

Contents

What Gauss missed about primes: $\sqrt{\quad}$, periodicity and symmetry, and ψ	
Visual presentation: before and after Gauss	2
What Gauss did know about prime numbers	4
Gauss vs. Gauss: prime counts and normal distributions	6
Gauss vs. Gauss vs. Riemann	8
Plot the differences	9
Gauss missed finer processing	11
Ways to combine small bins	14
A pictorial exposure of the mismatch in Gauss' prediction	19
Averaging on longer intervals	23
Average distance vs. average density	24
Prime density: averaging with Gaussian weighting	25
The first correction to prime density: averaging with log-Gaussian weighting	27
“Beats” in fluctuations	30
A wider view: wider log-Gaussian weighting	34
The main conjecture (the Riemann Hypothesis)	35
Oscillation in the after-the-main terms of asymptotic	36
The pattern in the terms of the trend	38
On luck: the trend	39
Back to narrow view	40
On the correlations of prime densities at x and at $C \cdot x$ (or C/x)	44
The earlier attempts on further corrections: wider log-Gaussian weighting (without removing “main oscillations”	48
The earlier attempts (in preparation): newer “primeDens”	50
Additive vs. multiplicative: the tale of two Landau	56
The average value of $1/\zeta'(\text{roots})$: surprising oscillations on the log-scale	56

What Gauss missed about primes: $\sqrt{\quad}$, periodicity and symmetry, and ψ

In construction! Most subtopics included, but no rhyme and reason...

The central communication device in our exposition is using carefully designed plots as stepping stones to help the reader to keep the balance in the presented progression of ideas. However, historically this tool is quite recent — its importance was not fully realized in Gauss' time.

Visual presentation: before and after Gauss

But I want to focus on what seems to me the biggest mathematical step that was taken at that time: the invention of graphing. [...] In particular, the idea of tracking some measurable quantity like position, velocity, temperature or brightness over an interval of time or for all the points of an object and making a geometric representation of all these numbers — of this function of time and space — was brand new. The idea is due to Nicole Oresme.^a

Mumford's coursenotes on *Applied Mathematics 18*, 2006

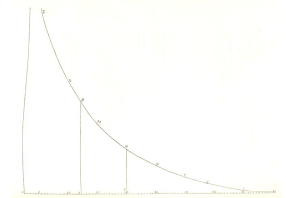
^aOresme lived in 1323–1382.

One of the first examples of mathematical visualization was the usage of bar graphs in Oresme's work, in 14th century to show dependence of speed on time in uniformly accelerated motion. However, in this visualization the independent variable (time) is discrete. The path to visualization of "continuous" phenomena took much longer time.

As a traditional story goes,¹ the idea of a *coordinate system on a plane* took form when sleepless Rene Descartes observed a fly on a ceiling, and realized that its position can be described by a pair of numbers.² As the first — trivial — application, denoting these numbers x and y , one can "find connection" between the circle or radius R centered at the origin and the equation $x^2 + y^2 = R^2$.

³

While Descartes paved the way for translation of geometric concepts into a numeric form, going in the other direction ("visualization") took a few more decades. In 1669, Huygens (In a correspondence with his brother Lodewijk) created the first graph of a continuous distribution function (as on the right).



¹ N.B. (???) Ref???

² Although this approach was not spelled explicitly in his original publication of 1639, and appeared only in the translation into Latin a few years later. It seems that the particular detail of using two perpendicular axes did not appear until a century later, in the posthumous edition of Colin McLaurin's *A Treatise on Algebra* (1748). Same year it was recommended by Euler in the textbook *Introductio in analysin infinitorum*.

Moreover, this fly (if it existed!) may have just helped to crystallize ideas already nascent in Apollonius of Perga's work of about 2 thousand years before Descartes.

³ N.B. (???) Apollonius: would attach coordinate-like structure to |a particular curve|. However: no negative numbers, and

Carl B. Boyer, "Apollonius of Perga" (1991):

There appear to be no cases in ancient geometry in which a coordinate frame of reference was laid down a priori for purposes of graphical representation of an equation or relationship, whether symbolically or rhetorically expressed. Of Greek geometry we may say that equations are determined by curves, but not that curves are determined by equations. <https://hsm.stackexchange.com/questions/5514/when-did-mathematician-start-to-draw-figures-from-equation/5515#5515> specific to conic sections <https://hsm.stackexchange.com/a/2382/7216>

A few more decades passed, and in 1694 Leibniz (who was a student of Huygens) proceeded again “in the direction of visualization”, using geometry to clarify a non-geometric concept. To motivate the notion of *a function*, he considered a curve on a plane, and the positions of particular points on the plane, and the slope of the curve at such a point. Nowadays we would say, for example, that “the y -coordinate and the slope are functions of x -coordinate”.

50 years later, Euler flipped the direction once more. In 1748, in Fig. 2 of Vol. 2 of *Introductio in analysin infinitorum*, he introduces the *idea of plotting*: given a function, can make a curve: the graph of this function.

The other major inventions were made already after Gauss’ time.

One invention was the lowering of the threshold of “working with ‘scales’ of the dependent or independent variable”. Even before that, it was very probable that a working mathematician would easily recognize that when $f(x)$ changes in many orders of magnitude, then plotting $\log f(x)$ has a good chance to give much easier to comprehend results than plotting $f(x)$ itself. (Same for the case when the independent variable x changes a lot, but we pay what happens on the interval $[1, 2]$ of x a similar attention to what happens on the interval $[10,000, 20,000]$. Then it makes sense to plot $f(\exp X)$.)

The idea of doing plotting on *logarithmic paper* was absolutely straightforward⁴: one *essentially* plots $\log f(x)$, or $f(\exp X)$, or $\log f(\exp X)$, but makes the grid related to *the original* variables f and x (as opposed to $\log f$ and $X := \log x$). This way things get much easier to people who do not necessarily understand well the properties of exponential and logarithmic functions. The mathematical side of this was discovered in 1844 by Lalanne.⁵ However, the usage of these methods of plotting started to proliferate only in the last decade of 19th century — after Lalanne’s textbook was published.⁶

The result of this proliferation was that several decades after this, people who needed to tell apart particular forms of functional dependency (with the simplest case being $y = Ax^B$ above) would have a “dictionary” (or a “zoo”) of different plots in front of them. For example, in the second half of 20th century, in Soviet Union everybody with math-heavy needs would have a copy of “Bronshtein–Semendyayev” book — and at this time,^{7 8} the back of the book was filled with such a zoo of graphs and recipes what to pay attention to tell them apart.

Remark 1: For us, data visualization is a tool allowing us to establish another “stepping stone” to balance on when proceeding along a chain of mutually-dependent ideas. However, one should not forget other usages of data visualization in the period of time we consider: to make a hard-to-swallow piece of information easier palatable to the public.

<https://daily.jstor.org/florence-nightingale-data-visualization-visionary/>⁹

⁴ **N.B. (???) Give an example**

⁵ For him, what was important is that these “log-log” plots of functions $y = Ax^B$ are straight lines. So one can use such plots to *recognized* dependencies of this form — as well as ways to easily find A and B .

Lalanne also investigated other ways to mark coordinate axes non-linearly to make *other classes of functions* easier to recognize. d’Ocagne developed Lalanne’s ideas after 1880 under the name of nomography — but this is not directly related to our story.

⁶ By the way, he called this textbook *Méthodes graphiques pour l’expression des lois empiriques ou mathématiques* (1880), which may be the first time when these methods of visualization were called by a “graph”-like word.

⁷ The more recent Russian editions removed this chapter. It is still present in the Spanish translation of 1973. Also see, for example, 2.16.2.5 (7) of the English edition Bronshtein I., Semendyayev K., Musiol G., Muehlig H. Handbook of mathematics (5ed., Springer, 2007).

⁸ **N.B. (???) Check! Give examples?**

⁹ **N.B. (???) Florence Nightingale**

What Gauss did know about prime numbers

... and I have frequently (since I lack the patience for a continuous count) spent an idle quarter of an hour to count another chiliad^a here and there; although I eventually gave it up without quite getting through a million.

Gauss' letter to Johann Franz Encke. 1849 (Later he discusses how he reached 3,000,000.)

^aThe name Gauss uses for a count of primes in an interval of length 100.

The aim of this chapter is to investigate what Gauss could (or must?) have understood about the distribution of prime numbers in alternative reality when he can gain a short-time access to the simplest graph plotting approaches of today (and the necessary computational resources). It looks like he himself did not realize the extra power that visualization of data can give us. On the other hand, he was *very* much interested in the question of the distribution of prime numbers!

Gauss (who was born in a poor family) had a major luck: his elementary school teacher assistant Johann Christian Martin Bartels¹⁰ found access¹¹ to the Duke of Brunswick Carl Wilhelm Ferdinand — who was famous for its wide interests, including science. Bartels was impressed by Gauss, and was only 7 years older; as E. T. Bell writes, “they established a warm friendship which lasted out Bartels’ life”. Bartels recommended Gauss to the Duke — and the Duke took interest in his education. In addition to a stipend,¹² in 1791, when Gauss was 14 years old, he received a collection of mathematical books, including a book with a table of natural logarithms of numbers up to 2,200 and a table of primes up to 10,009.

It may have been the juxtaposition of these tables which lead to the first important breakthrough by Gauss investigating the distribution of prime numbers: he noticed that¹³ the “average distance” between prime numbers¹⁴ near n has “a very strong similarity” to the natural logarithm $\log n$ of n . He discovered this already in¹⁵ 1792 or 1793.

About a year later¹⁶ he understood that this means that if one believes this property of average distance, then the number of primes between m and n should be “similar” to $\int_m^n dx/\log x$ (which one can denote $\text{li } n - \text{li } m$ with $\text{li } n := \int_2^n dx/\log x$). In particular, this estimates the count of primes below n as $\text{li } n$ or¹⁷ $1 + \text{li } n$.

This was only the very beginning. As Goldshstein writes, “Evidently Gauss considered the tabulation of primes as some sort of pastime and amused himself by compiling extensive tables on how the primes distribute themselves in various intervals of length 1000.” To do this, he was using larger and larger tables of primes or tables of prime decomposition.¹⁸

¹⁰Who later became the scientific advisor of Lobachevsky too!

¹¹**N.B. (???) In the movie “Measuring the Earth” he is (???) tutoring a relative of the Duke.**

¹²Below, we include paraphrases of the Mazur–Stein book *Prime Numbers and the Riemann Hypothesis* and Tschinkel’s paper on Gauss.

¹³In fact, already in 1791 he found a much coarser (and less interesting) pattern: the number of primes between 1 and n “is about” $n/\log n$. This pages from his notebook appear in the volume 10-1 of the Göttingen collection.

¹⁴We investigate this notion (and “a very strong similarity” below) later. As a very coarse example, the primes near 100 are 89, 97, 101, 103, 107, 109 with the average distance $(109 - 89)/5 = 4$. Note that $\log 100 \approx 4.6$ — which is not very close to the average distance 4, but is “similar” in a suitable sense.

¹⁵This is discussed in his letter to Johann Franz Encke of 1849 (in a book form). In this letter Gauss discusses the Legendre’s formula $1/(\log n - A)$ with $A \approx 1.15$. Already in 1794 Gauss knew a much better approach involving $\int_2^n dx/\log x$.

¹⁶**N.B. (???) Ref?**

¹⁷As the difference between 4 and ≈ 4.6 in Footnote 14 shows, the expected error of approximation makes the difference between these two formulas negligible.

¹⁸Usually these tables were compiled by other people. However, since he could find some errors in these tables, apparently he have been duplicating at least some calculations himself.

126000

These included (see the letter in Footnote 15) *Vega's Tables* (available to Gauss in 1796, and listing primes up to 400,031), then in 1811 a friend made Gauss a present of *Chernau's cribrum* (an excerpt is on the right¹⁹) reaching 1,020,000. Finally, the Burckhardt's factorization tables (published in 3 parts in 1814–17 and organized similarly to the cribrum) allowed Gauss to investigate primes up to the limit of these tables: to 3,000,000.

N.	D.	N.	D.	N.	D.	N.
1	—————	203	7. 11. 11. 149	401	11. 11421	601
7	7. 47. 383	9	61. 2069	3	41. 3083	7
11	—————	11	—————	7	19. 6653	11
13	—————	17	7. 13. 19. 73	9	83. 1523	13
17	23. 5479	21	113. 1117	13	7. 18059	17
19	—————	23	—————	19	167. 757	19
23	—————	27	—————	21	—————	23
29	193. 653	29	—————	27	7. 18061	29
31	—————	33	—————	31	23. 23. 239	31

Gauss converted all these sources to the two different formats: up to 1,000,000, he lists 1,000 counts of primes, dividing this million into subintervals of length 1,000 (as on the right²⁰). For intervals of length 100 above 1 million, he instead collects them into a group of 100 (covering primes in an interval of length 10,000), and lists (see the table below right) how many such “hundreds”²¹ have a particular count of primes (usually from 0 to 16). Each such group makes one of 10 columns in a small table which covers a range of length 100,000. For every such table (there are 20), Gauss also shows the “expected count of primes” obtained by integration of $dx/\log x$ over this range.

1	168	51	89	101	81	151	85	20
2	135	52	97	102	93	152	90	20
3	127	53	89	103	87	153	88	20
4	120	54	92	104	80	154	77	20
5	119	55	90	105	91	155	84	20
6	114	56	93	106	82	156	85	20

For example, there are 4 tables on p. 439 of *Gauss's Nachlass*; the third table (on the right) covers the range 1,600,000–1,700,000. The column marked as²² 163 covers the interval 1,620,000–1,630,000; it shows 710 primes, with, for example, 5 “hundreds” (of 100 “hundreds” in this interval) having exactly 10 primes.²³

1600000 ... 1700000

	161	162	163	164	165	166	167	168	169	170	
0											1
1						2					2
2	1	3	1		1	1	2			2	11
3	3	3	3	4	4	2	2	4	4		29
4	7	4	9	7	7	10	4	4	10	6	68
5	10	11	8	12	11	11	13	12	18	14	120
6	18	22	15	19	15	11	14	19	10	16	159
7	22	15	14	21	25	18	22	24	24	18	203
8	14	23	25	15	17	16	17	15	13	19	174
9	8	12	18	12	12	21	12	8	14	13	130
10	7	2	5	8	4	6	11	7	4	9	63
11	7	3	1	1	3	1	3	2	2	3	26
12	2	1	1	1				4			9
13	1				1				1		3
14		1									1
15						1					1

In this range of length 100,000 Gauss found²⁴ 7,012 primes—with the logarithmic integral having the value $\approx 6,985.13714$. Hence his prediction gives the error of less than 0.4% for this interval.

However, the counts in these tables vary a lot. For example, in the range 1,600,000–1,700,000 there are 10 columns each matching a range of length 10,000—and in these ten ranges the counts of “hundreds” containing exactly 10 primes are correspondingly 7, 2, 5, 8, 4, 6, 11, 7, 4, 9.

719 694 710 692 692 700 716 702 675 712 7012

$$\int \frac{dx}{\log x} = 6985,13714$$

¹⁹Every page listed in a particular thousand (omitting numbers divisible by 2, 3 or 5). For example, the excerpt above shows 126,421 as prime and shows $126,419 = 167 \cdot 757$.

²⁰This shows, for example, that between 153,000 and 154,000 there are 88 primes.

²¹Gauss calls them “chiliads”. Compare with the epigraph of this section on p. 4.

²²**N.B. (???) Gauss' way to numerate columns changes on this page. On the preceding page, he keeps only one digit—and the column marked 2 in the table for 1,300,000–1,400,000 stands for the interval 1,320,000–1,330,000. So there is an extra shift by 1.**

²³The counts of primes in the “hundreds” of this range go from 2 to 12. The numbers of “hundreds” with such counts are 1, 3, 9, 8, 15, 14, 25, 18, 5, 1, 1.

²⁴The correct value is 7,014. Compare with the table on p. 9.

Gauss' errors go up to an undercount of 30—but only in one range of length 100,000. The next smaller errors are the undercounts of 6 and 7. (These 3 largest errors are among the first 4 ranges of length 100,000 in Gauss' notes! Did he improve later?) There are also 3 undercounts of 4. (The only overcount is of 2 primes.)

Remark 2: Gauss' estimates were never published when he was alive.

Legendre published his $n/(B \log n - A)$ in 1798, and his $n/(\log n - A)$ in 1808 (with $A \approx 1.08366$).
<https://www.mathpages.com/home/kmath032.htm>

In 1850, Tschebycheff proved that Legendre's conjecture cannot be true unless 1.08366... is replaced by 1. (That $A=1$ follows easily from modifying $\int_2^n dx/\log x$ via substitution $x=n e^{-t}$ into $\log^2(n)/n \cdot \int_2^n dx \cdot (1/\log x - 1/\log n)$ and getting $\int_0^L te^{-t} dt \cdot \log(n)/(\log n - t)$ with a large L . The weight of te^{-t} is 1, and the average is 1; hence one should take the value of $\log(n)/(\log n - t)$ at $t=1$, which is asymptotically 1.)

In his letter to Encke mentioned above, Gauss shows that the numerical evidence does not support the value 1.08366. (From the Gauss logarithmic integral conjecture it would immediately follow that this asymptotic must hold with the constant being 1.)

Gauss vs. Gauss: prime counts and normal distributions

Is this needed at all?

As we explained above, the numbers in the body of Gauss' tables answer the questions²⁵:

How many of 100 “hundreds” in an interval of length 10,000 have a given count of primes?

Most probably, these questions were not important to Gauss by themselves. — Then they were just ways to organize calculations of the count of primes in *longer intervals*, of length 100,000, so that they are easy to double-check.

However, here we want to inspect how Gauss could have dealt with these numbers if he had a real interest. In his paid job he dealt with very similar questions a lot, and in the process developed the basics of mathematical statistics

The answer to the framed question above is usually small (rarely above 20) and varies a lot when we look at different intervals of length 10,000. This is due to the pool of primes in a range of length 10,000 being quite small (about 700), so one cannot expect them to be distributed into 100 bins (each of length 100) “very uniformly”.

However, if one looks into intervals much longer than 10,000, then the counts become more organized. For example, for the range 1,000,000–3,000,000, these counts become

Count of primes in a “hundred”	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
How many “hundreds”	2	41	169	594	1,443	2,664	3,621	4,022	3,143	2,216	1,249	539	211	67	13	4	1	1

(At least these numbers grow — up to count 7 — then start to decrease to 0. In other words, plotting this would give a “bell-like curve”. See below!)

Most probably, Gauss would not hesitate a moment given such a table²⁶ — it was him who understood the importance of “bell-like distributions” and used them in his paid job — to achieve unheard-of results!²⁷ It would be immediate to realize that when counts of certain “events” are distributed “as above”, *and there is a simple reason for this*, there is a good chance for this reason to be one of:

- If these “events” are random and independent, one gets a Poisson distribution.²⁸
- The counts may be the results of rounding a continuous *normal random variable* (a variable with *normal distribution* — discovered by Gauss).

²⁵One question per a possible count of primes. They go from 0 to 17.

²⁶I cannot find any trace of him actually investigating this, though.

²⁷**N.B. (???) Story about Stevenson?**

²⁸While Poisson published his description in 1847, this distribution was known from 1711 due to de Moivre's work.

- . Since the counts are non-negative, they may be the result of rounding *an exponent* of a normal random variable. (This is called a *log-normal variable*.)

So a question appears:

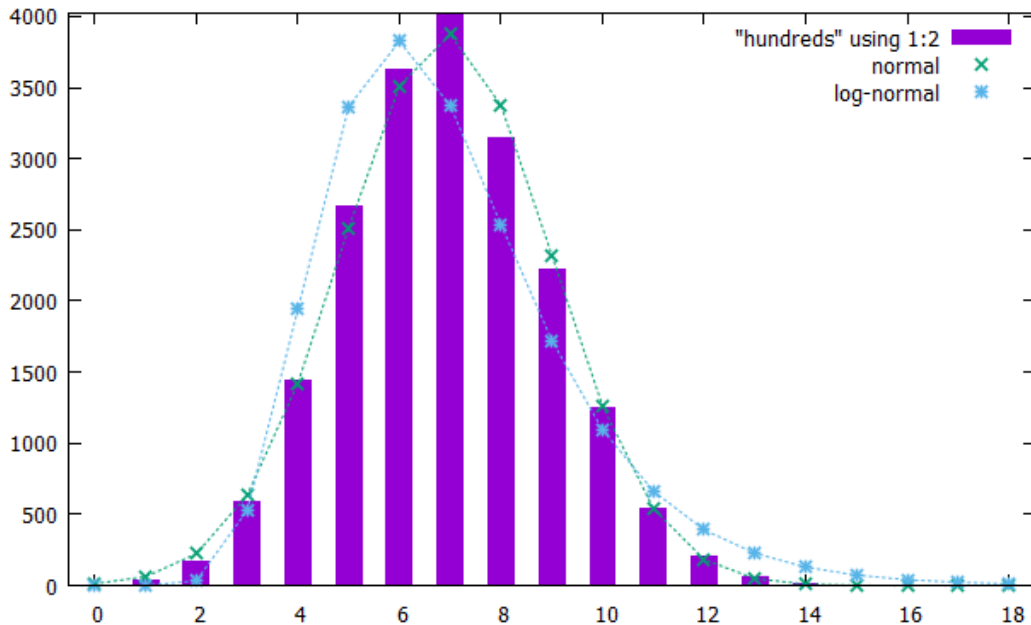
Are the counts in the table above distributed as in one of these 3 cases?

For the Poisson distribution, the check is very simple: the average value of the count of primes is $M \approx 6.915900$, and the average²⁹ distance to this “middle count” is $\sigma \approx 2.037873$. For the Poisson distribution, the second number should be close to the square root of the first one—which is ≈ 2.6298 . But in fact, it is $\approx 23\%$ smaller. **Conclusion:** the counts in the table above are close to “the middle count” 6.915900 significantly more often than if “the event of being prime *behaved similarly* to random independent events!

For each of the other two cases, we can calculate the average (expected) value for every entry in the table above. This becomes

Primes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Actual	2	41	169	594	1,443	2,664	3,621	4,022	3,143	2,216	1,249	539	211	67	13	4	1	1
Normal	13.68	62.27	223.72	634.56	1421.22	2513.65	3511.02	3873.12	3374.37	2321.80	1261.64	541.37	183.43	49.07	10.36	1.73	0.23	0.02
log-Normal	0.00	0.08	34.22	527.84	1940.07	3366.90	3824.39	3371.25	2536.25	1722.64	1093.94	664.38	391.76	226.64	129.56	73.57	41.65	23.58

Since it seems that Gauss did not use visualizations, he probably could easily analyze such data in a tabular form. However, for us (spoiled by civilization) this becomes way easier with plots:



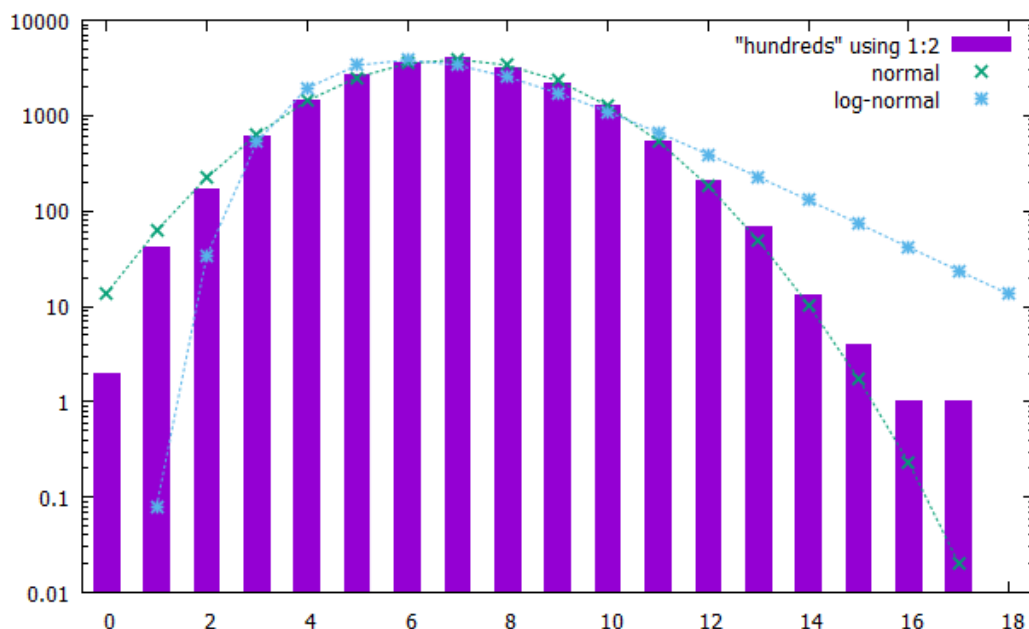
This way, one can immediately see that the plot with blue stars does not match well the bar-plot, while the plot with green crosses matches the bar-plot at least “pretty well”.

Conclusion: the counts of primes in “hundreds” follow pretty well the normal distribution with the average³⁰ $M \approx 6.915900$, and the variation $\sigma \approx 2.037873$. (One may say “close-to-normally-distributed as 6.915900 ± 2.037873 .”) However, when one looks more carefully, one can notice certain “systematic differences” with the normal distribution!

²⁹... in fact, “average quadratic”.

³⁰**N.B. (???) Relation to the average distance between primes. Mixing distributions with parameters differing by Δ gives a result which is in many respects Δ^2 -away from “mixing the parameters” instead.**

Indeed, the numbers in our table change too much in the order of magnitude for the differences to be easily seen. Switching to the log-scale graphs gives:



This time the green crosses are on a parabola³¹ —and one can immediately see that while this parabola gives a really good match near the middle of the graph,³² it systematically undershoots the bar-plot on the right, and systematically undershoots it on the left! So: **Conclusion:** between 1,000,000 and 3,000,000, comparing to “a normal distribution of ‘the count of primes in hundreds’”:

- The number of subintervals of length 100 which have 3 primes or less is smaller “than expected”.
- The number of subintervals of length 100 which have 12 primes or more is larger “than expected”.

I presume that Gauss did not possess approaches to “explain” the graph above—neither for why the counts in his tables follow a normal distribution so closely, nor why there are mismatches for number of “hundreds” with very few primes or with too many primes. Only recently,³³ Rubinstein and Sarnak found that (at least) very similar questions may be answered assuming a couple of conjectures which the majority of mathematicians would consider very plausible!

If we cannot spell out connection to Rubinstein–Sarnak better, why keep this?

Gauss vs. Gauss vs. Riemann

In the preceding section we inspected what Gauss could have done with certain intermediated steps of his calculations (which fill the bulk of his tables, but most probably were not important to him). However, the principal aim of these calculations was to check his conjecture about the average distance between primes. He splits the interval 1,000,000–3,000,000 into 20 subintervals of length 100,000, counts primes in every subinterval $[A, B]$, and compares the result with $\int_A^B dn/\log n$. This

³¹... but only approximately. — This is due to our rounding of a normal random variable to an integer.

³² Although this is much easier to inspect on the preceding plot!

³³ N.B. (???) In 1995?

gives the following table (of the counts in the Gauss tables — with errors in red, the actual counts, the “expected” counts³⁴, and the differences):

10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
7210	7194	7081	7098	7028	6971	7012	6931	6955	6902	6874	6857	6849	6787	6766	6804	6762	6714	6744	6705
7216	7224	7083	7105	7029	6972	7014	6931	6957	6904	6872	6857	6849	6791	6770	6808	6765	6717	6747	6707
7212.99	7165.94	7123.34	7084.48	7048.78	7015.79	6985.14	6956.54	6929.74	6904.54	6880.78	6858.30	6836.98	6816.71	6797.39	6778.96	6761.32	6744.43	6728.22	6712.64
3.01	58.06	-40.34	20.52	-19.78	-43.79	28.86	-25.54	27.26	-0.54	-8.78	-1.30	12.02	-25.71	-27.39	29.04	3.68	-27.43	18.78	-5.64

One can see how the “expected” counts steadily decrease (due to the “average distance” $\log n$ — in the denominator — increasing), changing by ≈ 500 . The actual counts do not have such a steady decrease — but since they stay within the distance 60 of the “expected” count, they “decrease on average”.

We know no comment of Gauss on the behaviour of the numbers in the last row. These tables also do not show whether Gauss compared these numbers with what happens with primes in the first million of numbers.

My conjecture is:

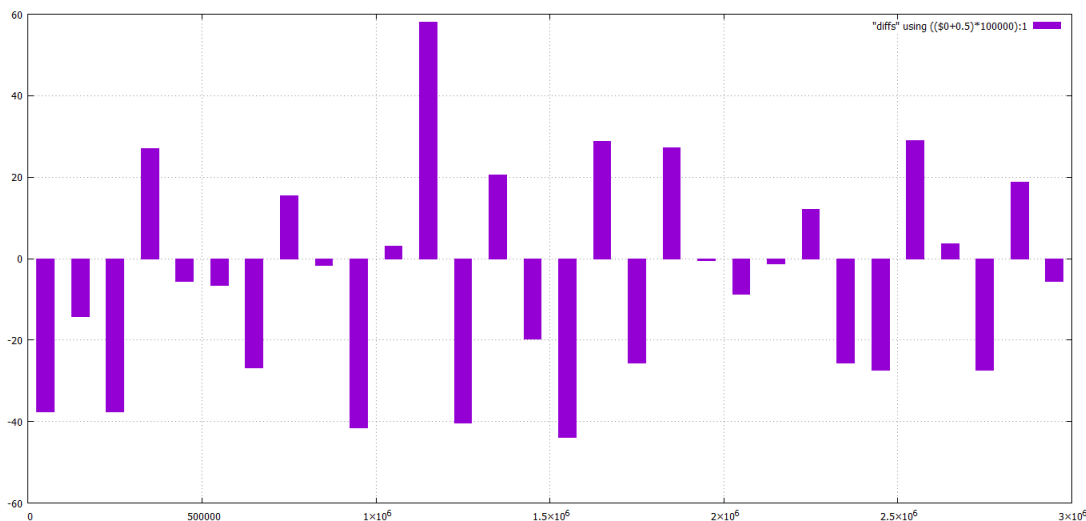
Gauss did not inspect the full range. — And this is how he missed the Riemann Hypothesis!

In the rest of this chapter, we investigate what one can do with numerical data Gauss had — and very similar data requiring more bookkeeping/visualization (and fewer errors!) than what was feasible for Gauss —

Plot the differences

Recall that the last row of the last table shows “the mismatch of prediction”. This is the difference between the actual count of primes in intervals of length 100,000, and the count predicted based on the rule “the average distance between primes near n is $\log n$ ”.

Here (in addition to fixing Gauss’ errors — what we already have done in the preceding section) we extend this row to the left,³⁵ and plot the results.



What Gauss doubtlessly could extract from these data is that

On average, the prediction of the average distance being $\log n$ works *very well*.

³⁴ These are very easy to check for consistency — and Gauss’ errors do not exceed a few hundredth. We give the correct (rounded) values.

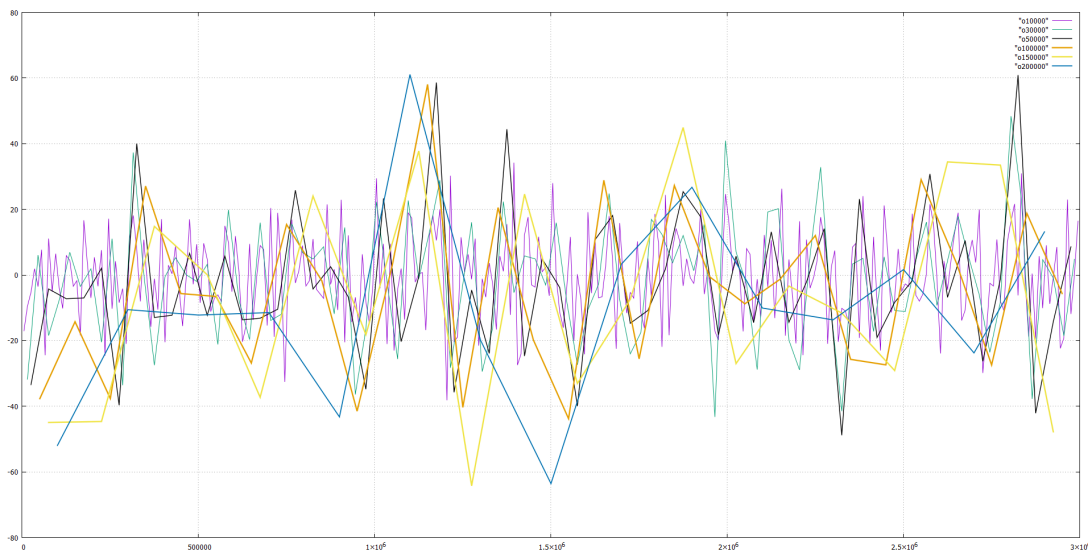
³⁵ Clarify just in case: the logarithmic integral for the leftmost bar is taken over $[1.5, 100,000]$. This choice of 1.5 does not affect the graph significantly.

However, *with what we know now* two other features may be noticed (although they are *almost suppressed* by the “noise”):

- In the right half, the bars are on average about 1.5 times shorter than in the left half.³⁶
- With one exception (of the interval 1,100,000–1,200,000) undercounts³⁷ overwhelm overcounts.

As we said, these features are “almost invisible” on this plot. However, if Gauss would notice them and suspected that they may be real (and not due to “unfortunate coincidence of fluctuations”) he could have relatively easily recheck this by subdividing into subintervals of smaller (or larger) length than 100,000.

For example, here we combine the plots for lengths of subintervals in 10,000, 30,000, 50,000 and 100,000. — And since Gauss knew the stats on all subintervals of length 10,000, he could rework his tables to find these data very quickly:



The first effect becomes more convincing here — at least on the plots with longer subintervals (although on the plots with shorter subintervals it is not easily recognizable³⁸).

The second effect is clearly visible on the left. On the right it is hidden³⁹ ⁴⁰ by “the fluctuations” of the plot. Although these effects are confirmed “only by tendencies”, if Gauss paid attention, he must have recognized that these tendencies need a lot of coincidences to be explained by fluctuations only.

Conclusion: Gauss could have started a more detailed examination of the question:

What are the tendencies in “the average distance between primes” correcting the main term $\log n$?

While Gauss was an absolute ace in statistical analysis (which is needed to handle “the fluctuations” on the plots above), it is not clear if the data in his table is enough to confirm the two effect above. But if he could do this:

- The first effect would lead him to the Riemann Hypothesis.
- The second effect would (could?!) lead to realization that the von Mangoldt function $\psi(n)$ leads to much easier to describe results that “just counting primes”.

³⁶N.B. (???) In fact, this is a random coincidence. On the plots below (where???) where we investigate “what happens between” the shown datapoints, one can see that *on average*, these “added” values in the right half are much higher.

³⁷The negative values in the plot above.

³⁸N.B. (???) Why?

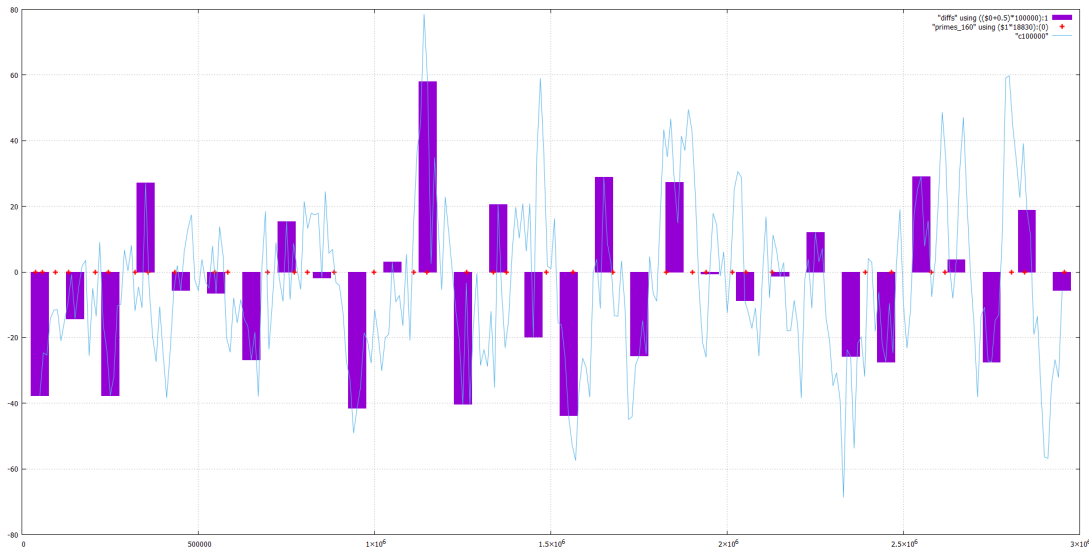
³⁹Nowadays we know that this effect is smaller on the right.

⁴⁰N.B. (???) But the noise should also be smaller on the right?!

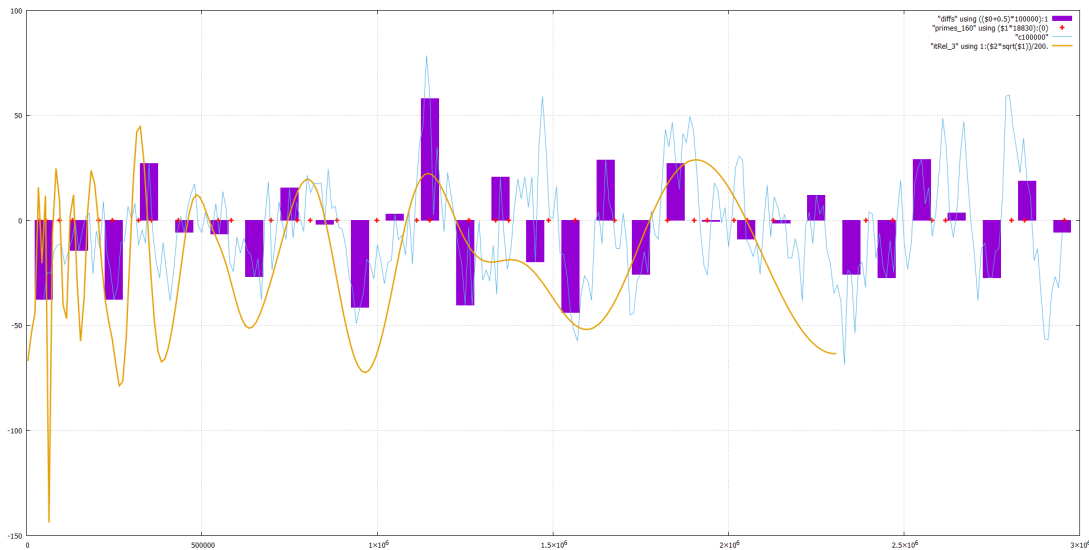
This would teleport Gauss into the level of understanding at the beginning of 20th century—*without any need for complex analysis.*

Gauss missed finer processing

This shows a more detailed picture of what happens in 100,000-long bins, as well as prime numbers multiplied by⁴¹ 18,830:



To slightly clarify the red crosses, the yellow line on this plot is the cyan line “very strongly averaged and suitably rescaled”⁴²;



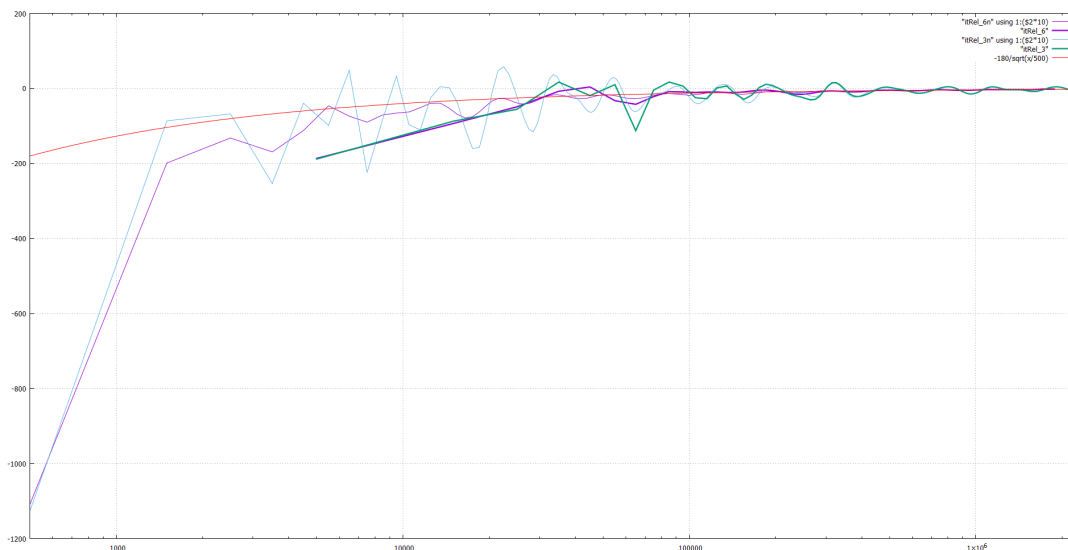
With this line it is much easier to see the tendency for the graph to be higher where the red crosses are denser, and be lower where the crosses are more rare. (For example, observe the peaks on the extreme left, and the “hills” on the yellow line near the crosses corresponding to twin primes: 11 and 13, 17 and 19, 29 and 31, 41 and 43, 59 and 61, 71 and 73 etc.)

⁴¹N.B. (???) Explain!

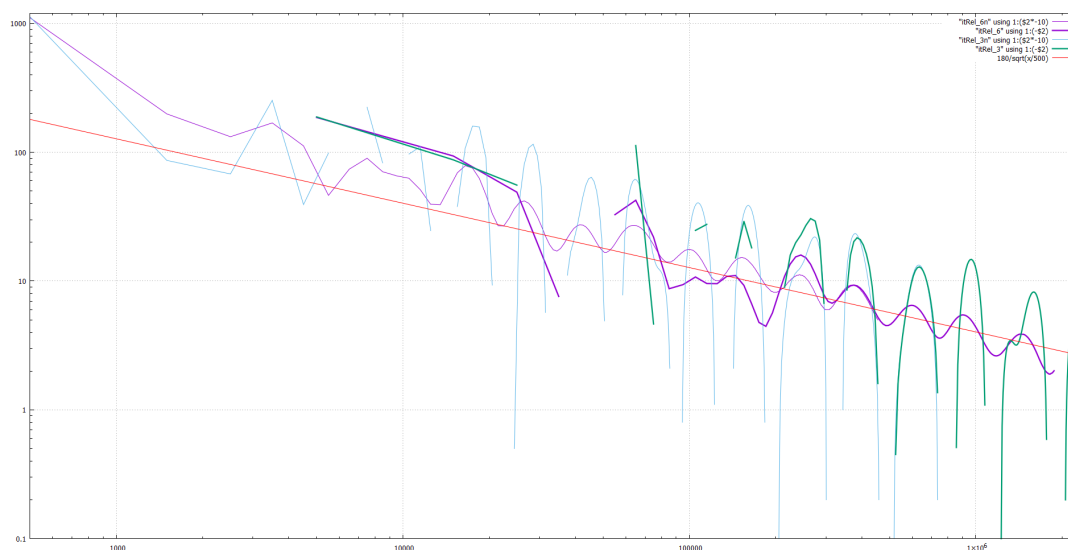
⁴²The only purpose of rescaling is to make the oscillation easier to track.

Using a much more delicate processing of the Gauss' mismatches allows one to *expose the trend* in these mismatches hidden by the fluctuations in the data above. To do this, one should do a *repeated averaging* of the 200 "coarse" mismatches in the Gauss notebooks.^{43 44}

Doing such averaging gives this:



The fitted "sqrt-hyperbola" on this graph was found using the log-log-version of the plot above (with the *y*-axis inverted):



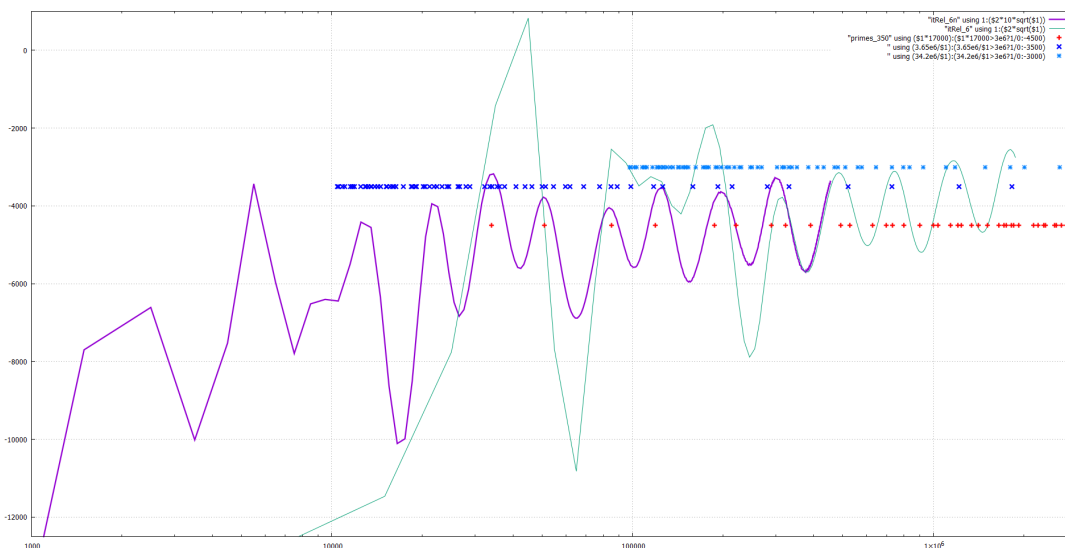
The thick lines are using Gauss' bins of length 10,000, the thin ones are for the bins of length 1,000. For the green plots, one can see a very good match between the thin and the thick plots on the extreme right of the thin plot (near 400,000). On the other hand, on the extreme left of the thick plot (for the bins centered at 5,000, 15,000 and 25,000) the thick green curve is way below the thin one. Since smaller bins allow much more control over cancellation of fluctuations, this shows that when

⁴³Below, we correct Gauss' errors, as well as extend Gauss calculations of mismatches to (one hundred of) the 10,000-long bins in the the first million (thick lines on the plots below). We also show what happens when one calculates mismatches with Gauss' predictions in 1,000-long bins Gauss used in the first million.

⁴⁴**N.B. (???) Discuss "effective widths", 0.3, 0.6, repeated averaging, de Moivre etc.**

the bin is so wide comparing to the value of x , our method of averaging undershoots the correct value. This also suggests that one should not trust the spike on the extreme left of the thin graphs.⁴⁵

Here we rescale the plots above with “stronger averaging” (those with the relative width 0.6) with a factor⁴⁶ \sqrt{x} :



The thick line uses Gauss’ bins of length 1,000, the thin one — of length 10,000. The red crosses are as above — and now it is crystally clear that:^{47 48}

If the density of primes near x is “above (or below) the average”, so is the density at $\sim 17,000x$.

Additionally, we mark numbers of the form $3.5 \cdot 10^6/p$ and $34.2 \cdot 10^6/p$ with p a small prime (in dark blue, and in light blue). Again, the density of numbers of such a form near x correlates very well with the average prime density near x . This demonstrates that for $C \approx 3.5 \cdot 10^6/p$ and $C \approx 34.2 \cdot 10^6$:

If the density of primes near x is “above (or below) the average”, so is the density at $\sim C/x$.

⁴⁵And on the rest of the graphs, the red hyperbola gives a good approximation to the non-oscillating part of “the trend”

⁴⁶Here again, this factor makes the features of the graph easier to see. — However, the numbers on the vertical axis are not easy to interpret.

⁴⁷N.B. (???) **What happened to 18,830? We suppressed high-frequency fluctuations a lot, so the match happens when shifting by $\log 17,000$. When not suppressing them so much, so one needs to be close to a multiple of several periods — so the conditions are different.**

For example, $\cos(14.135 \cdot t) + \cos(21.02 \cdot t)/15 + \cos(25.01 \cdot t)/50$ has maximum at $\sim 17,800$, but $\cos(14.135 \cdot t) + \cos(21.02 \cdot t)/5 + \cos(25.01 \cdot t)/7$ has maximum at ~ 18.1 . However, with 18,830 instead of 17,000, the match with the plot above is still “reasonably good” — but it works much better when we do not suppress the fluctuations so strong.

⁴⁸N.B. (???) **Of course, the periodicity of the oscillations on this graph makes the number 17,000 not very special: the same effect would appear when 17,000 is multiplied by a power of 1.553 — which corresponds to the period above.**

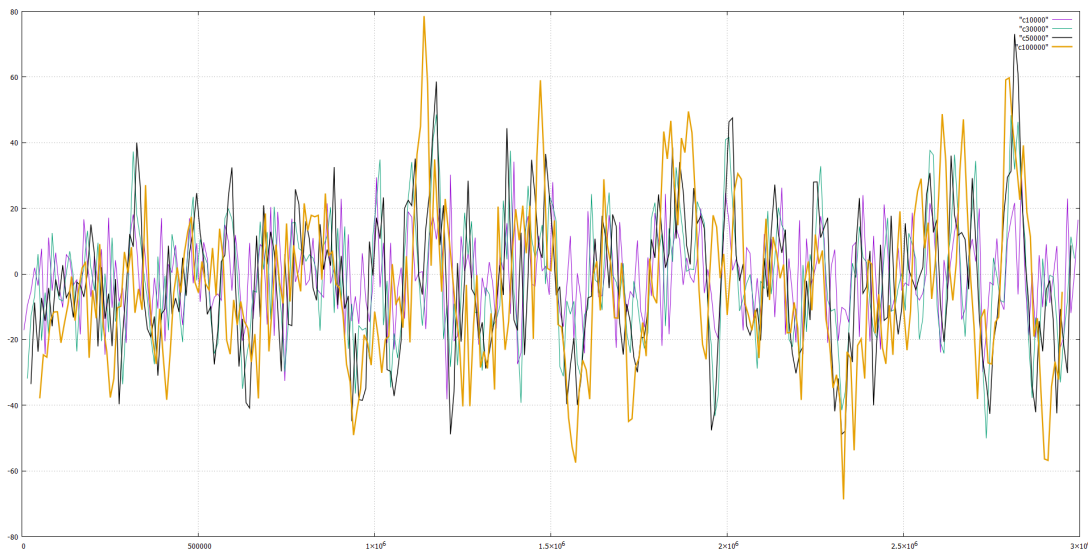
Since 10 periods of oscillation take us from $\sim 22,000$ to $\sim 1,800,000$, one can see that 22 periods of oscillation correspond to multiplication by $\sim 16,100$. This is close enough to 17,000 (for the purpose of this explanation).⁴⁹

⁴⁹N.B. (???) **Can we make these numbers more realistic?⁵⁰**

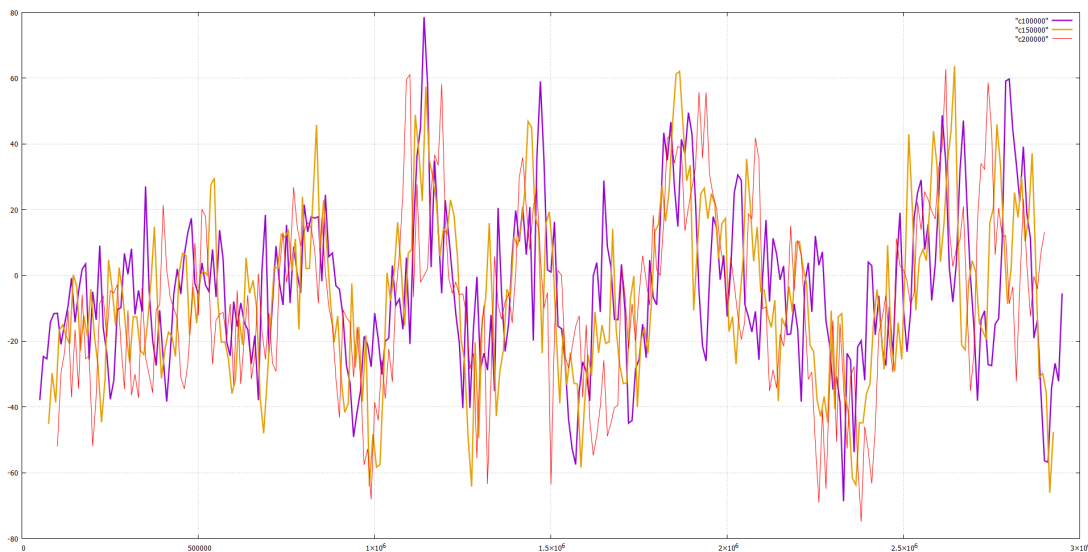
⁵⁰N.B. (???) **What the match between red crosses and the peaks of the plot shows is that although on the graph we may observe only the oscillations in the range about 10,000... 2,000,000, in fact, these same oscillations continue into the range 2... 150 as well.**

Ways to combine small bins

As above⁵¹, this shows the mismatches in bins of widths 10,000, 30,000, 50,000 and 100,000 (with midpoints in steps of 10,000).



This is likewise, but with widths 100,000, 200,000 and 300,000:

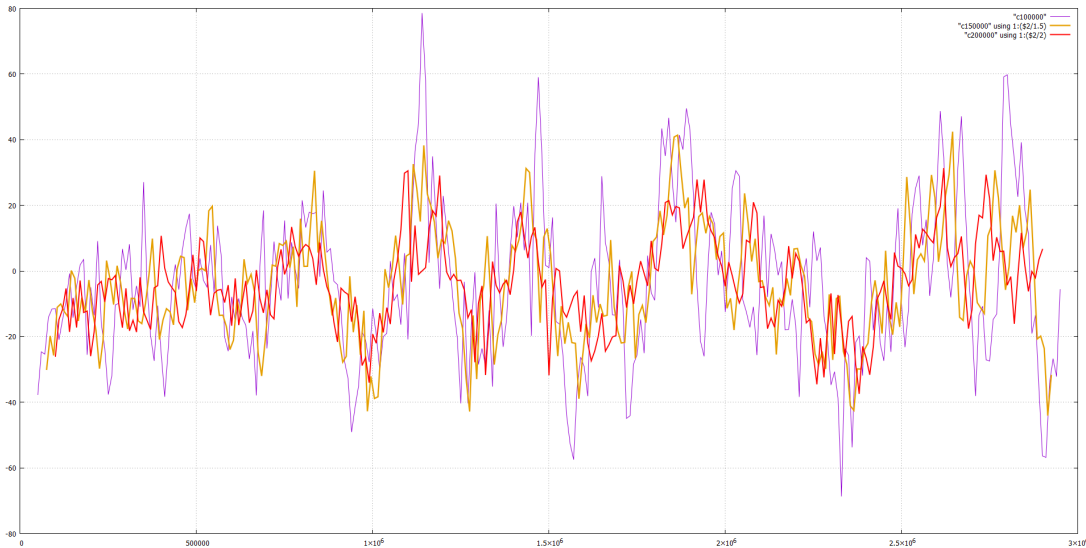


The first picture has very little “visual structure”: the fluctuations when we switch from a bin to a nearby bin are very large for so narrow bins — and these fluctuations obscure *the trend of these plots* too much. However, on the second picture one can actually trace “certain features” which appear with wider bins. **Question:** can one massage these plots to get yet more visible structure?

One approach is immediate: observe the plots in the latter picture on the far left. They all are mostly below $y = 0$, and the plots for wider bins are below the plots for narrower bins. And indeed, every bin of width 300,000 is made of 3 sub-bins of width 100,000 — so if the mismatch in every sub-bin is negative, the mismatch in the combined bin is going to be larger in magnitude.

⁵¹N.B. (???) Ref and ref!

Conclusion: to visualize slowly-changing trends, we need to renormalize the Gauss' mismatches. So we divide the mismatch by the (relative) width of the bin (taking width 100,000 as “the base”):



(If we would try to combine this with the similarly renormalized plots for bin widths $<100,000$, these added plots are going to be much higher, with the graph for width 10,000 going up to a few hundreds. — When we “preserved the trend” and widened the bins, we “dumped the fluctuations”!)

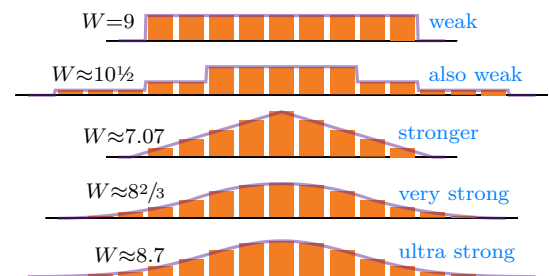
However, one can dump the fluctuations much better than what we did above. To explain this, note that the rescaling above means that the datapoints on yellow graph are midpoints for consecutive pairs of datapoints on the violet graph; likewise for consecutive triples on the violet graph vs. the red graph. (And if we add the plot for the bin width 10,000 to the same picture, the red plot *averages 30 consecutive datapoints* on the added plot.)

With this approach, the explanation for why the graph for wider bins is “smoother” is that *it averages out the fluctuations* in the graph for narrower bins. This leads to the following ideas for stronger dumping:

- What if we average together several graphs for several bin widths?
- What if we average together several datapoints on the red graph?

So: how would the result compare with “just using yet wider bins”?

Answer: it turns out that these two ideas are closely interrelated — and the relation goes through the notion of *the weighted average*. When we have several datapoints, and take an average of (say) 9 consecutive datapoints, *we sum them up with the weights $1/9$* (as in the top row on the right⁵²). When we *additionally* take average of (say) 5 consecutive datapoints, and the same for 15 consecutive datapoints — and then take the average of these three averages, this matches taking a weighted average with weights $1/45, 1/45, 1/45, 8/135, 8/135, 17/135, 17/135, 17/135, 17/135, 17/135, 8/135, 8/135, 1/45, 1/45, 1/45$ (as in the middle row on the right). (What is important for what follows is that these weights change by “jumps” — like on the top diagram.)



Furthermore, if we *average 5 consecutive averages*, each for 5 consecutive datapoints, this corresponds to weighted averaging using 9 weights which depend piecewise-linearly *and continuously*⁵³

⁵² We discuss the comments on the left and on the right these pictures below.

⁵³ Here we cheat a bit. To make this statement meaningful, one should also consider what happens when we replace 5 by a number going to ∞ . This is shown by the purple line on the plot.

on the position: they decay linearly from $1/5$ in the middle to $1/25$ at the ends. (As on the third row above right.) Moreover, if *we average the results again* (repeating this recipe “average 5 consecutive values” once more), the final result corresponds to weights which depend piecewise-quadratically on position — and this function has a continuous derivative! (As on the fourth row above right.) But this function has jumps in the second derivative. . .

Finally, the last row on the picture above corresponds to the recipe “take the average of 2 consecutive datapoints”—and we repeat this 19 times. This number 19 is chosen to make these weights *practically indistinguishable* from the preceding row—but we will see that the result gives a much better “suppression of fluctuations”.⁵⁴ For all practical purposes, these weights “have no jumps in derivatives”.

Given a particular way of taking weighted averages (which may “be scaled to different widths”, as above), there are two important questions:

- How strongly it “blurs” slowly changed inputs? (In other words: how far can it “change trends”?)
- How well it suppresses fluctuations in datapoints?

It turns out that the first question is may be answered by providing one number: the deviation. To avoid the trouble, we can use instead⁵⁶ “the equivalent width W of a rectangle”: for the top weights above right it is 9, for the next row it is ≈ 10.5 , then $5\sqrt{2} \approx 7.07$, then $5\sqrt{3} \approx 8.2/3$, and for the last one it is $2\sqrt{19} \approx 8.7$.

To get a taste of how the answer to the second question may look like, inspect what the filters⁵⁷ above do with datapoints with the quickest possible oscillations . . . , $-1, 1, -1, 1, -1, 1, -1, \dots$, as well as with . . . , $0, -1, 0, 1, 0, -1, 0, 1, 0, -1, \dots$ (which oscillate twice slower). The top filter suppresses them 9 times; the second one ≈ 8 times; the third (“the triangular one”) 25 times, the fourth 125 times, while the last filter kills the first input completely (already on the first step!) and suppresses the second input about $\sqrt{2^{19}} \approx 700$ times.

The general principle explaining this is⁵⁸:

The smoother is “the law governing the weights” (the purple lines above), the better is the suppression of flu

So to separate “trend” from “fluctuations”, we should better to filter using weighted averages with:

- The “equivalent width” W as small as possible. (To keep as much information about the trend as possible.)
- Use weights which are “as smooth as possible”. (To suppresses the fluctuations stronger.)

As we demonstrated above (and as Gauss should have known⁶⁰), one gets a very good balance between “mutating the trend” and “suppression of oscillation” when one uses “Gaussian weights” (as on the

⁵⁴Here we cheat again—twice. First, the width covered by the non-0 weights is much larger than what is drawn. However, the weights decrease so quick that we just cut off super-small weights from the picture. (And still, two smallest drawn orange rectangles are visible only with ultra-strong zooming.)

Second, with an odd number 19 of repetitions, one would have not one maximum, but two equal maxima in the middle. On the other hand, since $19 \gg 1$, these weights are indistinguishable from the corresponding Gaussian weights. What is drawn is the Gaussian weights⁵⁵ *shifted by $1/2$* to “move the maximum to a datapoint”. In other words, this shows “what happens when 19 is replaced by an even number”—but we rescale the width to $2\sqrt{19}$ (to match the weights on the previous row).

⁵⁵**N.B. (???) Explain!**

⁵⁶It is $2\sqrt{3}$ times larger than the (standard) deviation.

⁵⁷**N.B. (???) Explain!**

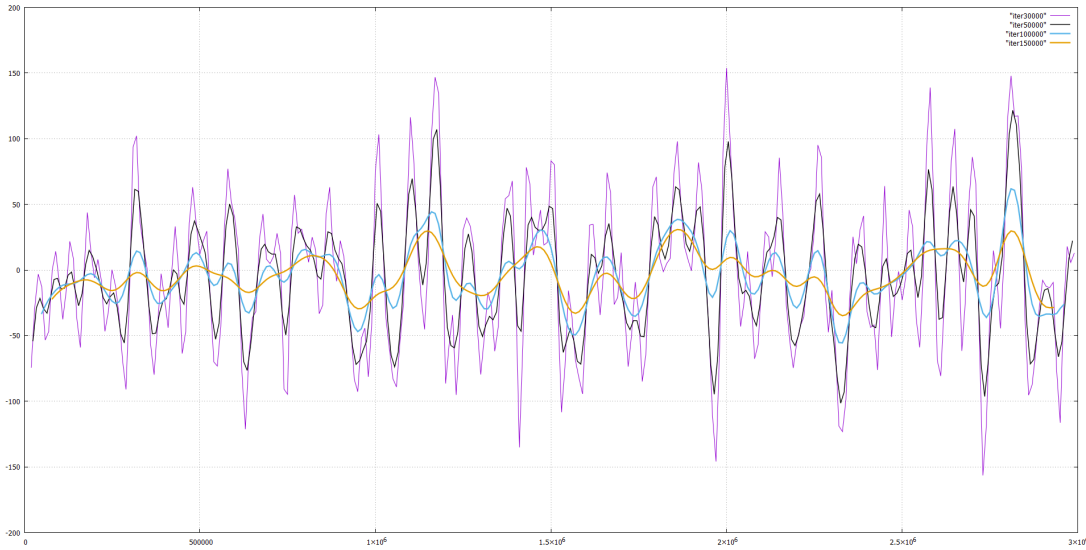
⁵⁸To explain this, one needs to inspect the Fourier transform⁵⁹ of the weights. (Compare with [this discussion](#).)

⁵⁹**N.B. (???) Ref!**

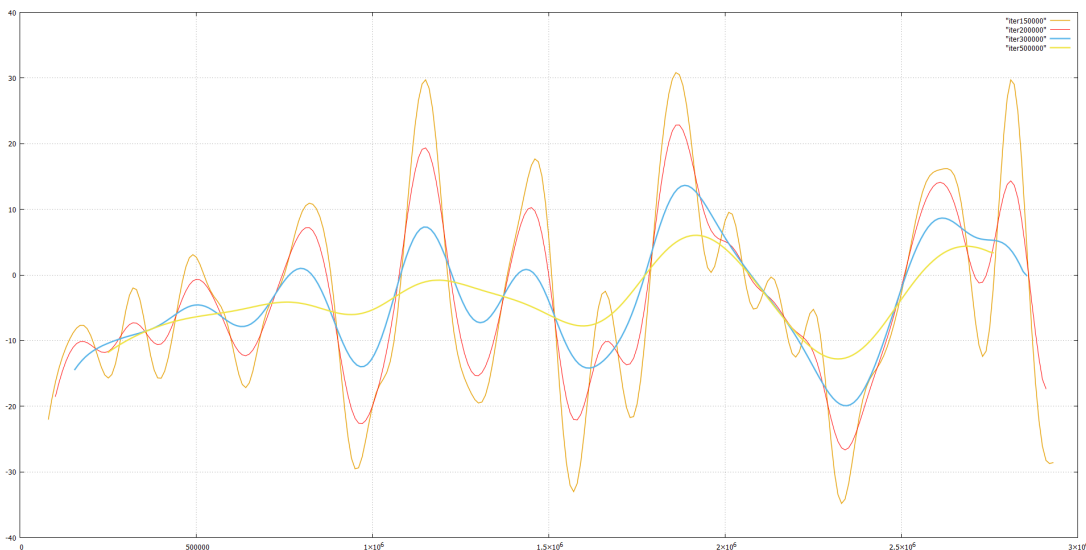
⁶⁰**N.B. (???) Did he?!**

last illustration above). They may be calculated by (repeated application of) taking averages of pairs of datapoints next to each other.⁶¹

This shows the Gauss' mismatches filtered as above with effective widths corresponding to taking 3, or 5, or 10, or 15 Gauss' bins of width 10,000:



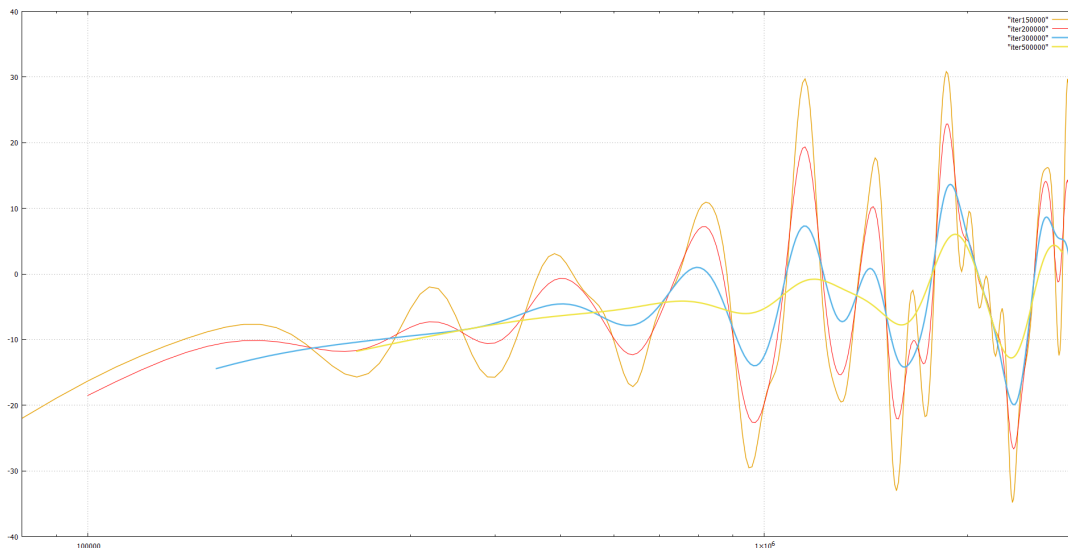
With these plots, the fluctuations are suppressed a lot—but still there is no clear evidence of simple corrections⁶² to Gauss' law on the density of prime numbers. However, the situation changes dramatically with yet stronger filtering (with widths of 15, 20, 30 or 50 Gauss' bins):



⁶¹N.B. (???) TBC

⁶²Except for undershooting up to $n \sim 700,000$ —but one could see it before too!

especially if one plots it in the log-scale of the x -axis:

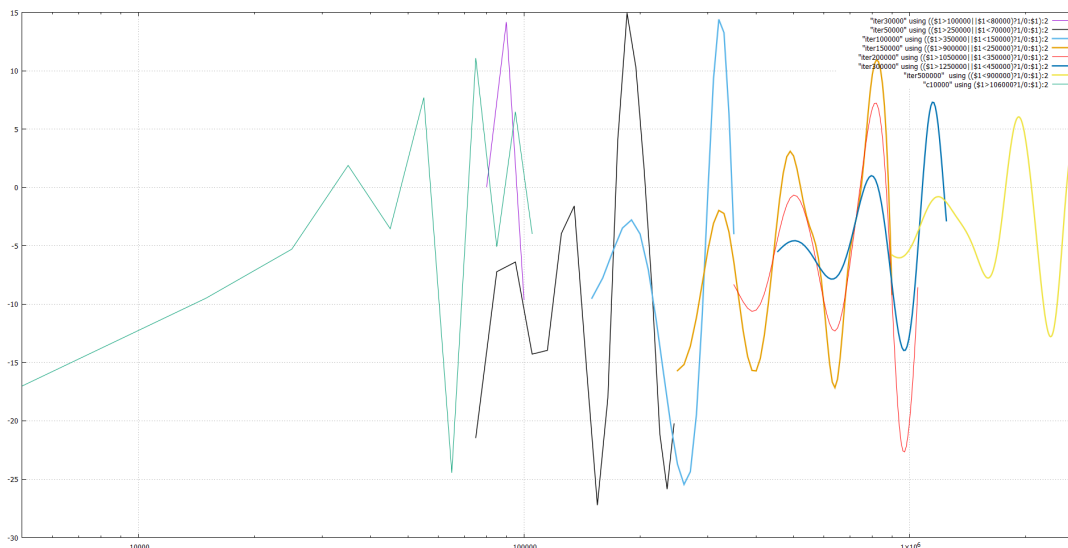


Here one can see a remarkable effect: there are clearly visible oscillation in the middle of the plot, and:

The oscillation in the middle have the same period as the oscillations on the right on the yellow graph.

Conclusion: using the effective width of 200 Gauss' bins dampens fluctuations well enough on the right of the plot — but in the middle it damps them too strong and makes these (interesting!) effects invisible.

This suggest using different “strengths of dumping oscillations” on different parts of the graph:



This combines the suitable parts of the graphs already shown above — but on the log-scale. For every value of x , we choose the graph with the smallest effective width on which this kind of oscillation is clearly visible. The widths are (these numbers are in units of Gauss' bins, 10,000):

Effective width	1	3	5	10	15	20	30	50
In the range of	0...10	8...10	7...25	15...35	25...90	35...105	45...125	90...300

Conclusion: To see this oscillation, the effective width of averaging should be $\sim 1/3$ of the current position.

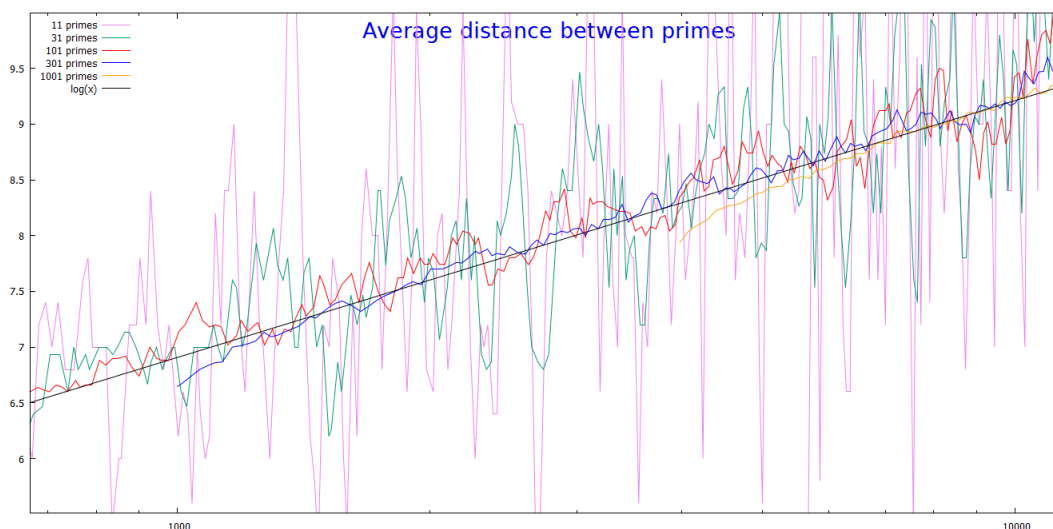
A pictorial exposure of the mismatch in Gauss' prediction

However, below we abandon the question:

Could all this be found *just by analysing the Gauss' tables?*

and switch to an alternative point of view. We assume that Gauss would investigate *exactly the same questions* of him as we discussed above—but that he had better tools of keeping the results that just a notebook. Moreover, we assume that he had “reasonable-for-today” computational resources and visualization tools.

In fact, one does not need to handle a lot⁶³ of prime numbers to get a nice confirmation of the law $\log n$ for the average distance between primes near n . For example, below we take a certain sequence of consecutive primes p_0, p_1, \dots, p_M and say that “near $x := (p_0 + p_M)/2$ the M -averaged distance between primes is $y := (p_M - p_0)/M$. We plot these distances (for different values of M (in the top-left corner) in the plot below. For this plot, one does not need to know primes above 25,000:



This and the following plots are made from the table `av-dist-primes` generated by the script `av-dist-primes.gp`. This table is visualized by

```
set title "Average distance between primes" offset 0,-3.8 font "bold,20" tc "blue"; set key left top
plot [*:1e10] "av-dist-primes" using 1:($2/10) w l lc "violet" title "11 primes", "" using 3:($4/30) w l title "31 primes", "" using 5:($6/101) w l title "101 primes", "" using 7:($8/301) w l title "301 primes", "" using 9:($10/1001) w l title "1001 primes", "" using 11:($12/10000) w l title "log(x)"
set title "Average distance between primes (relative to the trend log({/:Italic x}))" offset 0,-3.8 font "bold,20" tc "blue"; set key left top
corr(x,n) = log(x) # expected average gap between n+1 primes around x with the law gap=log(t)
plot [*:1e10] "av-dist-primes" using 1:($2/10/corr($1,10)) w l lc "violet" title "11 primes", "" using 3:($4/30/corr($3,30)) w l title "31 primes", "" using 5:($6/101/corr($5,101)) w l title "101 primes", "" using 7:($8/301/corr($7,301)) w l title "301 primes", "" using 9:($10/1001/corr($9,1001)) w l title "1001 primes", "" using 11:($12/10000/corr($11,10000)) w l title "log(x) trend"
```

For each M , the plot starts at the minimal possible value (corresponding to $p_0 = 2$). One can see that near its start, this curve has “a systematic dip”: is a bit below the line $y = \log x$ (which is a straight line due to the logarithmic scale of the x -axis)—but soon it gets really close to this line. Since smaller values of M correspond to averaging together a smaller number of distances between primes, it is not surprising that for these graphs the larger is M , the smaller are the fluctuations. (Since usually “averaging suppresses fluctuations”.)

However, the graphs with a very large M not start yet at a particular value of x —or it may start, but be in the range of that “systematic dip”. So if one wants to answer the question:

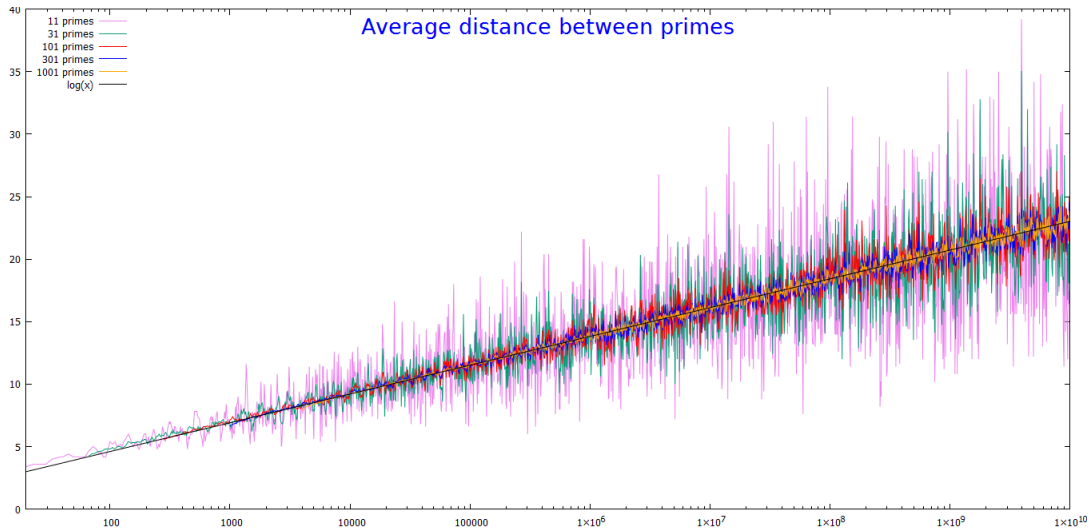
For a given value of x , which values of M would give the value of y with the “smallest error”?

⁶³On the other hand, making the plots below require *storing* the detailed info about *particular* prime numbers—in contrast to the approach of storing the “information about large groups of primes” which Gauss used.

one needs to balance the scale of fluctuations (larger for smaller M) with the scale of the “systematic dip”. Inspecting these plots, the preliminary approach would be:

Choose M such that p_M/p_0 is not too large.

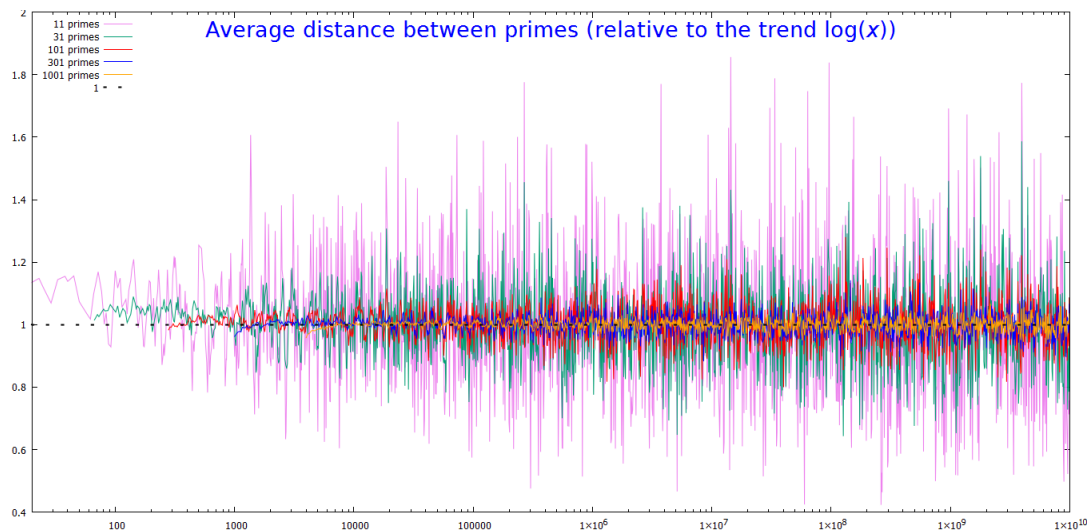
On this plot, the same graphs go for the whole range up to 10 billions:



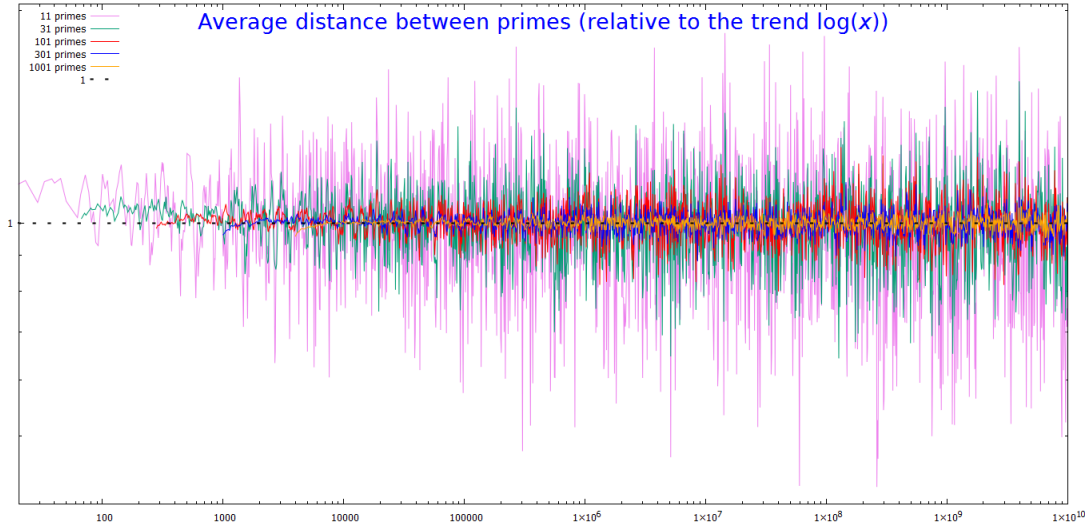
av-dist-primes

One can see that for smaller values of M , the initial dip is replaced by an “initial bump”.

This plot is as above, but we graph $Y := y/\log x$:



av-dist-primes-rel

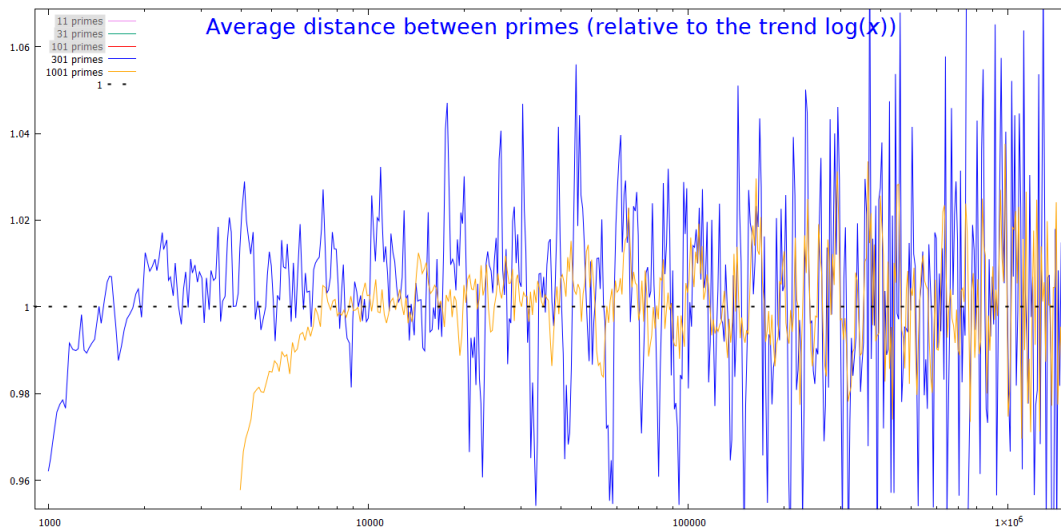


(version of the preceding picture with the vertical axis in log-scale).

On the other hand, the mismatches between the graphs on this plot and the line $Y = 1$ are not necessarily related to the errors in our conjectural law $\log n$ — there is an easier explanation! Indeed,⁶⁴ with this law the expected value of $p_M - p_9$ is *not* $M \log((p_M + p_9)/2)$, but the sum of the expected values of $p_1 - p_0, p_2 - p_1, \dots$ — which may be written as $\log((p_9 + p_1)/2) + \log((p_1 + p_2)/2) + \dots + \log((p_{M-1} + p_M)/2)$. And if $p_M/p_0 \gg 1$, then the latter may be noticeably different from $M \log((p_0 + p_M)/2)$.

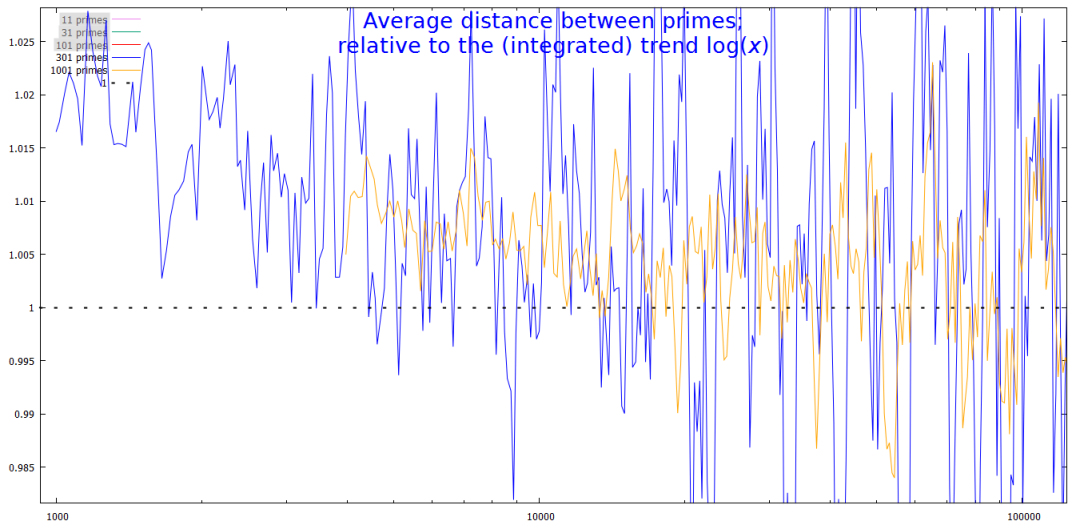
Conclusion: if the actual average distance between primes is $c(x) \log x$ (with c changing very slowly), then to find $c(x)$, we should plot not $(p_M - p_9)/(M \log((p_M + p_9)/2))$ (as we did above), but $M / \int_{p_0}^{p_M} dx / \log x$. (Indeed, then the expected value for M with this law is $\int_{p_0}^{p_M} dx / (c(x) \log x)$.)

To see the effects of this change, observe the relative position of two graphs (with $M = 300$ and $M = 1000$) before the change



⁶⁴N.B. (???) Compare with the code above.

and after the change:

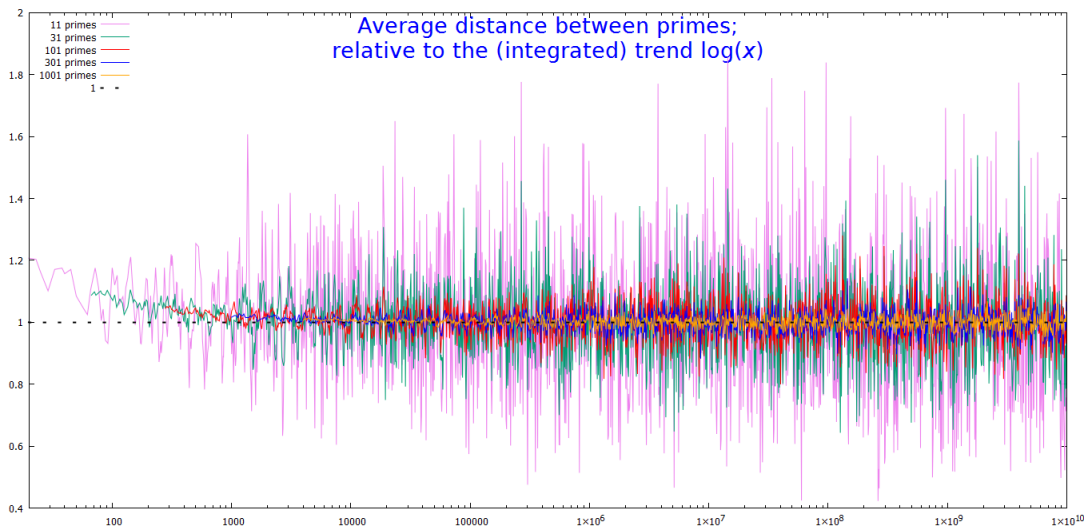


Before the change, there was a good match between the graphs—except for the beginning of the yellow graph. After the change the yellow graph passes “practically in ‘the middle’ of the fluctuations” of the blue graph.

In other words: the graphs with a larger M behaves very similar to “the graphs for a smaller M with dumped down fluctuations”. This means that if we want to write the average distance between primes as the trend $\cdot \log x + \text{fluctuations}$, then to approximate the value of this “trend”⁶⁵ for a particular value of x we are free to use the largest M which makes sense for this x .

In yet other words: for the purpose of finding the trend, choosing a larger M averages the trend over a longer interval (what makes the information about how “the trend” behaves less precise), while choosing a smaller M decreases the “degree of dumping down of fluctuations”—which “leaks” the fluctuations into the information about trend. These opposite tendencies of corruptions of information should be carefully matched against each other.

Here is how all the plots with different values of M look with the new approach:



One can clearly see that for smaller primes, the average distance between primes is significantly higher than what Gauss predicted.

⁶⁵possibly averaged over a certain interval

Conclusion: Write the “actual expected value” of the average distance between primes as $c(x) \log x$; then looking only at the leftmost pieces of the graphs in the picture above (the pieces “before the next graph starts”: first the violet piece, then the green, then the red, the blue, the yellow) exposes a result of a certain averaging of the function $c(x)$. This result is also partially contaminated by the “leaked fluctuations”.

Remark 3: In the rest of this chapter we inspect how “better calculation facilities” could help Gauss to “catch up with the progress in 19th century”. — And we restrict our attention to the analogues of *what Gauss already did* — without introducing any new ideas (like Riemann using complex analysis in his approach)!

Well, there is one possible exception: below we presume that it is clear that using weighted averages helps with dumping down fluctuations — and the smoother the kernel is, the better is the suppression of quickly changing fluctuations. (Note how this is similar to what we want the reader to understand⁶⁶ about the Fourier transform.) I do not know whether such arguments could have been automatic for Gauss. (He *did* have a lot of experience in processing of fluctuating time-dependent data.)

Averaging on longer intervals

On the plots above, we limited our averaging to 1,000 consecutive intervals between primes. The plot above shows that to keep fluctuations small in the right part of the plot, one should average *much more* intervals.

The plots above show that what one expects to get from using more consecutive intervals is:

- Increasing the total width of the combined intervals should dump down the fluctuations more, so allows to estimate the correcting function $c(x)$ above with higher precision.
- However, the plot would only give us the information about *a certain weighted average* of $c(x)$ on the combined interval. And for very wide intervals, the value of $c(x)$ on the left end is going to be significantly higher than at the middle x of the interval. Hence the average distance between primes gives information about averaged c , which (due to the concavity down of “the trend” of the plots above) is going to be higher than $c(x)$.

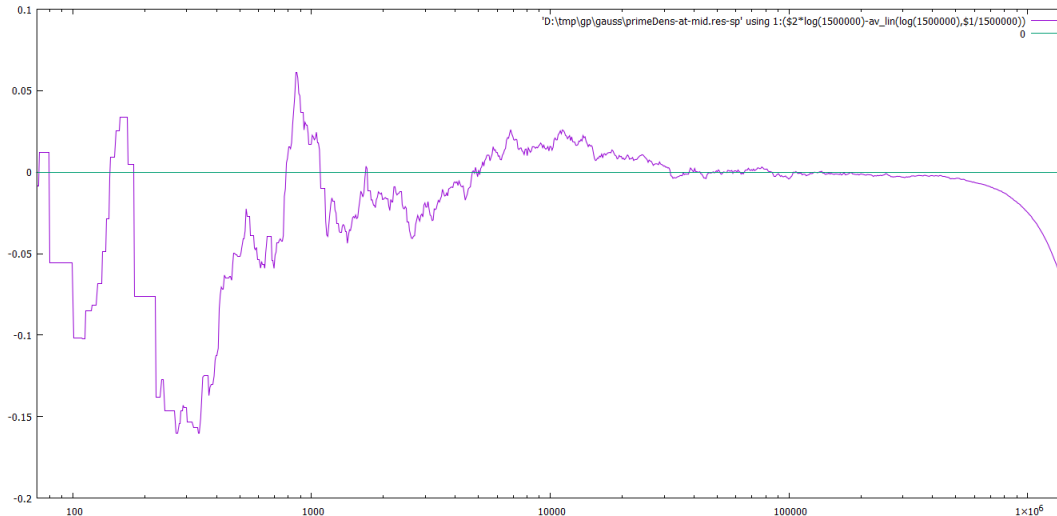
So we should expect that as fluctuations go down, a systematic error of “overshoot of c ” would appear. — However, doing plots as above with much more combined intervals between primes may require a lot of computational resources.

On the other hand, it is not hard to inspect what happens *at one particular point x* of the plots above when we change this count of consecutive intervals in wide extents. If we limit ourselves to the same bound $p < 3,000,000$ as Gauss, then for the midpoint $x \approx 1,500,000$ we can go up to the combined interval $[2, 2,999,999]$ of primes. This plot shows the relative mismatch between the actual average density⁶⁷ of primes on an interval and the Gauss’ prediction (in dependence on the $\frac{1}{2}$ -width

⁶⁶N.B. (???) Ref!

⁶⁷N.B. (???) Explain switch from “the average distance”, and how we deal with non-centered at 1.500,000 intervals.

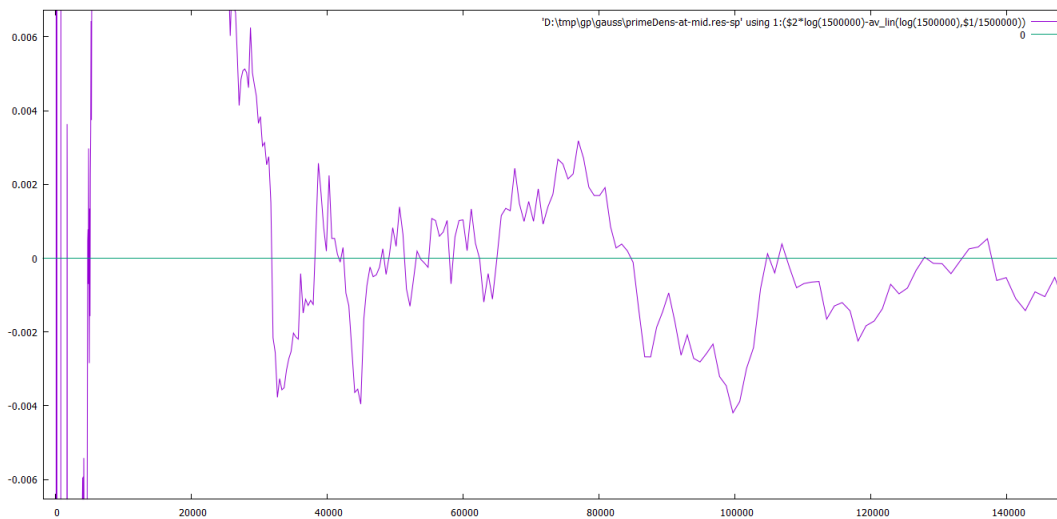
of the interval):



(The left end corresponds to the violet plot on the preceding graph, with 10 combined intervals.)

Indeed, as we increase the width, the fluctuations go down, and a “systematic drift” in the predicted direction appears. It is clear that the best balance between these two types of errors happens when the combined $\frac{1}{2}$ -width is 100,000–200,000.

On the third hand, zooming into non-logarithmic version of the graph shows this:



Conclusion: if *we average* the plot above in the range near 30,000–60,000, then the floctuations of the graph have a good chance to cancel each other — and the systematic error is quite small in this interval! In other words: using *several different “combined widths”*, then averaging the result may give us much higher precision of estimating $c(x)$.

Average distance vs. average density

It may be easier to think about average distance between primes, but when we start to discuss the approaches as at the end of the preceding section, it turns out that it is much easier to work with the inverse⁶⁸ value: “the average density of primes” in our interval.

⁶⁸More precisely, the *harmonic mean* of a value is inverse to the arithmetic mean of its inverse. Above, we used a “combined length” of the intervals between several primes — which corresponds to the *arithmetic* mean. Hence if we want to express *its inverse*, the density, in “local terms”, we can describe as the *harmonic* mean of inverse distances between consecutive primes.⁶⁹

The average distance between primes in an interval $[A, B]$ is similar to $(B - A)/N$; here if both ends are primes, then N is the count of primes in the interval—when one counts primes at ends with the coefficient $\frac{1}{2}$. So the inverse value is $D_{AB} := N/(B - A)$, and every prime number inside $[A, B]$ contributes $1/(B - A)$ into D_{AB} (with the primes at ends contributing $1/(2(B - A))$).

The convenience of the “average density” is that when one takes an average of several values D_{AB} with different A and B , this corresponds to considering every prime p with the coefficient $\mu(p)$; here μ is a certain function⁷⁰ (with compact support). (In fact, μ is the average of the corresponding functions $\mu_{A,B}$ we described above.)

For example, if we take a continuous family of intervals $[x_0 - t, x_0 + t]$ centered at x_0 , with t “distributed uniformly” on the interval $[0, T]$, then $\mu(x) = (T - |x - x_0|)/T^2$ inside $[x_0 - T, x_0 + T]$ and is 0 outside.

When instead of one function $\mu(x)$ we use several such functions in the form $\mu_{x_0}(x) = \mu_0(x - x_0)$ with different values of x_0 , then what we consider above becomes a function of x_0 which is a *convolution*⁷¹ with μ_0 . In other words, we consider the sequence of prime numbers as “an input signal”, and run “a low-pass filter”⁷² on this signal. The aim is to separate this “signal” into the (slowly-changing) “trend” $c(x)$ and into (high-frequencies) “fluctuations” (which are going to be suppressed by the filtering). As it is well-known,⁷³ one of the best filters is one based on the Gauss’ invention: the Gaussian function.

Generally speaking, the more smooth is the function μ , the better is the suppression of high frequencies in the corresponding filter. The important property of Gaussian filters is that they are in a certain sense “optimal” in this regard. So below we are going to use μ which is a Gaussian.

Prime density: averaging with Gaussian weighting

The “`lin`” flavors of plots below are generated by the script `prime-density-peacemeal001-extra.gp`. It counts the total weight of primes in a certain window $[(1 - \Delta)x, (1 + \Delta)x]$ with the Gaussian weight⁷⁴ centered at x with $\frac{1}{2}$ -width $w \cdot x$ (and $\Delta = W \cdot w$ with $W = 13$). Here the numbers x go in a geometric progression up to⁷⁵ about $2.66 \cdot 10^9$.

Remark 4: In fact, for every x , three “extra” numbers are emitted. These are not “weighted sums”, but “weighted *signed* sums”⁷⁷ with the sign at p depending on the $p \bmod 12$ (using 3 non-trivial characters mod 12).

⁷⁰For one interval $[A, B]$, the function μ takes value $1/(B - A)$ inside the interval, 0 outside, and $\frac{1}{2}$ at its ends.

⁷¹**N.B. (???) Explain!**

⁷²**N.B. (???) Explain!**

⁷³Although I do know whether Gauss understood this...



⁷⁴**N.B. (???) Did Gauss know about the usefulness of weighting with Gaussian kernel to suppress fluctuations?**

⁷⁵It is chosen to cover the the pair 71, 73 of primes after multiplication by the important constant⁷⁶ $34.2 \cdot 10^6$.

⁷⁶**N.B. (???) Explain!**

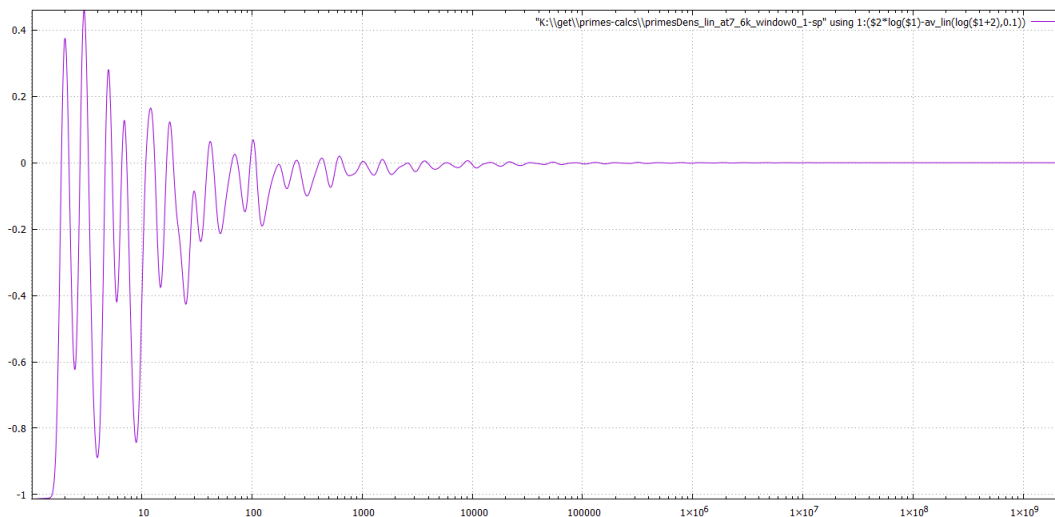
⁷⁷Such sums may be used in discussion of simpler topics: “prime races”. In these contexts, the “tricky” terms we discuss here (dealing with a slowly changing density $1/\log x$ and the resulting integral li) cancel out completely, and the terms we start to introduce in this section are “immediately visible”.

These topics are related not to the “literal” Riemann Hypothesis, but to the Generalized Riemann Hypothesis. The latter one is a bit harder to formulate, but it is much easier *to illustrate*.⁷⁸

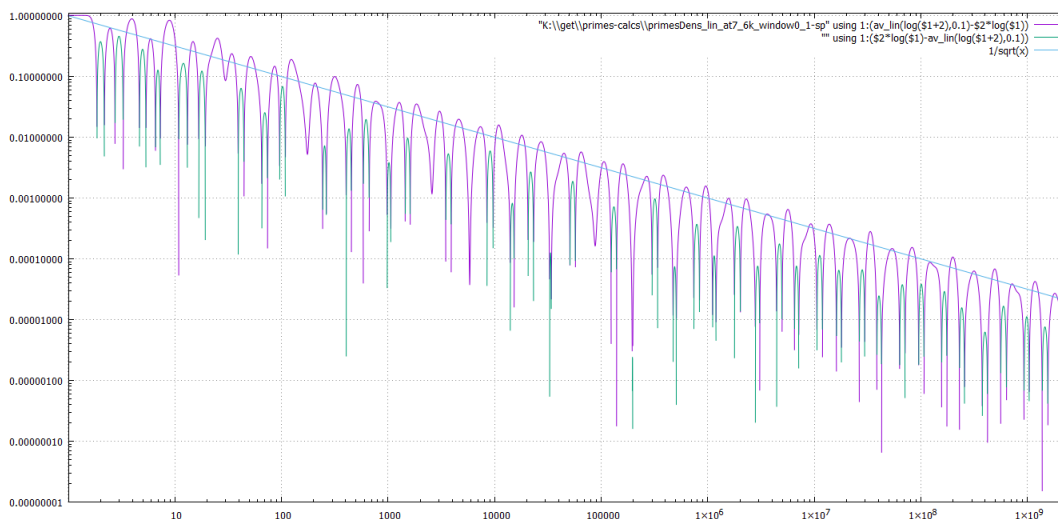
 For example, when we discussed prime races on our Math Circles, a couple of first-graders were able *to invent* the Generalized Riemann Hypothesis basing on a small amount of data we show them.  (However, we do not discuss “prime races” here!

⁷⁸**N.B. (???) Move parts of this to the body of the remark!**

This plot shows the relative error of the actual Gaussian-averaged density of primes near x comparing to the Gauss' prediction⁷⁹ $1/\log x$. Here the Gaussian has $1/2$ -width $w = 0.1$ (i.e., “the effective width” of our weighted average is⁸⁰ $1/5x$):



Already with this graph one can conjecture that this discrepancy decays at the approximately power law. This conjecture becomes much more convincing in the log-log-plot. For this, we plot the difference and the flipped-sign difference, and switch the y -axis to the log-scale:



Here we also plot $y = 1/\sqrt{x}$ to demonstrate that the match with the law $y = C/\sqrt{x}$ is reasonably good.

Conclusion: this method of averaging shows many effects:

- On average, there is a steady improvement of the precision of Gauss' prediction $1/\log x$.
- On average, the Gauss' prediction significantly overestimates⁸¹ the density of primes for small x .
- This over-estimation corresponds to an extra factor $1 - C/\sqrt{x}$ with $C \lesssim 1$.

⁷⁹To avoid systematic error, the Gauss' prediction should be also Gaussian-averaged. On the right of the plot, this decreases the Gauss prediction by $1/4,000$ — which would be very visible on the plots starting from x about $6 \cdot 10^6$.

⁸⁰This is somewhat similar to averaging using primes in the interval $[0.9x, 1.1x]$, but with a much stronger dumping down of fluctuations due to gradual decay of the Gaussian weight.

⁸¹This corresponds to an underestimation of the actual average distance between primes we have seen on the plot on p.22.

- When the fluctuations “on top of” that trend are dumped down by our method of averaging, the suppressed remains are very similar (on the right of the picture!) to periodic oscillations *on the log-scale* with ~ 5 periods when x grows 10 times.
- The left part of the picture shows that on top of these “log-periodic remains” there are “extra correcting” terms. — But these terms decay fast, so in the right half their contributions are hardly visible relative to the oscillations.

To find the value of C more precisely, and to inspect the last part of the conclusion, we need to suppress the fluctuations yet more. For this, we need to increase the $\frac{1}{2}$ -width w of averaging (which is 0.1 for the plot above). However, there are some difficulties for such an increase — some of them mathematical, some technical:

- “Mathematically”, the Gaussian weight μ has an infinite width (“support”). Practically, it decays so quick (“its tails are so narrow”), that usually the result of averaging does not change noticeably if one “cuts off the tails”. For example, cutting at $W \cdot w$ with $W = 6$ changes the total weight only by 2 billionth. With $W = 13$ (used above) the error is only about one unit in the 38th decimal place.⁸² (With $W = 10$, it is the 23rd decimal place; with $W = 4$ it is the 4th.)
- Using W such that $W \cdot w > 1$ includes negative values (and small positive values) into the range of averaging. In particular, we need to answer a tricky question: is -17 a prime?! From many points of view, it is useful to consider them as primes — so, theoretically, we should better check how the plots look like with both possible answers to this question.

Using $W = 10$ with $w = 0.1$ allows one to completely ignore this question (in particular, the difference is not going to be visible on the plots above). However, with $w = 0.25$ this may completely change how our plots behave.

- In addition to averaging the density of primes, to make plots as above, we need to average the Gauss’ prediction (as in [Footnote 79 on p. 26](#)). However, this prediction is not good near $x = 1$ — so we better avoid this region. — And this is not possible with $w \gtrsim 0.25$.
- Increasing w requires a complete recalculation of tables used for our plotting. The required time is proportional to w . Already with $w = 0.1$ the time of calculation is very significant.

The first correction to prime density: averaging with log-Gaussian weighting

There is an obvious solution to the problems listed above: “place” the prime number p not into the point $x = p$, but into $X = \log p$. In other words, given a “weight function μ ”, when we process a coordinate x_0 on the plot, count p not with the weight $\mu(p - x_0)$, but with the weight $\mu(\log p - \log x_0)$.

Among other benefits, this reduces the dependence on x_0 of the $\frac{1}{2}$ -width we use: on the plot above, it is $0.1x_0$. — And since 0.1 is “small enough”, the corresponding Gaussian weight corresponds closely to the log-flavor⁸³ (as above) with the width 0.1 — which does not depend⁸⁴ on x_0 .

⁸²We use this value, since it is the precision of our calculations anyway!

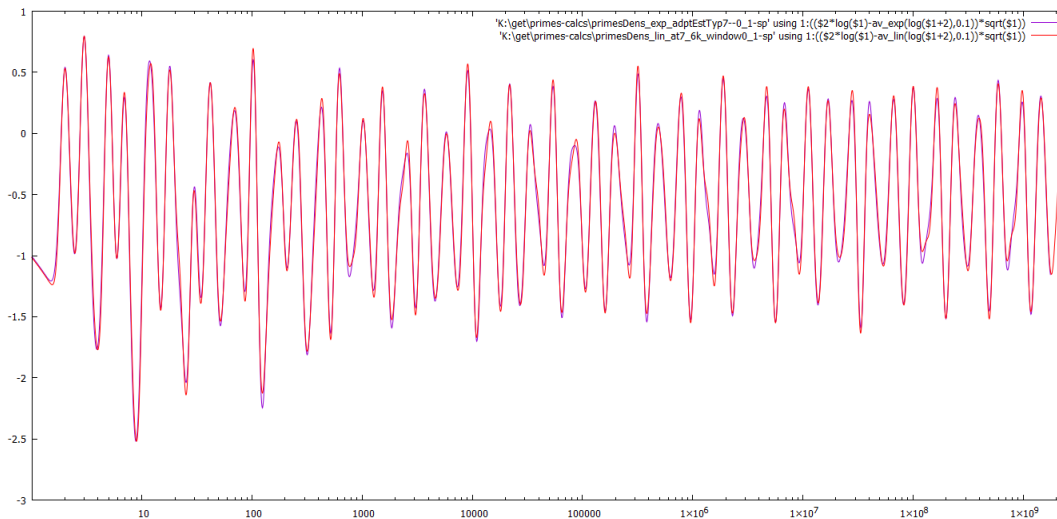
⁸³The “tails” of these distributions decay at very different rates. — However, this difference does not start to matter due to the tails being both very small where they differ a lot.

⁸⁴One very convenient corollary of this that given a (dense enough) table of averaging with constant $w = 0.1$, it is very easy to recalculate it into a table for larger w s. (Compare with [Footnote 99 on p. 35](#).)

The only — unfortunate — consequence is that the recalculation can be performed only on a more narrow interval. Practically, this results in the plots below “ending sooner” for larger values of w .

(Moreover, the precision of these plots decreases when going closer to the “right end” of these plots. This decrease is super-exponential in $\log x$ — but close to exponential. Where this is important, we may mark the zone where — the expected — precision decreases 5 times using dots.)

This illustrates the insignificance of passing to the log-flavor for small w :



The visible differences between these graphs are mainly due to the corresponding tables being calculated on slightly different grids. (As well as us not using “smoothed plotting” — the plots show the linear interpolation.)

The “exp” flavors of plots below are generated likewise (by something like `prime-density-peacemeal001-exp` but instead of Gaussian kernels we use log-Gaussian kernels. In other words, by a script like `prime-density-peacemeal001-exp.gp` (followed by “coarsification” step). It counts the total weight of logarithms of primes in a certain window $[(1 - \Delta) \log x, (1 + \Delta) \log x]$ with the Gaussian weight centered at $\log x$ with $\frac{1}{2}$ -width w (and $\Delta = W \cdot w$ with $W = 13$). Here the numbers x again go in a geometric progression up to the same value as above.

“Coarsification” is^{85 86}

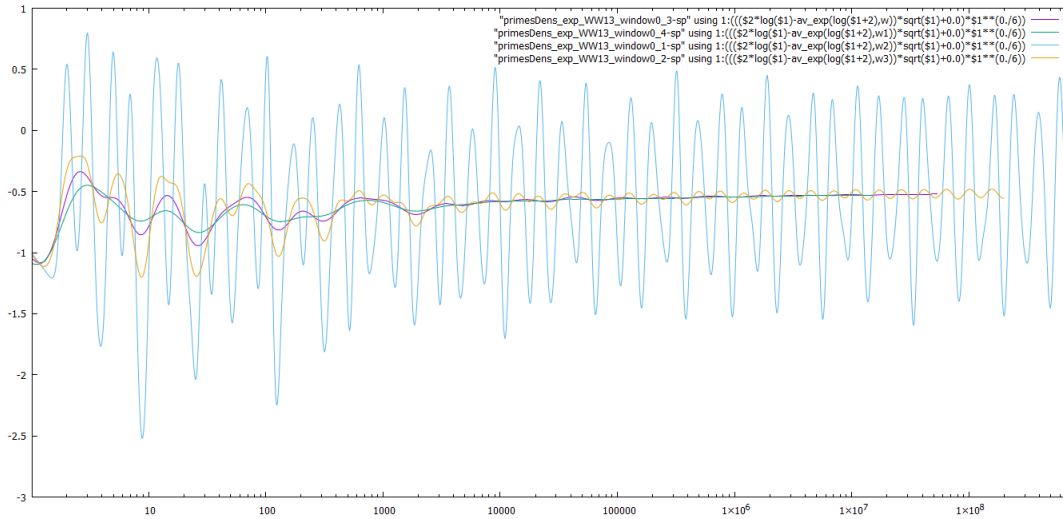
⁸⁵N.B. (???) TBC

⁸⁶N.B. (???) The coarsified tables contain extra columns: these give the Gaussian-averaged function $1/x$ (with an appropriate cut-off, and multiplied by t_n), with centers at the points $t_n := \log x + w^2/n$ for $n = 1, 2, 3, 5, 6, 7$. During plotting, we calculate $T \text{Av}_w(e^{ct}/t)|_T$ as $T/T + cw^2 \cdot e^{cw^2/2}((T + cw^2) \text{Av}_w(1/t)|_{T+cw^2})$.

Although a particular nature of cut-off does not matter for most of our plots (unless w is very large, it affects only the extreme lefts of our plots — which we ignore anyway) we use the Ramanujan’s cutoff (of 2.1.4 of “Twelve lectures” — with a misprint correction; this is (the Soldner constant; when calculating li , this cutoff is equivalent to a cut-off at 0 if one uses the principal-value regularization).

In the plots, we also clump the argument T to the Ramanujan’s cutoff if T is smaller than it.

The plot below uses log-Gaussian averaging with widths $w = 0.1, 0.2, 0.3$ and 0.4 to estimate the value C mentioned above:



(Recall that increasing w dumps down the high-frequencies in fluctuations stronger, and moving to the right decreases the role of the terms decaying much quicker than $1/\sqrt{x}$.)

Conclusion:

- All the effects discussed after the plots of the previous section are confirmed.
- The magnitude of the “periodic oscillation” of this discussion decays at very similar rate to the “correcting term” $-C/\sqrt{x}$.
- The constant C seem to be very close to $1/2$.
- Using such a constant $C = 1/2$ *still* leads to an overestimation of the density for small values of x . (This corresponds to the “extra correcting” terms of the discussion of the preceding section.)
- Inspecting the very far right of the yellow plot (for $w = 0.2$) shows *very strong* similitude to a (shifted) sinusoid. Such high w dumps down high (log-scale) frequencies very strongly — but still, this suggests that “the fluctuations” have only “a discrete log-scale spectrum” — at least for the range of “small frequencies”.

This gives a very strong suggestion of the presence of “spectral gaps” in the frequency spectrum⁸⁷ of fluctuations in the prime density (on the log-scale)! (Since this conclusion is related to the far right of the plot, this means that we ignore the more-rapidly “extra correcting terms” — so we consider only the “main terms” of the fluctuations.)

This suggests to write the prime density in the new form:

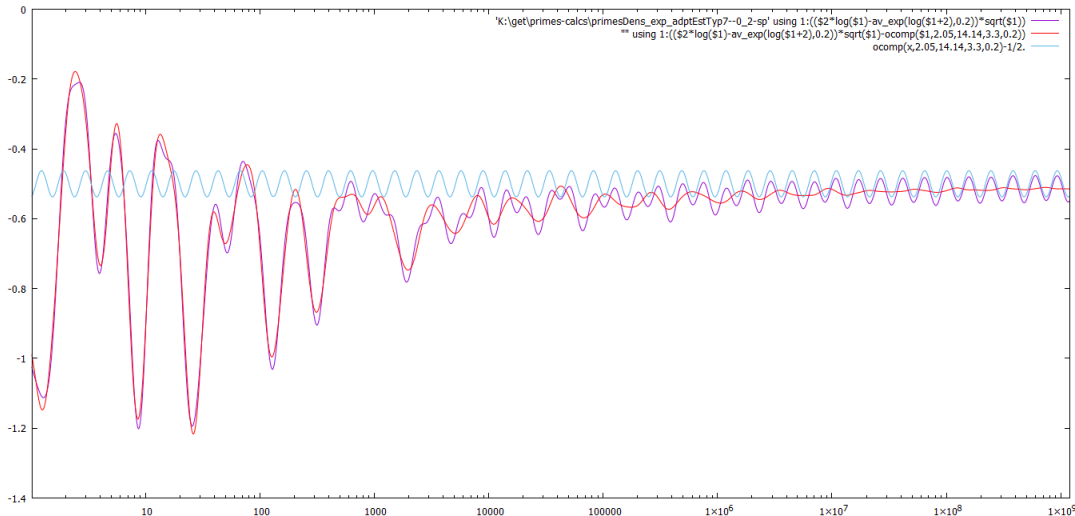
$$\boxed{\text{The average prime density near } x \text{ is } (1 - 1/2\sqrt{x} + \boxed{\text{Extra-terms}} + \boxed{\text{Fluctuations}}) / \log x.}$$

— and “extra terms” are small comparing to $1/\sqrt{x}$ (and — hopefully — decay quicker than this), while the fluctuations decay similarly to $1/\sqrt{x}$. (The plot above suggests that the “extra terms” give a negative contribution for small x .)

⁸⁷... i.e., in the Fourier transform in the log-scale. — And this is the same as Mellin transform.

“Beats” in fluctuations

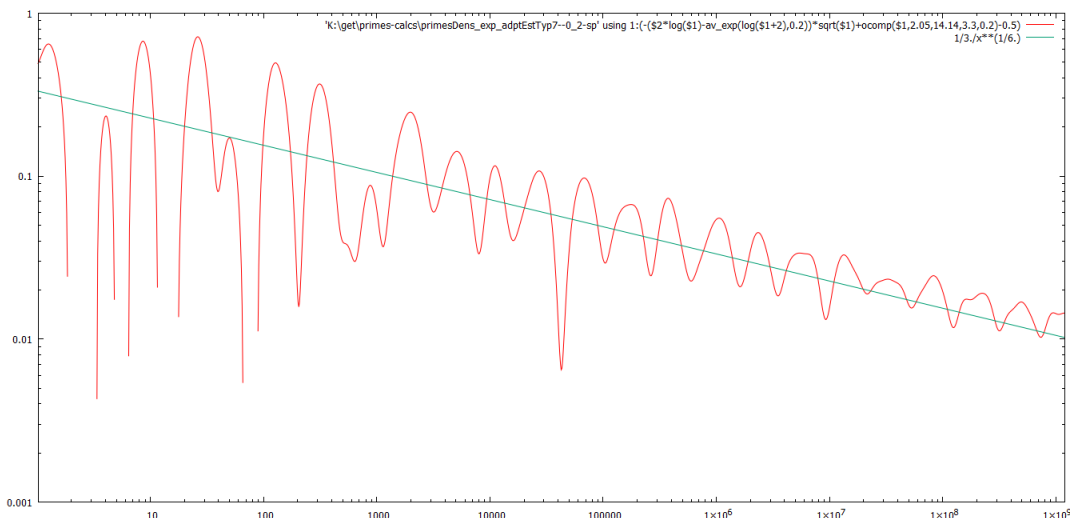
It is not very hard to fit a (cyan) sinusoid to the (violet) graph for $w = 0.2$ above⁸⁸:



Inspecting the difference between this graph and the fitted sinusoid (in red), one can see that “the remaining part” of the (suppressed) fluctuations is:

- Indeed much smaller than the sinusoid.
- Its main part on the right seems to be a (quickly decaying) oscillation *with twice the period* of the cyan sinusoid.

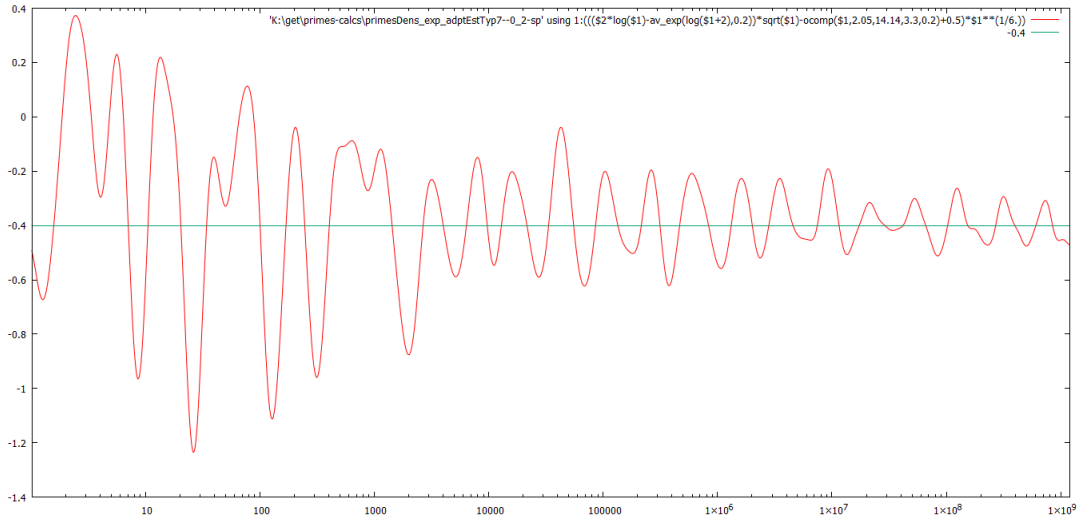
Moreover, using the log-scale on the vertical axis shows the asymptotic behaviour of the red graph above (with inverted sign):



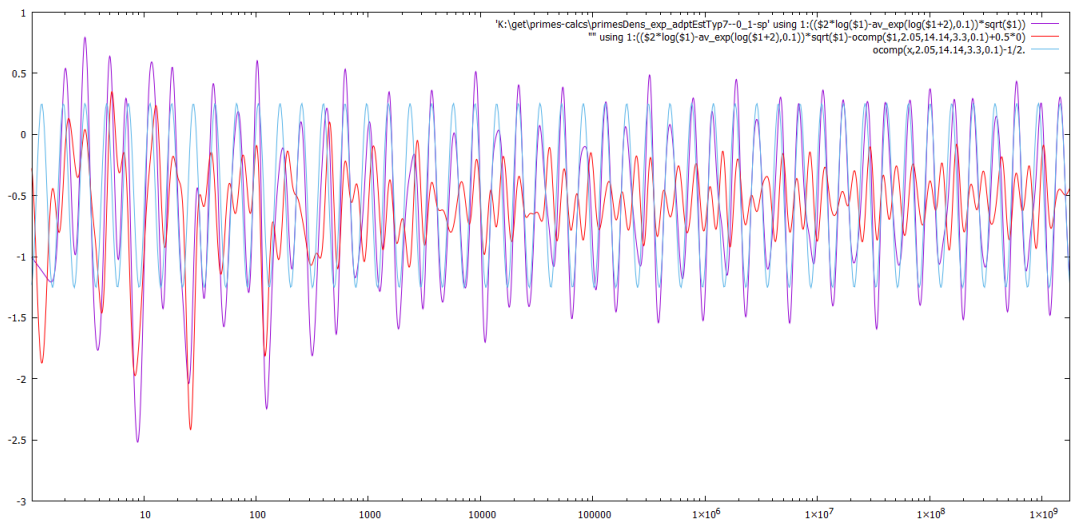
So “the remaining part” behaves as $\widehat{C}_3(\log x)/x^{1/6}$ with C_3 oscillating about $\sim -1/3$.

⁸⁸Here $\text{ocomp}(x, M, f, \varphi, w)$ gives the co-sinusoid with magnitude M , frequency f and phase φ suppressed by Gaussian averaging with $1/2$ -width w , at $\log x$.

Rescaling the preceding plot correspondingly, we get:⁸⁹



One can also subtract the fitted sinusoid (one we found above, just with a different coefficient due to the difference in w) from the density averaged with $w = 0.1$:



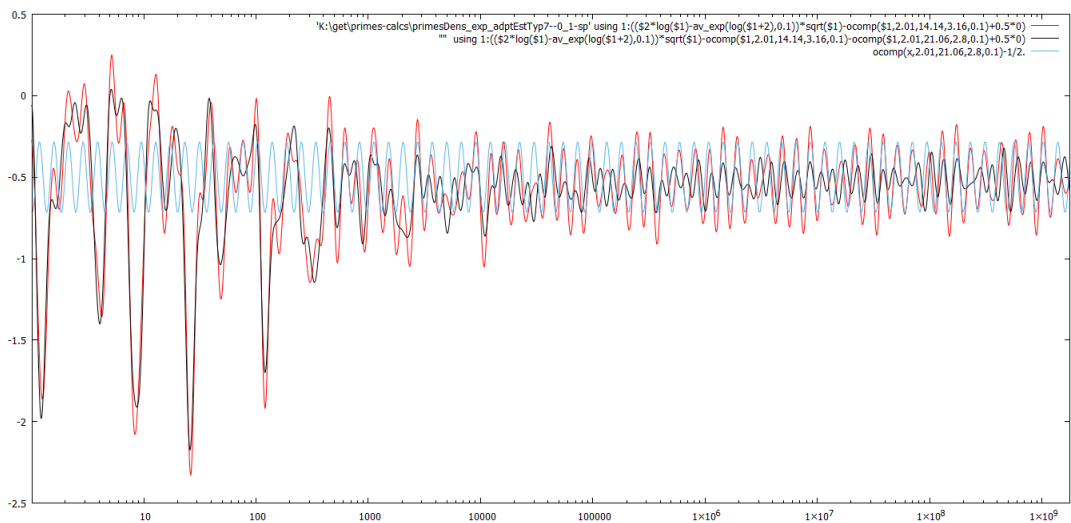
Here we see the major difference with the case of $w = 0.2$: the remainder does not decay. It seems to have a “very strong beat” with a higher frequency that the term we subtracted;⁹⁰ the ratio of frequencies seems to be close to $3 : 2$ (but with a hint of it being a tiny bit less: the mismatch in the phase on the right half of the picture suggests a correction of $\sim 0.7\%$, giving a frequency of ≈ 21.06). **Conclusion:** the “fluctuations” term contains strong signals with log-scale frequencies $f_1 \approx 14.14$ and $f_2 \approx 21.06$, and with another spectral gap (of unknown width) after f_2 .

To analyse the latter spectral gap, one may want to repeat what we did above: find a sinusoid which fits the “beat” of the red graph above; subtract it from the function we plot; and try to check whether the difference “has a strong beat” on a higher frequency. However, this is not what happens

⁸⁹**N.B. (???) Why av_exp2t does not work here?!**

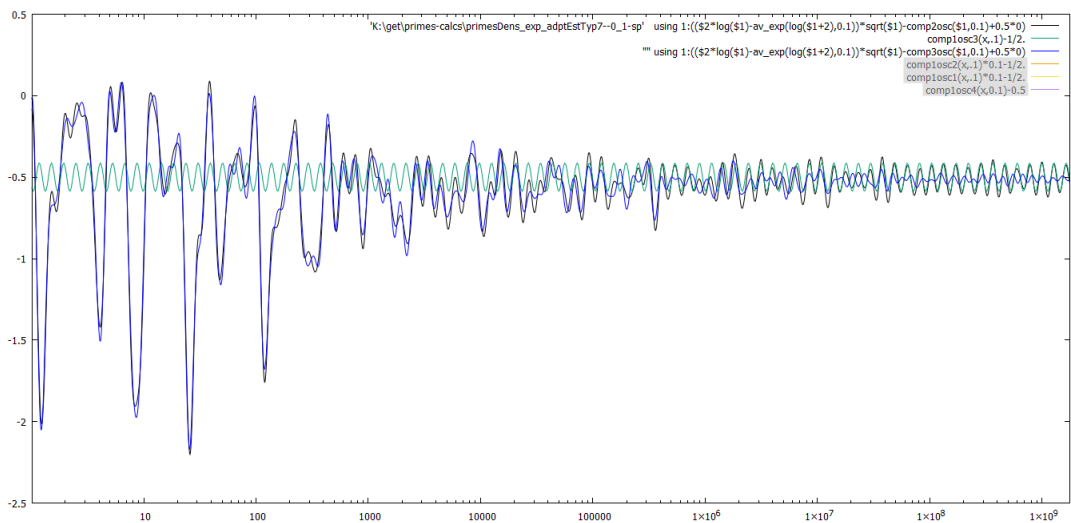
⁹⁰Hence with the $w = 0.2$ averaging, these oscillating terms are going to be suppressed *significantly* stronger. With (the frequency $3/2 \cdot 14.14$, the suppression is going to be ~ 150 times stronger than for the frequency 14.14 .) This is why they practically do not contribute into the plots with $w = 0.2$ above.

on the graph with⁹¹ $w = 0.1$:⁹²



The difference (in black) does not seem to have a component at a certain frequency “which dominates it”.

Fix: Tuning the phases of these two sinusoids, the (black) remainder acquires a strong visible beat frequency. On the plot below, we show a sinusoid fitting this (new) black plot, and the result of subtracting this sinusoid (in blue):⁹³



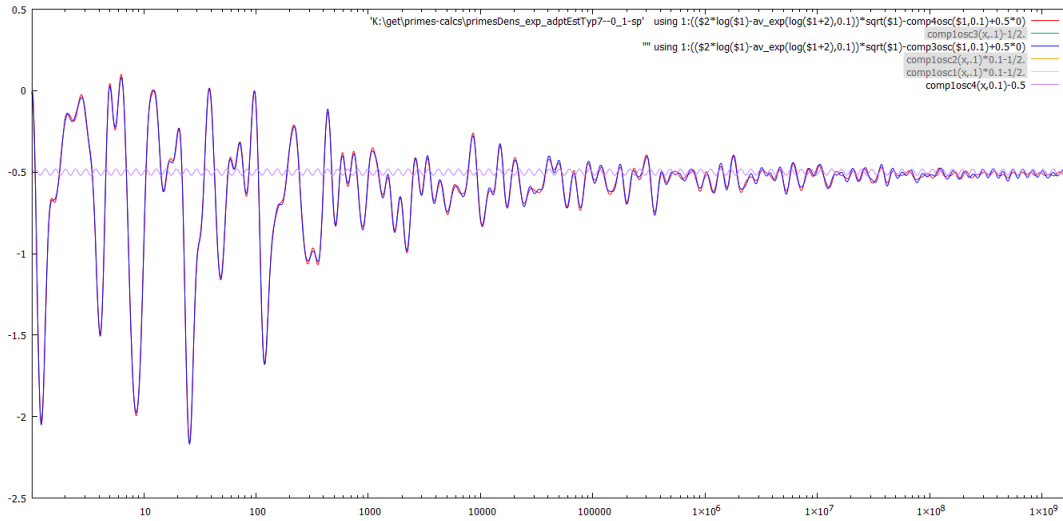
With these 3 main oscillating terms, the amplitude of “the remaining fluctuations of averaging with $w = 0.1$ ” decreases from about 0.8 to about 0.03 (in the right part of the graph).

⁹¹N.B. (???) Do we need to plot with $w = 0.1$ instead? But it seems one needs slightly different amplitude and phase: `ocomp($1,2.01,14.14,3.16,0.15)`?

⁹²N.B. (???) Here the amplitude and phase are corrected as above; this improves the “beat”—but it is on a smaller frequency?! Or maybe between 100,000 and 1,000,000 it is at ratio 6:5?

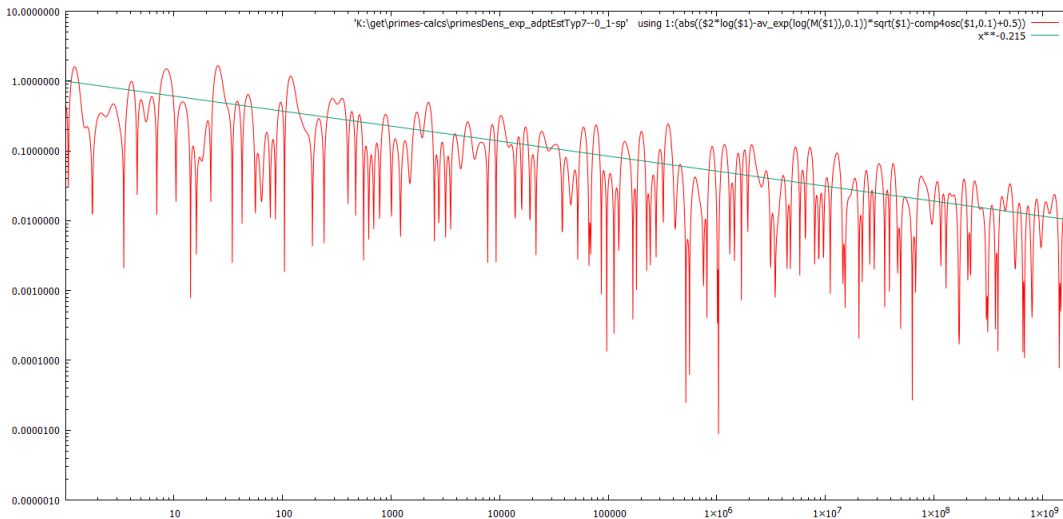
⁹³Here we use the magnitudes $M = 2$, and phases listed in T_EX comments above.

Adding a fourth term likewise (the purple sinusoid) reduces the amplitude of the remainder (in red) to about 0.011 (on the extreme right):



Again, zooming in on the right one can see that “the beat” of the red remainder seems to be slightly higher than the bit of the purple sinusoid.

However, even on the extreme right, it seems that “the remainder” after we remove 4 sinusoidal waves (drawn in red above) is mostly contributed by the fluctuations which decay quickly as we move to the right on the plot:



This shows⁹⁴ the variation of the red graph of the preceding picture from $-\frac{1}{2}$ in the log-log-scale, together with a similar power law.

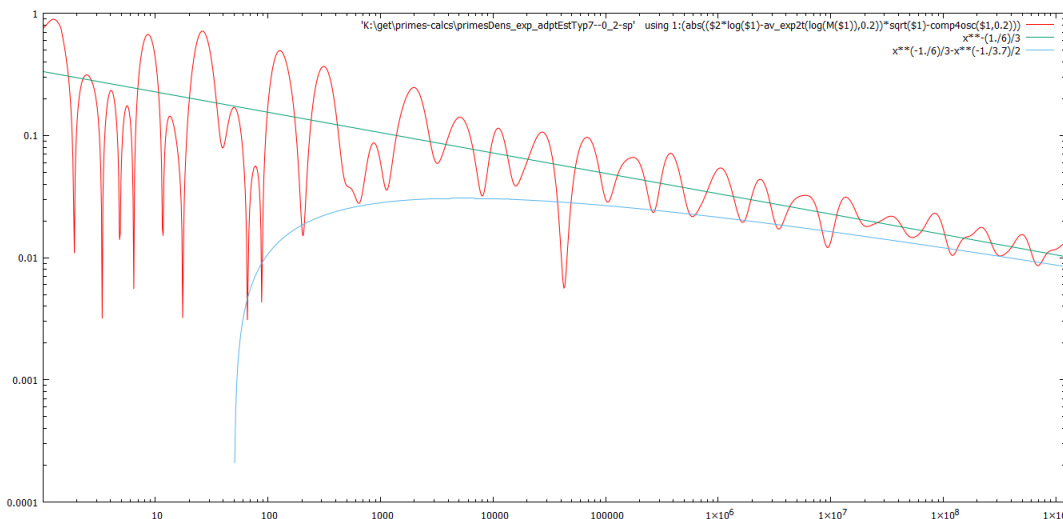
Conclusion: with our limit for the maximal primes we consider one cannot expect to get further with this method, and recognize more sinusoidal terms in the fluctuations we observe on these graphs. They are hidden by the decaying terms (with slower “beats”, so that the “next term” is suppressed by the Gaussian filtering much more than “moving to the right on our plots” decreases the decaying terms).

⁹⁴N.B. (???) Explain changing $\log(\$1+2)$ by $\log(M(\$1))$.

A wider view: wider log-Gaussian weighting

The plot above shows “the part of the error of Gauss’ approximation to the prime density near x which decays quicker than $1/\sqrt{x}$ ” — but with it, one cannot uncover the main features of this part: they are hidden by the quickly-varying parts of fluctuations. To expose these features, we need to suppress the high-frequency fluctuations further. So we need to use a larger w .

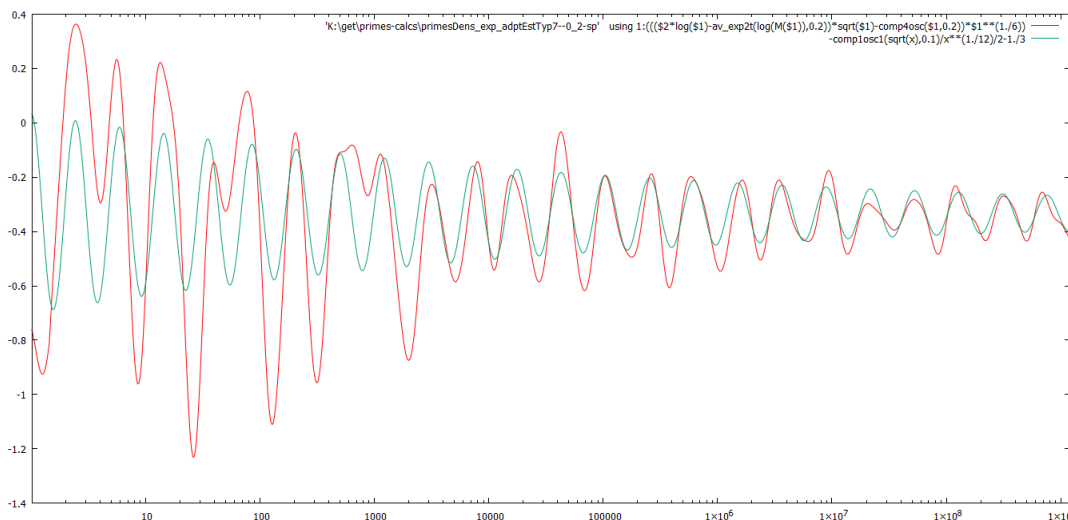
With $w = 0.2$ the plot above becomes: ⁹⁵



Conclusion: this part of the error is in many ways similar to the “part of the error decaying similar to $1/\sqrt{x}$ ”:

- It has a well-pronounced trend of the form $C_2/x^{2/3}$, with $C_3 \sim 1/3$.
- On top of this trend, the (averaged) fluctuations decay at approximately a power law.
- These averaged fluctuations have a very strong beat at a frequency on the log-scale about “2.5 periods per ‘an order of magnitude’”.

To get a better estimate of C_3 , we rescale the plot so that the vertical axis has the meaning of C_3 :



⁹⁵N.B. (???) Here we needed to switch to using `av_exp2t` to correctly average $1/2\sqrt{x}$.

and also plot *the (averaged) main oscillating term of the density of primes* (appropriately rescaled and with opposite sign) “slowed down 2 times” on the log-scale.⁹⁶ — And it shows an (incredibly) good match to the oscillations of the “red remainder” shown above.

The main conjecture (the Riemann Hypothesis)

One can come to the following conjecture about the error term of Gauss’ approximation

Prime density $\approx 1/\log x$:

Prime density $\cdot \log x - 1$ can be approximated very well by a linear combination of terms $1/x^{\alpha_n}$ and $\cos(f_{nm} \log x + \varphi_{nm})/x^{\alpha_n}$.

For the large values of x , the terms with the minimal possible α_n would dominate. In the previous section we have seen that these terms seem to have $\alpha_1 = 1/2$ and f_{11}, \dots, f_{14} close to 14.14, 21.06, 25.06, 30.4.

So far, the “trend” of Prime density $\cdot \log x$ we have seen is $1 - 1/2\sqrt{x} - C_3/x^{2/3} + \dots$, and the “discovered” fluctuations⁹⁷ on top of this trend are

$$\frac{2 \cos(f_{11} \log x + \varphi_{11}) + 2 \cos(f_{12} \log x + \varphi_{12}) + 2 \cos(f_{13} \log x + \varphi_{13}) + 2 \cos(f_{14} \log x + \varphi_{14}) + \dots}{\sqrt{x}} - \frac{\cos(1/2 f_{11} \log x + 1/2 \varphi_{11}) + \dots}{x^q}$$

with $q \sim 3/4$.

After the work of Riemann (made rigorous later by von Mangoldt¹⁰¹), we know that the green-framed conjecture above actually holds. However, all the values of α_n we observed by analysing our data are $1/2$ or more — and this is what neither Riemann nor von Mangoldt could prove (although $\alpha_n > 0$ has been proved independently by Jacques Hadamard and Charles Jean de la Vallée Poussin in 1896. In fact, up to now nobody could make a significant progress in improving the inequality $\alpha_n > 0$ to $\alpha_n \geq 1/2$ — and the latter estimate is confirmed so far by all numerical experiments.

This is the famous Riemann Hypothesis — which is a part of 6 Millennium Prize Problems. There are many formulations of this Hypothesis which are known to be equivalent between themselves. —

⁹⁶Recall that the first time we have been looking at “the frequency of the beat” in our plots, it was “about 5 periods per order of magnitude”. What the plot above shows is that these beats are 2x rescalings of each other.

⁹⁷Note that in the graphs above, we compare the *averaged density of primes* with these fluctuations *similarly averaged*.⁹⁸ So instead of $2 \cos$ above, we get a cosine with magnitude exponentially smaller: $2e^{-(wf_{nm})^2/2}$.

(Although since this sine wave enters our formulas actually multiplied by a power of x — so exponent of $\log x$ — one should add to f_{nm} above a small imaginary part. This leads to tiny corrections in magnitude and phase of the *correctly averaged* oscillations.)

⁹⁸In our tables, we count⁹⁹ (logarithms of) primes with a Gaussian weight on the log-scale. Since the density of integers on the log-scale near $\log x$ is x (and the “density of primes” is “relative to the density of integers!”), when matching our tables, one actually needs to average the oscillating terms below with an extra factor x .

⁹⁹More precisely, we Gaussian-blur $\sum_k \delta(z - \log p_k)$: we convolve this function with the Gaussian measure of the given $1/2$ -width w . (This is why we can easily recalculate from smaller w to a larger w' by convolving¹⁰⁰ with a Gaussian measure of $1/2$ -width $w_a := \sqrt{w'^2 - w^2}$.) (Since Gaussian measure has unbounded support, this leads to certain complications. Compare with Footnote 84 on p.27

¹⁰⁰There is a complication that between two convolution we discretize the result into a table with a certain step Δ (on the log-scale). However, using the Fourier transform shows that the resulting errors decay essentially as $e^{-(ww_a/w'\Delta)^2/2}$. So if Δ is reasonably small comparing to w , and $w'/w - 1 \ll 1$, the error can be made below the precision of calculations.

¹⁰¹**N.B. (???) Refs!**

And the simplest of these formulations is a much simplified form of our conjecture above on p. 35:

The averaged error term Prime density at x $- 1/\log x$ of Gauss's law "decays as" $1/\sqrt{x}$.

For example, here is one way to make this precise: one can average this error term on intervals¹⁰² of the form $[\varepsilon x, x]$ with any fixed $0 < \varepsilon < 1$, and one can understand¹⁰³ "decays as" as it being below $1/x^\alpha$ for sufficiently large x , provided $\alpha < 1/2$.

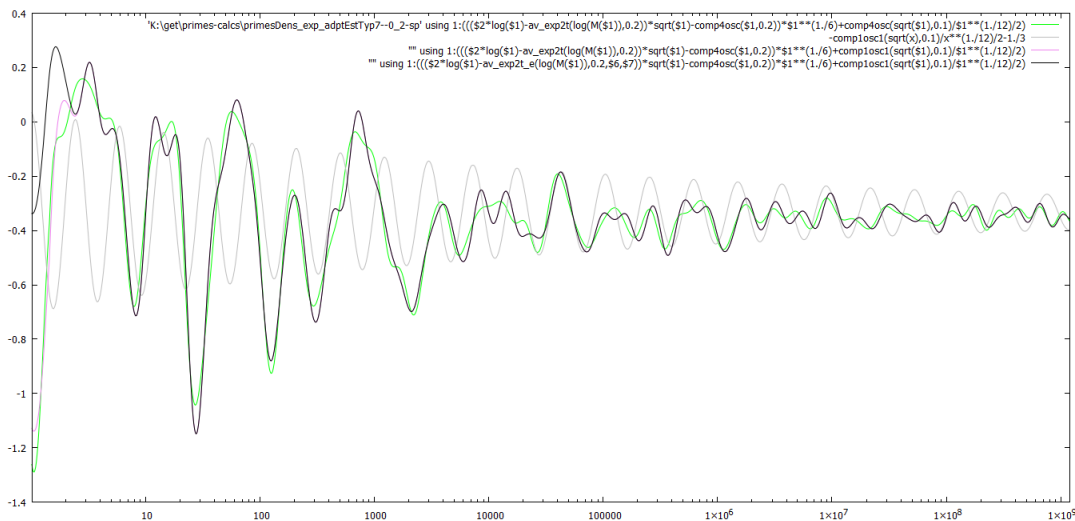
¹⁰⁵ ¹⁰⁶

Remark 5: Although the Riemann Hypothesis has its roots in the distribution of prime numbers, most frequently it is stated in a much more obscure form involving a sum $\zeta(s)$ of series of functions with a complex argument. Note that the form we wrote above seems to be much more elementary. (And with minimal knowledge of complex integration, it can be shown to be equivalent to the "standard formulation".)

Oscillation in the after-the-main terms of asymptotic

Do other sinusoids we discovered above appear in the remainder we discuss above?

On the next plot, the oscillating term above is in gray, and the result of subtracting this term from "the effective C_3 " is in violet. We also show the result of subtracting all 4 known oscillating terms (each "slowed down 2 times") instead of one (in green):



¹⁰² In other words, one considers the excess over $\int_{\varepsilon x}^x dx/\log x$ of the count of prime numbers in such an interval, divided by $(1 - \varepsilon)x$.

¹⁰³ There is a lot of leeway in possible formulations. For example, the stated weak form implies that this averaged error is actually much smaller: $O(\log x/\sqrt{x})$. (This was proven by von Koch in 1901 — with no improvements to this proven since then!)

On the other hand, there are reasons to believe that in fact the error term has the following *exact* speed of decay: $\log \log \log x)^2/2\pi\sqrt{x}$ (in the sense that \limsup and \liminf of the corresponding ratio are ± 1). This is Montgomery conjecture.¹⁰⁴ (H.L. Montgomery, The zeta function and prime numbers, Proceedings of the Queen's Number Theory Conference, 1979, Queen's Univ., Kingston, Ont., 1980, 1-31.)

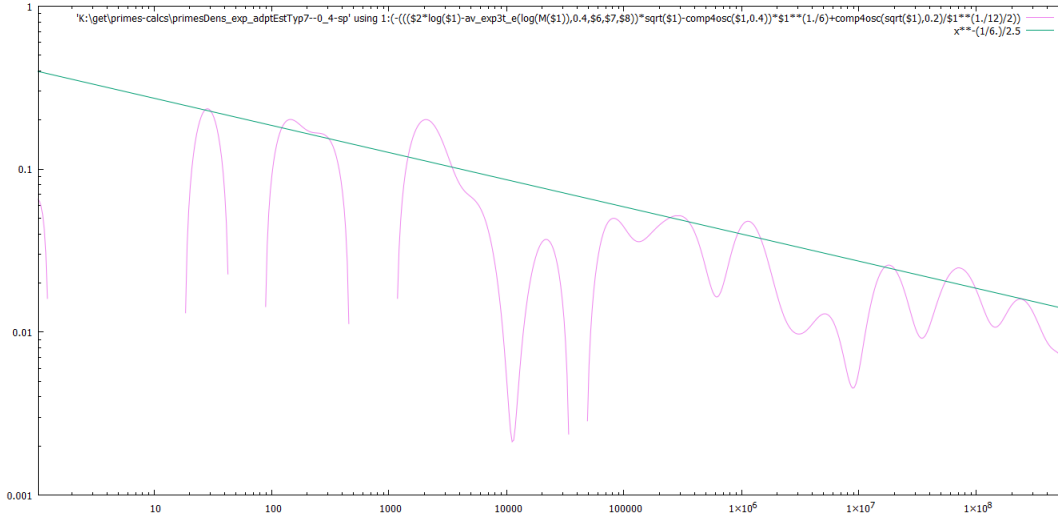
¹⁰⁴ **N.B. (???) Mention Rubinstein–Sarnak "enhancements"? Mention that local maxima of a sum of oscillating terms with independent frequencies appear where the phases of many such terms match. — But when we look for a match of phases of terms, the more terms, the rarer this is going to happen. So of coefficients are not in ℓ_1 , large magnitudes of the sum $s(x)$ are possible, but appear very rarely. This can gives a growth of $\max_{[0, X]} s(x)$ with such strange law as $\log \log \log x$.**

¹⁰⁵ **N.B. (???) Should we discuss the difficulties with double-summation in n and m ?**

¹⁰⁶ **N.B. (???) Should we discuss simplifications related to prime races?**

One can see that *at least* subtracting more terms decreases the remainder; while this does not show that these terms *actually appear* in the fluctuations, it is a significant confirmation.

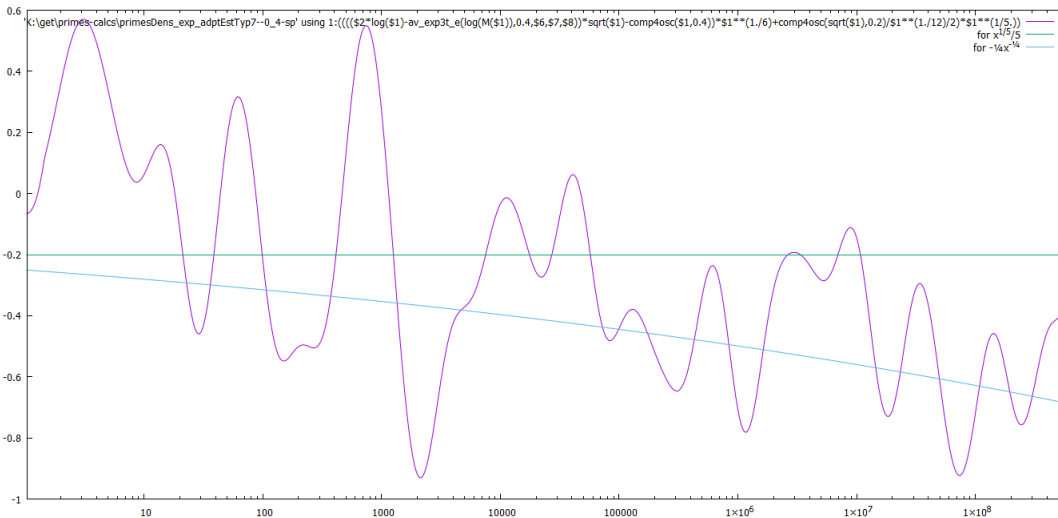
Doing likewise with $w = 0.4$, we plot the difference between $-1/3$ and the analogue of the green graph above:¹⁰⁷



in the log-log-plot. This suggests a power-law convergence, so one can expect that C_3 is $-1/3$ with a strong enough precision.

However, when we want to find *the next term after* $-C_3/x^{2/3}$ (with $C_3 = 1/3$), our approach fails. Indeed, this term may be described as giving C_3 dependence on x . For example, an extra term of the form $-1/4x^{3/4}$ would be equivalent to replacing C_3 by $\hat{C}_3(x) := 1/3 + 1/4x^{1/12}$. Since the plot above effective shows $\hat{C}_3(x) - 1/3$, to find the next term in the post-Gauss law¹⁰⁸ we need to find *which particular power law* decay appears in the preceding plot.

Multiplying the plot above by $x^{1/5}$ shows that it is not clear what particular kind of power decay happens on the plot above:¹⁰⁹



— but it seems to be somewhere *between* $-1/4x^{-1/4}$ and $-1/5x^{-1/5}$ (so the power $1/12$ is not confirmed). Taking into account the factor $x^{-2/3}$ at C_3 (which is the meaning of the y -axis of the plots before the

¹⁰⁷ **N.B. (???)** Here we needed to switch to using `av_exp3t` to correctly average $1/3x^{2/3}$. Otherwise on our graphs for C_4 the value is overestimated by 0.15 — but only on the extreme right.

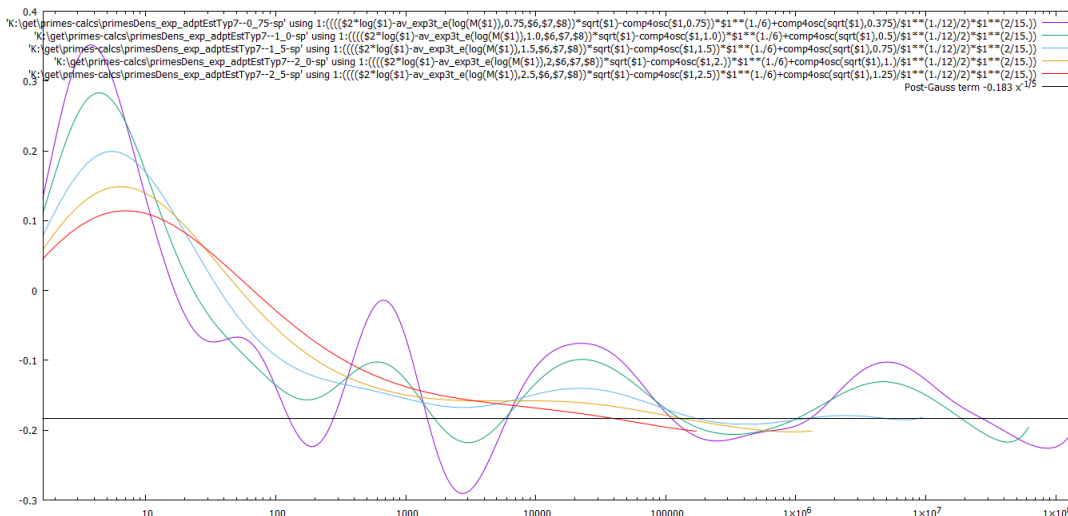
¹⁰⁸ **N.B. (???)** Discuss the name earlier!

¹⁰⁹ **N.B. (???)** after switching to $M()$, it seems the sloped line should be redone?

last), these would contribute “between $-1/4x^{11/12}$ and $-1/5x^{13/15}$ ” into the post-Gauss corrections to the prime density. (In particular, this does not follow¹¹⁰ the form of the other terms we found in this trend: $1 - 1/2\sqrt{x} - 1/3x^{2/3} + \dots$)

This exposes a beat which is *3 times slower* than the old beat: $1\frac{2}{3}$ periods per an order of magnitude.¹¹¹

In fact, the situation improves a lot with passing to $w = 0.75$ or more (up to $1\frac{1}{2}$):



Here we decreased the factor $x^{1/5}$ to $x^{1/8}$, and the straight line corresponds to a term $-0.18/x^{4/5}$ added to the trend of prime density. (Unfortunately, for $w \gtrsim 2$, the inherent uncertainty of averaging $1/\log x$ affects the part of the graphs we can calculate too much!¹¹²)

The pattern in the terms of the trend

The plots above suggest that the trend for Prime density $\times \log x$ behaves as

$$1 - 1/2\sqrt{x} - 1/3x^{2/3} - \widehat{C}_5(x)/x^{4/5}$$

with $\widehat{C}_5(x) \approx 0.183$ when x is in the range of a few millions. The last term is too similar to $1/5x^{4/5}$ following the pattern of the other terms to avoid **the question** of *the observed pattern*:

Can the real trend be $1 - 1/2\sqrt{x} - 1/3x^{2/3} - 1/5x^{4/5} - \widehat{C}_7(x)/x^\alpha$

with $\alpha > 4/5$ and \widehat{C}_7 not growing too quick? We know that $\widehat{C}_7(x)/x^{\alpha-4/5}$ is about -0.02 when x is in the range of millions.

In fact, 2, 3, 5 has a natural continuation as “prime numbers”. So to put a more precise meaning into the question above, it makes sense to inspect what are $\widehat{C}_n(x)$ and β in

The trend of Prime density $\times \log x$ is $1 - 1/2\sqrt{x} - 1/3x^{2/3} - 1/5x^{4/5} - 1/7x^{6/7} - \widehat{C}_n(x)/x^\beta$

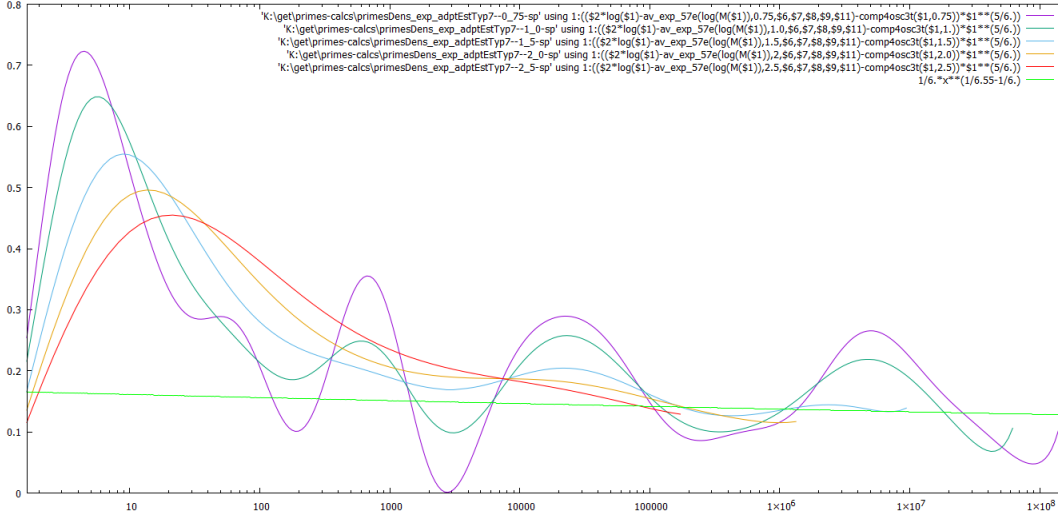
(to normalize β , suppose that $\widehat{C}_n(x)$ has a non-zero limit—or some such).

¹¹⁰ **N.B. (???)** Moreover, the appearance of the “old” beat “2.5 periods per order of magnitude” suggests that we should fine-tune the phase of the 4th sinusoid.

¹¹¹ **N.B. (???)** Explain!

¹¹² **N.B. (???)** We discussed this before!

This shows that the averaged $-\widehat{C}_m(x)/x^\beta$ behaves quite similar to $1/6x^{1-1/6.55}$ (the green line):



when x is in the range of a few millions.

This is too similar to $1/6x^{5/6}$ to avoid the following conjecture:

The trend of Prime density $\times \log x$ is $1 - 1/2\sqrt{x} - 1/3x^{2/3} - 1/5x^{4/5} + 1/6x^{5/6} - 1/7x^{6/7} - \widehat{C}_m(x)/x^\gamma$

(to normalize γ , suppose that $\widehat{C}_m(x)$ has a non-zero limit — or some such).¹¹³

On luck: the trend

Above we found the **Conclusion:** there is a very good fit with the trend

$$1 - 1/2\sqrt{x} - 1/3x^{2/3} - 1/5x^{4/5} + 1/6x^{5/6} - 1/7x^{6/7}$$

correcting the Gauss' density law.¹¹⁴

However, one should keep in mind that the last 3 terms are obtained by leaps of faith: there is a lot of cancellation between them, and one can replace them with other similar partially-cancelling-each-other terms which would give practically as good match as the expression above. One can prefer this particular choice of the last 3 terms only due to their similarity to the pattern of the terms before them.

Remark 6: There is very little chance to find the further terms by this method. The reason is that many following terms come in 4-tuples with signs of the form $+---+$ (for 10,11,13,14, then 15,17,19,21) with equal gaps in the first and the second pair — and these signs lead to a very strong cancellation inside such a 4-tuple of terms. After these two 4-tuples, two pair $+--$ follow (22, 23 then 26,29) — but they are significantly compensated by the following $--++$ (at 30, 31, 33, 34). **Conclusion:** the contribution of these terms is going to be very small — so it would not be possible to confirm it using our tables of the distribution of prime numbers. (These terms are going to be masked by fluctuations.)

While our guess for $-1/5x^{4/5}$ had a kind-of-weak, but still-convincing supportive evidence, it is not clear how one could find *any* kind of evidence for the terms following $-1/7x^{6/7}$.

Remark 7: On the other hand, it is the fact that the “following terms are mostly negligible” which led to a possibility of *guessing* the terms up to $-1/7x^{6/7}$.

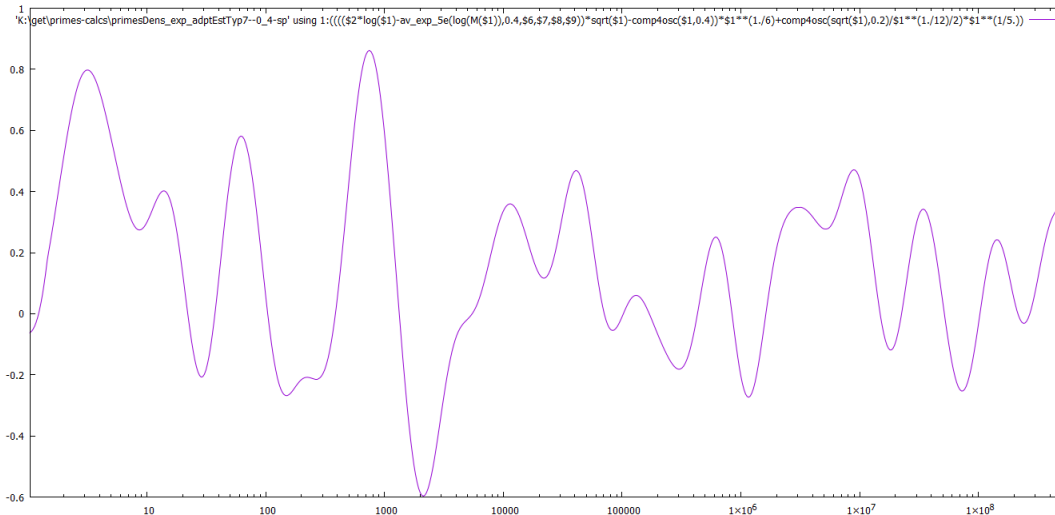
¹¹³ N.B. (???) TBC

Gauss used the function μ (in §81 of D.A.) but there is no indication he knew the Möbius inversion formula.

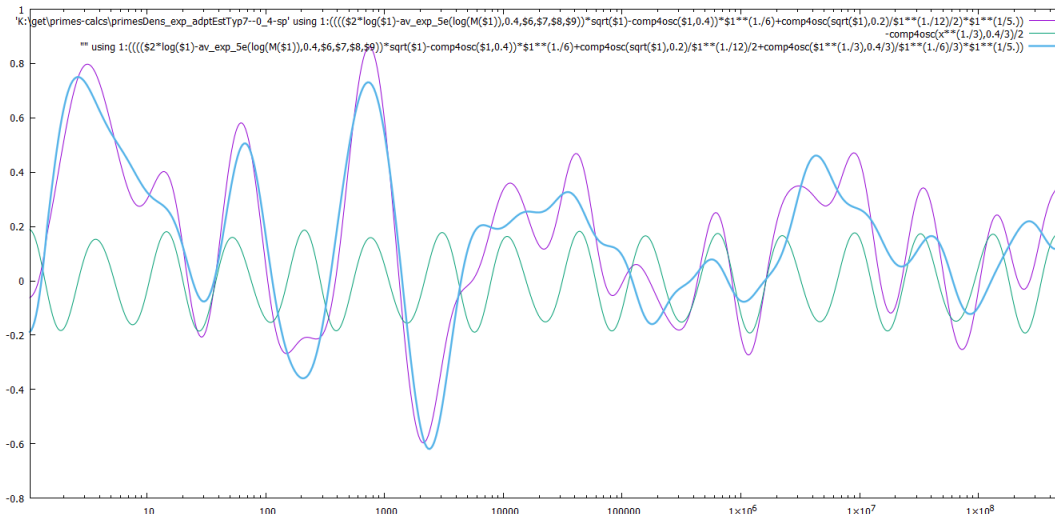
¹¹⁴ N.B. (???) At least with suitable correcting terms at the start, when $x \leq 64$. See the “compensated” plots below.

Back to narrow view

Going back to the case $w = 0.4$, here are the corrections on top of “the trend above with added first 4 frequencies of oscillation (both for $1/\sqrt{x}$ and $1/x^{3/4}$)”:

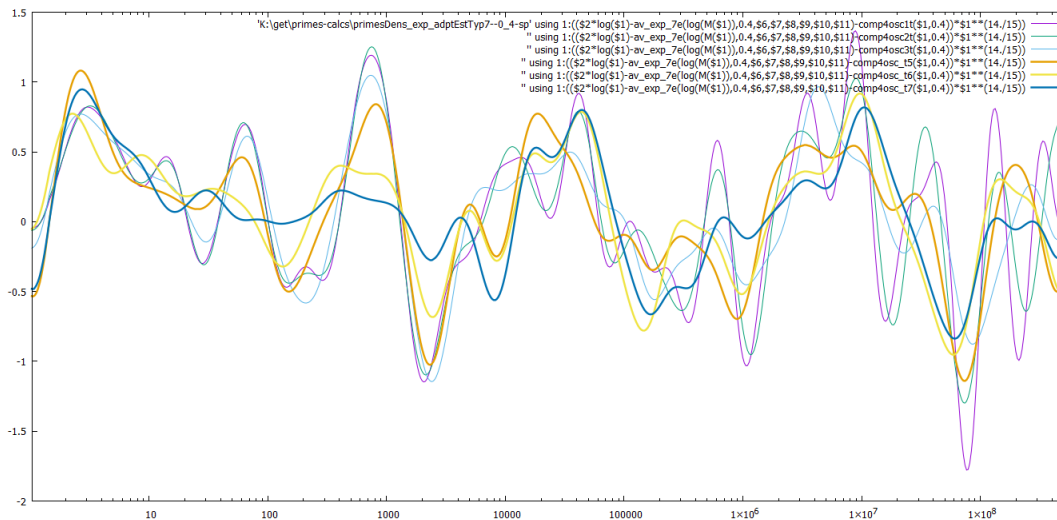


(Here we multiply by a suitable power of x to make fluctuations of similar magnitude for different x .) It probably is not so surprising now that “the beat” on the plot above is going 3 times slower than the first beat we observed (with “5 periods per order of magnitude”). This shows this beat, and the result of subtracting this beat (with amplitude $1/3x^{5/6}$ and the phase inverted):



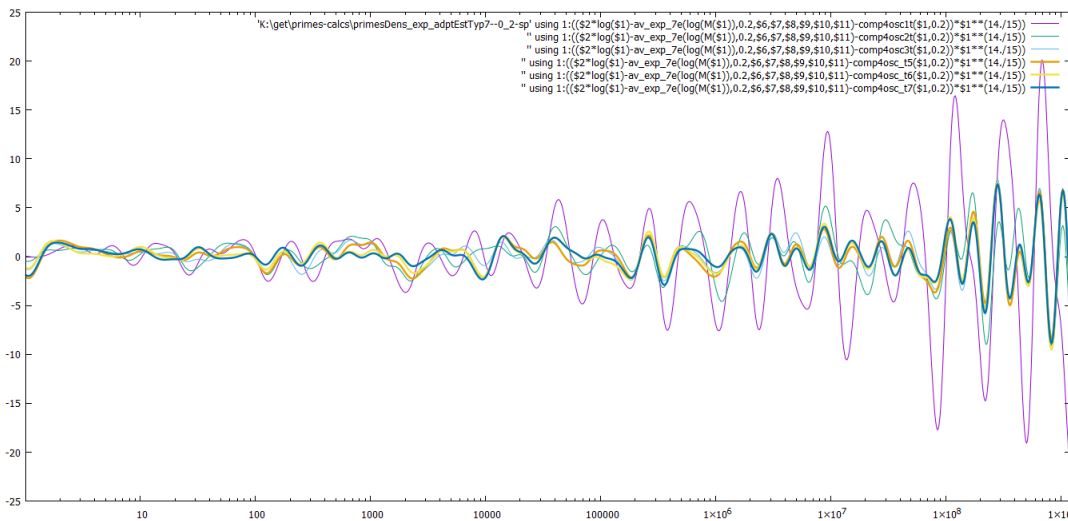
This time, “killing this beat” does not change the magnitude of fluctuations *a lot*—but the graph simplifies, and it is now easier to see yet slower beats on this graph.

This shows how “the remainders” w.r.t. the trend above behave when we cancel the first 4 oscillations in every “order of decay” of our asymptotic expansion¹¹⁵ for $\boxed{\text{Prime density}} \times \log x$:



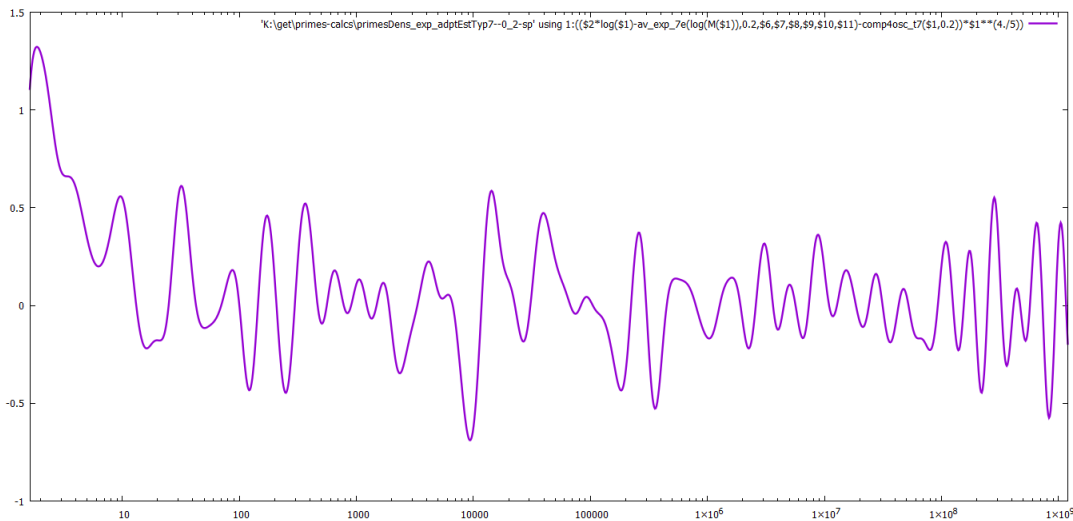
This shows that (in the range we investigated) this formula predicts the the weighted count of prime numbers (with the effective width $0.8 \cdot x$) around x with an error at most $x^{1/15}$. For example, the peak for $x \sim 10^7$ this is the error of about $1/5$ “extra primes”.

For averaging with the effective width $0.4 \cdot x$, using 4 oscillating terms is not enough, and (with the scaling used above) the error grows much quicker than $x^{1/15}$:



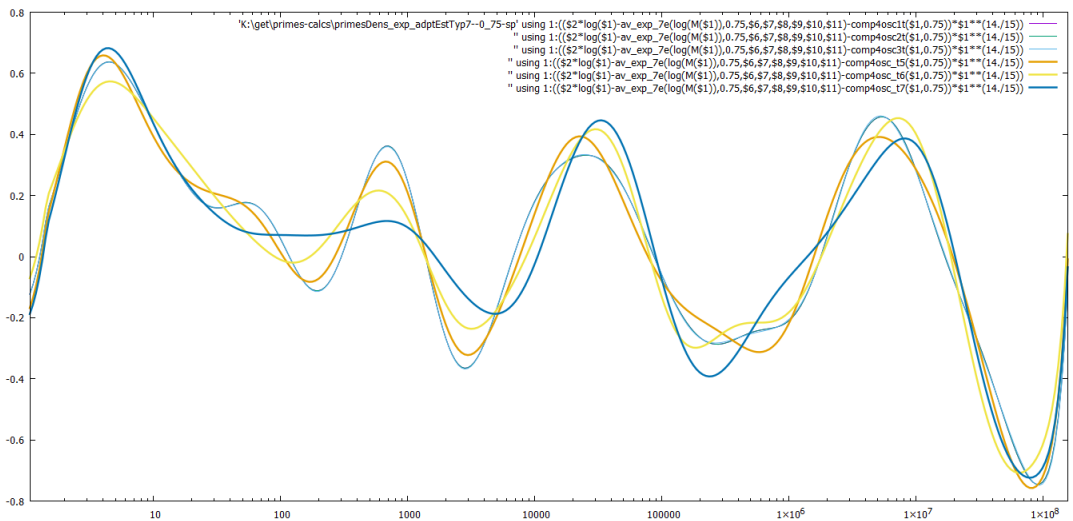
¹¹⁵ N.B. (???) Explain how the oscillations are matched to terms in the trend.

For example, near $x \sim 10^7$ the errors is about $\frac{3}{5}$ “extra primes”. Rescaling the best-approximation plot less aggressively gives:



— which suggests the mismatch growing similarly to $\frac{1}{2}x^{1/5}$ (in the range in question ¹¹⁶)

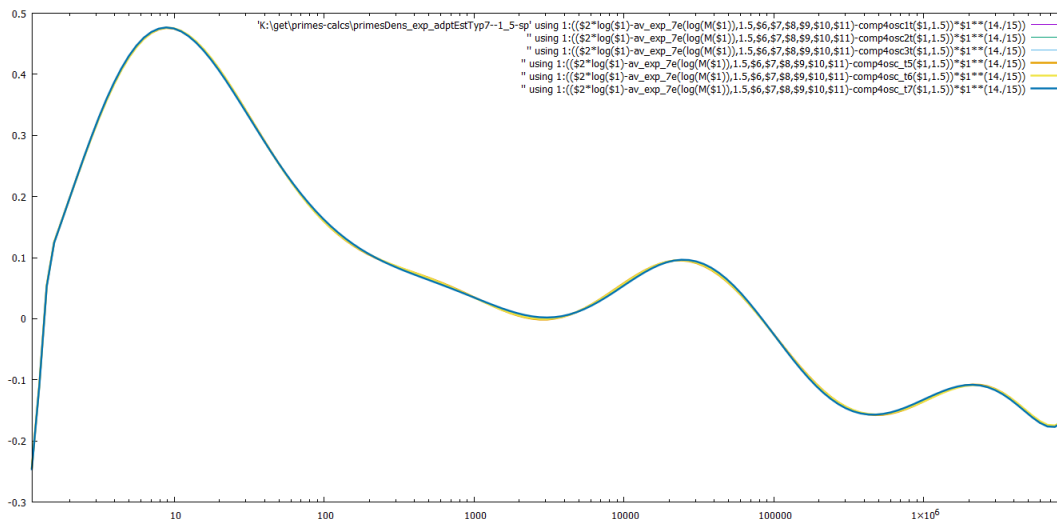
For averaging with the effective width 1.5 on the log-scale, compensating these oscillating terms is mostly excessive (the Gaussian dumping down has much stronger effect, so the difference between these graphs is negligible):



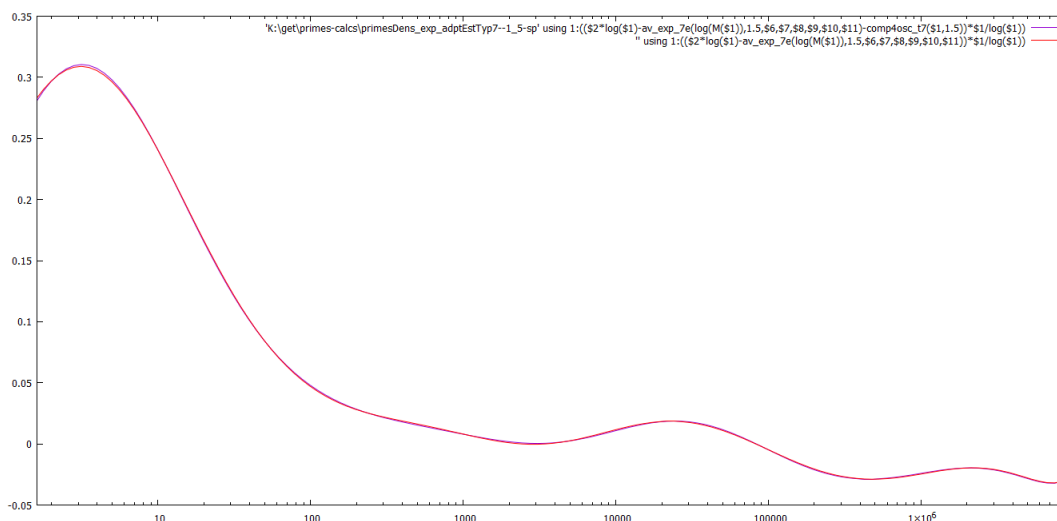
Still, near $x \sim 10^7$ the errors is about $\frac{1}{10}$ “extra primes”.

¹¹⁶ Given the decrease in magnitude about 10^6 – 10^8 , this may suggest a sum of two terms, one growing slower than $x^{1/5}$ — visible in the left half, the other growing quicker than $\frac{1}{2}x^{1/5}$ — visible on the far right.

For averaging with the effective width 3 on the log-scale, compensating these oscillating terms is completely unneeded:



(the difference between these graphs is almost invisible). The following plot shows the same data but scaled vertically so that it shows the corresponding log-Gaussian weighted count of primes:¹¹⁷



The maximal error for $x > 180$ is about 0.03 “extra primes” (in the range we investigate). (This is out of the total log-Gaussian-weighted count of primes “near 10^7 ” being above¹¹⁸ 6 millions with $w = 1.5$.)

Unfortunately, with $w \gtrsim 0.75$ there is practically no visible difference when we subtract “the term as above”: with such a width of averaging, the oscillations we subtract are more or less completely dumped down anyway!

¹¹⁷ Blue graph corresponds to using “our best” estimate of oscillating terms (with 4 frequencies per each of 6 rates of decay), red to using no oscillating terms at all.

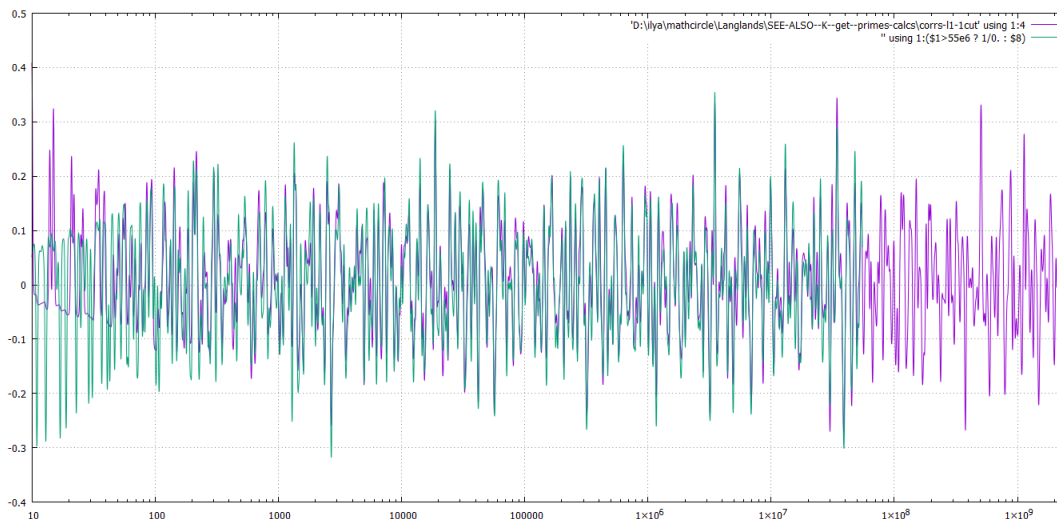
¹¹⁸ One should keep in mind that due to the nature of log-Gaussian weighting with half-width w , the count for a particular value of x involves “many” primes above x — “and mainly primes about $X \sim e^{wx}$ ”. For example, this X is close to the solution to the following the problem: find X such that the primes in the interval $[(1 - \Delta)X, (1 + \Delta)X]$ (with a fixed $\Delta \ll 1$) give the largest contribution into the weighted sum. (Compare with Footnote 98 on p. 35 and Footnote 86 on p. 28. This is somewhat similar to e^{-wx}, e^{wx} being centered not at x , but at $x \cosh w$.)

In particular, while the right end of the x -axis above is $\sim 10^7$, the right part of the graph “counts primes mostly of the order of magnitude of $\sim 10^8$ (and “a few times larger or smaller”). So the weighted count may be roughly estimated as $\sqrt{2\pi}wx e^{w^2/2} / \log(xe^{w^2})$.

On the correlations of prime densities at x and at $C \cdot x$ (or C/x)

On the plot on p.13 we already saw that *with a quite strong filtering of fluctuations*¹¹⁹ (and appropriate scaling), the density of primes at x correlates very good with the density at $17,000 \cdot x$ (at least when x is in $[2, 160]$). (This plot shows the *excess* $\mathcal{E}(x)$ in the density of primes over “its expected value”.)

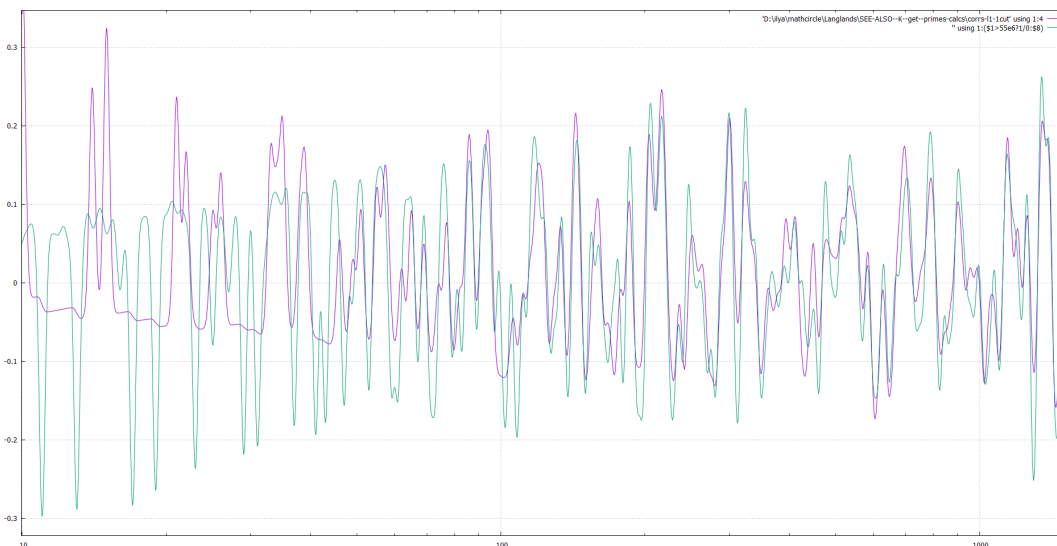
This shows the correlations between these two densities at different values of C ; this plot is for a much more narrow width of averaging¹²⁰ (effective $\frac{1}{2}$ -width $w = 0.01x$):



The green graphs shows the correlation for $x \leftrightarrow C \cdot x$, and the purple one for $x \leftrightarrow C/x$. (The possibility of the latter correlation to be strong is shown by blue marks on the plot we discussed above on p.13.)

Conclusion: the almost-perfect match between these two graphs (for $C \gtrsim 100$) shows that the correlation $x \leftrightarrow C \cdot x$ would remain as good as it is *if we formally extend the excess $\mathcal{E}(x)$ in the density of primes by putting $\mathcal{E}(1/x) := \mathcal{E}(x)$ for $x \geq 1$.*

(In fact, the matches between these graphs are observable already for $C \geq 32$:



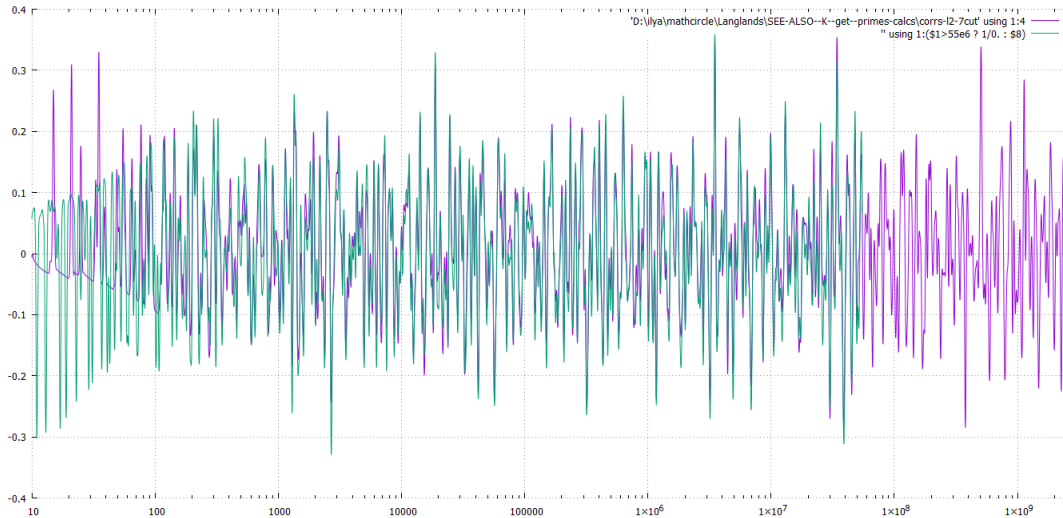
¹¹⁹ Recall that the width of Gauss filtering was $\sim 0.3x$.

¹²⁰ Before we used “averaging on top of Gauss’ bins of width 10,000”. This plot shows averaging without binning (or: “with bins of width $\ll 1$ ”).

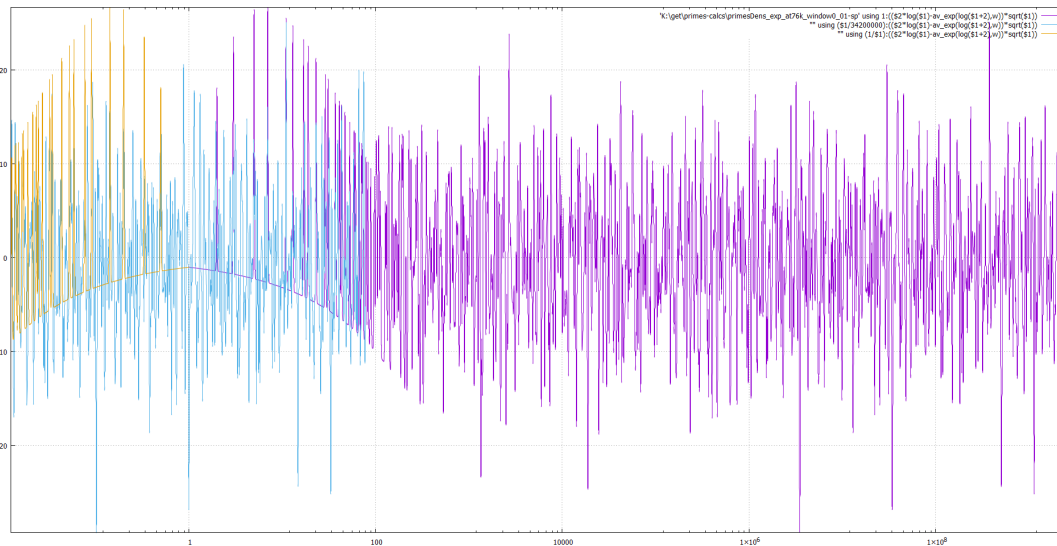
Observe that for $C \lesssim 100$ the peaks on the purple graph happen at products of two primes.¹²¹⁾

Here one can see that instead of strong (visual) correlation when $C \approx 17,000$ (and with very strong filtering of oscillations) we saw above¹²² we get an exceptionally high¹²³ correlation for $C \approx 18,830$ when we use very weak filtering. The other maxima visible on this plot are at $C \approx 3.5 \cdot 10^6$, $C \approx 34.3 \cdot 10^6$ and $C \approx 506 \cdot 10^6$.

These plots are similar, but exclude the region of very small primes (with the weight gradually linearly??? decaying in the interval $[2,7]$):



(The intent was that “the yet unaccounted terms” in prime density give a very strong contribution when $|\log x|$ is very small. So we attempted to exclude this. However, there is no visible improvement here, so this “trick” is not needed!)



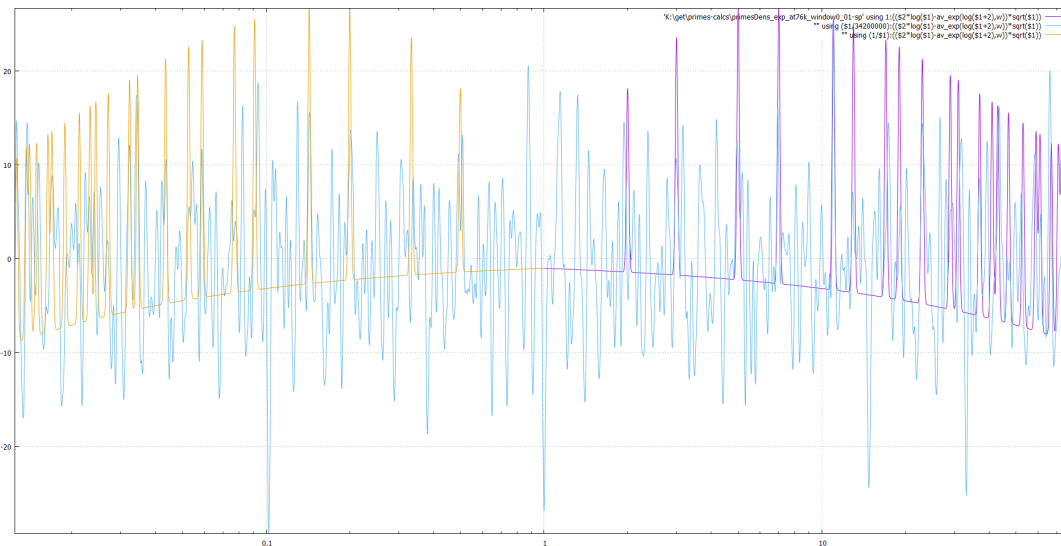
The purple graph shows $\mathcal{E}(x)$, the orange one is its reflection $\mathcal{E}(1/x)$, while the blue graph is a suitable translation of the right third of the purple to the left. Unfortunately, this plot does not visualize well

¹²¹⁾ On this plot, these products are blurred with the effective $\frac{1}{2}$ -width of 1.4%. Near 100, such products are (on average) separated by comparable distances $\log x / \log \log x \approx 3$, so the individual products do not contribute much into the graph.

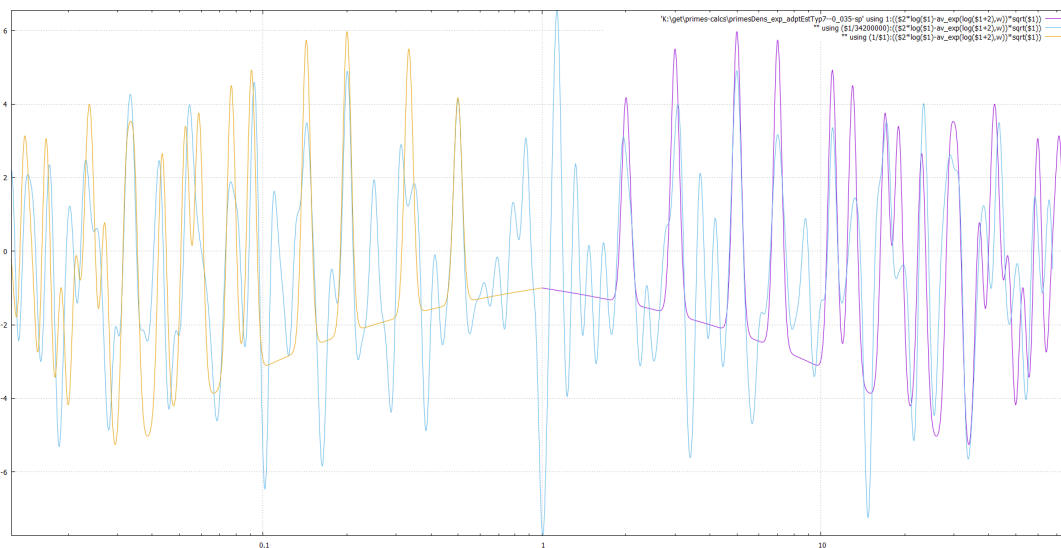
¹²²⁾ **N.B. (???) Ref!**

¹²³⁾ Although not very impressive in geometric terms: typically, “the angle” between function $\mathcal{E}(x)$ and $\mathcal{E}(C \cdot x)$ is close to 90° , and very few C correspond to the angle below 80° . For $C \approx 18,830$, the angle is about $\arccos 0.31 \approx 72^\circ$.

the similarity of 3 graphs in the left part of the picture. However, zooming into the left part shows that the (rare and narrow!) peaks on the purple and the orange graphs are match almost perfectly by the (narrow!) peaks on the purple graph:

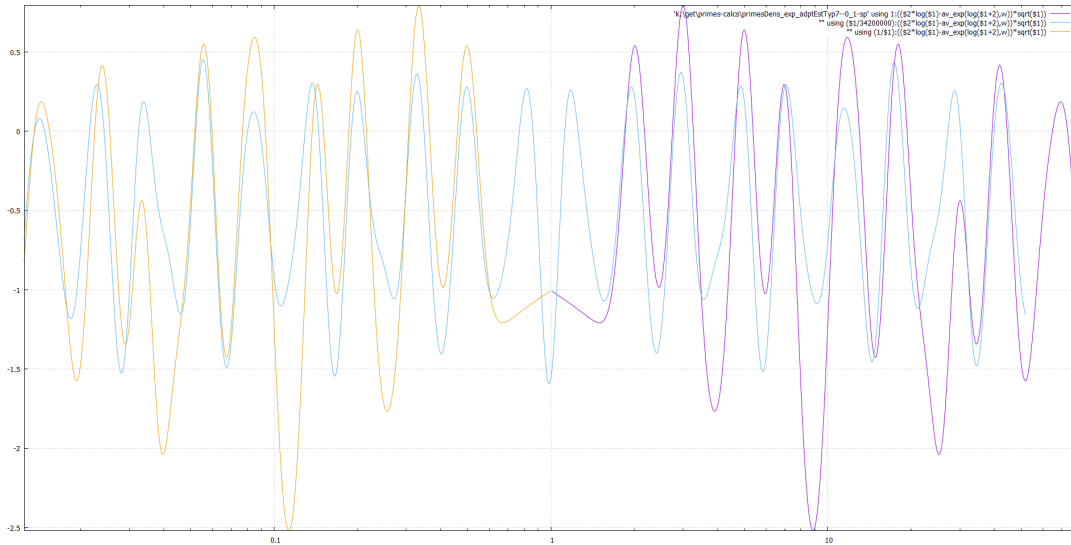


Although (due to the difference in frequency of the peaks) many peaks in the purple graph remain unmatched, it turns out that (contrary to expectations — given a good match in narrow zones only) the similarity of graphs *only increases* when one averages with a larger width (such as $w = 0.035x$ on this plot):

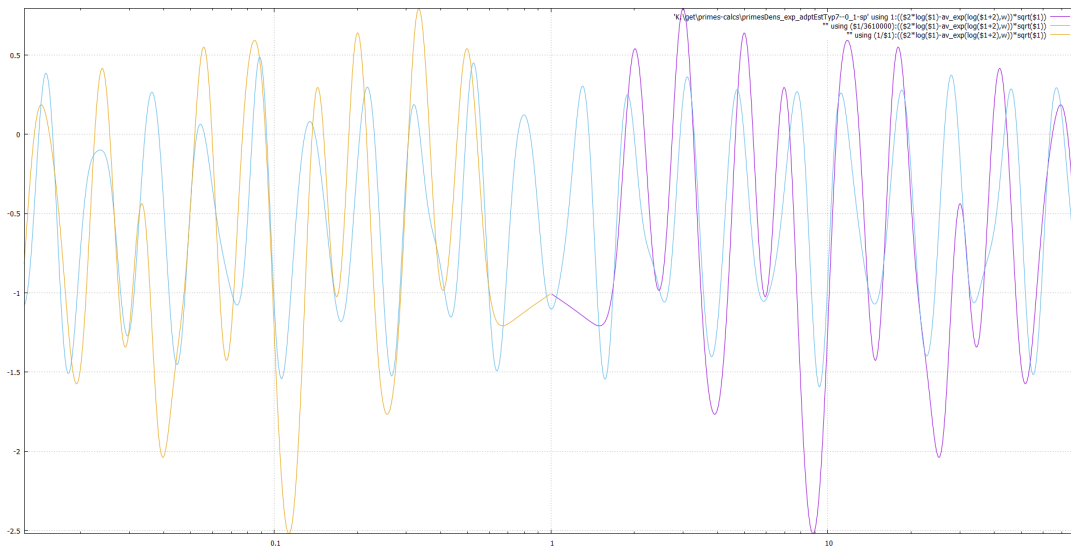


The “matched peaks” regain their heights, while “unmatched peaks” are dumped down!

With yet wider averaging (such as $w = 0.1x$ on this plot) all the peaks and trous match (except in the very center):



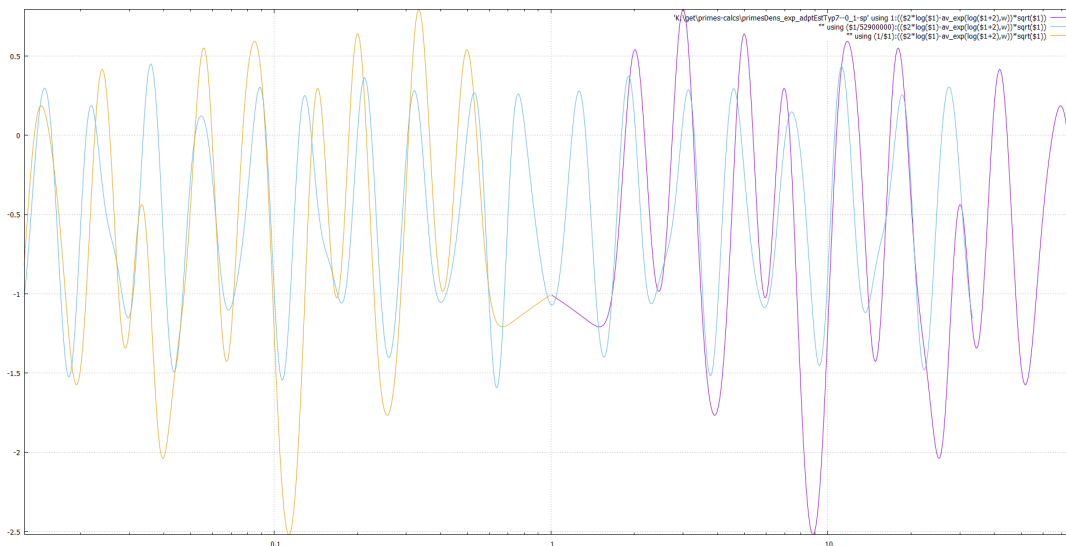
If we use other maxima on the plot of correlations¹²⁴ above, with the peak near 3,500,000 the match is nearly not as perfect (here we use $C = 3,610,000$):



with a decent match, and—again—an extra double-hill near 0.

¹²⁴ N.B. (???) Ref

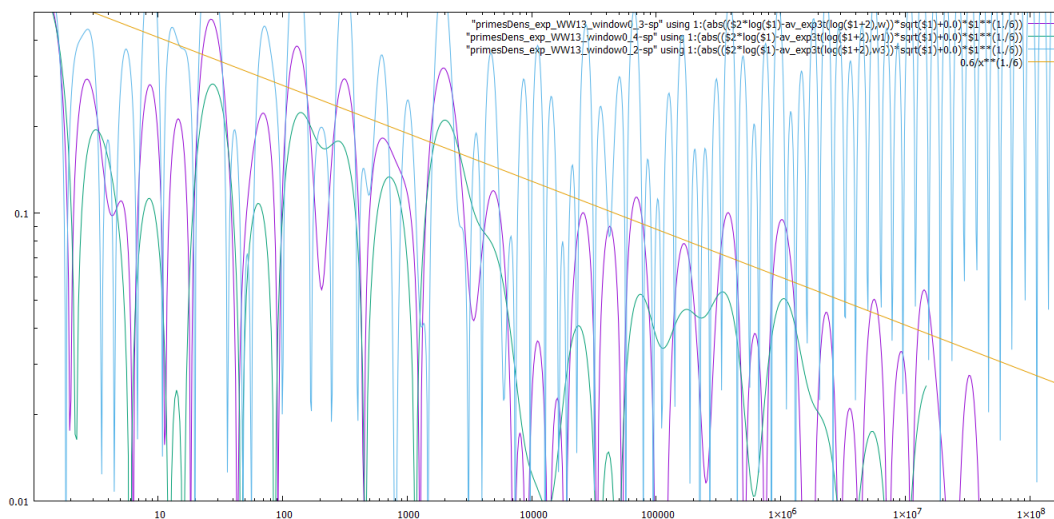
For the other maximum $C \approx 506 \cdot 10^6$, using $C = 529 \cdot 10^6$ gives the graph



but the “extra hills” near $x = 1$ are still very similar to what happens above.¹²⁵

The earlier attempts on further corrections: wider log-Gaussian weighting (without removing “main oscillations”

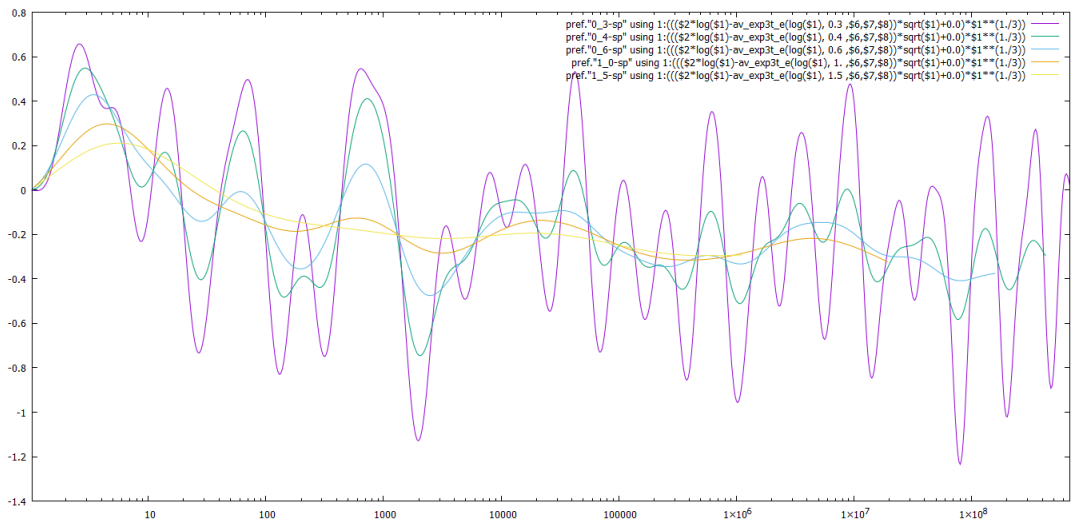
primeDens-av0_1_-0_2-0_4-0_4-abs-pow2_3



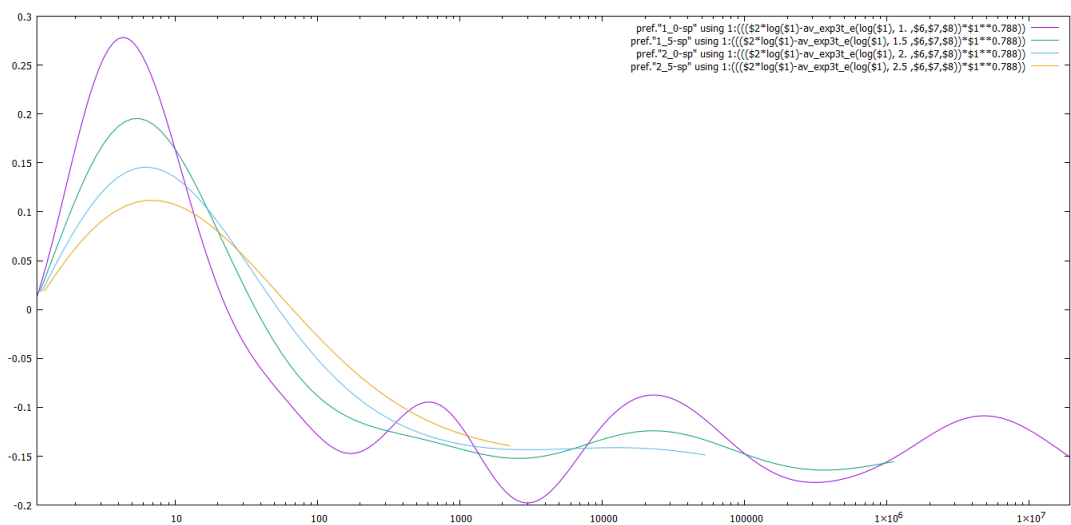
¹²⁵ **N.B. (???)** To explain this behaviour in the “very center”, the function $\mathcal{E}(x)$ may in fact be written as $\mathcal{E}_{\text{per}}(\log x) - \mathcal{E}_{\text{hill}}(\log x)$, with \mathcal{E}_{per} almost periodic (as the good matches when shifting the graphs — as above — suggest) and even, and $\mathcal{E}_{\text{hill}}$ having a double peak near 0 and exponentially decaying for large arguments (and even!).

Hence for a suitable C the function $\mathcal{E}(Cx)$ is close to $\mathcal{E}(x) + \mathcal{E}_{\text{hill}}(\log x)$, so \mathcal{E} has “an extra double hill” for $x \sim C$.

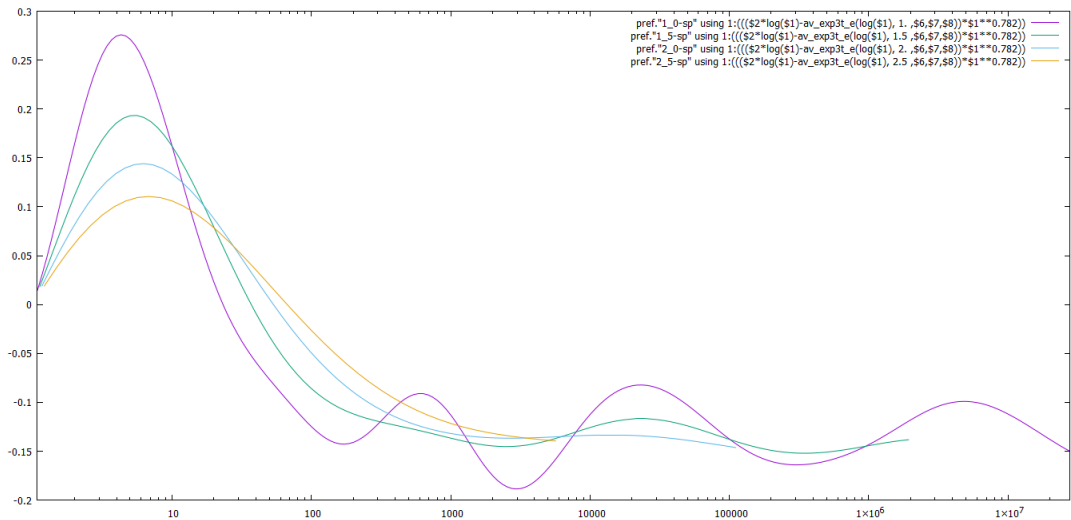
primeDens-av0_3-0_4-0_6-1-1_5-pow5_6



primeDens-av1-1_5-2-2_5-pow0_788

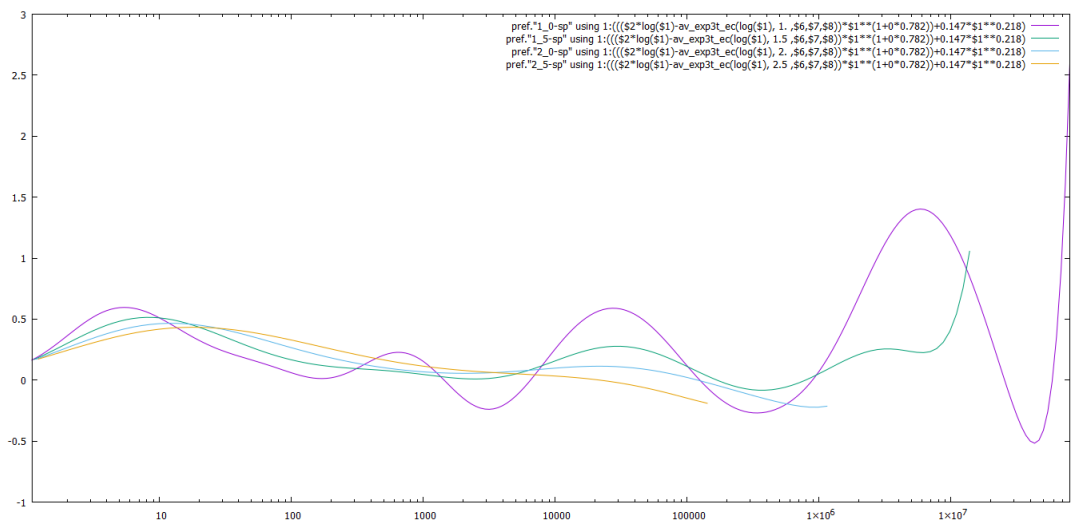


primeDens-av1-1_5-2-2_5-pow0_782



The earlier attempts (in preparation): newer “primeDens”

primeDens-av1-1_5-2-2_5-pow1-nocomp

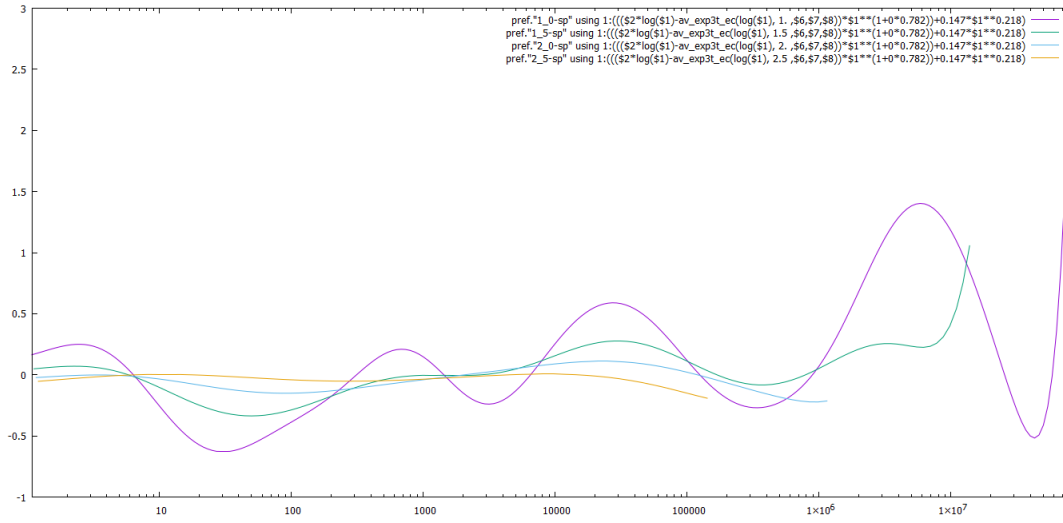


The “compensated” versions below try to modify the Ramanujan’s normalization of li by adding

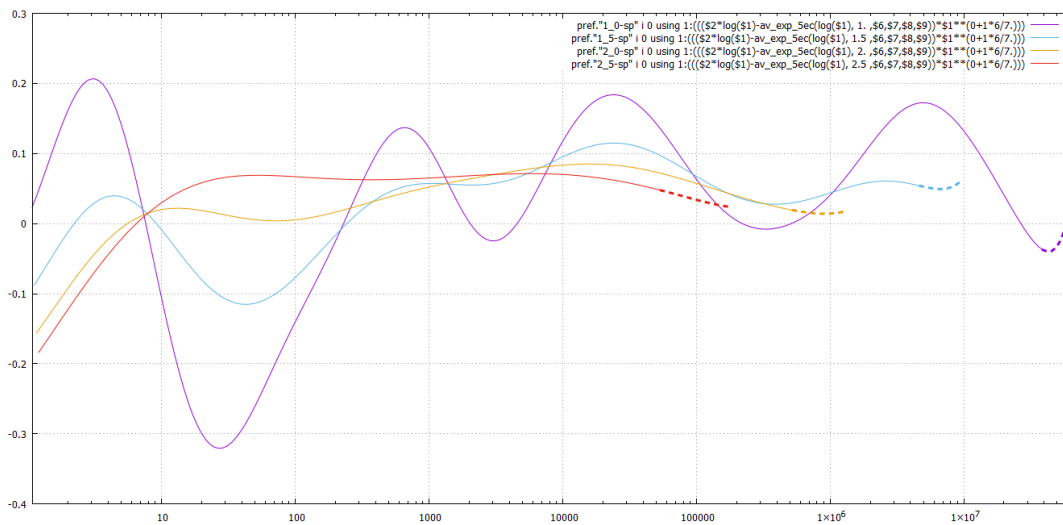
$$\text{comp}(x, w) = (-0.3 * \text{comp0}(x, w, 2) + 0.7 * \text{comp0}(x, w, 5) + 1.1 * \text{comp0}(x, w, 16) + 0.7 * \text{comp0}(x, w, 32) + 0.6 * \text{comp0}(x, w, 64))$$

$$\text{with } \text{comp0}(x, w, x0) = 1 / \sqrt{2 * \pi} / w * \exp(-(x - \log(x0)) ** 2 / (w ** 2)).$$

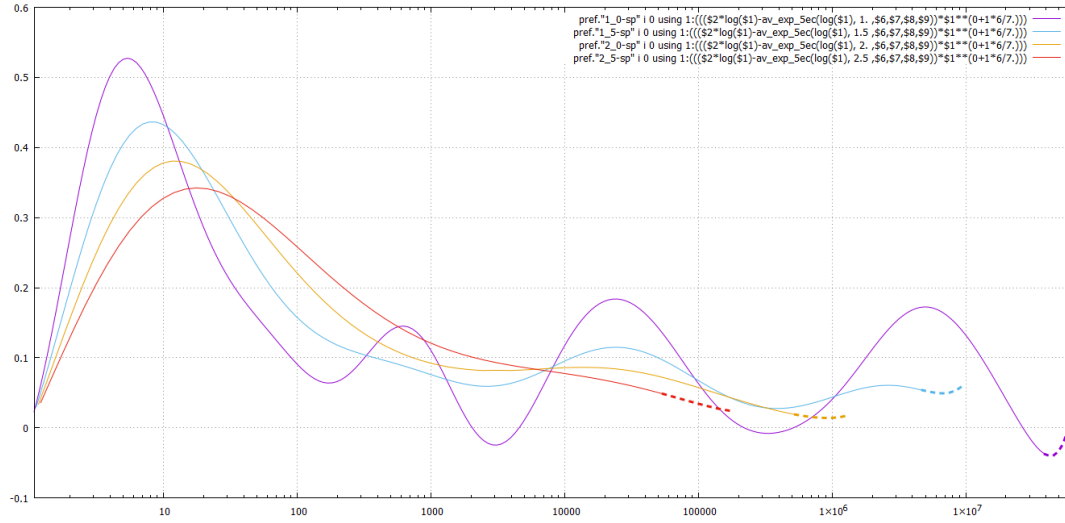
primeDens-av1-1_5-2-2_5-pow1-comp



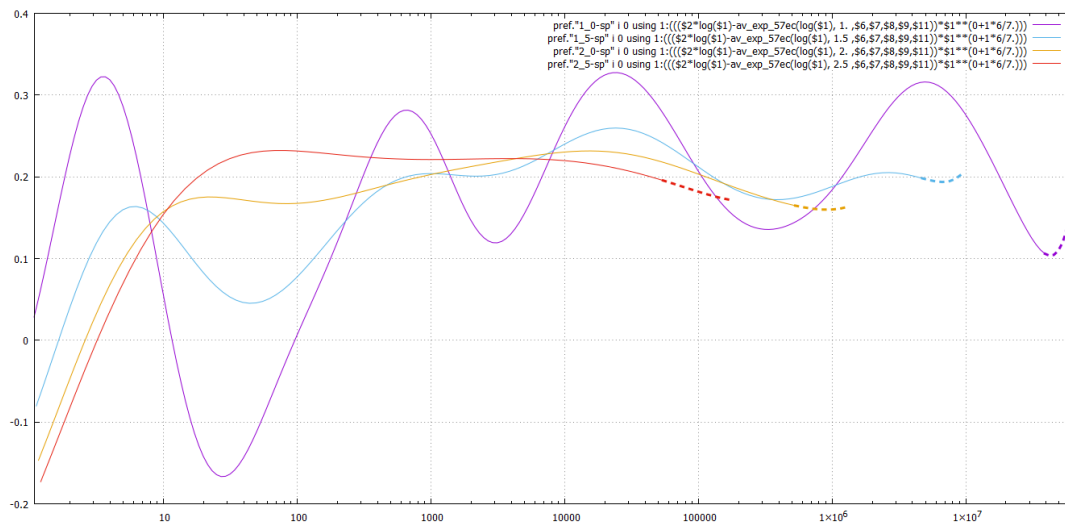
primeDens-5th-av1-1_5-2-2_5-pow6_7-comp



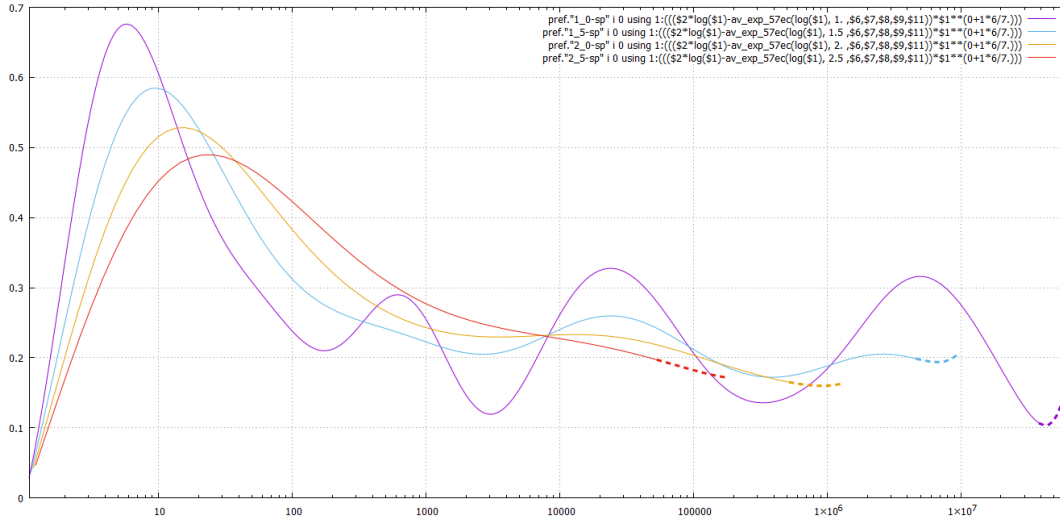
primeDens-5th-av1-1_5-2-2_5-pow6_7-nocomp



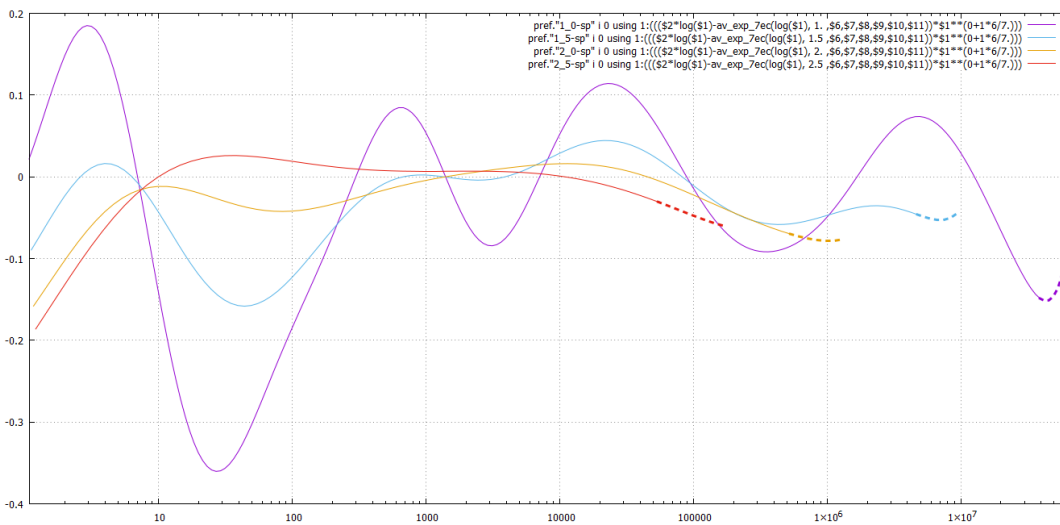
primeDens-5_7th-av1-1_5-2-2_5-pow6_7-comp



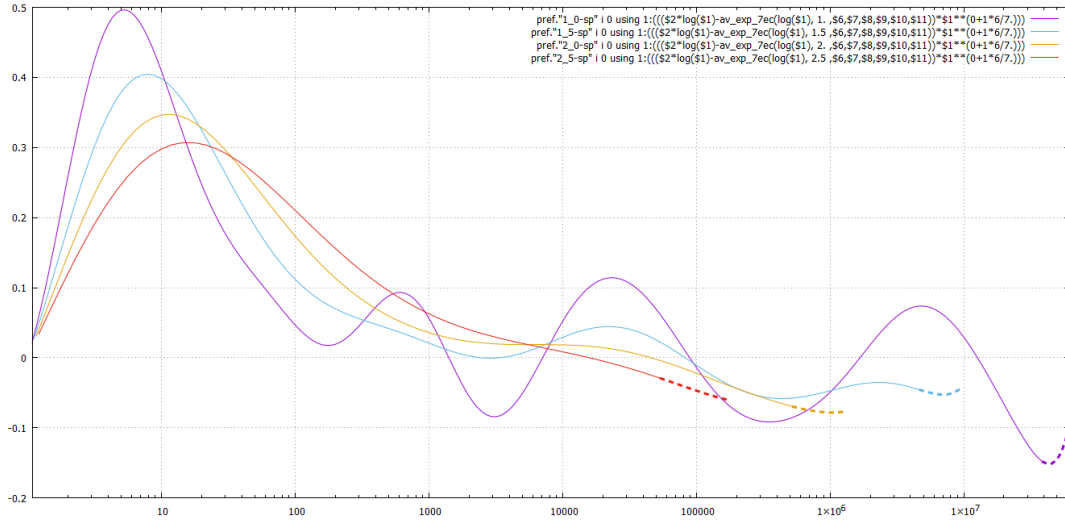
primeDens-5_7th-av1-1_5-2-2_5-pow6_7-nocomp



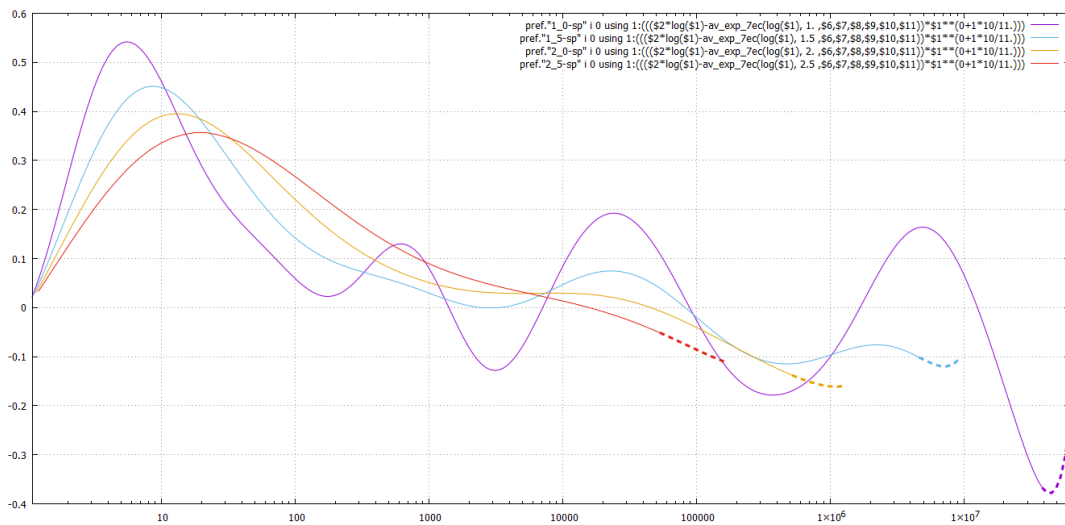
primeDens-7th-av1-1_5-2-2_5-pow6_7-comp



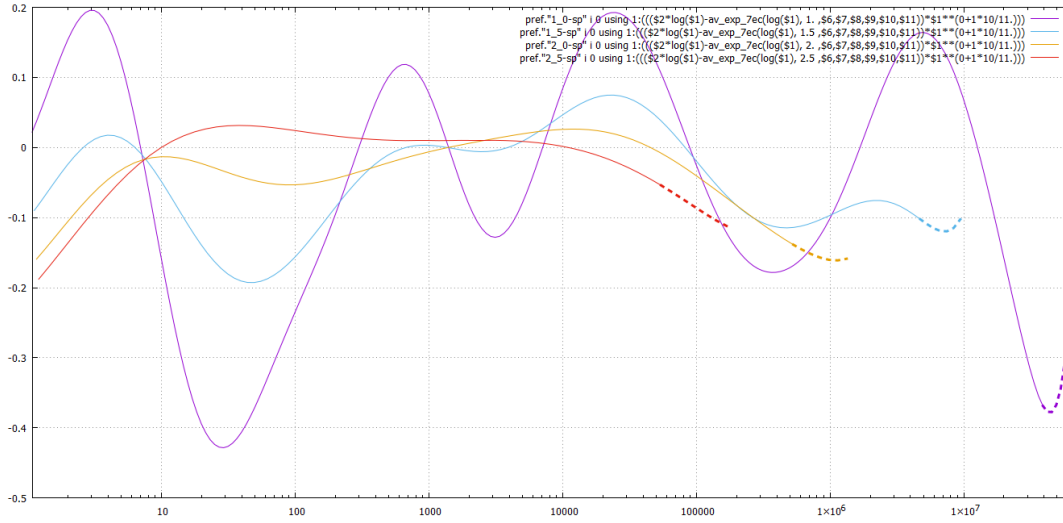
primeDens-7th-av1-1_5-2-2_5-pow6_7-nocomp



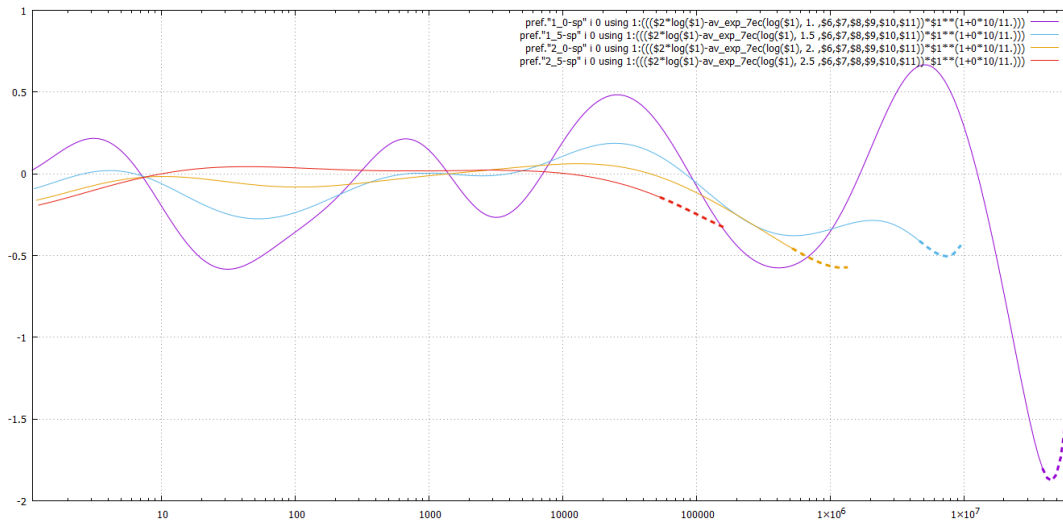
primeDens-7th-av1-1_5-2-2_5-pow10_11-nocomp



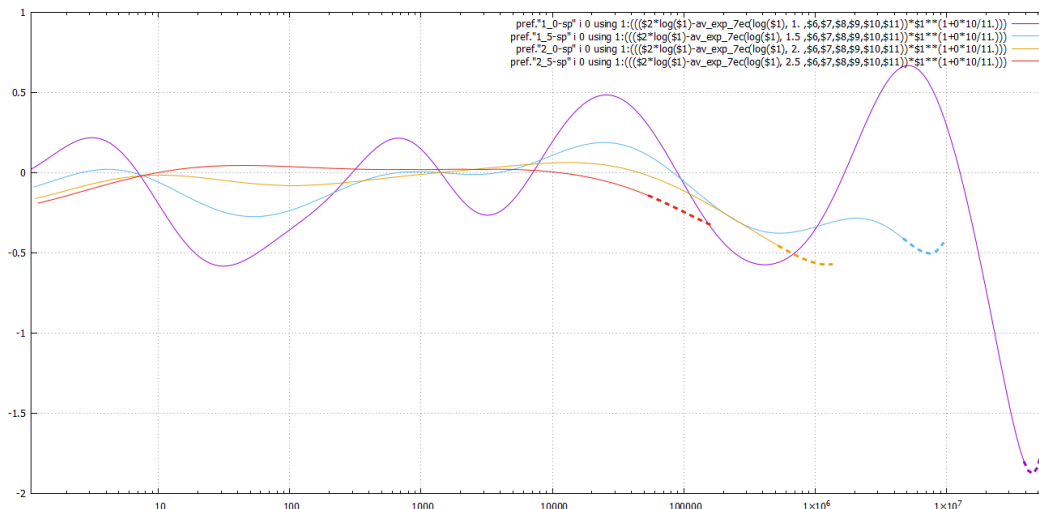
primeDens-7th-av1-1_5-2-2_5-pow10_11-comp



primeDens-7th-av1-1_5-2-2_5-pow1-comp



last-plot-of-summer



Additive vs. multiplicative: the tale of two Landau

“Multiplicative” number theory essentially studies correlations at x and at $x + \Delta x$ with $\Delta x \sim x$. “Additive” number theory does it¹²⁶ on smaller scales — such as $O(1)$. ζ -function. Fourier vs. Mellin.

Edmund Landau made Goldbach conjecture famous by including it into the list of “Landau’s problems” (on the 12th Congress of Mathematicians). However, not everybody was very enamored with it. For example, the epigraph¹²⁷ is made by his namesake, the outstanding physicist¹²⁸ Lev Landau.

N. Vavilov writes:

Например, Лев Давидович Ландау следующим образом комментировал гипотезу Гольдбаха: «Простые числа не нужно складывать, простые числа нужно умножать». Это бонмо неоднократно с одобрением цитировал Владимир Игоревич Арнольд.

Another recall is in Senderov–Spivak:

«Зачем складывать простые числа? – недоумевал великий физик Ландау. – Простые числа созданы для того, чтобы их умножать, а не складывать!».

The average value of $1/\zeta'$ (roots): surprising oscillations on the log-scale

This section seems to be unrelated to the rest of this chapter: the pictures below show the stats of $1/\zeta'$ at the roots of ζ -function on the critical strip. The knowledge of these stats may be useful to working with the Mertens function $M(n)$. (The script below uses the file `zprime-raw-full` contains these derivatives obtained via `zeta'(1/2+I*Z[i])`.)

Running `gp -q -f <zprime-av001.gp` etc. produces the tables used for plotting (like `zprime-av0_3--WW6`, for which $w = 0.3$ and $W = 6$ in the notations below). The logic of these tables goes like this:

¹²⁶ N.B. (???) Is not it reverted?!

¹²⁷ N.B. (???) ?

¹²⁸ His Nobel prize is for superconductivity — but his contributions into contemporary physics are extremely multifaceted.

For a zero $z_n = \frac{1}{2} + it_n$ of ζ , we put at $\log t_n$ the measure $1/|\zeta'(z_n)|$, then “average”¹²⁹ the total measure with Gaussian weights of the $\frac{1}{2}$ -width w (with the cut-off at $w \cdot W$). If the largest t_n in our tables is T , then we can calculate the weighted average up to $t_{\max} := T/\exp(w \cdot W)$. These calculations are performed for the values of t from 10 (which is a bit less than the first t_1 — so nothing interesting can happen below this value) to t_{\max} divided into up to 5,000 datapoints positioned uniformly in the log-scale.

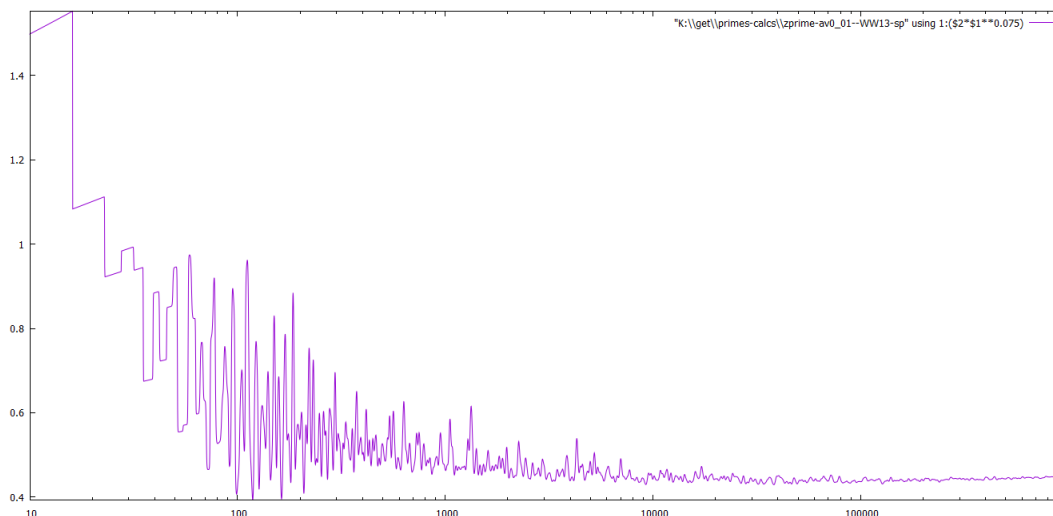
The (final) implementation:

```
do_av(pts=5000, w=0.01, WW=13, Min=10) = { \ \ 13
  my(Max=Z[#Z]*exp(-WW*w),st=(Max/Min)^(1/(pts-1)),pos,logpos, minK=1, minZ, maxK=1, maxZ, tot);
  for(pt=0,pts-1,
    pos=Min*st^pt; logpos=log(pos);
    minZ=pos*exp(-WW*w); maxZ=pos*exp(WW*w);
    while(minK<=#Z && minZ>Z[minK], minK++);
    while(maxK<=#Z && maxZ>Z[maxK], maxK++);
  \ \ print([minK,maxK]);
    tot = sum(k=minK, maxK, exp(-(logpos-logZ[k])^2/2/w^2)*[1,1/abs(Zprime[k])]);
  \ \ print(tot);
    filewrite(fileno, [pos, tot[2]/tot[1], tot[1]]);
    print1("; "pt);
  );
}
```

with a consequent fixing by `perl -wlpe "s/[\\[\\]]//g; s/\\sE/e/g" ! >!-sp`. (Plotting

```
plot "K:\\get\\primes-calcs\\zprime-av0_01--WW13-sp" using 1:($2*$1**0.075) w l
```

shows the behaviour of the averaged $1/|\zeta'(z_{\text{zero}})|$ in the right half of log-plot (for $t \in [10,000, 1,000,000]$) similar to¹³⁰ $4/(9 \cdot t^{0.075})$:



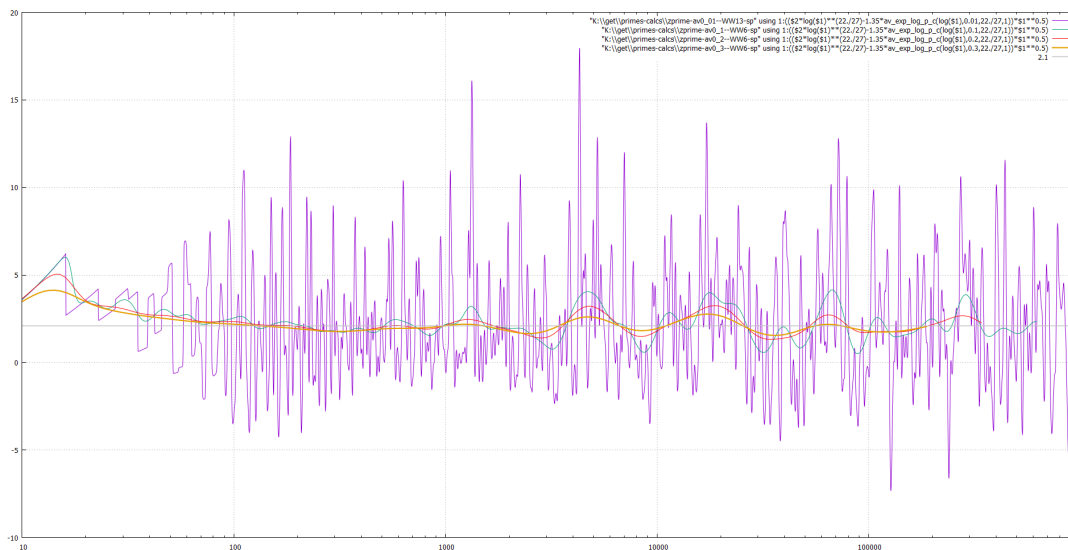
(This plot shows what $4/9$ should be replaced with to match our data.)

However, such very small powers of t are not easily distinguishable from $\log t \equiv d/d\lambda t^\lambda$. Doing such a replacement leads to a much more precise asymptotic: $(1.35 + 2.1/\sqrt{t})/\log^{22/27}(t)$ (with fluctuations

¹²⁹In other words, we take a weighted average of numbers $1/|\zeta'(z_n)|$ with a Gaussian weight (at $\log t_n$) centered at a particular value of $\log t$. We inspect how this average depends on t .

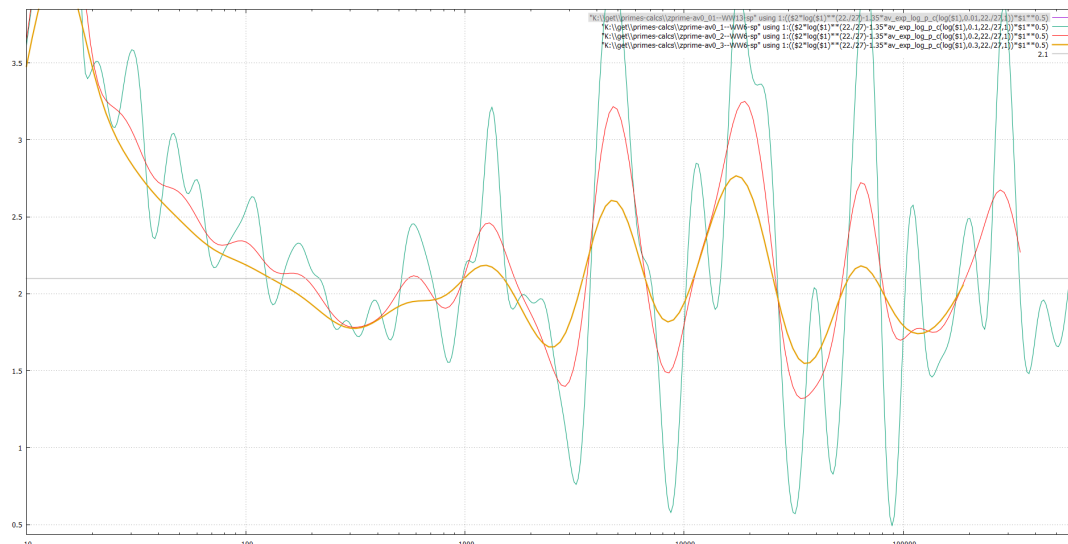
¹³⁰Instead of $4/9$ one gets ≈ 0.43635 with $w = 0.3$ (and $W = 6$).

about 2.1):



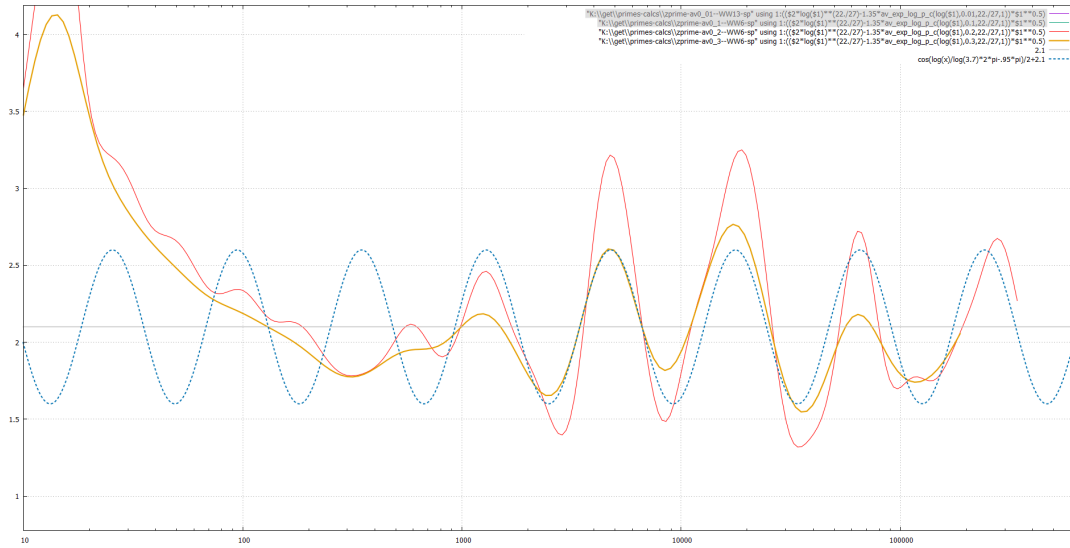
(The vertical axis of this plot again shows what 2.1 should be replaced with to match our data. We also plot what happens with wider Gaussians, of widths 0.1, 0.2 and 0.3 instead of 0.01.)

Already inspecting the plot above for width 0.01 one can see suspicious oscillation-like phenomena in the right half. Doing averaging with a 30 times wider log-Gaussian gives obvious oscillations *on the log-scale* — which seems to be totally unexpected!¹³¹ This zoomed-in plot (with width 0.01 omitted) shows that there are certain oscillatory terms added to 2.1 in the preceding formula:



¹³¹ N.B. (???) This is with a quite narrow cut-off of $W = 6$. It should better be redone! However, plotting $(\frac{3}{\psi(-1.7479+\log(x))})$ shows no trace of oscillation “in the third column” — even on the scale of 10^{-7} .

This overlays “the beat” over the plots with averaging width 0.2 and 0.3:

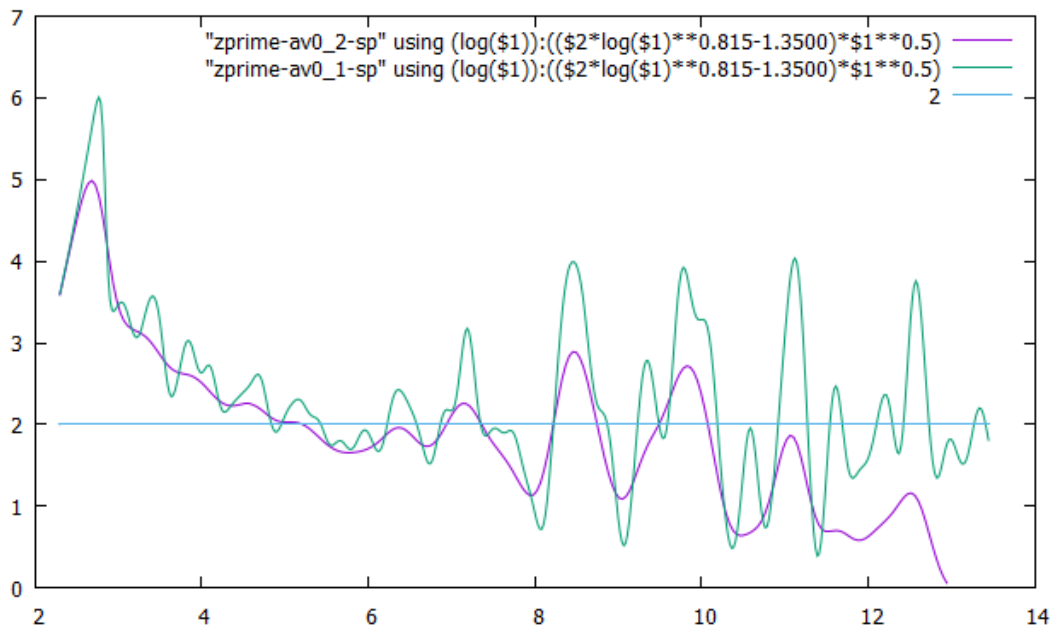


In other words, this shows that:

The averaged value of $1/|\zeta'(z_{zero})|$ near $\frac{1}{2} + it$ and $\frac{1}{2} + 3.7it$ correlate.¹³²

(At least in the interval $[500, 400,000]$.)

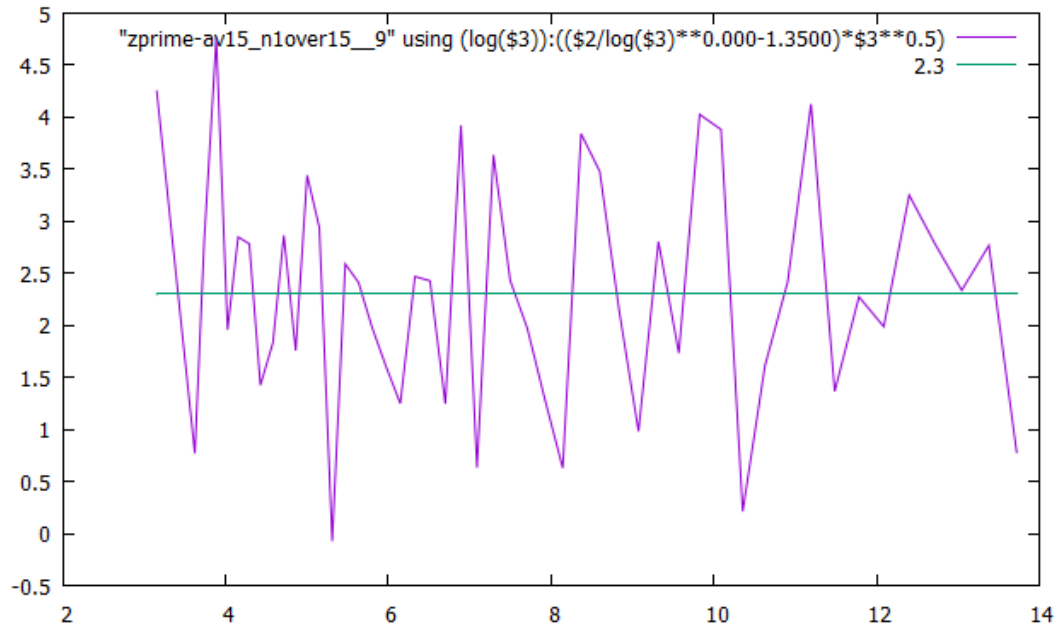
Obsoleted (before averaging powers of $\log x$): This is what happens with a bit narrower log-Gaussian averaging: zprime-av0_1_-0_2-w5



¹³² N.B. (???) This corresponds to a relation (of unknown nature!) between distribution of primes near x and near $x^\circ := x^{3.7}$.
 (Compare with a similar correlation for $x^\circ := 3.42 \cdot 10^7 \cdot x$ we investigated above. See Footnote 75 on p. 25.)

One more (very old) picture with lost provenance.

Based on the table `zprime-av15_n1over15__9` (obtained by an earlier version¹³³ of the process above—but the code to produce this table may have been lost...):



¹³³ N.B. (???) Double-check!