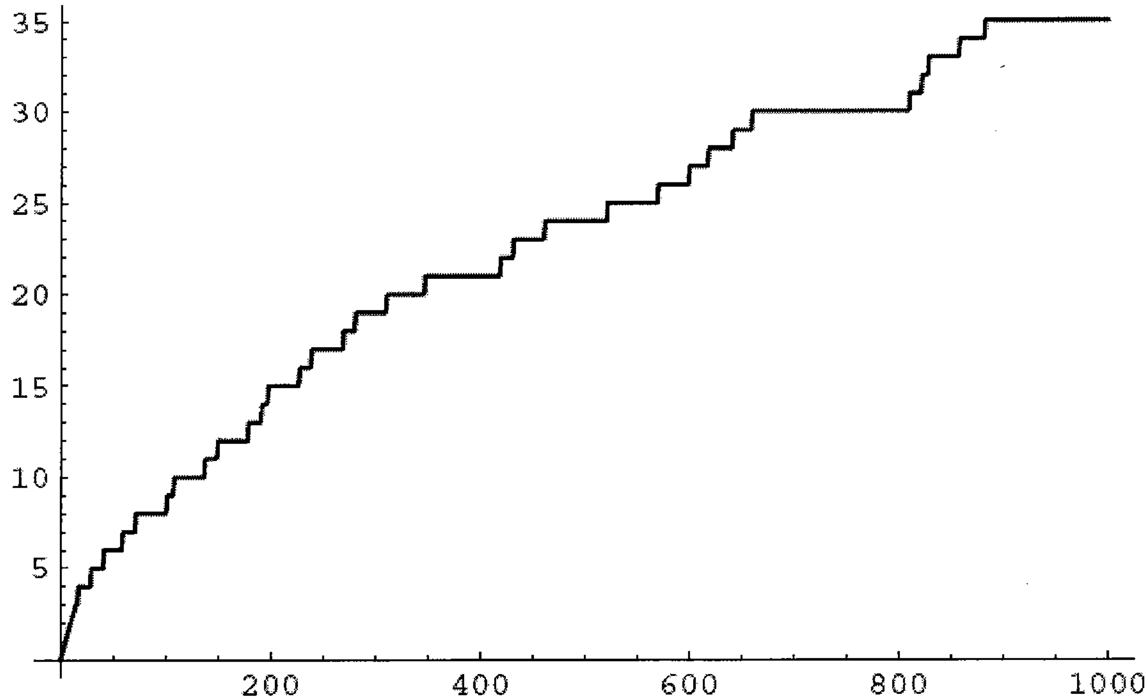
4. Base image from D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 8. Labels added for clarity.

li(x) only about 100 greater than π(x) at values between 1 million and 1.1 million. That's an excellent approximation. x/ln(x) is between 6 and 7 thousand off by this point.



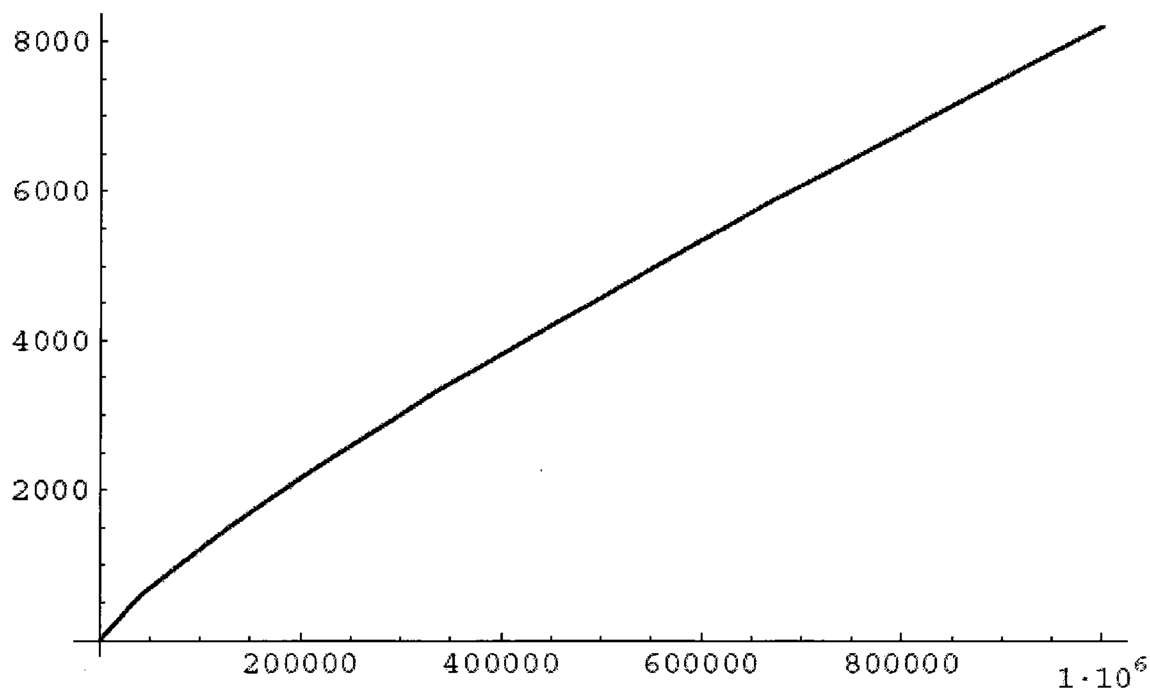
5. D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 11.

ß(x) increases in a jagged, punctuated fashion. There's a huge flat line between (659,661) and (821,823) where there aren't any twin primes. But there are 20 individual primes in that interval. Also note that there are 25 primes in the first 100 integers, but to get to 25 pairs of twin primes, it takes about 500 integers.



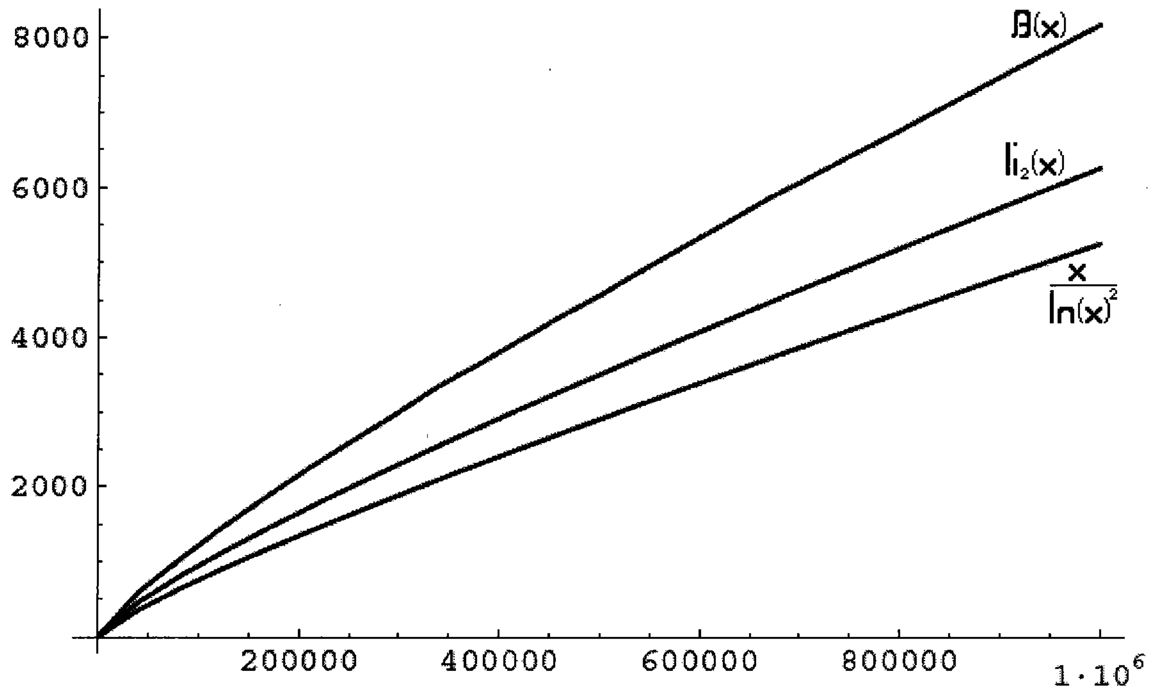
6. D.A. Goldston, *Are There Infinitely Many Twin Primes?*,
http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 11.

ß(x) increases much more slowly than π(x). Its rate of increase is also decreasing.



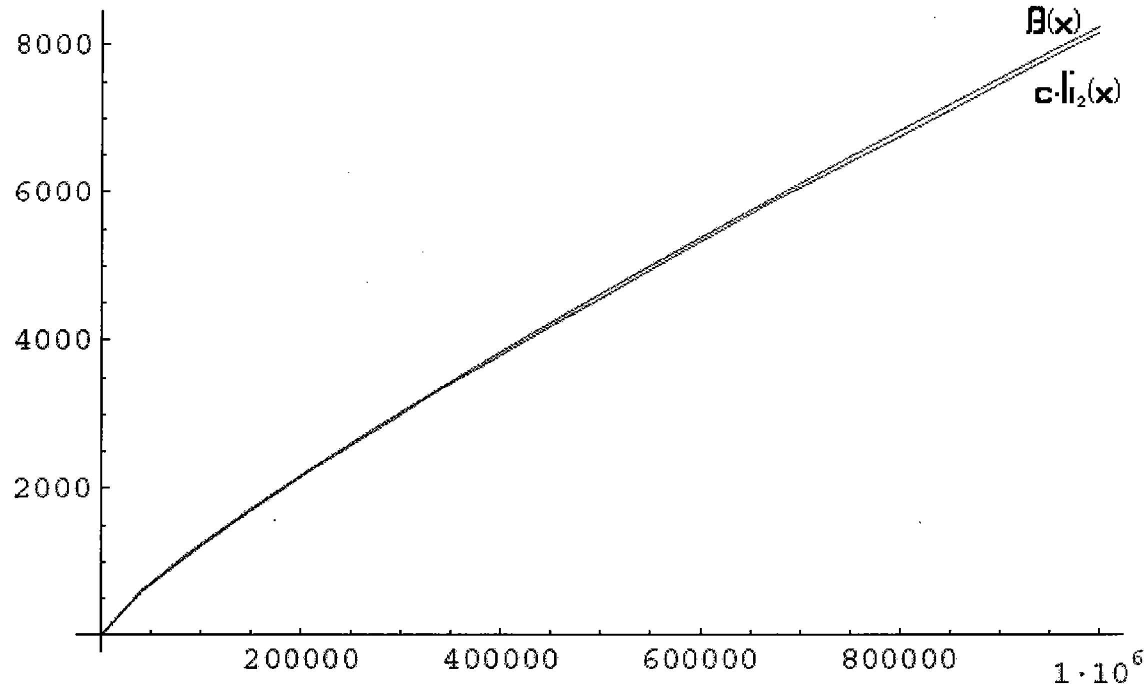
7. D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 12. Labels added for clarity.

This graph compares ß(x), $li_2(x)$, and $x/[\ln(x)]^2$. Both $li_2(x)$ and $x/[\ln(x)]^2$ are incredibly poor approximations of ß(x).



8. D.A. Goldston, *Are There Infinitely Many Twin Primes?*, http://www.math.sjsu.edu/~goldston/twinprimes.pdf, 13. Labels added for clarity.

c $li_2(x)$ ends up being a great approximation of ß(x). ß(x) is only slightly greater at values nearing one million.



Bibliography:

Goldston, D.A. *Are There Infinitely Many Twin Primes?* <http://www.math.sjsu.edu/~goldston/twinprimes.pdf>

http://mathforum.org/library/drmath/view/61378.html

http://mathworld.wolfram.com/TwinPrimes.html

http://mathworld.wolfram.com/PrimeGaps.html

http://www.math.utah.edu/~pa/Eratosthenes.html

http://primes.utm.edu/lists/small/10000.txt

http://primes.utm.edu/bios/page.php?id=949

Soundararajan, K. *Small Gaps Between Prime Numbers: The Work of Goldston-Pintz—Yildirim.* <http://www.ams.org/bull/2007-44-01/S0273-0979-06-01142-6/S0273-0979-06-01142-6.pdf>