What is Riemann's Hypothesis?

Barry Mazur William Stein

June 28, 2007

Contents

1 Foreword

The Riemann Hypothesis is one of the great unsolved problems of mathematics and the reward of \$1,000,000 of *Clay Mathematics Institute* prize money awaits the person who solves it. But—with or without money—its resolution is crucial for our understanding of the nature of numbers.

There are at least four full-length books recently published, written for a general audience, that have the Riemann Hypothesis as their main topic. A reader of these books will get a fairly rich picture of the personalities engaged in the pursuit, and of related mathematical and historical issues.

This is *not* the mission of the booklet that you now hold in your hands. We aim—instead—to explain, in as direct a manner as possible and with the least mathematical background required, what this problem is all about and why it is so important. For even before anyone proves this *hypothesis* to be true (or false!), just getting familiar with it and with some of the the ideas behind it, is exciting. Moreover, this hypothesis is of crucial importance in a wide range of computational mathematics, even if it is never proved, since it gives us an excellent sense of how long certain computer programs will take to run, which bolsters our confidence to initiate a computation that might take weeks or even months to complete.

Our "booklet" comes in two pieces: the actual text which is a mere *** pages, and a CD which has a gallery of illustrations that are meant to accompany the text. There are very few mathematical equations in our text. We believe that a complete *reading* of our booklet and *viewing* of the accompanying illustrations need not take more than *** hours. But for the more adventurous readers, we have included in the CD something we call *worksheets* that explain to the reader how to experiment with the parameters for the ranges of data illustrated, so as to get a vivid sense of how the numbers 'behave.'

So, what *is* the Riemann Hypothesis? Here is a *first formulation* of this hypothesis, put in one ridiculously condensed paragraph. We will be giving other, quite miraculous, formulations of this hypothesis as well, but the task of our booklet is to expand the following paragraph into a real *explanation* and to convince you of its importance and its beauty.

What *sort* of Hypothesis is the Riemann Hypothesis?

It has long been known that that for any real number X the number of prime numbers less than X (denoted $\pi(X)$) is approximately $X/\log X$ in the sense that the *ratio* $\frac{\pi(X)}{X/\log X}$ tends to 1 as X goes to infinity. *The Riemann Hypothesis* would give us a much more accurate "count" for $\pi(X)$ in that it will offer a specific smooth function $R(X)$ (hardly any more difficult to describe than $X/\log X$) and then conjecture that $R(X)$ is an *essentially square root accurate approximation* to $\pi(X)$; i.e., for any given exponent greater than $1/2$ (you choose it: $0.501, 0.5001, 0.50001$, for example) and for large enough X where the phrase "large enough" depends on your choice of exponent the error term i.e., the difference between $R(X)$ and the number of primes less than X in absolute value is less than X raised to that exponent (e.g. $< X^{0.501}, < X^{0.5001}$, etc.)

In this paragraph we have put in boldface three key phrases, **prime numbers**, **square-root accurate**, and **error term**. A famous mathematician, Raoul Bott, once said—giving advice to some young mathematicians—that whenever one reads a mathematics book or article, or goes to a math lecture, one should aim to come home with something very specific (it can be small, but should be *specific*) that has application to a wider class of mathematical problem than was the focus of the text or lecture. If we were to suggest some possible *specific* items to come home with, after read our booklet, our three key- phrases would head the list. As for words of encouragement to think hard about the first of these, i.e., prime numbers, we can do no better than to quote a paragraph of Don Zagier's classic 12-page exposition of all this mathematics, *The First 50 Million Prime Numbers*:

"There are two facts about the distribution of prime numbers of which I hope to convince you so overwhelmingly that they will be permanently engraved in your hearts. The first is that, [they are] the most arbitrary and ornery objects studied by mathematicians: they grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision."[1]

2 Thoughts about numbers: ancient, medieval, and modern

If we are to believe Aristotle, the early Pythagoreans thought that the principles governing Number are "the principles of all things," the elements of number being more basic than the Empedoclean physical elements *earth, air, fire, water.* To think about number is to get close to the architecture of " what is."

So, how far along are we in our thoughts about numbers?

René Descartes, almost four centuries ago, expressed the hope that there soon would be "almost nothing more to discover in geometry." Contemporary physicists dream of a "final theory." But despite its venerability and its great power and beauty, the pure mathematics of numbers may still be in the infancy of its development, with depths to be explored as endless as the human soul, and never a final theory.

Figure 2.1: Don Quixote and "his" Dulcinea del Toboso

Numbers are obstreperous things. Don Quixote encountered this when he requested that the “bachelor” compose a poem to his lady Dulcinea del Toboso, the first letters of each line spelling out her name. The “bachelor” found

> “a great difficulty in their composition because the number of letters in her name was 17, and if he made four Castilian stanzas of four octosyllabic lines each, there would be one letter too many, and if he made the stanzas of five octosyllabic lines each, the ones called *décimas* or *redondillas,* there would be three letters too few...”

”It must fit in, however, you do it,” pleaded Quixote, not willing to grant the imperviousness of the number 17 to division.

Seventeen is indeed a prime number: there is no way of factoring it as the product of smaller numbers, and this accounts—people tell us—for its occurrence in some phenomena of nature, as when last year the 17-year cicadas all emerged to celebrate a “reunion” of some sort in our fields and valleys.

Prime numbers, despite their *primary* position in our modern understanding of number, were not specifically doted over in the ancient literature before Euclid, at least not in the literature that has been preserved. Primes are mentioned as a class of numbers in the writings of Philolaus (a predecessor of Plato); they are not mentioned specifically in the Platonic dialogues, which is surprising to me given the intense interest Plato had in mathematical developments; and they make an occasional appearance in the writings of Aristotle, which is not surprising, given Aristotle’s emphasis on the distinction between the *composite* and the *incomposite*. “The incomposite is prior to the composite,” writes Aristotle in Book 13 of the Metaphysics.

But, until Euclid, prime numbers seem not to have been singled out as *the* extraordinary mathematical concept, central to any deep understanding of numerical phenomena, that they are now understood to be.

There is an extraordinary wealth of established truths about numbers; these truths provoke sheer awe for the beautiful complexity of prime numbers. But each of the important new discoveries we make give rise to a further richness of questions, educated guesses, heuristics, expectations, and unsolved problems.

3 What are prime numbers?

Primes as atoms. To begin from the beginning, think of the operation of multiplication as a bond that ties numbers together: the equation $2 \times 3 = 6$ invites us to imagine the number 6 as (a molecule, if you wish) built out of its smaller constituents 2 and 3. Reversing the procedure, if we start with a whole number, say 6 again, we may try to factor it (that is, express it as a product of smaller whole numbers) and, of course, we would eventually, if not immediately, come up with $6 = 2 \times 3$ and discover that 2 and 3 factor no further; the numbers 2 and 3, then, are the indecomposable entities (atoms, if you wish) that comprise our number. By definition, a **prime number** (colloquially, *a prime*) is a whole number, bigger than 1, that cannot be factored into a product of two smaller whole numbers. So, 2 and 3 are the first two prime numbers. The next number along the line, 4, is not prime, for $4 = 2 \times 2$; the number after that, 5, is. Primes are, multiplicatively speaking, the building blocks from which all numbers can be made. A fundamental theorem of arithmetic tells us that any number (bigger than 1) can be factored as a product of primes, and the factorization is *unique* except for rearranging the order of the primes. For example, if you try to factor the number 300, there are many ways to begin:

$$300 = 30 \times 10 \qquad \text{or} \qquad 300 = 6 \times 50$$

and there are various other starting possibilities. But if you continue the factorization ("climbing down" any one of the possible "factoring trees") to the bottom, where every factor is a prime number, e.g., in Figure 3.1, you always end up with the same collection of prime numbers:

$$300 = 2^2 \times 3 \times 5^2.$$

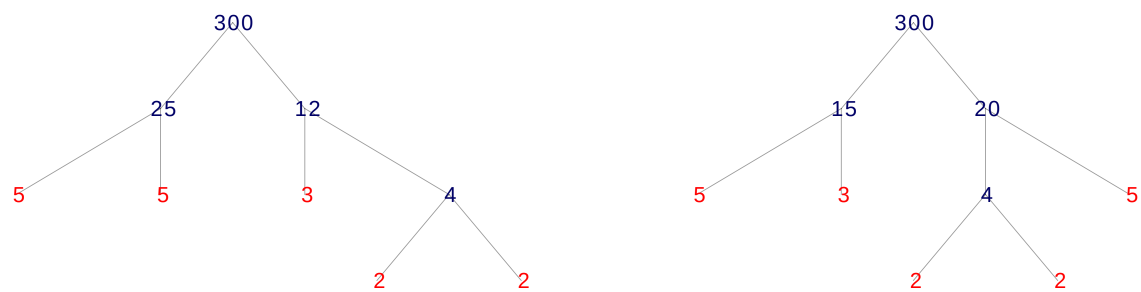

Figure 3.1: A factor tree that illustrates the factorization of 300 as a product of primes.

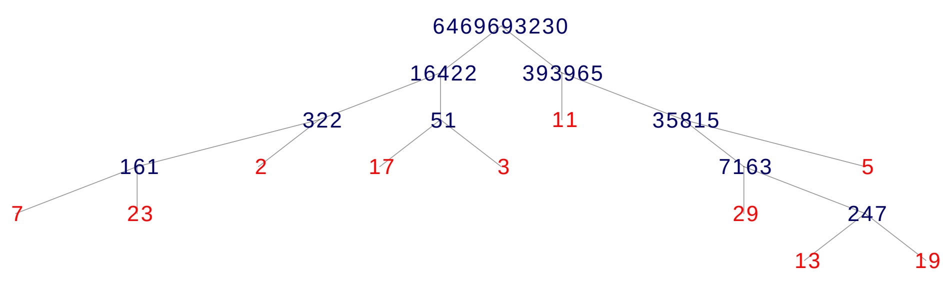

Figure 3.2: Factorization tree for 6469693230

The Riemann Hypothesis probes the question: how intimately can we know prime numbers, those *atoms* of multiplication? Anytime we visit a website and purchase something online, prime numbers having over 150 decimal digits are used to keep our bank transactions private. This ubiquitous use to which these giant primes are put depends upon a very simple principle: it is much easier to multiply numbers together than to factor them. If you had to factor, say, the number 143 you might scratch your head for a few minutes before discovering that 143 is 11×13. But if you had to multiply 11 by 13 you would do it straightaway. Offer two primes, say, P and Q each with more than 100 digits, to your computing machine and ask it to multiply them together: you will get their product $N = P \times Q$ with its 200 or so digits in nanoseconds. But present that number N to any current desktop computer, and ask it to factor $N-$ and the computer will fail to do the task. The safety of much encryption depends upon this guaranteed[1] failure!

If we were latter-day number-phenomenologists we might revel in the discovery and proof that

$$p = 2^{32582657} - 1$$

is a prime number, this number having 9,808,358 digits! There are, in fact, infinitely many prime numbers, as was proved in Euclid's *Elements* as follows. Suppose there are only finitely many primes $p_1, \ldots, p_n$. Let $n = p_1 p_2 \cdots p_n + 1$. Then n is divisible by some prime, but no p_i divides n, which is contrary to our assumption that $p_1, \ldots, p_n$ is the complete list of primes.

The number p displayed above is the largest prime we know, where by "know" we mean that we know it so explicitly that we can *compute* things about it. For example, the last two digits of p are 71. Of course p is not the largest prime number since there are infinitely many primes, e.g., the next prime q after p is a prime. But there is no obvious way to compute anything about q. For example, what is the last digit of q in its decimal expansion?

Eratosthenes, the mathematician from Cyrene (and later, librarian at Alexandria) explained how to *sift* the prime numbers from the series of all numbers: in the sequence of numbers,

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29,

for example, start by circling the 2 and crossing out all the other multiples of 2. Next, go back to the beginning of our sequence of numbers and circle the first number that is neither circled nor crossed out (that would be, of course, the 3); then cross out all the other multiples of 3. This gives the pattern: go back again to the beginning of our sequence of numbers and circle the first number that is neither circled nor crossed out; then cross out all of its other multiples. Repeat this pattern until all the numbers in our sequence are either circled, or crossed out, the circled ones being the primes.

In Figure **??** we use the primes 2, 3, 5, and 7 to sieve out the primes up to 100.

Especially if you have had little experience with math, may I suggest that you actually follow Eratosthenes' lead, and perform the repeated circling and crossing-out to watch the primes emerge, intriguingly staggered through our sequence of numbers,

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

[1] Nobody has ever *proved* that there is no fast way to factor integers.

4 Questions about primes that any person might ask

We become quickly stymied when we ask quite elementary questions about the spacing of the infinite series of prime numbers.

For example, *are there infinitely many pairs of primes whose difference is* 2*?* The sequence on the page seems to be rich in such pairs

$$5-3=2, \quad 7-5=2, \quad 13-11=2, \quad 19-17=2,$$

and we know loads more such pairs (for example, there are 1,177,209,242,304 such pairs less than 1,000,000,000,000,000) but the answer to our question, *are there infinitely many?*, is not known. *Are there infinitely many pairs of primes whose difference is* 4*?* Answer: equally unknown. *Is every even number greater than* 2 *a sum of two primes?* Answer: unknown. *Are there infinitely many primes which are* 1 *more than a perfect square?* Answer: unknown.

Is there some neat formula giving the next prime? More specifically, *If I give you a number* N*, say* $N =$ *one million, and ask you for the first number after* N *that is prime, is there a method that answers that question without, in some form or other, running through each of the successive numbers after* N *rejecting the nonprimes until the first prime is encountered?* Answer: unknown.

Here is a curious question that you can easily begin to check out for small numbers. We know, of course, that the *even* numbers and the *odd* numbers are nicely and simply distributed: after every odd number comes an even number, after every even, an odd, there is an equal number of odd number as even numbers less than any given odd number; and there may be nothing else of interest to say about the matter. Things change considerably, though, if we focus our concentration on *multiplicatively even* numbers and *multiplicatively odd* numbers.

A **multiplicatively even** number is one that can be expressed as a product of *an even number of* primes; and a **multiplicatively odd** number is one that can be expressed as a product of *an odd number of* primes. So, any prime is multiplicatively odd, the number $4 = 2 \cdot 2$ is multiplicatively even, and so is $6 = 2 \cdot 3$, $9 = 3 \cdot 3$, and $10 = 2 \cdot 5$; but $12 = 2 \cdot 2 \cdot 4$ is multiplicatively odd. Here is some data:

Table 4.1: Multiplicatively even and odd numbers $\leq N$

N	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Multiplicatively even	1	1	2	2	3	3	3	4	5	5	5	5	6	7	8
Multiplicatively odd	1	2	2	3	3	4	5	5	5	6	7	8	8	8	8

Now looking at this data, a natural, and simple, question to ask about the concept of multiplicative *oddness* and *evenness* is:

Is there some $N \geq 2$ for which there are more multiplicatively even numbers less than or equal to N than multiplicatively odd ones?

A *negative* response to this question implies the Riemann Hypothesis. In contrast to the list of previous questions, the answer to this one is known: alas, there is such an N. In 1960, Lehman showed that for $N = 906,400,000$ there are 708 more multiplicatively even numbers up to N than multiplicatively odd numbers (Tanaka found in 1980 that the smallest N with this property is $N = 906,150,257$). For more details, see P. Borwein, *Sign changes in sums of the Liouville Function.*

WILLIAM: Do you think it might be worth putting in a graph for this data? I would have said that this is our "second" formulation of RH except for the curous fact that if one can prove the above, that would imply both RH *and* the fact that the zeroes are all simple. For a reference here, there is a nice short paper of Norbert Wiener *Notes on Polya's and Turan's hypothesis concerning Liouville's factor* on this (page 765 of volume II of Wiener's Collected Works; also: G. Polya *Vershiedene Bemerkungen zur Zahlentheorie* Jahresbericht de Deutschen Mathematiker-Vereinigung, **28** (1919) 31-40

These are questions that have been asked about primes (and we could give bushels more), questions expressible in simple vocabulary, that we can't answer today. We have been studying numbers for over two millenia and yet we are indeed in the infancy of our understanding.

5 How many primes are there?

Slow as we are to understand primes, at the very least we can try to count them. You can see that there are 10 primes less than 30, so you might encapsulate this by saying that the chances that a number less than 30 is prime is 1 in 3. This frequency does not persist, though; here is some more data: There are 168 primes less than a thousand (so we might say that among the numbers less than a thousand the chances that one of them is prime is roughly 1 in 6).

There are 78,498 primes less than a million (so we might say that the chances that a random choice among the first million numbers is prime have dropped to roughly 1 in 13).

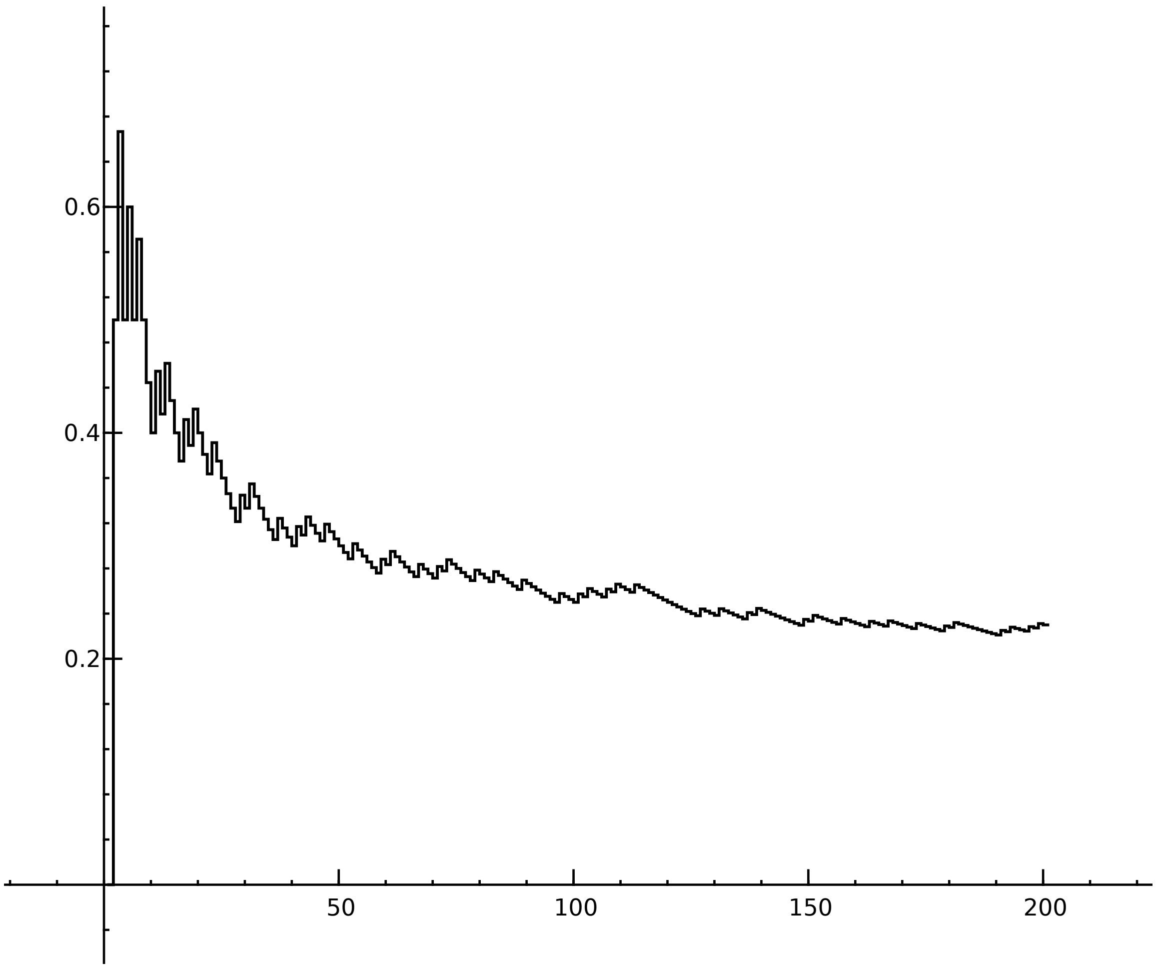

Figure 5.1: Proportion of primes up to $N = 200$

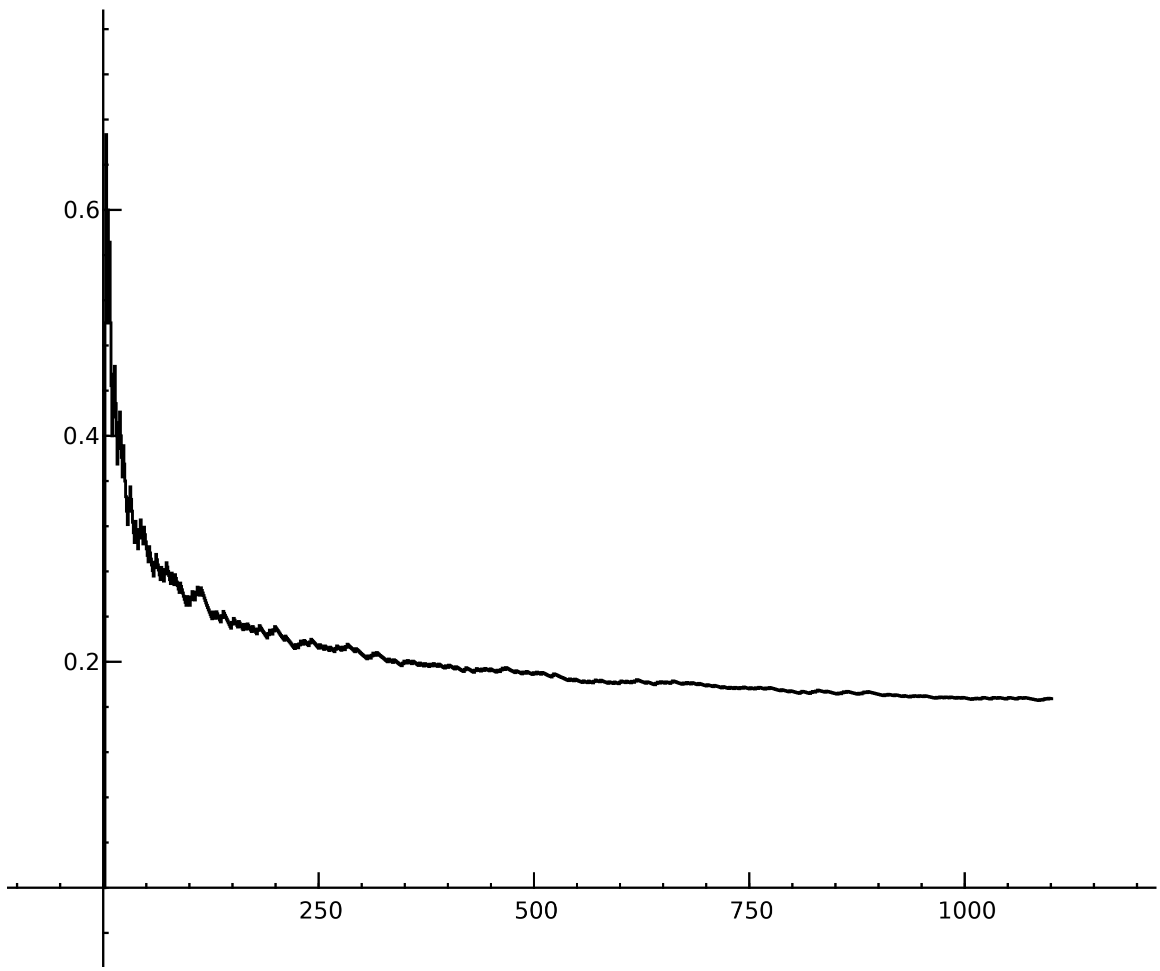

Figure 5.2: Proportion of primes up to $N = 1100$

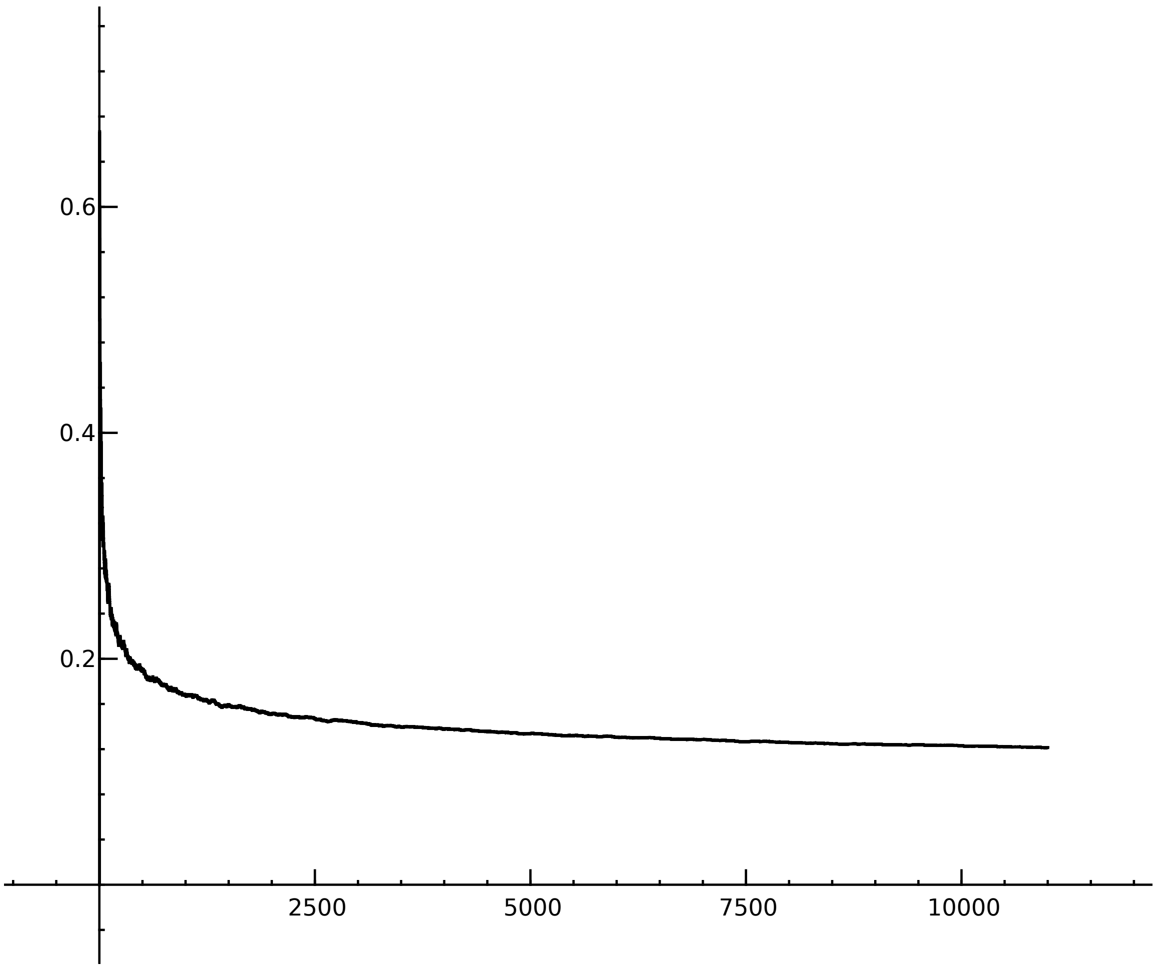

Figure 5.3: Proportion of primes up to $N = 11000$

There are $455,052,512$ primes less than ten billion; i.e., 10,000,000,000 (so we might say that the chances are down to roughly 1 in 22).

Primes, then, seem to be thinning out. Return to the sifting process we did earlier, and take a look at a few graphs, to get a sense of why that might be so. There are a hundred numbers less than or equal to 100, a thousand numbers less than or equal to 1000, etc.: the shaded bar graph that looks like a regular staircase, each step the same length as each riser, climbing up at, so to speak, a 45 degree angle, counts all numbers up to and including N.

Following Eratosthenes, we have sifted those numbers, to pan for primes. Our first move was to throw out roughly half the numbers (the even ones!) after the number 2. The cross-hatched bar graph in this figure which is, with one hiccup, a regular staircase climbing at a smaller angle, each step twice the lengths of each riser, illustrates the numbers that are left after one pass through Eratosthenes' sieve, which includes, of course, all the primes. So, the chances that a number bigger than 2 is prime is *at most* 1 in 2. Our second move was to throw out a good bunch of numbers bigger than 3. So, the chances that a number bigger than 3 is prime is going to be even less. And so it goes: with each Eratosthenian move in our sieving process we are winnowing the field more extensively, reducing the chances that the later numbers are prime.

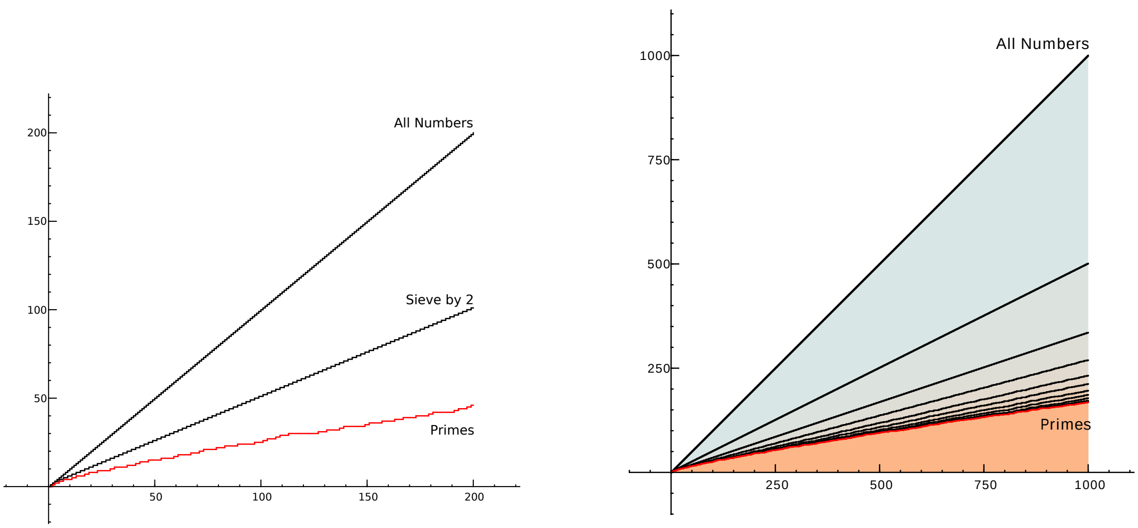

Figure 5.4: Sieving by removing multiples of 2 up to 100 and Sieving for primes up to 1000

The red curve in these figures actually counts the primes: it is the beguilingly irregular *staircase of primes.* Its height above any number N on the horizontal line records the number of primes less than or equal to N, the accumulation of primes up to N. Refer to this number as $P(N)$. So $P(2) = 1$, $P(3) = 2$, $P(30) = 10$; of course, if you believed some of the data above you could plot a few more values of $P(N)$, like $P(\text{ten billion}) = 455,052,512$.

Let us accompany Eratosthenes for a few further steps in his sieving process. Here is a graph of all whole numbers after we have removed the even numbers greater than 2, and the multiples of 3 greater than 3 itself:

From this graph you can see that if you go "out a way" the likelihood that a number is a prime is less that 1 in 3. Here is what Eratosthenes sieve looks like after sifting $2, 3, 5$, and 7,

This data may begin to suggest to you that as you go further and further out on the number line the percentage of prime numbers among all whole numbers tends towards 0% (it does).

To get a sense of how the primes accumulate, we will take a look at the staircase of primes for $N = 25$, and $N = 100$.

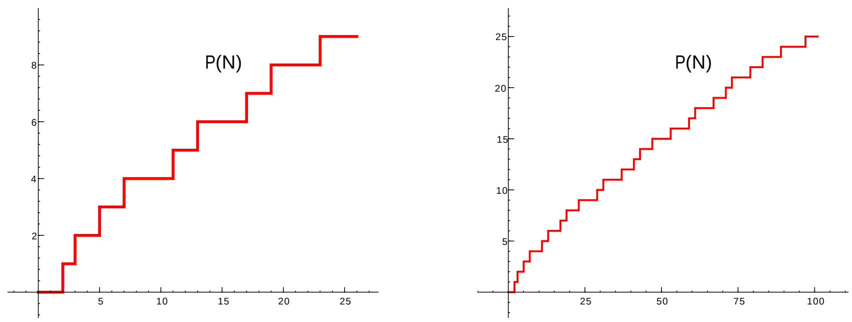

Figure 5.5:

6 Prime numbers viewed from a distance

The striking thing about these figures is that as the numbers get large enough, the jagged accumulation of primes, those quintessentially discrete entities, becomes smoother and smoother, to the eye. How strange and wonderful to watch, as our viewpoint zooms out to larger ranges of numbers, the accumulation of primes taking on such a smooth, elegant, shape.

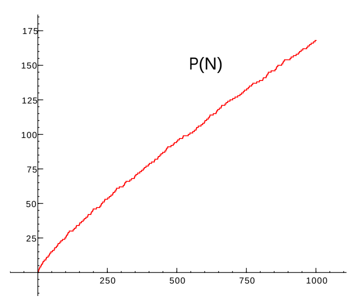

Figure 6.1:

But don't be fooled by the seemingly smooth shape of the curve in the last figure above: it is just as faithful a reproduction of the staircase of primes as the typographer's art can render, for there are about $8,000$ tiny steps and risers in this curve, all hidden by the thickness of the print of the drawn curve in the figure. It is already something of a miracle that we can approximately describe

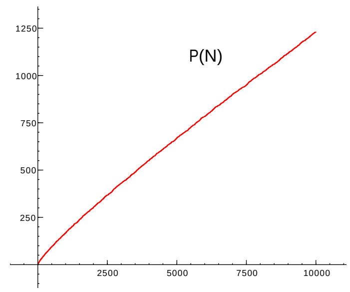

Figure 6.2:

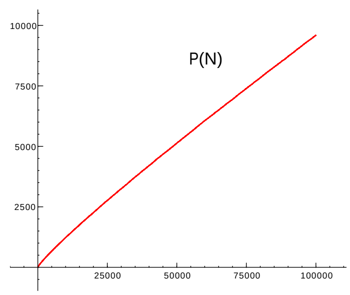

Figure 6.3:

the build-up of primes, somehow, using a *smooth curve*. But *what* smooth curve?

That last question is *not* rhetorical. If I draw a curve with chalk on the blackboard, this can signify a myriad of smooth (mathematical) curves all encompassed within the thickness of the chalk-line, all–if you wish–reasonable approximations of one another. So, there are many smooth curves that fit the chalk-curve. With this warning, but very much fortified by the data of figure 2b, let us ask: *what is a smooth curve that is a reasonable approximation to the staircase of primes?*

7 "Pure" and "Applied" Mathematics

At this point, we hope that we don't wear down your patience, for we would like to indulge in a digression about the role of question- asking (as above) in mathematics. Loosely speaking, there are two types of mathematics: *pure* and *applied*. Now, we probably put more theoretical math in the *applied* category than most of our colleagues would. This is because we want a given piece of mathematics to be categorized as "pure" or "applied" turning on whether the intention behind it, is that it eventually be used elsewhere (either within or outside the discipline of Mathematics)—in which case we'll call it *applied*—or it is to explain and get at the root of some concept—in which case we'll call it *pure*. Of course, there is a great overlap. Moreover, many questions in mathematics are "hustlers" in the sense that, at first view, what is being requested is that some simple task be done (e.g., *to find a smooth curve that is a reasonable approximation to the staircase of primes*)—once successfully done, it will, presumably, have many useful by-products—that is, it will be a piece of mathematics that we might classify as *applied*. But, the fuller motivation behind the question is *pure*: to strike behind the mask of the phenomenology of the mathematical situation, and get at the hidden fundamentals that actually govern the phenomena.

The particular issue before us is, in our opinion, twofold, both applied, and pure: can we curve-fit the "staircase of primes" by a well approximating smooth curve? The story behind this alone is marvelous, has a cornucopia of applications, and we will be telling it below. But our curiosity here is driven by a question that is pure, and less amenable to precise formulation: are there mathematical concepts at the root of, and more basic than (and "prior to," to borrow Aristotle's use of the phrase,) *prime numbers*–concepts that account for the apparent complexity of the nature of primes?

[[TODO (william): Insert a discussion here about how believing RH is crucial to doing many algebraic number theory calculations. E.g., complexity analyses for algorithms like factoring depend on RH. Also, standard class group algorithms are vastly faster if we assume RH (or even more!), so number theorists do so all the time when doing numerical experiments (though of course they try to remove the hypothesis...). See `http://pari.math.u-bordeaux.fr/dochtml/html.stable/`, e.g., for some sense of what this is like in the trenches... It says "Warning. Make sure you understand the above!"]]

8 A probabilistic "first" guess

The search for such approximating curves began, in fact, two centuries ago when Carl Friedrich Gauss defined a certain beautiful curve that, experimentally, seemed to be an exceptionally good fit for the staircase of primes. Let us denote Gauss's curve $G(X)$; it has an elegant simple formula comprehensible to anyone who has had a tiny bit of calculus [If you make believe that the chances that a number N is a prime is inversely proportional to the number of digits of N you might well hit upon Gauss's curve]. In a letter written in 1849 Gauss claimed that as early as 1792 or 1793 he had already observed that the density of prime numbers over intervals of numbers of a given rough magnitude X seemed to average $1/\log X$.

Gauss was an inveterate computer: he wrote in his 1849 letter that there are $216,745$ prime numbers less than three million (This is wrong: the actual number of these primes is 216816). Gauss's curve predicted that there would be $216,971$ primes– a miss, Gauss thought, by 226 (but actually he was closer than he thought: the correct *miss* is a mere 161; not as close as the recent US elections, but pretty close nevertheless). So you shouldn't be too surprised if Figures 4 and 2b [[fix references]] look the same! Gauss's computation brings up two queries: will this spectacular "good fit" continue for arbitrarily large numbers? and, the (evidently prior) question: what counts as a good fit?

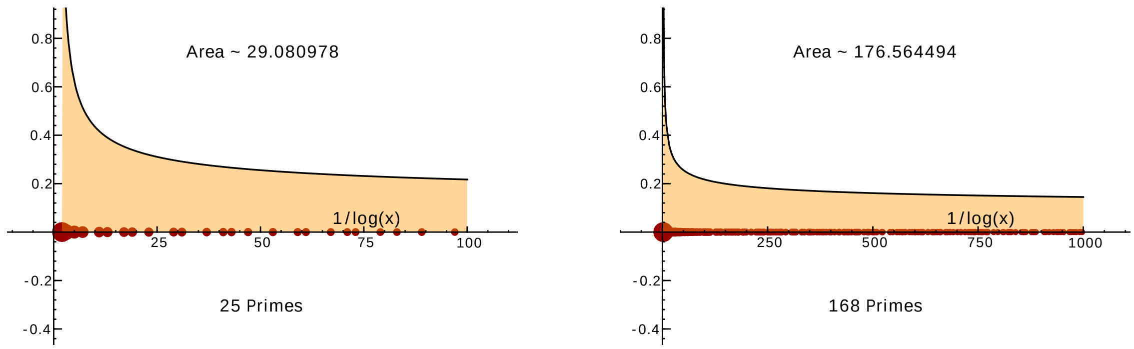

Figure 8.1: The expected tally of the number of primes $< X$ is equal to the area underneath the graph from 1 to X.

9 What is a "good approximation"?

If you are trying to estimate a number, say, around ten thousand, and you get it right to within a hundred, let us celebrate this kind of accuracy by saying that you have made an approximation with *square-root error* ($\sqrt{10{,}000} = 100$). Of course, we should really use the more clumsy phrase "an approximation with at worst *square-root error*." Sometimes we'll simply refer to such approximations as *good approximations.* If you are trying to estimate a number in the millions, and you get it right to within a thousand, let's agree that—again—you have made an approximation with *square-root error* ($\sqrt{1{,}000{,}000} = 1{,}000$). Again, for short, call this a *good approximation.* So, when Gauss thought his curve missed by 226 in estimating the number of primes less than three million, it was well within the margin we have given for a "good approximation."

More generally, if you are trying to estimate a number that has D digits and you get it almost right, but with an error that has no more than, roughly, half that many digits, let us say, again, that you have made an approximation with *square-root error* or synomymously, a *good approximation.*

This rough account almost suffices for what we will be discussing below, but to be more precise, the specific *gauge of accuracy* that will be important to us is not for a mere *single* estimate of a *single* error term,

$$\text{Error term} \;=\; \text{Exact Value} \;\text{-}\; \text{Our "good approximation"}$$

but rather for *infinite sequences* of estimates of error terms. Generally, if you are interested a numerical quantity $q(X)$ that depends on the real number parameter X (e.g., $q(X)$ could be $\pi(X)$, "the number of primes $< X$") and if you have an explicit candidate "approximation," $q_{\text{approx}}(X)$, to this quantity let us say that $q_{\text{approx}}(X)$ is **essentially a square-root accurate approximation to** $q(X)$ if for *any* given exponent greater than 1/2 (you choose it: 0.501, 0.5001, 0.50001, ... for example) and for large enough X— where the phrase "large enough" depends on your choice of exponent—the **error term**—i.e., the difference between $q_{\text{approx}}(X)$ and the *true* quantity, $q(X)$, is, in absolute value, less than $q_{\text{approx}}(X)$ raised to that exponent (e.g. $< X^{0.501}$, $< X^{0.5001}$, etc.).

Readers who know calculus and wish to have a technical formulation of this definition of *good approximation* might turn to the endnote [2] for a precise statement.

Remark 9.1. To get a feel for how basic the notion of *approximation to data being square root close to the true values of the data* is—and how it represents the "gold standard" of accuracy for approximations, consider this fable.

Imagine that the devil had the idea of saddling a large committee of people with the task of finding values of $P(N)$ for various large numbers N. This he did in the following somewhat ridiculous manner, having already worked out which numbers are prime and which composite, himself. Since the devil is, as everyone knows, *into the details,* he has made no mistakes: his work is entirely correct. He gives each committee member a copy of the list of all whole numbers between 1 and one of the large numbers N in which he was interested, the devil having put check-marks by the numbers on that list that are prime numbers, and no mark by the composite numbers. Now each committee member would count the number of primes by considering each number, in turn, on their list to figure out whether or not there is a check beside it, and tally up the ones that have checks. But since they are human, they will indeed be making mistakes, say 0.001% of the time. Assume further that it is just as likely for them to make the mistake of thinking that a number *has* a check-mark that does not have one, as the contrary: that a number hasn't got a mark, when it does. If many people are engaged in such pursuit, some of them might over-count $P(N)$; some of them might under-count it. The average error (over-counted or undercounted) would be proportional to $\sqrt{N}$.

10 What is Riemann's Hypothesis? (*second formulation*)

Recall that in our first formulation, the Riemann Hypothesis approximated $\pi(X)$, the number of primes $< X$, by the function $X/\log X$, while Gauss's guess was in terms of the area in the region from 2 to X under the graph $1/\log X$, a quantity sometimes referred to as $\mathrm{Li}(X)$ (pronounced *Lī*-of-X)[2]. The two functions $X/\log X$ and $\mathrm{Li}(X)$ are relatively close: they are *essentially square-root accurate approximations* one of the other. Therefore it is easy to see that $X/\log X$ is an essentially square-root accurate approximation to $\pi(X)$ if and only if $\mathrm{Li}(X)$ is such. Here is a graph of the three functions $X/\log X$, $\mathrm{Li}(X)$, and $\pi(X)$ in the range ***.

William: CAN WE PUT A GRAPH OF THIS HERE?

Note that $\mathrm{Li}(X)$ is a "closer fit" to $\pi(X)$ (at least in this range) so it is natural to give this second formulation of Riemann's Hypothesis—perfectly equivalent to the first formulation that we made in section 1:

[2] "Li" is short for *logarithmic integral*, $\int_2^X dx/\log x$.

The Riemann Hypothesis (second formulation)

For any real number X the number of prime numbers less than X is approximately $\mathrm{Li}(X)$ and that this approximation is essentially square root accurate.

Even though Riemann's Hypothesis has not yet been proven (or disproven) we know that both functions $X/logX$ (as in our "first formulation") and $\mathrm{Li}(X)$ (as in this "second formulation") are respectably close approximations to $\pi(X)$. We know, for example, that for any number greater than 1 that you give (we will illustrate this by imagining that you gave the number 1.000000001) if you take X sufficiently large (and how large "sufficiently large" is depends on the number that you gave) $\pi(X)$ is no greater than 1.000000001 times $\mathrm{Li}(X)$, and $\mathrm{Li}(X)$ is no greater than 1.000000001 times $\pi(X)$. The same is true if $\mathrm{Li}(X)$ is replaced by $X/\log X$. These facts are known as *The Prime Number Theorem*, a very hard-won piece of mathematics! (See the Appendix for a discussion of this theorem.)

The elusive Riemann Hypothesis, however, is much deeper than the Prime Number Theorem, and takes its origin from some awe-inspiring, difficult to interpret, lines in Bernhard Riemann's magnificent 8-page paper, "On the number of primes less than a given magnitude," published in 1859.

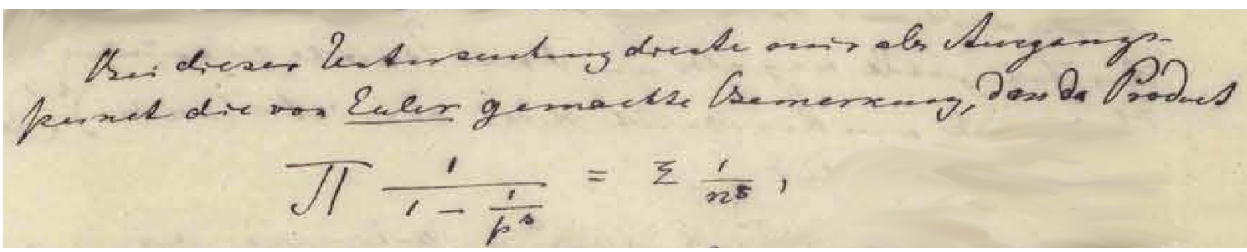

Figure 10.1: From Riemann's Manuscript

PUT FIGURES K,L HERE;– these are some of the jpg files of Riemann's manuscript that Jim Carleson sent to me.

Riemann's hypothesis, as it is currently interpreted, turns up as relevant, as a key, again and again in different parts of the subject: if you accept it as *hypothesis* you have an immensely powerful tool at your disposal: a mathematical magnifying glass that sharpens our focus on number theory. But it also has a wonderful protean quality– there are many ways of formulating it, any of these formulations being provably equivalent to any of the others.

This Riemann Hypothesis remains unproved to this day, and therefore is "only a hypothesis," as Osiander said of Copernicus's theory, but one for which we have overwhelming theoretical and numerical evidence in its support. It is the kind of conjecture that Frans Oort might label a *suffusing conjecture* in that it has unusually broad implications: many many results are now known to follow, if the conjecture, familiarly known as RH, is true. A proof of RH would, therefore, fall into the *applied* category, given our discussion above. But however you classify RH, it is a central concern in mathematics to find its proof (or, a counter-example!).

11 The *information* contained in the staircase of primes

We have borrowed the phrase "staircase of primes" from the popular book *The Music of Primes* by Marcus du Sautoi, for we feel that it captures the sense that there is a deeply hidden architecture to the graphs that compile the number of primes (up to N) and also because—in a bit—we will be tinkering with this carpentry. Before we do so, though, let us review what this staircase looks like, for different ranges.

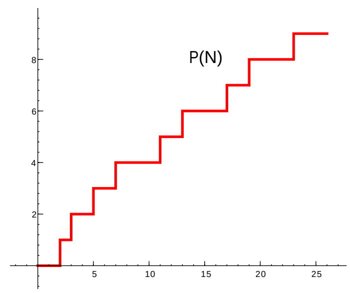

Figure 11.1:

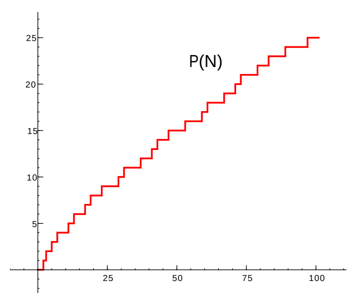

Figure 11.2:

The mystery of this staircase is that the *information* contained within it is—in effect—the full story of where the primes are placed. This story seems to elude any simple description. Can we "tinker with" this staircase without destroying this valuable information?

12 Tinkering with the carpentry of the staircase of primes.

For starters, notice that all the *risers* of this staircase have unit length. That is, they contain no numerical information except for their placement on the x-axis. So, we could distort our staircase by changing (in any way we please) the height of each riser; and as long as we haven't brought new risers into—or old risers out of—-existence, and have not modified their position over the x-axis, we have retained all the information of our original staircase.

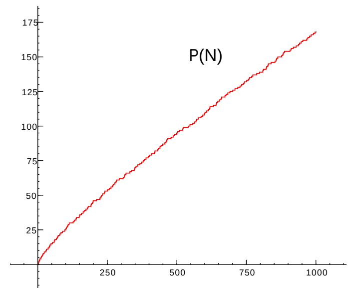

Figure 11.3:

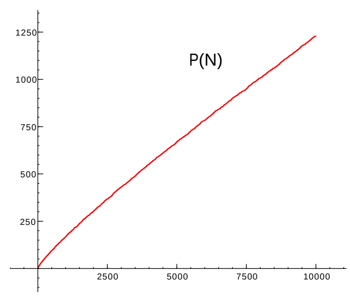

Figure 11.4:

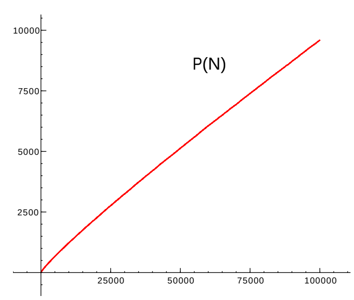

Figure 11.5:

A more drastic-sounding thing we could do is to judiciously add new steps to our staircase. At present, we have a step at each prime number p, and no step anywhere else. Suppose we built a staircase with a new step not only at $x = p$ for p each prime number but $x = 1$ and $x = p^n$ where p^n runs through all powers of prime numbers as well. Such a staircase would have, indeed, many more steps than our original staircase had, but, nevertheless, would retain much of the quality of the old staircase: namely it contains within it the full story of the placement of primes *and their powers.*

A final thing we can do is to perform a distortion of the x-axis (elongating or shortening it, as we wish) in any specific way, as long as we can perform the inverse process, and "undistort" it if we wish. Clearly such an operation may have mangled, but hasn't destroyed information irretrievably.

We shall perform all three of these kinds of operations eventually, and will see some great surprises as a result. But for now, we will perform distortions only of the first two types. We are about to build a new staircase that retains the precious information we need, but is constructed according to the following architectural plan.

- We first build a staircase that has a new step precisely at $x = 1$, and $x = p^n$ for every *prime power* p^n with $n \geq 1$. That is there will be a new step at $x = 1, 2, 3, 4, 5, 8, 9, 11, \ldots$
- Our staircase starts on the ground at $x = 0$ and height of the riser of the step at $x = 1$ will be $\log(2\pi)$. The length of the riser of the step at $x = p^n$ the prime power p^n will not be 1 (as was the length of all risers in the old staircase of primes) but rather: the step at $x = p^n$ will have the height of its riser equal to $\log p$. So for the first few steps listed in the previous item, the risers will be of length $N = \log(2\pi), \log 2, \log 3, \log 2, \log 5, \log 2, \log 3, \log 11, \ldots$ These vertical dimensions might lead to a steeper ascent but no great loss of *information.*

  Although we are not quite done with our architectural work, here is what our staircase looks like, so far:

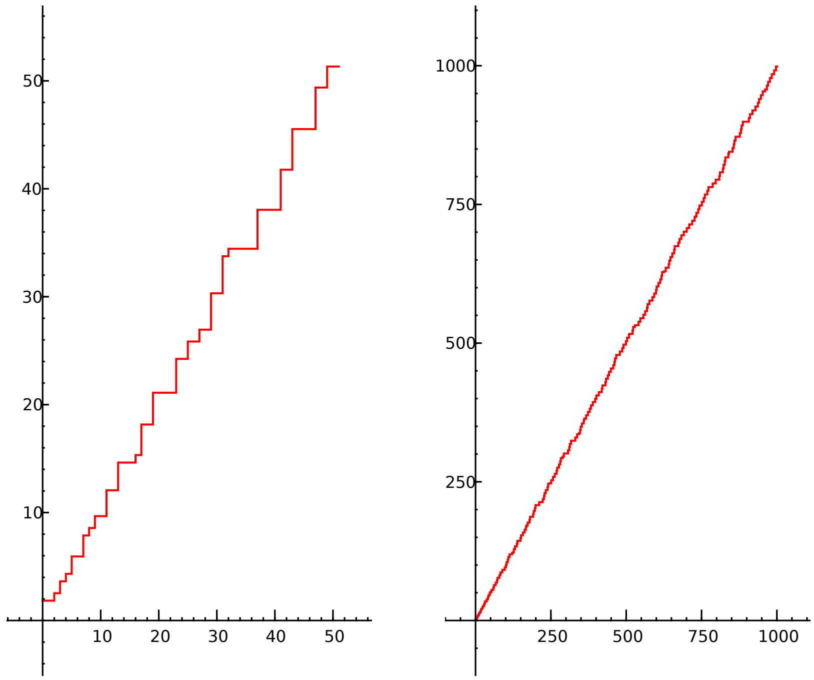

Figure 12.1: Illustration of the newly constructed staircase that counts weighted prime powers.

Notice that this new staircase looks, from afar, as if it were nicely approximated by the 45 degree straight line, i.e., by the simple function x. In fact, we have—by this new architecture—yet a third *equivalent* way of formulating Riemann's hypothesis:

> **The Riemann Hypothesis (third formulation)**
>
> This new staircase is essentially square root close to the 45 degree straight line, i.e., to the simple function x.

13 To our readers of Part I.

We hope that you now understand our motivation for distorting our staircase, as we have just done: the statement of the Riemann Hypothesis—admittedly as elusive as before—has, at least, been expressed elegantly and simply, given our new staircase that approximates (conjecturally with *essential square root accuracy*) a 45 degree straight line.

We have offered, so far, three equivalent formulations of the Riemann Hypothesis, all having to do with the manner in which the prime numbers are situated among all whole numbers. In doing this, we hope that we have convinced you that—in the words of Don Zagier—primes seem to obey no other law than that of chance and yet exhibit stunning regularity. This is the end of part I of our booklet, and is largely the end of our main mission, to explain—in elementary terms—*what is Riemann's Hypothesis?*

For readers who have studied Differential Calculus and who are happy with complex numbers, we shall go further and show that the pursuit of Riemann's hypothesis may provide a key to deep structure of primes, and to the nature of the laws that primes do obey.

Part II

14 How Calculus manages to find the slopes of graphs that have no slopes

Differential Calculus, initially the creation of Newton and/or Leibniz, acquaints us with *slopes* of graphs,

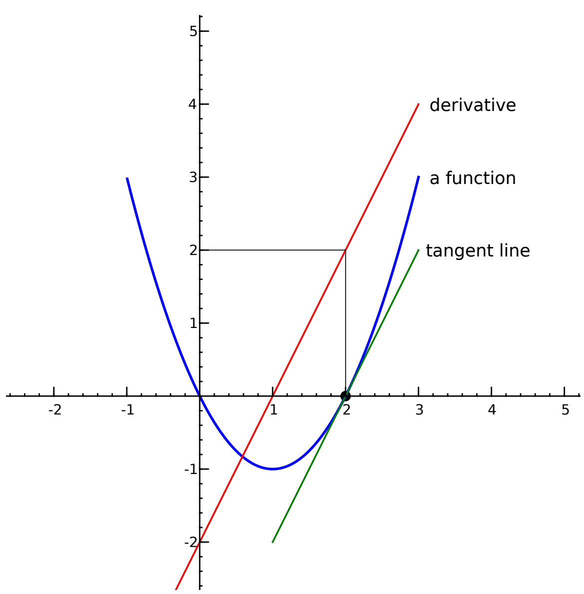

Figure 14.1: A picture of a graph of a function (blue), a slope at a point (green), and the graph of the derivative (red).

Calculus explains to us how to calculate those slopes, and finally, shows us the power that we then have to answer problems we could not answer if we couldn't compute those slopes.

Usually, in elementary Calculus classes we are called upon to compute slopes only of smooth graphs—i.e., graphs that actually *have* slopes at each of their points, such as in the illustration just above. What could Calculus possibly do if confronted with a graph that has *jumps?* For example:

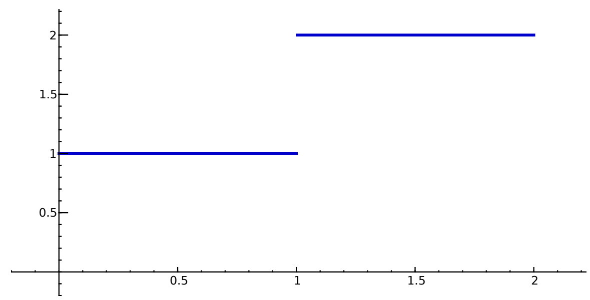

Figure 14.2: A picture of a graph that is 1 up to a point and then 2 after that point.

The most comfortable way to deal with the graph of such a function is to just approximate it by a nice smooth function

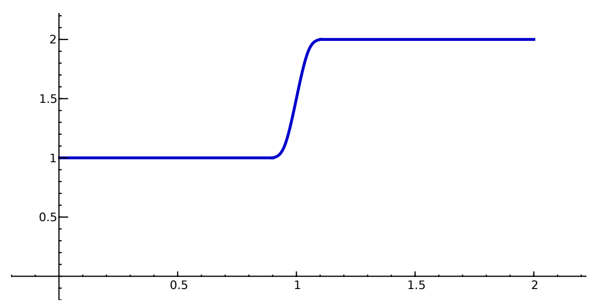

Figure 14.3: A picture of a smooth graph approximating the graph that is 1 up to some point x and then 1 after that point, the smooth graph being flat mostly.

and then take the *derivative* of that smooth function. Of course, this is just an approximation, so we might try to make a better approximation, which we do in each successive graph in the the figure.

Notice, what is happening: as the approximation gets better and better, the derivative will be zero mostly, with a blip at the point of discontinuity, and the blip will get higher and higher. Now, early mathematicians (Newton, Leibniz)—in replacing approximate speeds by instantaneous velocities

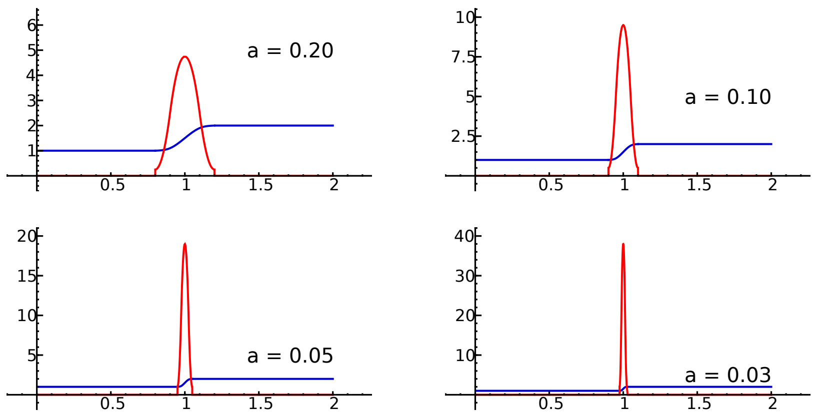

Figure 14.4: A picture of the derivative of a smooth graph approximating the graph that is 1 up to some point and then 0 after that point. In each case, the blue graph is 1 until $1 - \varepsilon$ and 2 after $1 + \varepsilon$.

by *passing to limits*—had to wait a while before later mathematicians (e.g., Weierstrass) gave a rigorous foundation for what they were doing. Happily however, this process we are discussing, of approximating discontinuous functions more and more exactly by smooth functions, and taking their derivatives to get the blip-functions as we have just seen in the graphs above led mathematicians (notably the French mathematician, Laurent Schwartz) to provide a mathematically rigorous foundation (that we needn't go into here) called the *theory of distributions.* Thanks to this theory you can, so to speak, *go all the way* to form what you can legitimately call the *derivative* of the function that is 0 before the point x and 1 afterwards; if we were to think of this as a *function* we would be in trouble: it would have ∞ as a value at x, and—worse—it would be no use to us. But the modern remedy is to formulate a conceptual context in which it has a genuine existence, and can be genuinely used in exactly the same way as the derivatives of smooth functions are used. The only change here is that the derivative of this discontinuous function has a status, not as a function, but rather as a *distribution* (For people who are interested in pursuing this, see [***].
Note: These "blip-functions" as we have been referring to them are more officially called *Dirac δ-functions*, the adjective "Dirac" being in honor of the physicist who first worked effectively with this concept, the "δ being the symbol he assigned to these objects, and the noun "function" should properly be in quotation-marks for—properly speaking the Dirac δ-function is not—as we have explained above—a bona fide function but rather a *distribution*.

The curious aspect of the Dirac δ-function that has its blip at the point x is that it is flatly *zero* in any interval in the t-line that doesn't include the point $t = x$. We say that this Dirac δ-function as its **support** at the point x.

So what happens when we take the derivative—in the sense of distributions—of a complicated staircase? For example:

Well, we would have blip-functions (alias: Dirac δ-functions) at each of points of discontinuity of $\Psi(x)$; that is, at $x =$ any power of a prime.

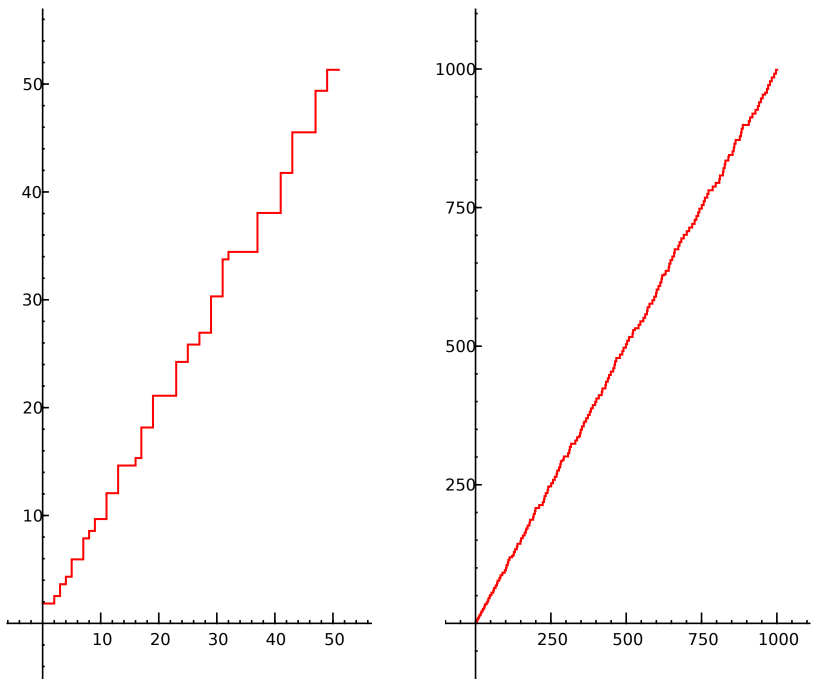

Figure 14.5: Illustration of the staircase $\Psi(x)$ constructed in Part I that counts weighted prime powers.

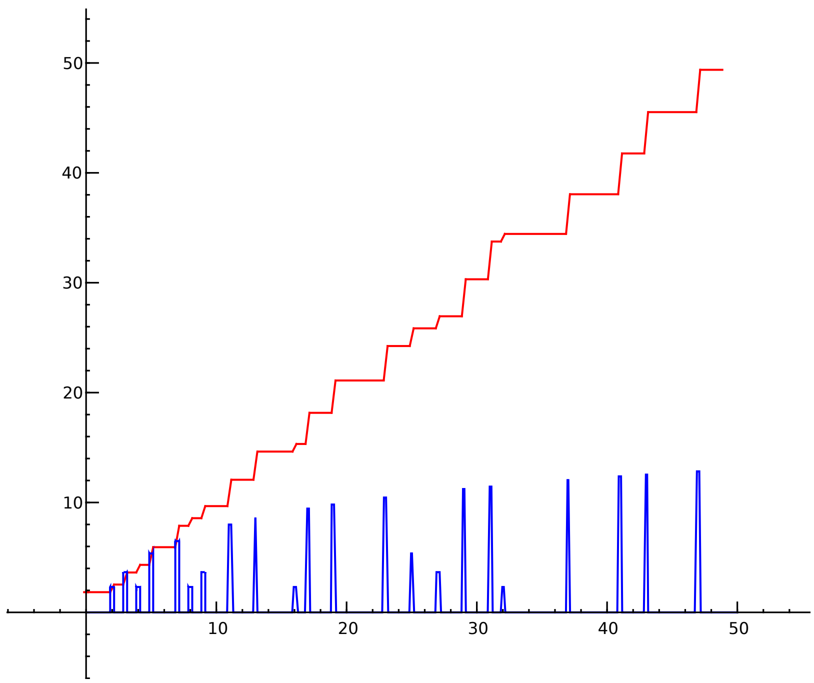

Figure 14.6: Continuous approximation to the staircase $\Psi(x)$ (in red) along with a plot (in blue) of the derivative of this approximation.

As we have hinted above, we lose no information if we further modify our staircase by distorting the x-axis, replacing x by e^t (as long as we *remember what it is we have done!*) so—for reasons that we will *only* see after the next section—we reconfigure our distribution $\Psi'(x)$ replacing it with the distribution

$$\Phi(t) := \Psi'(e^t)/e^{\frac{t}{2}}.$$

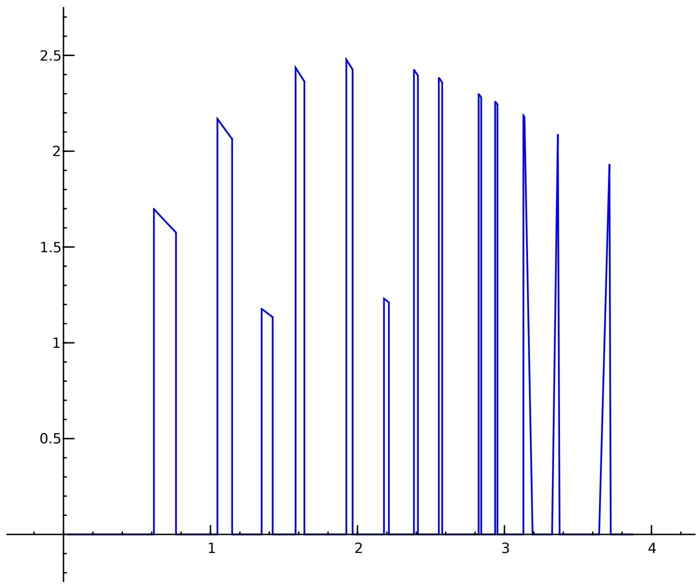

Figure 14.7: An approximation to the distribution $\Phi(t)$

This distribution $\Phi(t)$ has a "blip" at $t = 0$ and also at each positive multiple of $\log p$ as p runs through the primes. It has its *support* at the discrete set of points $t = 0$ and $t =$ multiples of $\log p$ as p runs through the primes.

Distributions with discrete support we will refer to as **spike distributions** because graphs of functions that approximate them look like they have *spikes* around the points in their support.

The spike distribution $\Phi(t)$ will be the main actor, taking center stage, for the rest of Part II.

To be sure, none of the crucial valuable information of placement of primes is beyond retrieve, for the blips constituting of this distribution (its *support*) would allow us to reconstruct the position of the prime numbers among all numbers.

But we must explain our motivation for subjecting our poor initial staircase to this series of brutal acts of distortion to end up with the distribution $\Phi(t)$. For this we embark on a discussion of **spectral analysis**.

15 What do primes, heat, and sound have in common?

[[todo (william)]] Aside. Here we want a really good, really concrete, nevertheless concise, discussion of Fourier transforms, and spectra, with illustrations from the one of the web-pages that does a Fourier analysis of the sounds made by musical instrument, or those that analyze—say—vowel sounds in English. We might talk a bit about white noise and pure tones. But all this is kept within very strict page limits, and—as always—written in a language completely comprehensible to anyone.

It may pay to make a digression to discuss the 1822 contribution of J.-B. Fourier, in a treatise (*Théorie analytique de la Chaleur* that, apparently, Lord Kelvin referred to as a "mathematical poem." Fourier's viewpoint is now used throughout the scientific world.

When applied to *acoustics* Fourier's theory will aid in analyzing and/or synthesizing any sound: it will analyze the sound wave into its component pure frequencies. Also–if you are so disposed—you can synthesize that sound wave by recomposing it, given the information of its constituent frequencies, and their amplitudes. If, say, the sound emitted as a "square wave" in the figure below,

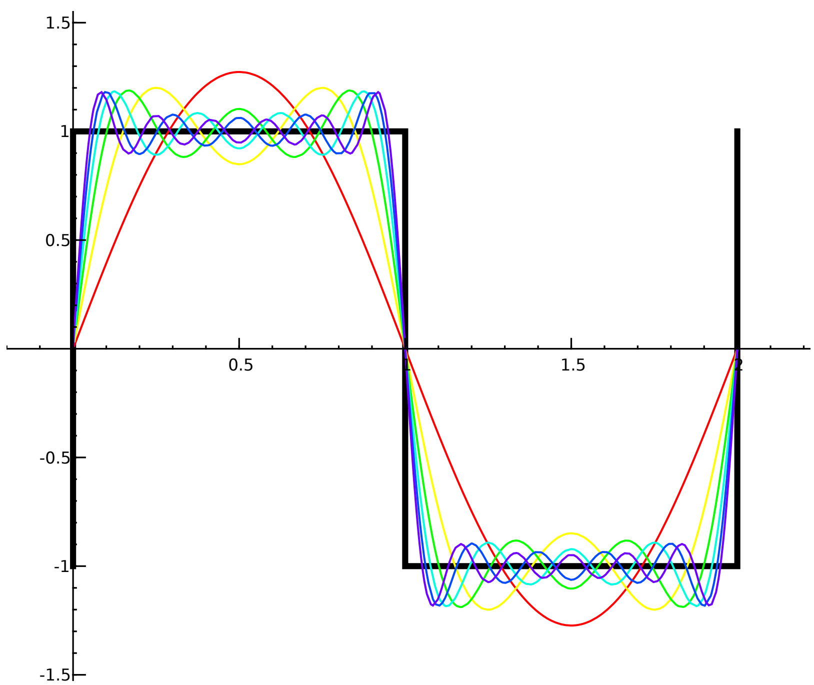

Figure 15.1: The square-wave with its various Fourier approximations

the first picture in the figure gives the so-called *fundamental* of the square wave; that is, the pure tone (a sine wave) whose frequency is the fundamental frequency of the square wave. The subsequent pictures show how you can achieve better and better approximations to that square wave sound by adding more and more of its constituent higher frequency pure tones, these higher components being referred to as the *harmonics* of the square wave sound.

Fourier's idea has settled in as the very language of optics, and electromagnetic phenomena: any source of light is understood (analyzed) as comprised of its **spectrum,** i.e., the sum of pure frequencies–with their corresponding phases and amplitudes–that make up that light source. Such a spectral analysis of, say, the light emitted from the Hydrogen atom, is both

• the *signature* of the hydrogen atom, in the sense that it is so characteristic of hydrogen that form the spectrum alone you can tell if it is, or isn't hydrogen that you are dealing with, and

• a crucial *key* to the deeper internal structure of hydrogen: the spectral lines are distanced, one from another, in proportion to the differences of squares of reciprocal whole numbers: a critical clue, historically leading to the Bohr atom and beyond ....

The beauty of Fourier's modus operandi is that it all arises from a single *model* wave (a sine wave). From this "model wave" both the *fundamental* and all the higher frequency *harmonics* of any spectrum are fashioned; the harmonics are all simply sine waves as well, but modified appropriately in their *frequency, phase,* and *amplitude.* To synthesize any sound wave, say, you need only know this numerical data (*frequency, phase,* and *amplitude*) of the fundamental, and of each of the higher harmonics.

16 Spike distributions that have discrete spectrum

Aside. Here we talk about distributions with discrete support; a 'spike distribution' is then a sum of δ-distributions supported on a discrete set of points. We talk about the curious situation where the Fourier transform of one of these is again one, with a few clean examples. We note that any spike distribution that has the property that its Fourier transform is again a spike also has the curious "error-correcting virtue" that if you ever forget a finite number of points of its support you can retrieve them. We need really interesting SAGE worksheets here.

17 The Spike distribution $\Phi(t)$; the Riemann Hypothesis (*fourth formulation*)

Here we should show a series of better and better approximations to the Fourier transform of $\Phi(t)$, noting that there are spikes at (or around, at least) the real numbers 14.13... etc.

We then blurt out:

There is an infinite sequence of real numbers $\theta_1, \theta_2, \theta_3, \ldots$ such that the distribution $\Phi(t)$ is *equal to* **the distribution defined as the summation**
$$1 + \sum_{i=}^{\infty} \cos(\theta_i t)$$
...

Aside. Here we will motivate the above equivalent statement of RH (that $\phi(t)$ is a sum of the $\cos(\theta t)$'s for a discrete set Θ of θ's. Discuss this amazing feature. Discuss that if Θ exists at all, i.e., if RH is true, Θ has the "error-correcting virtue." We begin to compute it in a "worksheet" given knowledge of the primes, and we wonder whether we can find any other (nonbanal, of course) discrete set Θ' that has the property that the sum of the $\cos(\theta' t)$'s for $\theta' \in \Theta'$ is discrete.

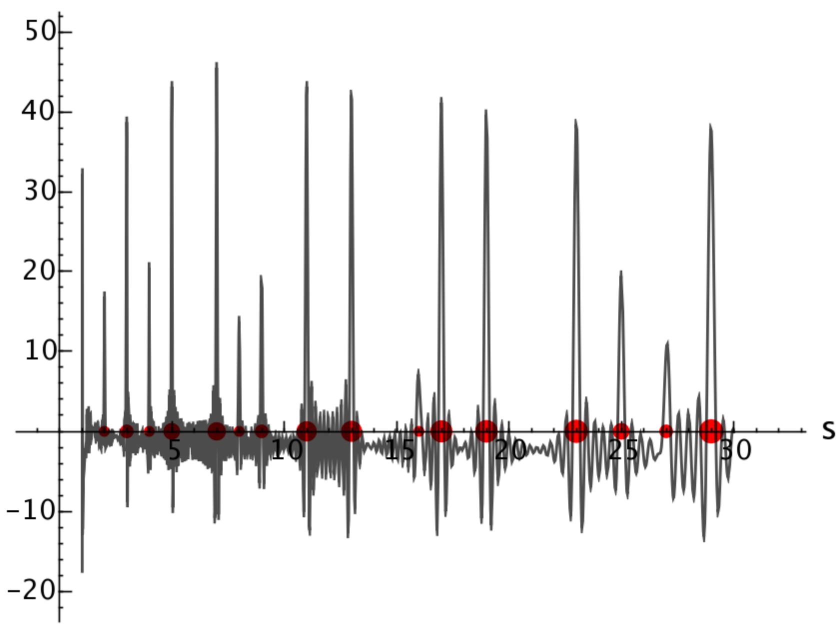

Figure 17.1: Illustration of $-\sum_{i=1}^{1000} \cos(s\theta_i)$, where $\theta_1 \sim 14.13, \ldots$ are the first 1000 frequencies. The red dots are at the prime powers p^n, whose size is proportional to $\log(p)$.

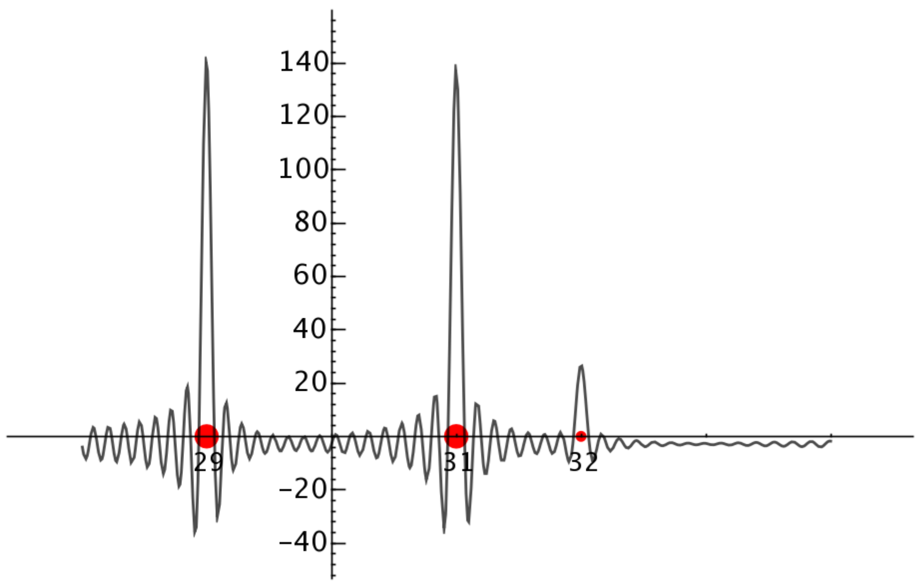

Figure 17.2: Illustration of $-\sum_{i=1}^{1000} \cos(s\theta_i)$ in the neighborhood of a twin prime. Notice how the two primes 29 and 31 are separated out by the Fourier series, and how the prime power 2^5 appears.

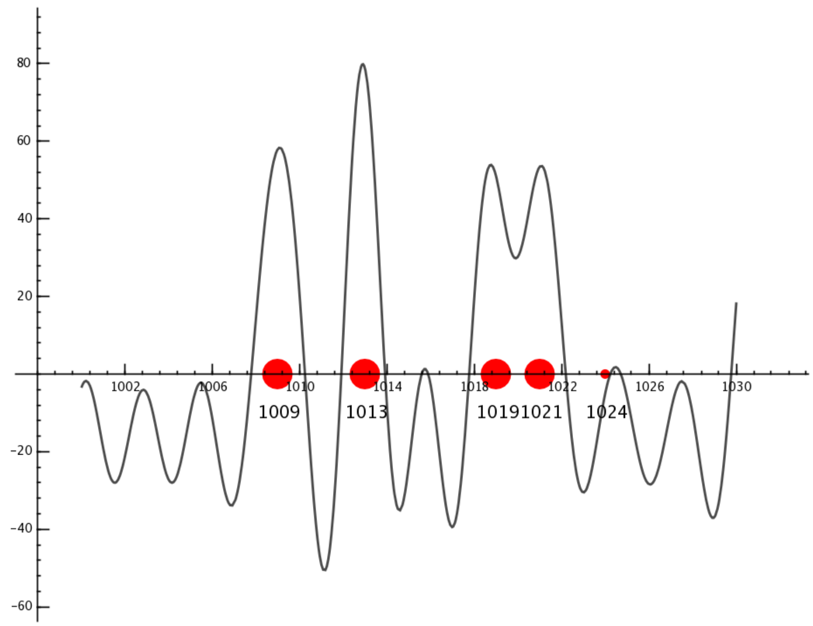

Figure 17.3: Fourier series illustration from 1000 to 1030

18 The undistorted staircase: returning to $P(N)$

We have been dealing in this Part II of our book with $\Phi(t)$ a distribution that—we said—contains all the essential information about the placement of primes among numbers. We have given a clean restatement of Riemann's hypothesis, the fourth restatement so far, in term of this $\Phi(t)$. But $\Phi(t)$ was the effect of a series of recalibrations and reconfigurings of the original untampered-with staircase of primes. A test of whether we have strayed from our original problem—to understand this staircase—would be whether we can return to the original staircase, and "reconstruct it" so to speak, soley from the information of $\Phi(t)$—or equivalently, assuming the Riemann hypothesis as formulated in the previous section—can we construct the staircase of primes $P(N)$ solely from knowledge of the sequence of real numbers $\theta_1, \theta_2, \theta_3, \ldots$

The answer to this is yes, and is given very beautifully by Bernhard Riemann himself in his famous 1859 article cited above.

Think of Riemann's smooth curve $R(X)$ as the *fundamental* approximation to $P(N)$. Riemann offered much more than that in his wonderful 1859 article.

WILLIAM: More facsimile photos of Riemann 's manuscript

Bernhard Riemann provided an exact formula for $P(N)$ (a formula reminiscent of Fourier's analysis of functions as constituted out of sines) that analyzes and/or synthesizes the staircase of primes.

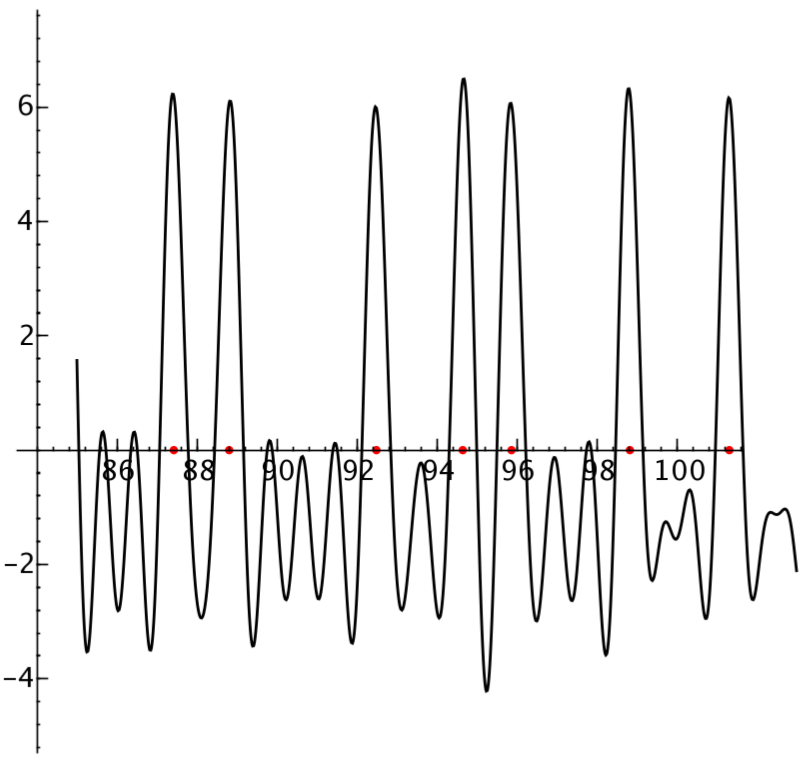

Figure 17.4: The Fourier transform of the distribution supported on prime powers, i.e., which is $\delta\frac{\log(p)}{p^{n/2}}$ at the prime power p^n.

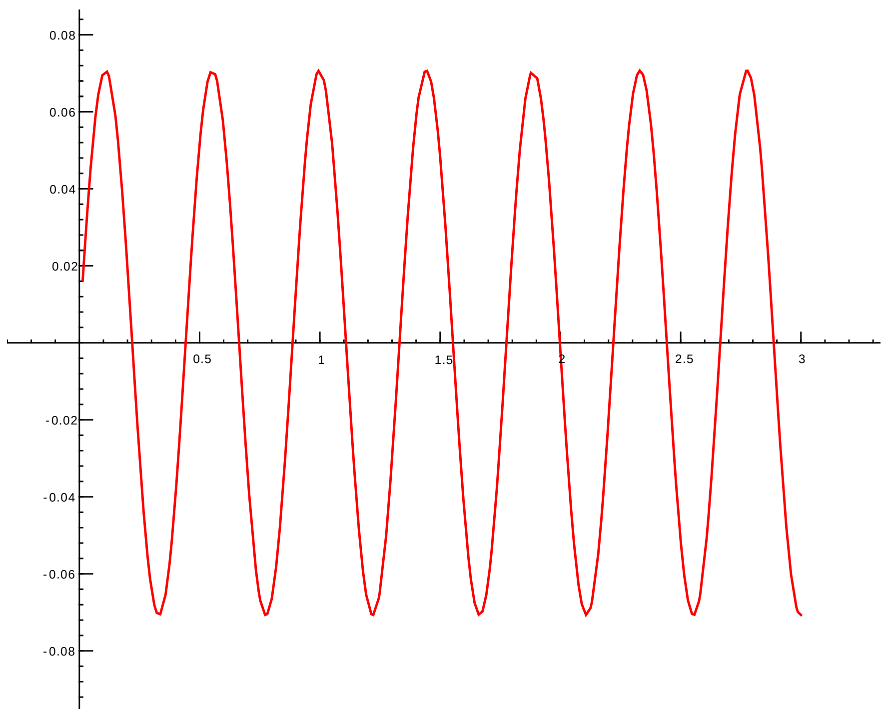

Figure 17.5: Illustration of $\dfrac{\theta \sin(t\theta) + \frac{1}{2}\cos(t\theta)}{\theta^2 + 1/4}$, for $\theta \sim 14.13$, which is the wave component of the first Fourier approximation to $\Phi(t)$

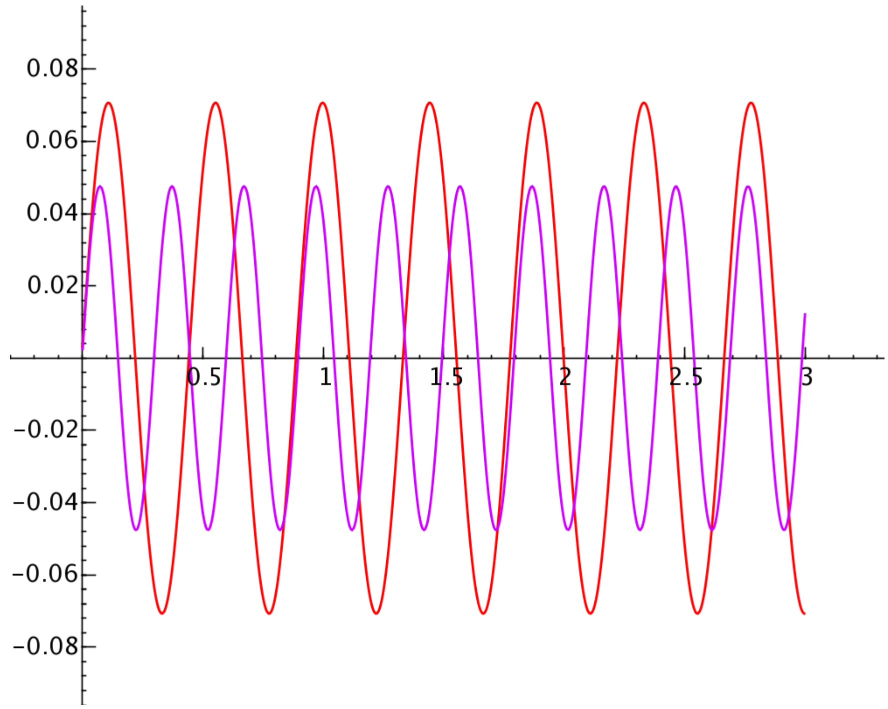

Figure 17.6: The first two wave components that appear in the Fourier series of $\Phi(t)$, are $\frac{\theta\sin(t\theta)+\frac{1}{2}\cos(t\theta)}{\theta^2+1/4}$ for $\theta \sim 14.13, 21.02$

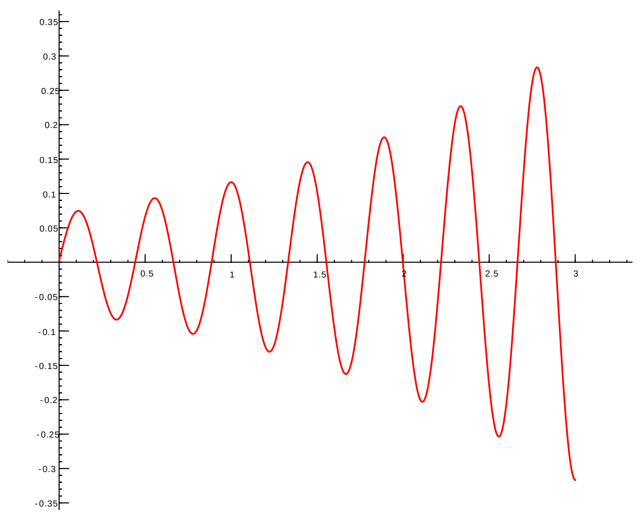

Figure 17.7: The first term of the Fourier series of $\Phi(t)$ is $\dfrac{e^{t/2}\theta\sin(t\theta)+\ \frac{1}{2}e^{t/2}\cos(t\theta)}{\theta^2+1/4}$, where $\theta \sim 14.134725$

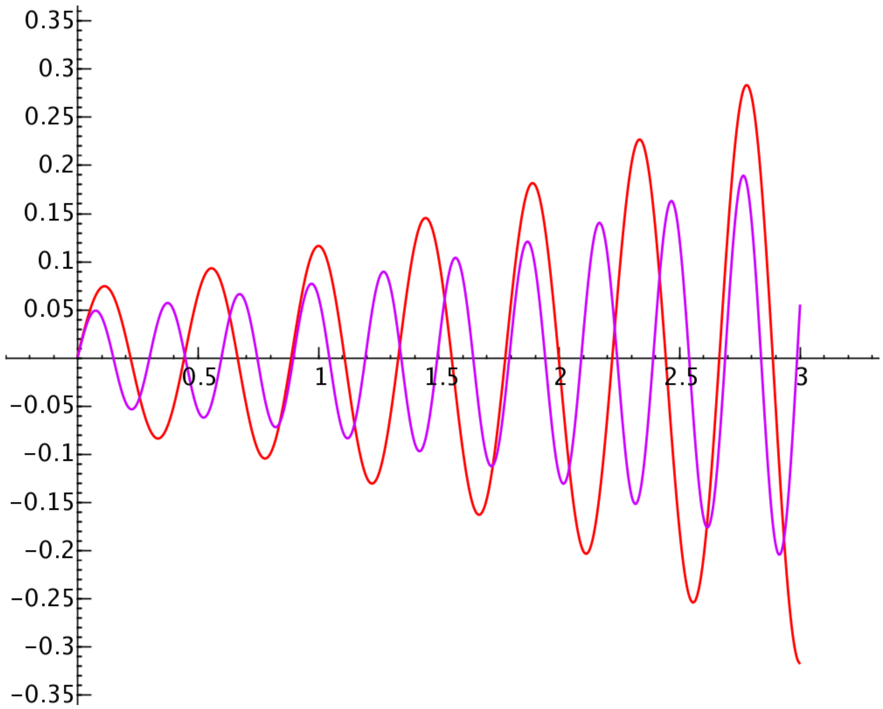

Figure 17.8: The first two components of the Fourier series of $\Phi(t)$ (including the exponential factors) are $e^{t/2} \cdot \frac{\theta \sin(t\theta) + \frac{1}{2}\cos(t\theta)}{\theta^2 + 1/4}$ for $\theta \sim 14.13, 21.02$

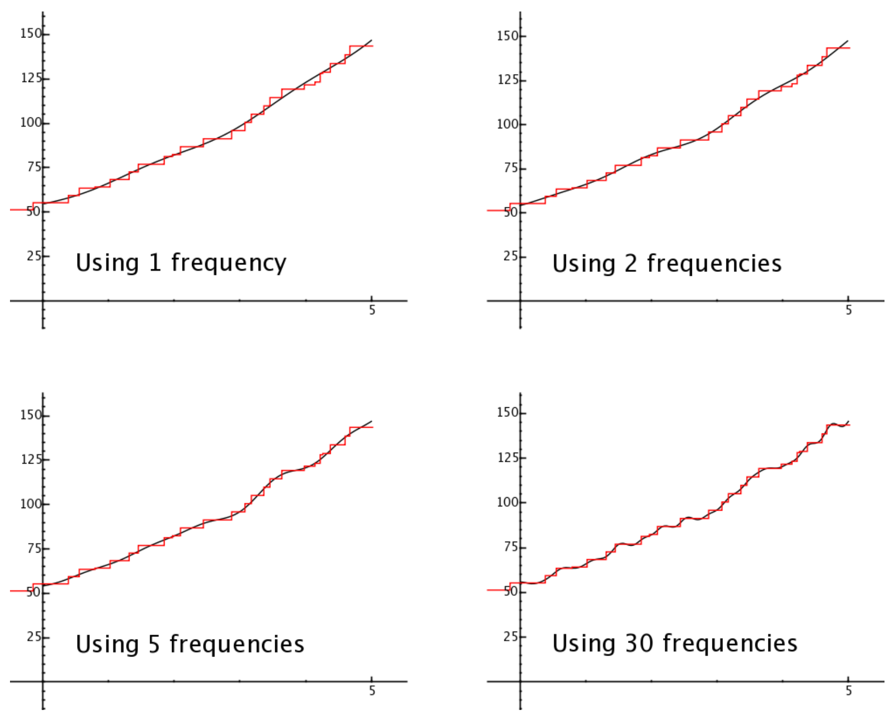

Figure 17.9: Fourier approximations to the step function Ψ obtained using 1, 2, 5, and 30 frequencies.

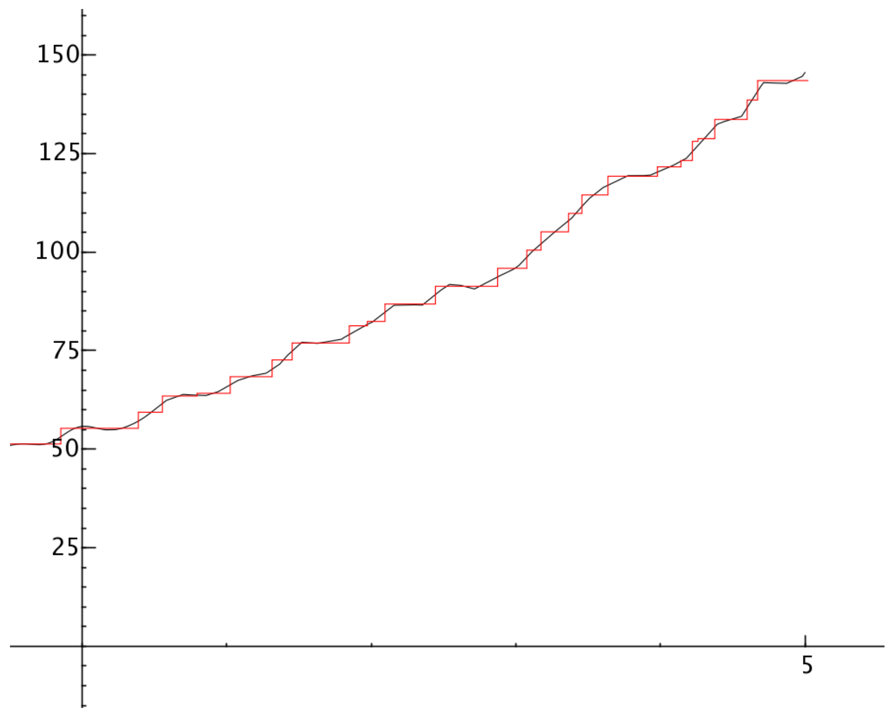

Figure 17.10: Fourier approximation to the step function Ψ obtained using 30 frequences and graphed from 4 to 5.

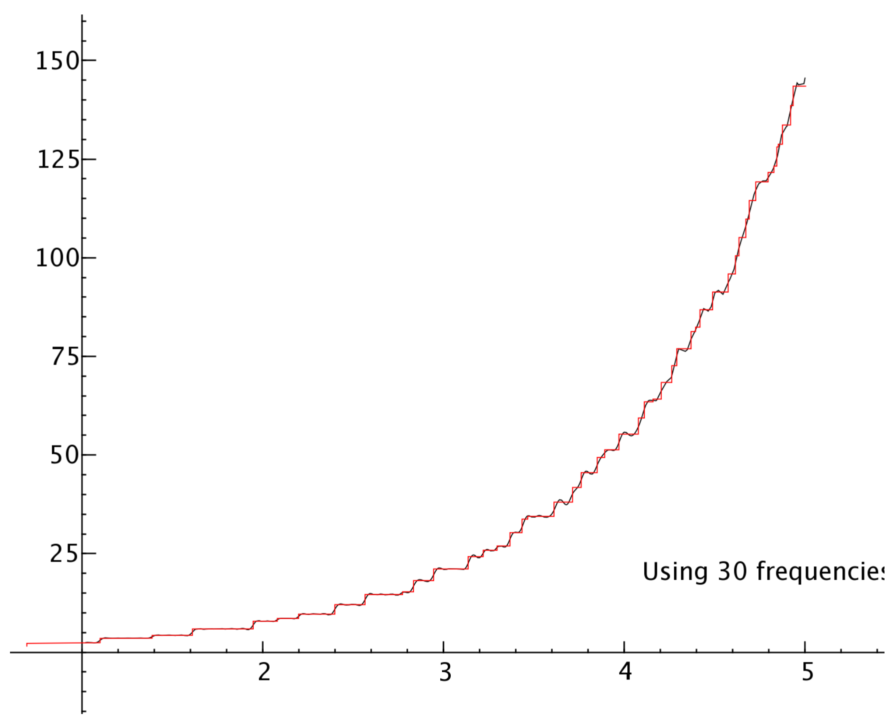

Figure 17.11: Fourier approximation to the step function Ψ obtained using 30 frequences.

Riemann started with a specific smooth function, that we will refer to as $R(X)$, a function the Riemann offered, just as Gauss offered his $Li(X)$, as a candidate smooth function approximating the staircase of primes[3]

Riemann's $R(X)$ seems to be a better approximation to $P(N)$ than Gauss's $Li(X)$:

WILLIAM: maybe a comparison of the three graphs $\pi(X)$, $Li(X)$ and $R(X)$ showing $R(X)$ closer?

Just as Gauss's $Li(X)$ has—in its favor— a probabilistic plausibility argument that might suggest that it is a good approximation to the number of primes, Riemann's $R(X)$ has, in fact, a more refined probabilistic plausibility argument.

NOTE: maybe a comparison of the three graphs $\pi(X)$, $Li(X)$ and $R(X)$ showing $R(X)$ closer?

Not only did Riemann provide a "fundamental" (that is, a smooth curve that is an astoundingly close to $P(N)$) but he viewed this as just a starting point, for he gave the recipe for providing an infinite sequence of corrective terms—call them Riemann's *harmonics*; we will denote the first of these "harmonics" $C_1(X)$, the second $C_2(X)$, etc. Riemann gets his first corrected curve, $R_1(N)$, from $R(N)$ by adding this first harmonic to the fundamental,

$$R_1(N) = R(N) + C_1(N),$$

he gets the second by correcting $R_1(N)$ by adding the second harmonic

$$R_2(N) = R_1(N) + C_2(N),$$

and so on

$$R_3(N) = R_2(N) + C_3(N),$$

and in the limit provides us with an exact fit.

The Riemann Hypothesis, if true, would tell us that these correction terms $C_1(X), C_2(X), C_3(X), \dots$ are all *square-root small*, and all the successively corrected smooth curves

$$R(X), R_1(X), R_2(X), R_3(X), \dots$$

are good approximations to $P(N)$.

[3]Gauss's guess, usually denoted $\mathrm{Li}(x)$, is $\mathrm{Li}(x) = \int_2^\infty dx/\log(x)$, while Riemann's guess is

$$R(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} \mathrm{Li}(x^{\frac{1}{n}}),$$

where $\mu(n)$ is the Moebius function.

The elegance of Riemann's treatment of this problem is that the corrective terms $C_i(N)$ are all *modelled on* the fundamental $R(N)$ and are completely described if you know the sequence of real numbers $\theta_1, \theta_2, \theta_3, \ldots$ of the last section.

PUT FIGURES N,O,P, Q HERE;–this should be Williams graphs of R_k's, for $k = 1, 29, 50, 100$ to show the emergence of a kind of stepwise shape of the smooth curves in question.

(2) He provided an extraordinary recipe that allows us to work out the harmonics,

$$C_1(X), C_2(X), C_3(X), \ldots$$

without our having to consult, or compute with, the actual staircase of primes. As with Fourier's modus operandi where both *fundamental* and all *harmonics* are modeled on the sine wave, but appropriately calibrated, Riemann fashioned his higher harmonics, modeling them all on a single function, namely his initial guess $R(X)$.

Aside. We use Riemann's function $R(X)$, and discuss how to reconstruct the *original* staircase of primes by the R_k's. [[todo: put riemann R_k worksheets 6,8,36,37 here.]]

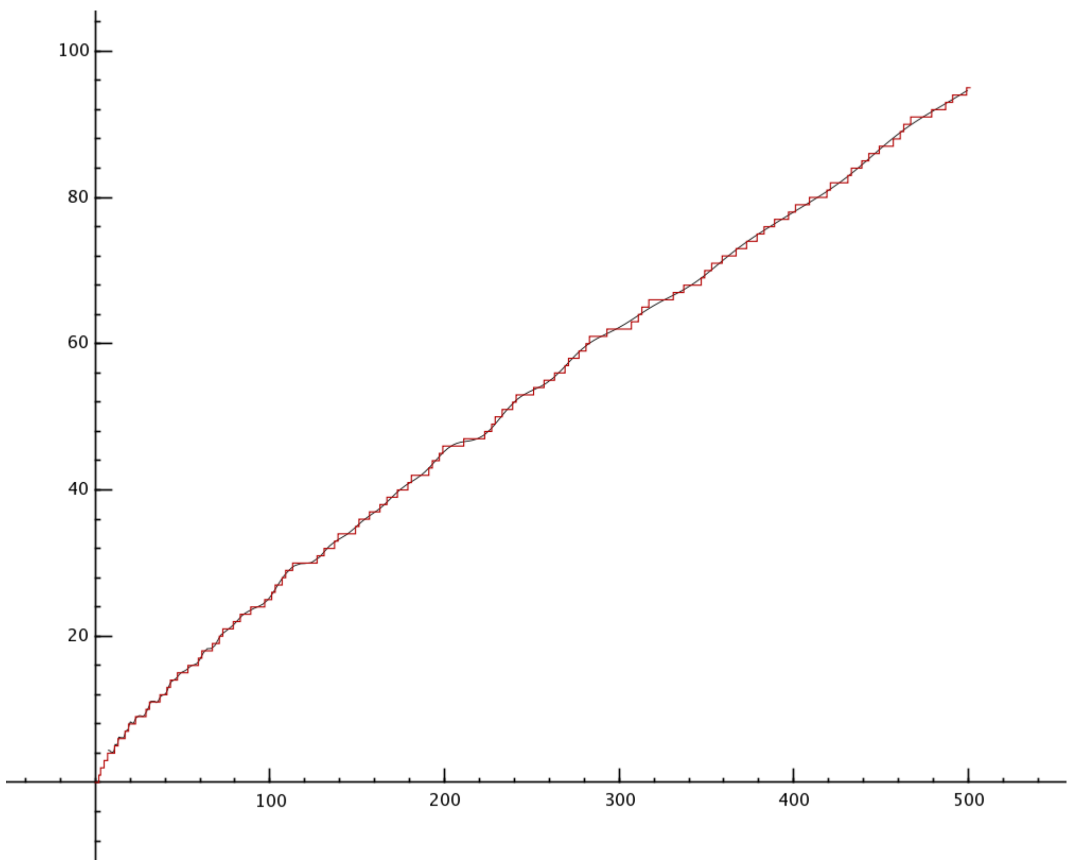

Figure 18.1: The function R_{10} approximating the staircase of primes up to 500

[* All this needs to be written***]**

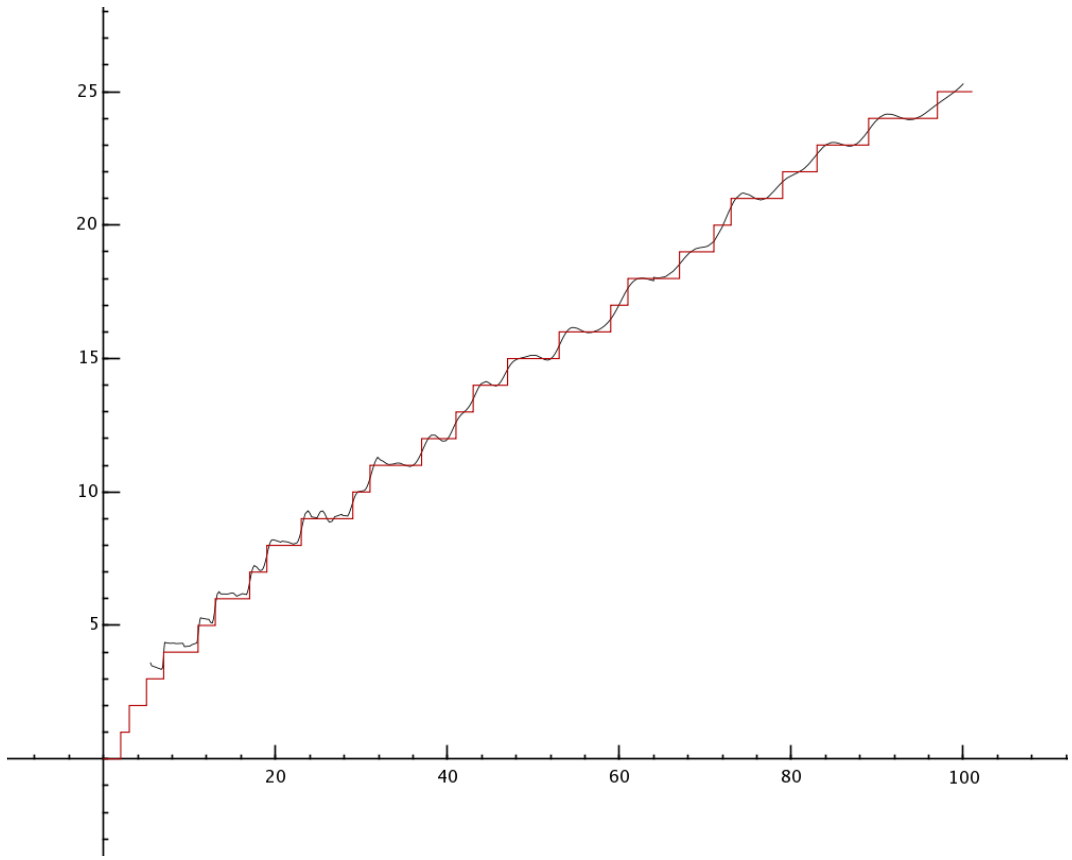

Figure 18.2: The function R_{29} approximating the staircase of primes up to 100

19 William: some alternative text about Riemann

From Riemann's smooth curve to the staircase of primes. Approximations are all well and good, but exact formulas are even better. Can we get a better formula than Riemann's $R(N)$ by adding to it a clever, simple to calculate, correction term? And slightly more ambitiously,

Can we achieve an exact formula for the precise accumulation of primes, $P(N)$,–a formula rather than a mere approximation–by making a clean, infinite, sequence of explicit corrections to Riemann's curve?

The answer to this is given very beautifully by Bernhard Riemann himself in his famous 1859 article cited above.

Think of Riemann's smooth curve $R(X)$ as the *fundamental* approximation to $P(N)$. Riemann offered much more than that in his wonderful 1859 article.

(1) He provided an exact formula for $P(N)$ (reminiscent of Fourier) that analyzes and/or synthesizes the staircase of primes, starting with his "fundamental," $R(X)$ and correcting it successively by stipulating an infinite sequence of correction terms– call these correction terms Riemann's *harmonics*; we will denote the first of these "harmonics" $C_1(X)$, the second $C_2(X)$, etc. Riemann gets his first corrected curve, $R_1(N)$, from $R(N)$ by adding this first harmonic to the fundamental,

$$R_1(N) = R(N) + C_1(N),$$

he gets the second by correcting $R_2(N)$ by adding the second harmonic

$$R_2(N) = R_1(N) + C_2(N),$$

to the first corrected and so on. He provided an extraordinary recipe that allows us to work out these harmonics,

$$C_1(X), C_2(X), C_3(X), \ldots$$

without our having to consult, or compute with, the actual staircase of primes. As with Fourier's modus operandi where both *fundamental* and all *harmonics* are modeled on the sine wave, but appropriately calibrated, Riemann fashioned his higher harmonics, modeling them all on a single function, namely his initial guess $R(X)$.

The Riemann Hypothesis, if true, would tell us that these correction terms $C_1(X), C_2(X), C_3(X), \ldots$ are all *square-root small*, and all the successively corrected smooth curves

$$R(X), R_1(X), R_2(X), R_3(X), \ldots$$

are good approximations to $P(N)$, and in the limit provide us with an exact fit.

For people who know complex numbers, for $k = 1, 2, \ldots$ the $k-$th corrective term is given by

$$C_k(X) = -R(\sqrt{X}X^{i\theta_k})$$

where these mysterious exponents $\theta_1, \theta_2, \theta_3, \ldots$ are the same real numbers discussed in the previous section: *the spectrum of the distribution* $\Phi(t)$.

PUT Williams graphs of R_k's, for various K ($k = 1, 29, 50, 100$?) to show the emergence of a kind of stepwise shape of the smooth curves in question.

Best of all, Riemann provided an extraordinary recipe that allows us to work out the sequence of real numbers harmonics, $\theta_1, \theta_2, \theta_3, \ldots$ and therefore the corrective terms

$$C_1(X), C_2(X), C_3(X), \ldots$$

Part III

20 Further alternative text here

Aside. Here we point out that even though the *mere existence* of the θ's implied the square root approximation info that we began our discussion with, we computed the θ's only assuming RH and *knowing* the placement of primes. So if we are ever to use the θ's to tell us more precise things about the placement of primes, we need an independent way of computing them. Also, if we had an independent way of computing them, we simply have We then move towards discussion of the Riemann zeta function.

[[Insert color pics here.]]

In the previous sections we gave a description of the spectrum of primes that provided the key to calculating the tinkered-with staircase of primes. Suppose you didn't want to tinker with the carpentry of that staircase. ... Half a century after Gauss's initial exploration, Bernhard Riemann produced a slight variant of Gauss's approximating curve,[1] which has some heuristic reason to be an ever-so-small improvement in the approximating accuracy [one can explain Riemann's curve also by consideration of probability]. Riemann's curve, which we will denote $R(X)$, seems to predict the actual number of primes to an astounding accuracy. For example, there are precisely $50,847,534$ prime numbers under a billion and Riemann's curve would predict a number that undercounts this $50,847,534$ by a tiny 79. If you want to compare Gauss's curve, Riemann's curve, and the actual staircase of primes, for the range of numbers less than or equal to $50,000$, there is no problem: Figure 4, given the thickness of the printed curve there, is a picture of *any* of these three curves, they are so close! But despite Figure 4, appearances can be deceiving, for –for all we know– both curves, Gauss's and Riemann's, may eventually stray far from the precious staircase of primes that they are supposed to be approximating. Riemann's curve $R(X)$ gives us an estimate for $P(N)$, the number of primes less than or equal to N.

For a number N, how close, really, is Riemann's estimate, $R(N)$, to $P(N)$, the true number of primes?

With this new piece of vocabulary, " approximation with square-root error, " alias "good approximation," we can frame a clean question that might make more precise what we have seen visually as the miraculous convergence, in Figure 4, of the graphs of Riemann's curve and the true staircase of primes:

(*) *Is $R(N)$ a good approximation to the actual accumulation of primes, $P(N)$?*

We have shaped this question inspired by a certain amount of things, which we haven't gone into in this lecture, but which are already mathematically proven; namely, $R(N)$ is known to be *no better* an approximation to the accumulation of primes, than an approximation with square-root error. On the other hand, $R(N)$ is also known not to be wildly off the mark: the ratio of $R(N)$ to $P(N)$ tends to 1; the essential mathematical content of this– but not the way we have just expressed it– was already guessed by Gauss when he was only fifteen, and finally proved by Hadamard and independently de la Vallée-Poussin, close to a century later, and so is a bona fide theorem, not merely a hypothesis!

Part III

21 The Riemann Zeta-Function; and Riemann's Hypothesis (fifth version)

Here we develop a tiny bit of the traditional route of exposition to RH. Namely, in terms of the successive nontrivial *zeroes of Riemann's zeta-function***.** There are infinitely many of these nontrivial[4] zeroes of the ζ-function. If you want to see tables of the first 100,000 of these zeroes lovingly calculated to an accuracy within $3 \cdot 10^{-9}$, consult Andrew Odlyzko's tables:

www.dtc.umn.edu/ odlyzko/zeta_tables/. The first three of these zeroes are:

$$\rho_1 = \frac{1}{2} + 14.134725\ldots i \qquad \rho_2 = \frac{1}{2} + 21.022040\ldots i \quad \rho_3 = \frac{1}{2} + 25.01\ldots i \quad \text{etc..]}$$

I hope that it is OK if I don't explain this at all, for it does require calculus but we can think of these "zeroes" as being a certain precious infinite sequence of quantities, that contain within them the secret of the exact determination of the pattern of prime numbers. (You can squint at the way Riemann wrote out the zeta-function by trying to make out the formula on the first page of his manuscript, in figure 5a.)

You will notice the curious persistence of the $\frac{1}{2}$'s in the exponents of the three zeroes listed above. Riemann called attention to this phenomenon, which we now know holds for all 1029.9 billion zeroes that have been tabulated as of Feb. 18, 2005! The $\frac{1}{2}$ in the exponent of a correction term is what guarantees that the correction term is indeed "square- root small." We are ready, then, for another equivalent formulation of the Riemann Hypothesis, this being–in fact–the more traditional way of expressing it:

The Riemann Hypothesis (fifth formulation)

The Riemann Hypothesis is equivalent to the statement that all the nontrivial zeroes of the Riemann zeta-function lie on the line $\frac{1}{2} + iy$ in the complex plane

That a simple geometric property of these zeroes (lying on a line!) is directly equivalent to such profound (and more difficult to express) regularities among prime numbers suggests that these zeroes and the parade of Riemann's corrections governed by them–when we truly comprehend their message–may have lots more to teach us, may eventually allow us a more powerful understanding of arithmetic. This infinite collection of complex numbers, i.e., the nontrivial zeroes of the Riemann zeta function, plays a role with respect to $P(N)$ rather like the role the *spectrum* of the Hydrogen atom, plays in Fourier's theory. Are the primes themselves no more than an epiphenomenon, behind which there lies, still veiled from us–a yet-to-be-discovered, yet-to-be-hypothesized, profound conceptual key to their perplexing orneriness. Are the many innocently posed, yet unanswered, phenomenological questions about numbers–such as in the ones listed earlier– waiting for our discovery of this deeper level of arithmetic? Or for layers deeper still? Are we, in fact, just at the beginning?

These are not completely idle thoughts, for a tantalizing analogy relates the number theory we have been discussing to an already established branch of mathematics–due, largely, to the work of Alexander Grothendieck, and Pierre Deligne–where the corresponding analogue of Riemann's hypothesis has indeed been proved...

22 Some fragments to be either used or discarded

(2) If $f(x)$ and $g(x)$ are real-valued functions of a real variable x such that for any $\epsilon > 0$ both of them take their values between $x^{1-\epsilon}$ and $x^{1+\epsilon}$ for x sufficiently large, then say that $f(x)$ and $g(x)$ are **good approximations of one another** if, for any positive ϵ the absolute values of their difference is less than $x^{\frac{1}{2}+\epsilon}$ for x sufficiently large. The functions $\mathrm{Li}(x)$ and $R(x)$ of end-note **(1)** are good approximations of one another.

(3) People who know that these correction terms are index by the nontrivial zeroes of the Riemann zeta-function may well ask how I propose to order them if RH is false; the following prescription will do: order them in terms of (the absolute value of) their imaginary part, and in the unlikely situation that there is more than one zero with the same imaginary part, order zeroes of the same imaginary part by their real parts, going from right to left.

23 Glossary

Aside. Here we will be giving all the connections with the standard literature and coventional terminology that we restrained ourselves from giving in the text itself. Our aim is not even to mention complex numbers in the text, so no Riemann zeta function either, but in this glossary we can talk about $\zeta(s)$ its zeroes, and these being the *frequencies* mentioned in the text. For the moment the list of entries is the following but it will expand.

$P(N) = \pi_0(N)$, $Q(N) = \psi_0(N)$, $\log, \exp$, δ, **distributions, RSA cryptography, Mersenne prime,** $Li(x)$, **random walk, spectrum, harmonic, fundamental, frequency, phase, amplitude, band-pass, complex numbers, complex plane, Riemann Zeta function, zeroes of zeta.**