The Origin of the Prime Number Theorem

Dominic Klyve*

December 18, 2018

Introduction

At least since the time of the ancient Greeks and Euclid's *Elements*, mathematicians have known that there are infinitely many primes. It seems that it wasn't until the eighteenth century, however, that they turned their attention to the task of estimating the number of primes up to an arbitrary value x. The study of this count of primes, often denoted with the function $\pi(x)$, continues to be an active topic of research today.[1] Although good approximations to $\pi(x)$ have been found, mathematicians continue to find better bounds on the "error" provided by these approximations. In this project, we will look at the work of the first two mathematicians who made a careful study of values of $\pi(x)$, and compare their conjectures about how the function behaves.

Part 1: Legendre's book

We look first at the work of Adrien-Marie Legendre (1752–1833). Legendre worked in several fields of mathematics, but in this project we are most concerned with his *Essai sur la Théorie des Nombres* (Essay on Number Theory) [Legendre, 1798] – the first number theory textbook ever written. Legendre's book covered a large number of topics, only one of which was prime numbers. In the seventh section of the fourth chapter of his more than 500-page book, Legendre wrote about his discoveries and conjectures concerning the enumeration of primes[2].

On a very remarkable law observed in the enumeration of prime numbers

Although the sequence of prime numbers is extremely irregular, one can however find, with a very satisfying precision, how many of these numbers there are from 1 up to a given limit x. The formula that resolves this question is

$$y = \frac{x}{\log(x) - 1.08366}[\ldots].$$

*Department of Mathematics, Central Washington University, Ellensburg, WA 98926; dominic.klyve@cwu.edu.

[1] The Greek letter π makes a "p" sound, and stands for 'prime".

[2] All translations from Legendre's book are the work of the author of this project.

Comparing this formula with the enumeration given in the most extensive tables, such as Wega's [sic: Vega's], gives the following results:

It is impossible that a formula could represent a sequence of numbers of such a great extent, and one subject to frequent anomalies, [completely] faithfully. In order to further confirm such a remarkable law, we add that having searched, using a process we will reveal soon, how many prime numbers there from from 1 to $1,000,000$, we have found that there are 78527 of them, expect [possibly] for a small error, which could be due to the length of the calculations[3]. But in setting $x = 1,000,000$, the preceding formula gives $y = 78543$. There is therefore no doubt, not only that the general law is represented by a function $\frac{x}{A \log x + B}$, but that the coefficients A and B are indeed values very nearly $A = 1 \ldots$, $B = -1.08366$. It remains to demonstrate this law *a priori*, and that is an interesting question which we will discuss below.

(For those who don't read French, the word "Nombre" in Figure 1 means "number", "par la formule" means "according to the formula", and "par les Tables" means "according to the Tables", meaning the tables of prime numbers collected by people such as Slovene mathematician Jurij Vega (1754–1802), who Legendre mentioned in the quotation above. Note that in these counts, the number 1 is included as a prime, although it is not counted as such in most modern sources.)

Task 1 Use a calculator to check the first five values that Legendre gave "according to the formula". Are they correct?

Task 2 Why do you think Legendre may have decided to use the log function in his formula? (Hint: look at the ratio of numbers that are prime less than 10000, 20000, 40000, and 80000; try the same thing for $100,000$, $200,000$, and $400,000$.)

Task 3 Where do you think Legendre may have gotten the constant 1.08366? (It seems that he never gave the reason; brainstorm one or more possibilities here.)

Task 4 One way to determine whether a formula is reasonable to is look at the size and sign (positive or negative) of its errors as the input grows larger. Make a list of the errors in Lagrange's formula for each value that he calculated. Do you see any pattern in the errors as x grows? Do you think his formula is a good guess for the number of primes up to any value x?

Part 2: Gauss' letter

Unbeknownst to the mathematical world, one person had tried to answer the question of the number of primes up to x before Legendre. Carl Friedrich Gauss (1777–1855) had done this in the early 1790's,

[3] Legendre's count here is not terribly accurate; there are, in fact, only 78499 primes up to a million if we include 1 as prime.

though he didn't publish any of his work. We know about his work because of a letter he wrote to his colleague Johann Encke (1791–1865)[4] many years later (on Christmas Eve, 1849), in which Gauss described his calculations and his own estimate for the number of primes up to x, in addition to comparing his estimate to that of Legendre.[5].

> My distinguished friend:
>
> Your remarks concerning the frequency of primes were of interest to me in more ways than one. You have reminded me of my own endeavors in this field which began in the very distant past, in 1792 or 1793, after I had acquired the Lambert supplements to the logarithmic tables[6] Even before I had begun my more detailed investigations into higher arithmetic, one of my first projects was to turn my attention to the decreasing frequency of primes, to which end I counted the primes in several chiliads and recorded the results on the attached white pages. I soon recognized that behind all of its fluctuations, this frequency is on the average inversely proportional to the logarithm, so that the number of primes below a given bound n is approximately equal to
>
> $$\int \frac{dn}{\log n},$$
>
> where the logarithm is understood to be hyperbolic.

(Note that " the logarithm is understood to be hyperbolic" is Gauss' way of saying that the logarithm above is the natural (base-e) logarithm.)

Task 5 What's a chiliad?

Task 6 Give an estimate for how long it would take to count the number of primes in a chiliad, say starting at $700,000$. Justify your estimate.

Task 7 Use the counts "According to the Tables" to compare Gauss estimate to to Legendre's. Which do you think is a better estimate, and why?

[4]Though here we see only Encke's correspondence with Gauss on counting primes, he was primarily an astronomer. Johann Franz Encke became well known in the nineteenth century after calculating the orbit a comet far shorter than any known that the time. He went on to make an accurate calculation of the distance from the Earth to the Sun, and later made a detailed study of Saturn's rings.

[5]The text of this letter is taken from the English translation of L.J. Goldstein, in *A History of the Prime Number Theorem.* [Goldstein, 1973]

[6]John Lambert (1728–1777) published *Zusätze zu den logarithmischen und trig. Tabellen* (Additions to the logarithmic and trig. tables) in 1770. I have been unable to determine to which table of logarithms they served as additions.)

Part 3: Comparing the estimates

Gauss statement of his estimate below a given bound is a bit confusing: if n is the given bound, then n can't vary under the integral. A clearer way to write his guess (and one we often use today) is that the number of primes is approximately equal to

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

The Li function is often referred to as the "logarithmic integral." The idea behind this is a bit surprising. If we assume that each integer n has a $1/\log n$ probability of being prime, and further assume that being prime is a random event with this probability, then the total number of primes we expect up to n is the sum of all these probabilities – and the integral is simply a "smoothing out" of the discrete sums.

We would like to see how Gauss integral compares with Legendre's estimate.

Task 8 Although it's not possible to find an analytic anti-derivative of $\frac{dt}{\log t}$, it is possible to use integration by parts to expand the definite integral defining $\mathrm{Li}(x)$ into a "main term" that almost matches Legendre's estimate (there are small constants that are different), and a smaller integral. Do this, then explain how the two estimates differ.

The notation $f(x) \sim g(x)$ is used to describe that the two functions grow at roughly the same rate, and is read as "$f(x)$ is asymptotic to $g(x)$". Formally we write $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Task 9 Prove or disprove: $\mathrm{Li}(x) \sim \frac{x}{\log x}$.

Task 10 Compute the value given by Gauss' estimate for some of the values in the table. Which estimate do you think is better, and why?

Gauss continued his letter to Enke:

Thus (for many years now) the first three million [integers] have been counted and checked against the integral.

A small excerpt follows:

Below	Here are prime[7]	$\int \frac{dn}{\log n}$ + Error	Your formula[8] + Error
500000	41556	41606.4+50.4	41596.9+40.9
1000000	78501	79627.5+126.5	78672.7+171.7
1500000	114112	114263.1+151.1	114374.0+264.0
2000000	148883	149054.8+171.8	149233.0+350.0
2500000	183016	183245.0+229.0	183495.1+479.1
3000000	216745	216970.6+225.6	217308.5+563.5

I was not aware that Legendre had also worked on this subject; your letter caused me to look in his *Theorie des Nombres*, and in the second edition I found a few pages on the subject which I must have previously overlooked (or, by now, forgotten). Legendre used the formula

$$\frac{n}{\log n - A},$$

where A is a constant which he sets equal to 1.08366. After a hasty computation, I find in the above cases the deviations

$$-23,3 + 42,2 + 68,1 + 92,8 + 159,1 + 167,6$$

These differences are even smaller than those from the integral, but they seem to grow faster with n so that it is quite possible they may surpass them.

Task 11 As you did in Task 4, Gauss wanted to see if he could find a pattern in the size and sign of the errors given by Legendre's formula. Describe in your own words what Gauss concluded.

Task 12 Gauss noted that Legendre's errors grow faster than his own. Do you think this is significant? Why or why not?

To make the count and the formula agree, one would have to use, respectively, instead of A = 1.08366, the following numbers:

TABLE C
1,09040
1,07682
1,07582
1,07529
1,07179
1,07297

It appears that, with increasing n, the (average) value of A decreases; however, I dare not conjecture whether the limit as n approaches infinity is 1 or a number different from 1. I cannot say that there is any justification for expecting a very simple limiting value.

Task 13 What do you notice about the values of A that Legendre would need to make his formula precise (Table C)? What does this suggest to you about a possible limiting value of A?

Part 4: Afterword

In a sense, all of the characters in this story were correct. If we let $\pi(x)$ represent the number of primes up to x, the following marvelous theorem has been proven.

Theorem 1 (The Prime Number Theorem) *The number of primes not greater than x, $\pi(x)$ satisfies the asymptotic relationship*

$$\pi(x) \sim \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}.$$

The proof of this beautiful fact was, however, difficult to find and very deep. Indeed, it was not until 1896 (many years after the death of Gauss) that two mathematicians, Jacques Hadamard (1865–1963) and Charles Jean de la Vallée Poussin (1866–1962), independently proved the result. Even then, their proofs were both proven using advanced and subtle results in complex analysis that were not known in Gauss' day.

References

Larry J Goldstein. A history of the prime number theorem. *The American Mathematical Monthly*, 80 (6):599–615, 1973.

A Legendre. Essai sur la th\(\acute {\text {e}}\) orie des nombres, 1798.

Notes to Instructors

This Primary Source Project provides students with a short inroduction to the first writings on what would become the Prime Number Theorem. In particular, it examines the published work of Legendre and the unpublished work of Gauss counting primes and trying to find a function that estimates the count well. No proof of the Prime Number Theorem is provided (or even hinted at), but students can gain experience in comparing the magnitude and rate of growth of various error bounds. In completing the project, students also engage in mathematical activities including looking for patterns, justifying estimates and understanding asymptotic behavior.

PSP Content: Topics and Goals

In this project students have an opportunity to decide which of the conjectured functions (that of Gauss or of Legendre) seems to provide a better estimate. The project allows for independent comparison of the bounds, and provides some motivation for Gauss conjecture.

Student Prerequisites

This project is appropriate for a course in number theory. Students should be familiar with the definition of a prime number, and should have taken a second course in calculus (sufficient to have covered integration by parts).

Suggestions for PSP Implementation

- **Day 0.** Introduce project. This might be a nice chance to explore with students what they think will happen to the frequency of primes in the integers as numbers grow larger.

  Day 0 Homework: Read Introduction and Part 1; work on Tasks 1–4.

- **Day 1.** Students work together in groups to discuss their answers to the the homework question, and then to work on Part 2 and Part 3, Tasks 8 and 9.

  Day 1 Homework: Tasks 10–13.

- **Day 2.** Go over student answers as a class, and summarize the prime number theorem. Students can read Part 4 in groups, or the instructor can simply cover the material.

LaTeX code of the entire PSP is available from the author by request to facilitate preparation of reading guides or other assignments related to the project. The PSP itself can also be modified by instructors as desired to better suit their goals for the course.

Commentary on Selected Student Tasks

Task 3 asks students to hypothesize an origin of Legendre's constant of 1.08366. To the best of my knowledge, this has never been fully explained. The instructor should consider giving full credit to any reasonable guess. Moreover, I would be grateful if you'd share any particularly strong hypotheses with me!

In Task 8, students should discover that $\int \frac{dx}{\log x}$ is asymptotically $\frac{x}{\log x} + \int dx \log^2 x$. If you like, you can ask students to repeat the integration by parts process a few more times. As it turns out,

$$\mathrm{Li}(x) \sim \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2}{\log^3 x} + \cdots + \frac{(n-1)!}{\log^n x} + \cdots .$$

Approaching this, however, can be a bit thorny. The series doesn't converge, and is a good approximation only n is not too large and x is not too small. This is an interesting problem to consider in a first or second course in analytic number theory, but is probably not withing the scope of a first number theory class.

Recommendations for Further Reading

Acknowledgments

The development of this project has been partially supported by the TRansforming Instruction in Undergraduate Mathematics via Primary Historical Sources (TRIUMPHS) project with funding from the National Science Foundation's Improving Undergraduate STEM Education Program under grant numbers 1523494, 1523561, 1523747, 1523753, 1523898, 1524065, and 1524098. Any opinions, findings, and conclusions or recommendations expressed in this project are those of the author and do not necessarily represent the views of the National Science Foundation. For more information about TRIUMPHS, visit:
`http://webpages.ursinus.edu/nscoville/TRIUMPHS.html`

.

Limite x	Nombre y par la formule.	Nombre y par les Tables.	Limite x
10000	1230	1230	100000
20000	2268	2263	150000
30000	3252	3246	200000
40000	4205	4204	250000
50000	5136	5134	300000
60000	6049	6058	350000
70000	6949	6936	400000
80000	7838	7837	
90000	8717	8713	

Figure 1: Original table from Legendre's *Essay on Number Theory*