PRIMES IN SHORT INTERVALS: HEURISTICS AND CALCULATIONS

ANDREW GRANVILLE AND ALLYSA LUMLEY

Abstract. We formulate, using heuristic reasoning, precise conjectures for the range of the number of primes in intervals of length y around x, where $y \ll (\log x)^2$. In particular we conjecture that the maximum grows surprisingly slowly as y ranges from $\log x$ to $(\log x)^2$. We will show that our conjectures are somewhat supported by available data, though not so well that there may not be room for some modification.

1. Introduction

We are interested in estimating the maximum and minimum number of primes in a length y sub-interval of $(x, 2x]$, denoted by

$$M(x,y) := \max_{X\in(x,2x]} \pi(X+y)-\pi(X) \text{ and } m(x,y) := \min_{X\in(x,2x]} \pi(X+y)-\pi(X),$$

respectively. It is widely believed that $m(x,y)=0$ for $y \ll (\log x)^2$ though we do not know the precise value of the implicit constant. However there has been little study of how $m(x,y)$ subsequently grows, or of how $M(x,y)$ behaves for $y \ll (\log x)^{2+o(1)}$. In this article we will conjecture a series of guesstimates for $M(x,y)$ and $m(x,y)$ in different ranges, comparing these estimates to what relevant data we can compute, and discussing some of the issues that prevent us from being too confident of these guesses.

The starting point for our investigations came from a comparison of two known observations:

Based on the (conjectured) size of admissible sets we believe that

$$M(x,y) \sim \frac{y}{\log y}$$

for y fixed and also for y in some range depending on x, perhaps as large as $y \ll \log x$.[1] On the other hand based on a modification of Cramér's probabilistic model [3] for the distribution of primes (which in turn is based on Gauss's observation that the primes have density $\frac{1}{\log x}$ around x), we believe that

$$M(x,y) \sim \sigma_+(A)\frac{y}{\log x}$$

for $y = (\log x)^A$ with $A > 2$, for some constant $\sigma_+(A) > 1$, for which $\sigma_+(A) \to 1^+$ as $A \to \infty$.

Thanks are due to James Maynard for some helpful remarks on both the content and the exposition, and to Kevin Ford and Drew Sutherland for making various data available.

[1] That is, there exists a constant $c > 0$ such that if x is sufficiently large then $M(x,y) \sim \frac{y}{\log y}$ for all $y \le c \log x$.

Therefore it seems that in both ranges, $M(x,y)$ is roughly linear in y: In particular,

$$M(x,y) \sim \frac{y}{\log\log x} \text{ for } y \text{ a little smaller than } \log x,$$

whereas, if $c_+ := \sigma_+(2)$ then

$$M(x,y) \sim c_+ \frac{y}{\log x} \text{ for } y \text{ a little bigger than } (\log x)^2.$$

If true then $M(x,y)$ has quite different slopes, $\frac{1}{\log\log x}$ vs. $\frac{c_+}{\log x}$, in these two different ranges, and so there is a substantial change in behaviour of $M(x,y)$ as y grows from around $\log x$ to slightly beyond $(\log x)^2$. Our main goal is to investigate what happens in-between, though also to give heuristic support for the claims above.

At the end-points of this in-between interval, the above claims suggest that

$$M(x,\log x) \sim \frac{\log x}{\log\log x} \text{ whereas } M(x,(\log x)^2) \asymp \log x,$$

so $M(x,y)$ does not seem to get much bigger as y grows from $\log x$ to $(\log x)^2$; indeed it grows by only a factor of $\log\log x$. This is very different from before and after this interval: As y goes from 1 to $\log x$ we expect $M(x,y)$ to grow by a factor of $\asymp \frac{\log x}{\log\log x}$, and as y goes from $(\log x)^2$ to $(\log x)^3$ to grow by a similar factor of $\asymp \log x$ (and indeed for any subsequent interval of multiplicative length $\log x$). This does not seem to have been previously observed.

Based on an appropriate heuristic we conjecture that if $1 < A < 2$ then

$$M(x,(\log x)^A) \sim \frac{1}{2-A} \cdot \frac{\log x}{\log\log x};$$

more precisely that if $\log x \le y = o((\log x)^2)$ then

$$M(x,y) \sim \frac{\log x}{\log\left(\frac{(\log x)^2}{y}\right)}. \tag{1}$$

We will provide data with x up to 10^{12} to support this claim, though it should be noted that although this is as far as we have been able to compute, these x are still small enough that secondary terms are likely to have a significant impact. For this reason we also look at

$$M(x,2y)/M(x,y)$$

because we expect that, as $x \to \infty$ this looks much like 1 in this range, and 2 outside this range. However we will compare the data for this ratio to a more precise conjecture.

In this article we will argue that there are four ranges of y in each of which we expect different behaviour for $M(x,y)$, namely:

$$y \ll \log x;\ \log x \ll y = o(\log x)^2;\ y \asymp (\log x)^2;\ \text{ and } y/(\log x)^2 \to \infty \text{ with } y \le x.$$

We will present these separately in the introduction though there is significant overlap in the theory; and when it comes to presenting data for a given value of x up to which we can compute, it is often unclear where one y-interval should end and the next begin.

1.1. **Guesstimates for very short intervals: $y \ll \log x$.** We believe that if $y \leq \log x$ then

$$M(x,y) \sim \frac{y}{\log y}. \tag{2}$$

In fact we have a more precise conjecture than this for $y \leq (1-o(1)\log x$: A set of integers A is *admissible* if for every prime p there is a residue class mod p that does not contain any element from the set. Let $S(y)$ be the maximum size of an admissible set A which is a subset of $[1,y]$.[2] We believe that if $y \leq (1-o(1))\log x$ then[3]

$$M(x,y) = S(y). \tag{3}$$

These two conjectures are consistent since it is believed that $S(y) \sim \frac{y}{\log y}$. The data seems to confirm the conjecture (3) for $x = 10^k$ for $k = 9, 10, 11$ and 12:

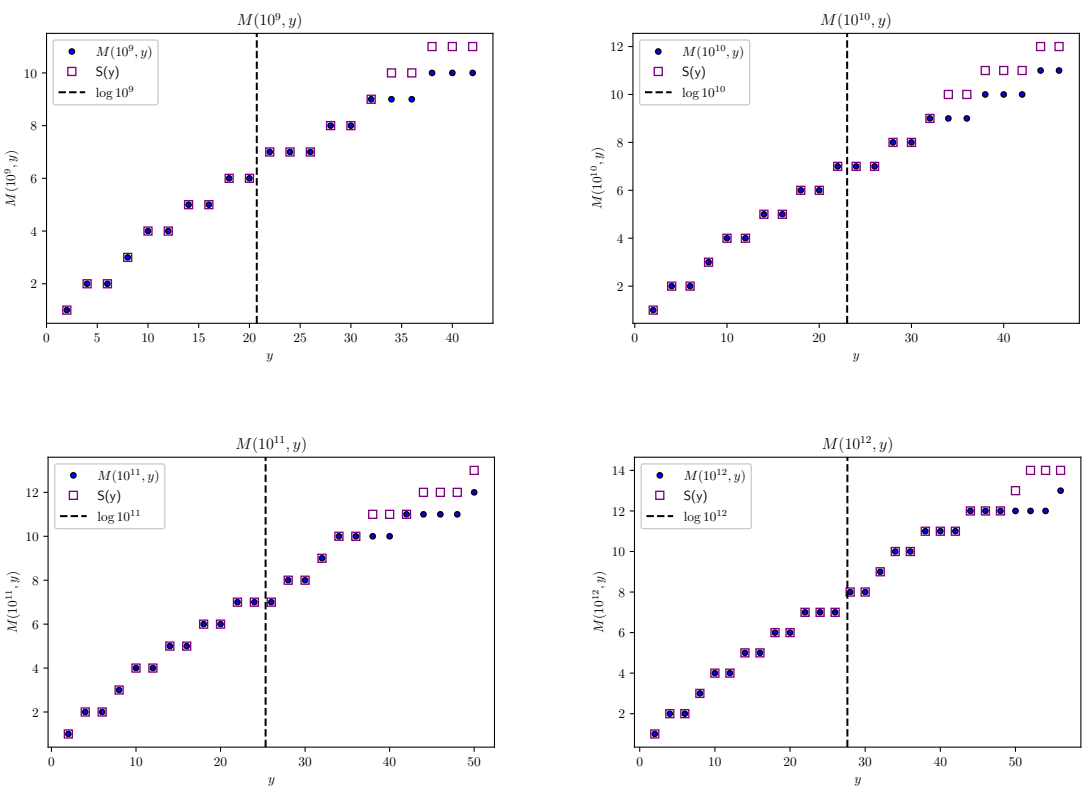

FIGURE 1. $M(x,y)$ vs. $S(y)$ for $x = 10^k, k = 9, \ldots, 12$ and $y \leq 2\log x$. We conjecture that $M(x,y) = S(y)$ up to the dashed line at $y = \log x$

In these graphs, for each y (the horizontal axis), a colored-in dot represents $M(x,y)$, and an empty box represents the value of $S(y)$. It appears that $M(x,y) = S(y)$ for

[2]We say that A, and any translate of A, has *length* $\leq y$.

[3]The "$o(1)$" here can be interpreted as saying that for any fixed $\epsilon > 0$, if x is sufficiently large then (3) holds for all $y \leq (1-\epsilon)\log x$.

y up to about $\frac{3}{2}\log x$, and then $M(x,y) = S(y)-2, S(y)-1$ or $S(y)$ for y between $\frac{3}{2}\log x$ and $2\log x$ for these values of x.

More on this in section 2.

1.2. **Intermediate length intervals:** $\log x \le y = o((\log x)^2)$**.** In this range we believe that (1) holds:

$$M(x,y) \sim L(x,y) \text{ where } L(x,y) := \frac{\log x}{\log\left(\frac{(\log x)^2}{y}\right)}.$$

However, when comparing this prediction to the data, it is not obvious how to interpret "$o((\log x)^2)$" for a given x-value. We have made the rather arbitrary choice of $\frac{1}{2}(\log x)^2$ as the upper bound for the y-range. We have also taken $\frac{1}{2}\log x$ as a lower bound which reflects our uncertainty as to whether things can really be predicted so precisely, though we have marked $\log x$ with a dashed line.

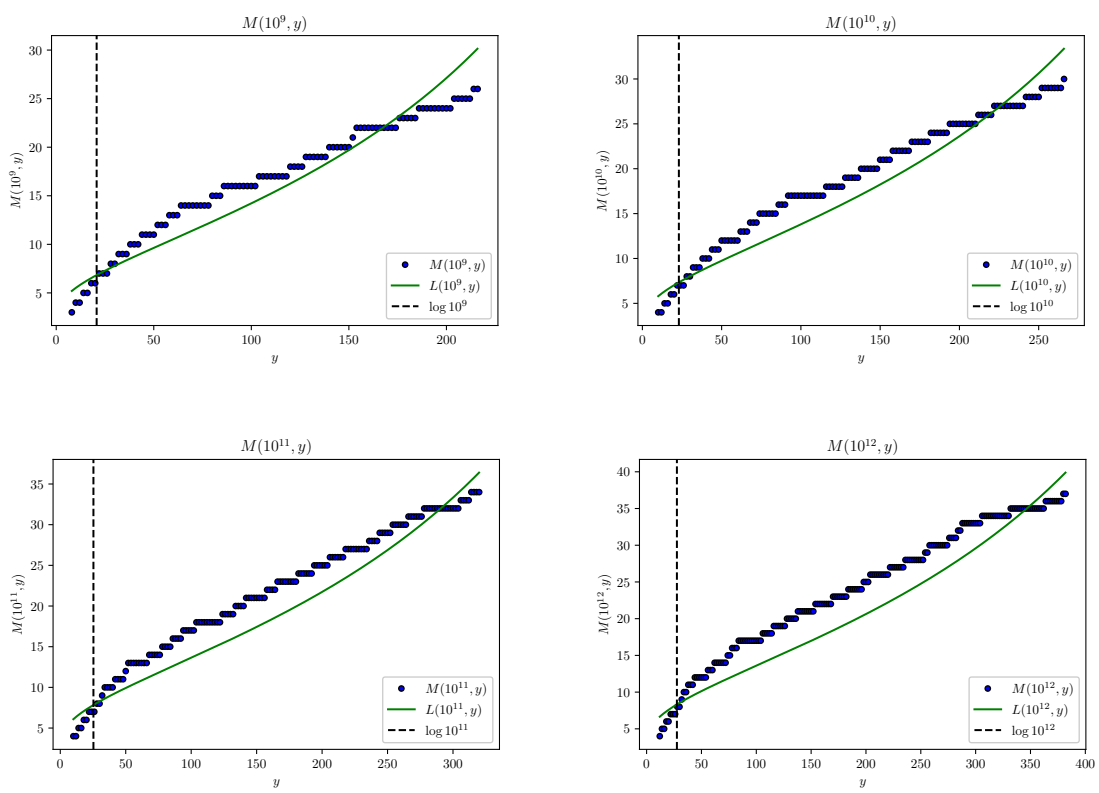

FIGURE 2. $M(x,y)$ vs.$L(x,y)$ for $x = 10^k, k = 9,\dots,12$ and $\frac{1}{2}\log x \le y \le \frac{1}{2}(\log x)^2$. Dashed line at $y = \log x$, which is the end of the range of the $M(x,y) = S(y)$ conjecture.

Here, for each y (the horizontal axis), a colored-in dot represents $M(x,y)$, and the continuous curve $L(x,y)$ (our prediction in (1)). Our prediction and the data seem

to co-incide at $y = \log x$ (where the dashed line is), and again at a point that seems to be slowly increasing (towards $\frac{1}{2}(\log x)^2$) as x grows. The graph indicates that our prediction provides a pretty good approximation to the data in the whole range, though it is concave up whereas the data itself appears to yield a curve that is concave down. We have no explanation for that.

1.3. **The maximum on longer intervals: $y \asymp (\log x)^2$.** Here we mean that $y = t(\log x)^2$ for some fixed value of t. In this range we will need two implicit functions: For every given $t > 0$ consider the equation

$$u(\log u - \log t - 1) + t = 1.$$

We will show that for every $t > 0$ there is a unique solution $u_+(t)$ with $u_+(t) > t$. If $0 < t < 1$ there is no solution in $u \in (0, t)$, so we let $u_-(t) = 0$. If $t > 1$ then there is a unique solution $u_-(t)$ with $0 < u_-(t) < t$. We believe that there exist constants $c_-, c_+ > 0$ such that if $y = t(\log x)^2$ then

$$m(x, y) \sim u_-(c_- t) \log x \text{ and } M(x, y) \sim u_+(c_+ t) \log x. \tag{4}$$

We will see below that $c_\pm$ are constants that can be defined in terms of sieving intervals. We know that $c_+ \geq 1.015\ldots$ and $c_- \leq \frac{e^\gamma}{2} = 0.890536\ldots$, and perhaps both of these inequalities should be equalities.[4] Here is the data for $M(x, y)$ in this range:

[4]We will assume that $c_+ = 1.015\ldots$ and $c_- = 0.8905\ldots$ throughout for the purpose of comparing our conjectures to our data. We will explain the significance of $1.015\ldots$ below.

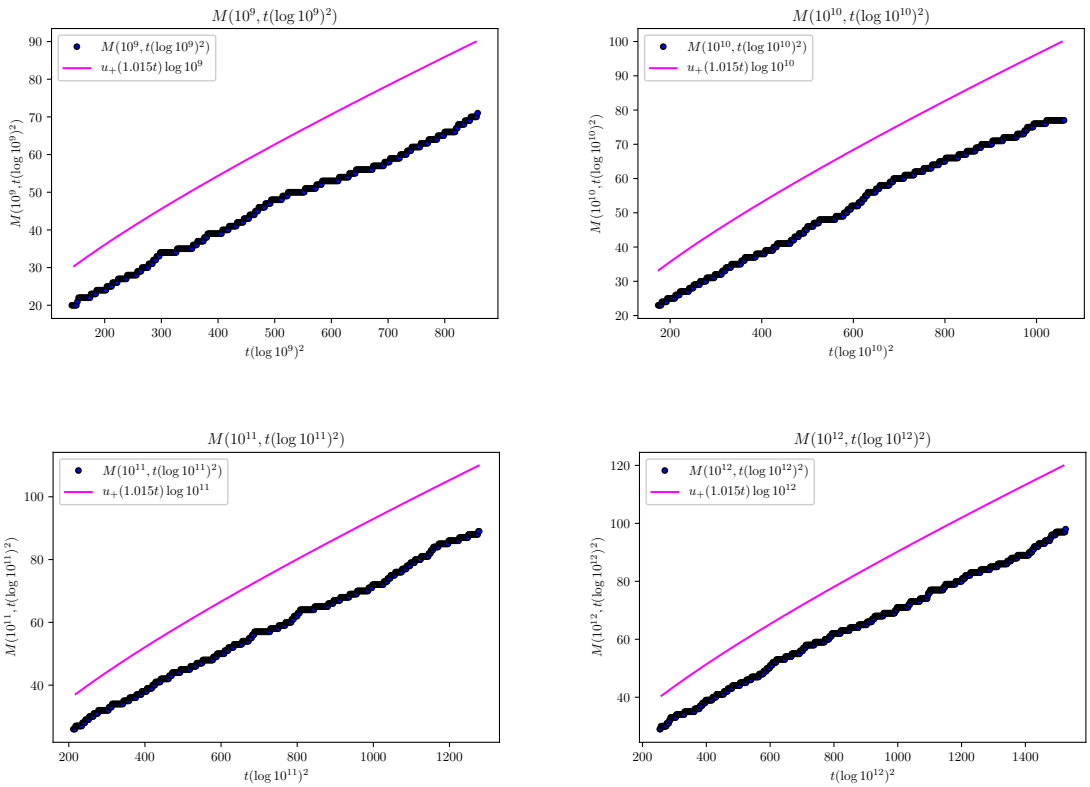

FIGURE 3. $M(x, y)$ vs. $u_+(1.015t)\log x$ where $y = t(\log x)^2$ for $x = 10^k, k = 9, \ldots, 12$ and $\frac{1}{3}(\log x)^2 \le y \le 2(\log x)^2$.

Here, for each y (the horizontal axis), a colored-in dot represents $M(x, y)$, and the red curve represents our prediction $u_+(1.015t)\log x$ where $y = t(\log x)^2$. It appears that this prediction is too large by a factor of about 35% (and if c_+ is larger than 1.015 then the red curve will be even further above the data). However we believe this is a consequence of only calculating up to $x = 10^{12}$. In this range for y it is already well-known that data for (a function of) the minimum does not yet satisfy the standard conjectures:

1.4. **The minimum on longer intervals: $y \asymp (\log x)^2$.** The prediction (4) implies that if $c\_t < 1$ then $m(x, t(\log x)^2) = 0$ but not if $c\_t > 1$. That is, we conjecture the following lower bound for the maximal gap between consecutive primes:

$$\max_{x<p_n\le 2x} p_{n+1} - p_n \sim c_-^{-1}(\log x)^2 \gtrsim 2e^{-\gamma}(\log x)^2;$$

and it is feasible that we have equality here. This is larger than Cramér's original conjecture (that this maximal gap is $\sim (\log x)^2$). As we will discuss, Cramér's reasoning is flawed by failing to take account of divisibility by small primes (a point originally made by the first author back in [7] and recently re-iterated by the in-depth analysis of Banks, Ford and Tao in [1].) However the data does not really support either

conjecture, as the largest gap between consecutive primes that has been found is about $.9206(\log x)^2$ (a shortfall of around 22% from $2e^{-\gamma} \approx 1.1229\cdots$).

p_n	$p_{n+1} - p_n$	$(p_{n+1} - p_n)/\log^2 p_n$
113	14	.6264
1327	34	.6576
31397	72	.6715
370261	112	.6812
2010733	148	.7026
20831323	210	.7395
25056082087	456	.7953
2614941710599	652	.7975
19581334192423	766	.8178
218209405436543	906	.8311
1693182318746371	1132	.9206

FIGURE 4. (Known) record-breaking gaps between primes

In [1] the authors graphed how the maximal gap between primes grows on a graph compared to $2e^{-\gamma}(\log x)^2, (\log x)^2$ and the more precise $(\log x)(\log x - \log\log x)$ which we discuss in section 9.5. Here we graph the first two functions and the best fit functions of the form $\log x(a \log x + b \log\log x + c)$ where $a = 1$ or $2e^{-\gamma}$.

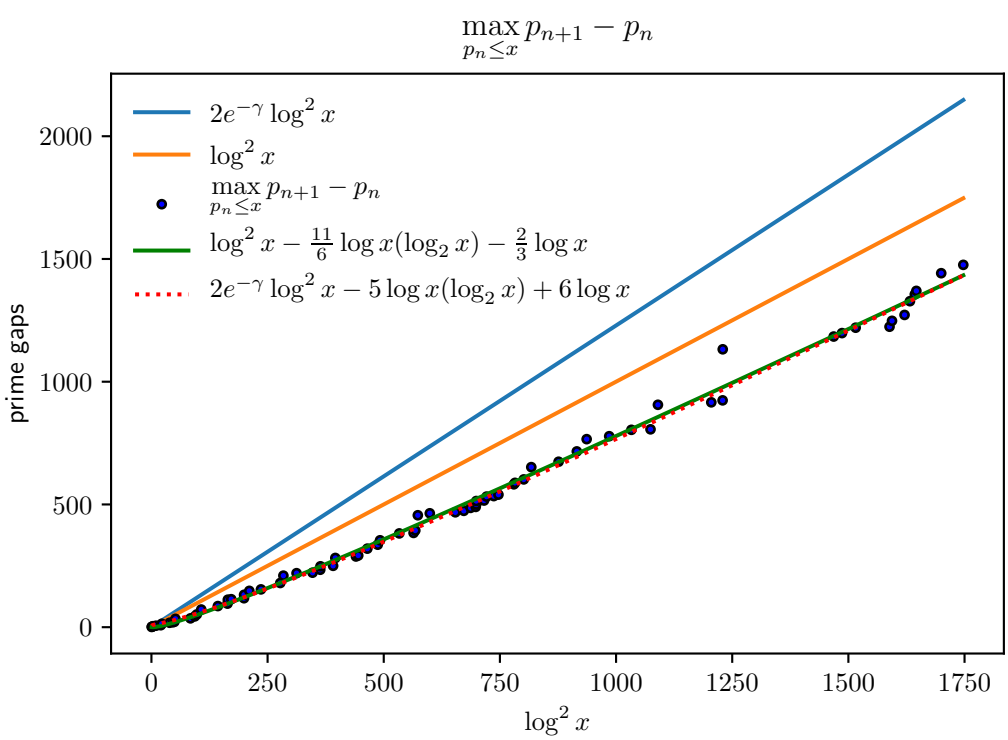

FIGURE 5. $\max_{p_n \le x} p_{n+1} - p_n$ vs. Conjectured approximations

The data for the largest gap between consecutive primes is substantially smaller than our two predictions. No one has suggested a good reason for this shortfall, though in appendix A we explain how at least some of this shortfall is due to the use of asymptotic estimates for primes and sieves, for relatively small values.

In Figure 5, we have also graphed at the best fit to the data of curves of a certain form, and the fit is tight. This suggests that we should be looking harder at possible secondary terms and reasons why they might occur (though we leave that for another occasion).

If (4) really does hold then $m(x,y) \sim u_-(c_-t)\log x$ for $y = t(\log x)^2$, where $u_-(c_-t) = 0$ when $c_-t \le 1$, but $u_-(c_-t) > 0$ for $c_-t > 1$. It is of interest to compare this prediction for $m(x,y)$ to the data, and we will assume that $c_- = \frac{e^\gamma}{2} = 0.8905\ldots$ for the purpose of comparison:

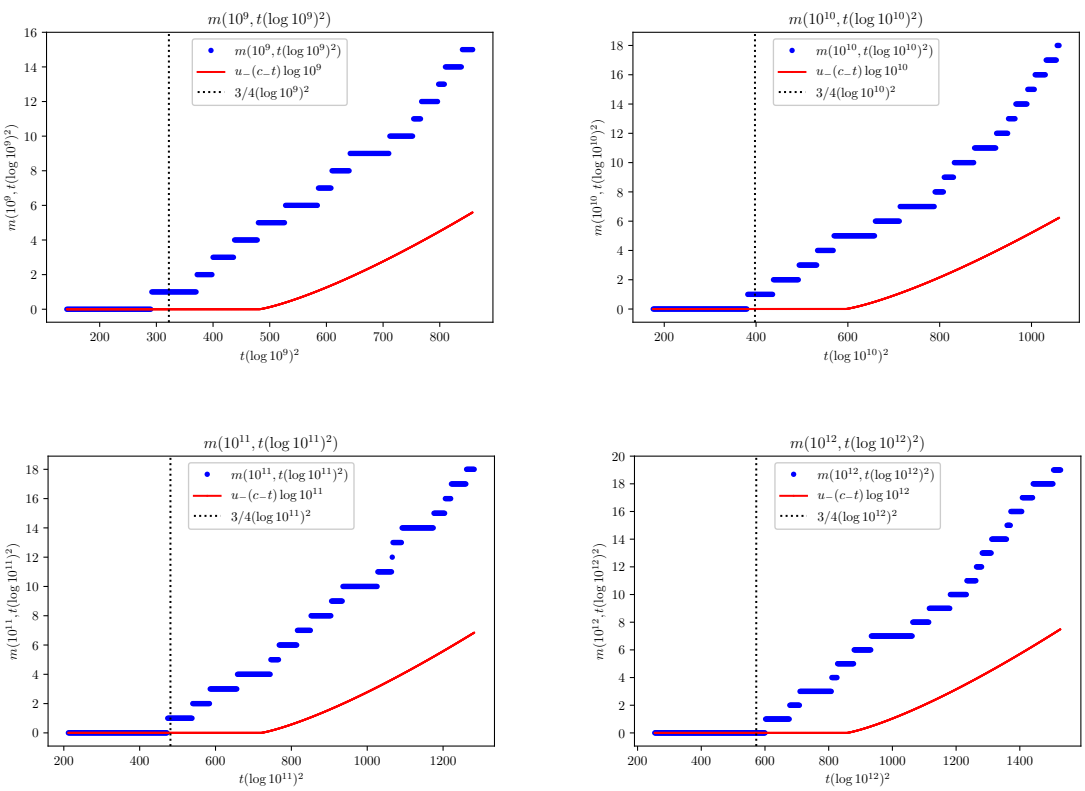

FIGURE 6. $m(x,y)$ vs. $u_-(0.8905t)\log x$ where $y = t(\log x)^2$ for $x = 10^k, k = 9,\ldots,12$ and $\frac{1}{3}(\log x)^2 \le y \le 2(\log x)^2$.

For these values of x it appears that the smallest y for which $m(x,y) > 0$ is at about $y = \frac{3}{4}(\log x)^2$, which is significantly smaller than in the prediction (though the ratio $y/(\log x)^2$ appears to be growing slowly with x). This confirms what we saw in the previous two figures when studying $G(x)$. We plotted the maximum $M(x,y)$ vs

our prediction in this same range in Figure 3 and that data there appears to have a similar shape to our prediction. However it is not obvious here whether the data for the minimum, $m(x, y)$, has a similar shape to our prediction.

It is perhaps of interest to see the data for the minimum and maximum on the same graph, compared to the expected number of primes in a interval of length y between x and $2x$:

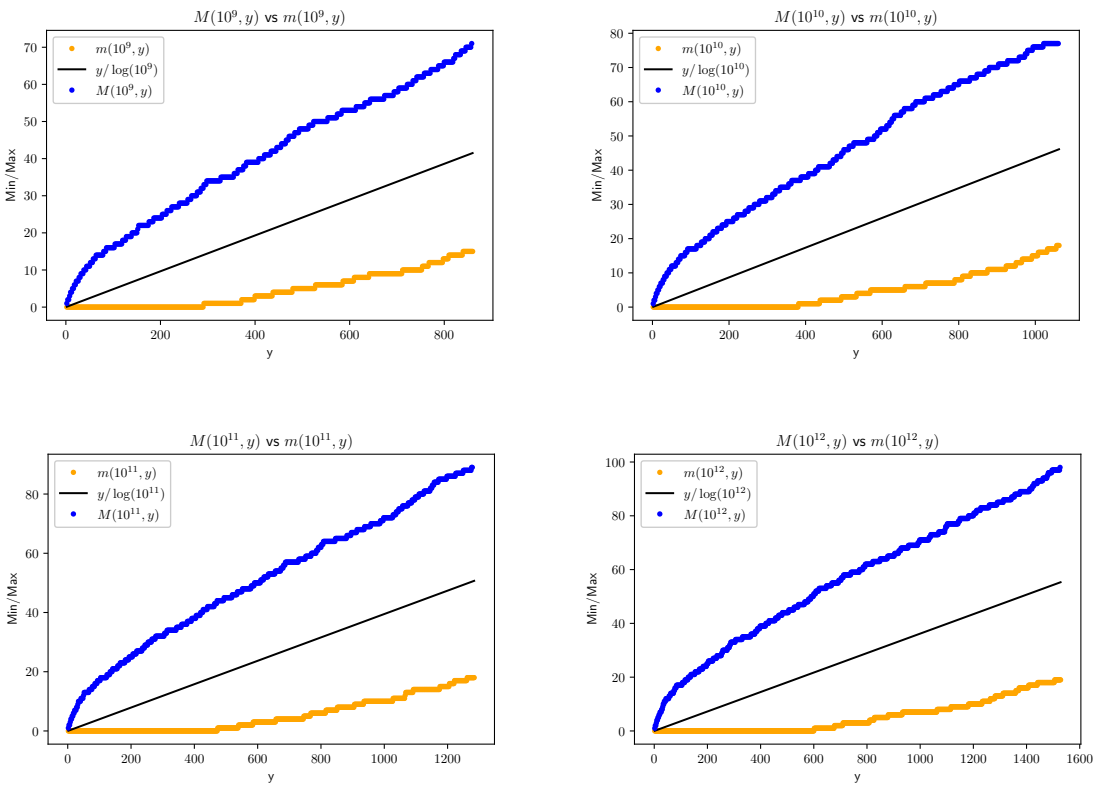

FIGURE 7. $m(x, y)$ vs. $\frac{y}{\log x}$ vs. $M(x, y)$ where $y = t(\log x)^2$ for $x = 10^k, k = 9, \ldots, 12$ and $2 \leq y \leq 2(\log x)^2$.

We now compare our predictions in the range $\frac{1}{3}(\log x)^2 \leq y \leq 2(\log x)^2$ with the data, for both the maxima and the minima, on the same graph, to get a better sense of how well these fit:

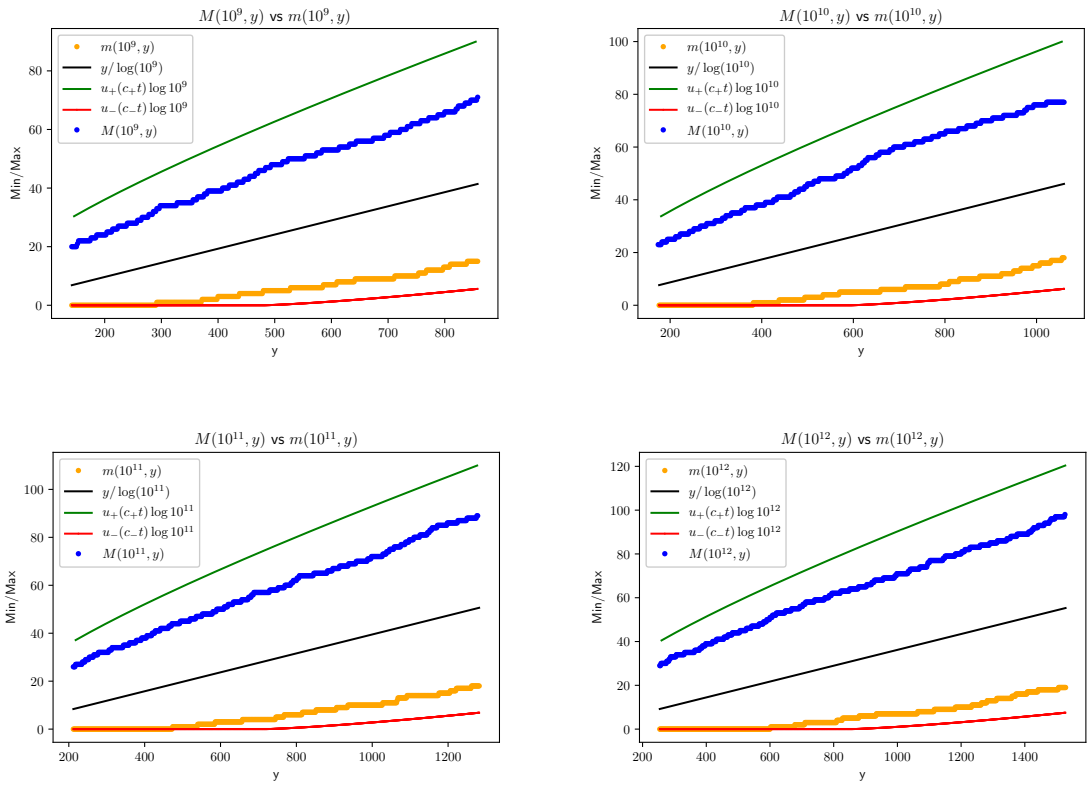

FIGURE 8. $u_-(c_-t)\log x$ vs $m(x,y)$ vs. $\frac{y}{\log x}$ vs. $M(x,y)$ vs $u_+(c_+t)\log x$ in ascending order, where $y = t(\log x)^2$ for $x = 10^k, k = 9, \dots, 12$ and $\frac{1}{3}(\log x)^2 \le y \le 2(\log x)^2$.

We do not know what conclusions to draw from this data!

1.5. **Long intervals:** $y/(\log x)^2 \to \infty$**.** We believe that there exist continuous functions $0 < \sigma_-(A) < 1 < \sigma_+(A)$ for which $\sigma_-(A), \sigma_+(A) \to 1$ as $A \to \infty$, such that if $y/(\log x)^2 \to \infty$ then

$$m(x,y) \sim \sigma_-(A)\frac{y}{\log x} \text{ and } M(x,y) \sim \sigma_+(A)\frac{y}{\log x} \tag{5}$$

writing $y = (\log x)^A$. Moreover we should take

$$c_- = \sigma_-(2) \text{ and } c_+ = \sigma_+(2)$$

above. We will obtain these conjectures from a discussion of sieve theory.

At first sight these conjectures seem to be inconsistent with Selberg's result that

$$\pi(x+y) - \pi(x) \sim \frac{y}{\log x}$$

for almost all x, assuming that $y/(\log x)^2 \to \infty$ (which he proved assuming the Riemann Hypothesis). However the "almost all" in the statement allows for exceptions and in

1984, Maier [13] exhibited, for all $A > 2$, constants $\delta_+(A), \delta_-(A) > 0$ for which there is an infinite sequence of integers x_+ and x_- with

$$m(x_-, y_-) \lesssim \delta_-(A)\frac{y_-}{\log x_-} \text{ and } M(x_+, y_+) \gtrsim \delta_+(A)\frac{y_+}{\log x_+}$$

where $y_\pm = (\log x_\pm)^A$. As far as we know it could be that $\sigma_-(A) = \delta_-(A)$ and $\sigma_+(A) = \delta_+(A)$ for each A, as we will discuss below.

1.6. **Another statistic.** The data in sections 1.1 and 1.2 seem to support our conjectures for $M(x, y)$ in the range $y = o((\log x)^2)$, but the data in sections 1.3 and 1.4 for larger y are less encouraging. For this reason it seems appropriate to return to the question of how $M(x, y)$ grows as a function of y in the range $y \asymp (\log x)^2$, and so we examine the ratio

$$r_+(x, y) := M(x, 2y)/M(x, y).$$

Our *asymptotic predictions* suggest that this looks like $2 + o(1)$ if $y \leq \frac{1}{2}\log x$ and if $y/(\log x)^2 \to \infty$, and $1 + o(1)$ if $\log x \leq y = o((\log x)^2)$. For $y \asymp (\log x)^2$ our prediction for $M(x, y)$ is more complicated; indeed if $y = t(\log x)^2$ then we predict that this looks like

$$\rho_+(t) := u_+(2c_+t)/u_+(c_+t)$$

and we now compare this new statistic to the data:

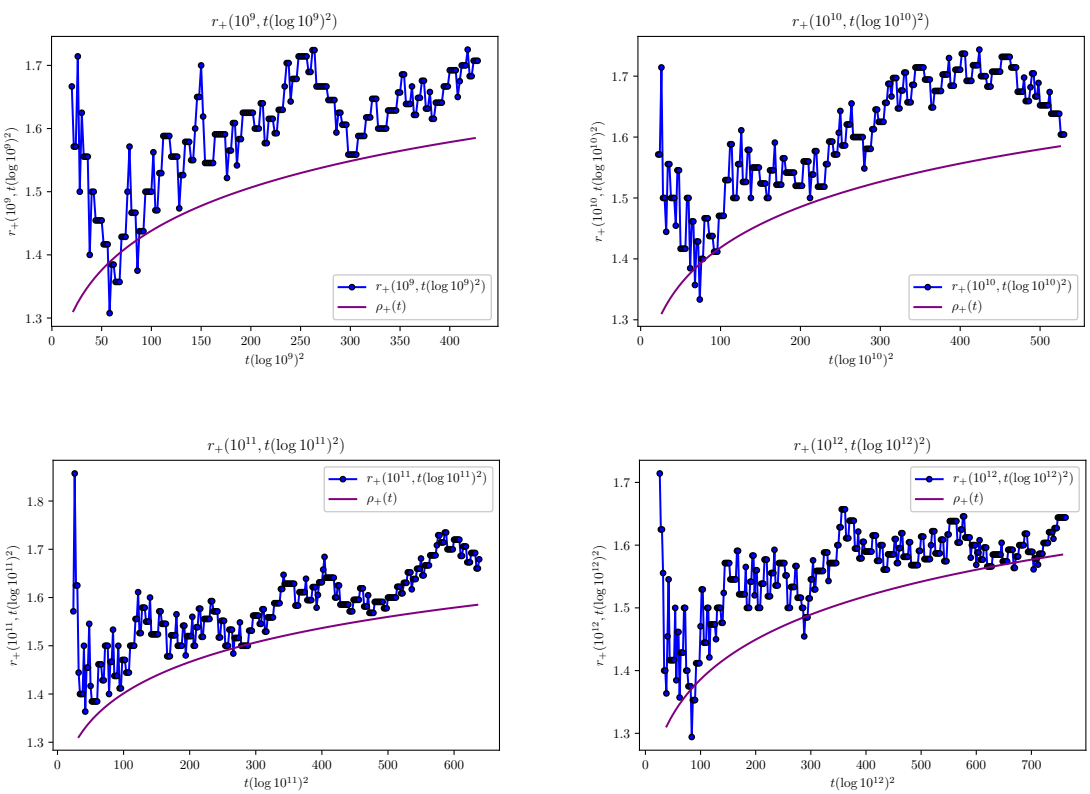

FIGURE 9. $M(10^k, 2y)/M(10^k.y)$ for $k = 9, \ldots, 12$ and $y \leq (\log(10^k))^2$.

We can see the shape of our prediction looks correct but it is a little on the low side. What is encouraging is that the fit seems to get better as k grows.

2. Some historical comparisons

2.1. **Small and Large gaps between consecutive primes.** Following up the 2013 breakthrough by Yitang Zhang [21] on small gaps between primes, Maynard [15] and Tao [20] proved that there are shortish intervals that contain m primes for any fixed m. Their remarkable work implies that

$$M(x,y) \gg \log y,$$

which unfortunately is far smaller than what is conjectured here, in all ranges of y.[5]

Similarly Ford, Green, Konyagin, Maynard and Tao [5], following up on [4, 16], recently showed that

$$m(x,y) = 0 \text{ for some } y \gg \frac{\log x \log\log x \log\log\log\log x}{\log\log\log x},$$

and they believe their technique (which consists of looking only at divisibility by small primes) can be extended no further than y as large as $(\log x)(\log\log x)^{2+o(1)}$ which is far smaller than what is conjectured (here and previously).

2.2. **Unusual distribution of primes in intervals.** As discussed in section 1.5, Maier [13] proved that there can be surprisingly few or many primes in an interval of length $(\log x)^A$ with $A > 2$. His proof can be easily modified to express his result in terms of certain sieving constants: Define

$$S(x,y,z) := \#\{n \in (x, x+y] : \ (n, P(z)) = 1\}$$

where $P(z) := \prod_{p \le z} p$, and let

$$S^+(y,z) := \max_x S(x,y,z) \text{ and } S^-(y,z) := \min_x S(x,y,z).$$

For each fixed $u \ge 1$ we define

$$\sigma_+(u) := \limsup_{z\to\infty} S^+(z^u, z) \Big/ \prod_{p\le z}\left(1 - \frac{1}{p}\right) \cdot z^u$$

$$\text{and } \sigma_-(u) := \liminf_{z\to\infty} S^-(z^u, z) \Big/ \prod_{p\le z}\left(1 - \frac{1}{p}\right) \cdot z^u.$$

We will discuss what we know about the constants $\sigma_-(u)$ and $\sigma_+(u)$ in the next section, although we state here that we believe that

$$S^+(z^u, z) \sim \sigma_+(u) \prod_{p\le z}\left(1 - \frac{1}{p}\right) \cdot z^u \text{ and } S^-(z^u, z) \sim \sigma_-(u) \prod_{p\le z}\left(1 - \frac{1}{p}\right) \cdot z^u. \tag{6}$$

[5]Before Zhang we could only say for $y \ll \log x$, that $M(x,y) \ge 1$, and after Zhang only that $M(x,y) \ge 2$, so this really was a Great Leap Forward.

Maier's proof in [13] can be modified to show that for $y = (\log x)^A$ and $z = \epsilon \log x$ we have

$$M(x,y) \gtrsim S^+(y,z) \cdot \frac{e^\gamma \log z}{\log x}$$

which implies that there exist arbitrarily large x $(= x_+)$ for which

$$M(x,y) \gtrsim \sigma_+(A) \frac{y}{\log x}.$$

Analogously that there are arbitrarily large x $(= x_-)$ for which

$$m(x,y) \lesssim \sigma_-(A) \frac{y}{\log x}.$$

If, as we believe, (6) holds then these estimates are true for all x. In (5) we have conjectured that these bounds are "best possible"; paraphrasing, we are postulating that Maier's observation about the effect of small prime factors is the key issue in estimating the extreme number of primes in intervals with lengths significantly longer than $(\log x)^2$. In fact our conjectures come from firstly sieving by small primes, and secondly looking at the tail probabilities of the binomial distribution that comes from a probabilistic model which takes account of divisibility by small primes.[6]

We will study in Appendix B how well some (relatively small) data for the full distribution compares to reality.

3. Sieve methods and their limitations

Let $\mathcal{A}$ be a set of integers (of size y) to be *sieved* (in our case the integers in the interval $(X, X+y]$), such that

$$\#\{a \in \mathcal{A} : \ d|a\} = \frac{g(d)}{d} X + r(\mathcal{A}, d)$$

where $g(d)$ is a multiplicative function, which is more-or-less 1 on average over primes p in short intervals (in our case each $g(p) = 1$), and the error terms $r(\mathcal{A}, d)$ are small on average (in our case each $|r(\mathcal{A}, d)| \leq 1$). The goal in sieve theory is to give upper and lower bounds for

$$S(\mathcal{A}, z) := \{n \in \mathcal{A} : \ (n, P(z)) = 1\}.$$

This equals $G(z)y$ "on average" where $G(z) := \prod_{p \leq z}(1 - \frac{g(p)}{p})$. In 1965, Jurkat and Richert [12] showed that if $y = z^u$ then

$$(f(u) + o(1)) \cdot G(z)y \leq S(\mathcal{A}, z) \lesssim F(u) \cdot G(z)y, \tag{7}$$

where $f(u) = e^\gamma(\omega(u) - \frac{\rho(u)}{u})$ and $F(u) = e^\gamma(\omega(u) + \frac{\rho(u)}{u})$, and $\rho(u)$ and $\omega(u)$ are the Dickman-de Bruijn and Buchstab functions, respectively. One can define these functions directly by

$$f(u) = 0 \text{ and } F(u) = \frac{2e^\gamma}{u} \text{ for } 0 < u \leq 2$$

[6] In [17] Maynard asks similar questions for integers that are the sum of two squares. He proved unconditionally that there are intervals $(X, X+y]$ which contain $\gg \frac{y}{(\log x)^{1/2}} + y^{1/10}$ integers that are the sum of two squares.

(in fact $F(u) = \frac{2e^\gamma}{u}$ also for $2 < u \le 3$) and

$$f(u) = \frac{1}{u}\int_1^{u-1} F(t)dt \text{ and } F(u) = \frac{2e^\gamma}{u} + \frac{1}{u}\int_2^{u-1} f(t)dt \text{ for all } u \ge 2.$$

Iwaniec [11] and Selberg [19] showed that this result is "best possible" by noting that the sets

$$\mathcal{A}^\pm = \{n \le x : \lambda(n) = \mp 1\}$$

where $\lambda(n)$ is Liouville's function (so that $\lambda(\prod_p p^{e_p}) = (-1)^{\sum_p e_p}$) satisfy the above hypotheses, with

$$S(\mathcal{A}^-, z) \sim f(u) \cdot G(z)\#\mathcal{A}^- \text{ and } S(\mathcal{A}^+, z) \sim F(u) \cdot G(z)\#\mathcal{A}^+. \tag{8}$$

Since our question (bounding $S(x, y, z)$) is an example of this *linear sieve* we deduce that

$$f(u) \le \sigma_-(u) \le 1 \le \sigma_+(u) \le F(u),$$

and we expect that all of these inequalities are strict. However, in [9], it is shown that if there are infinitely many "Siegel zeros",[7] then, in fact,

$$\sigma_-(u) = f(u) \text{ and } \sigma_+(u) = F(u) \text{ for all } u \ge 1.$$

Given that eliminating Siegel zeros seems like an intractable problem for now, we are stuck. However in this paper we are allowed to guess at the truth, but we really know too few interesting examples to even take an educated guess as to the true values of $\sigma_-(u)$ and $\sigma_+(u)$. It is useful though to note the following:

Lemma 1. *$\sigma_+(u)$ is non-increasing, $\sigma_-(u)$ is non-decreasing, and $\sigma_+(u), \sigma_-(u) \to 1$ as $u \to \infty$*

Proof of Lemma 1. Select x so that $S(x, z^B, z) = S^+(z^B, z)$ is attained. For $A < B$, partition the interval $(x, x + z^B]$ into z^{B-A} disjoint subintervals of length z^A, and select the subinterval with $\#\{n \in (X, X + z^A] : (n, P(z)) = 1\}$ maximal. Therefore

$$\begin{aligned} S^+(z^A, z) &\ge \max_{\substack{X = x + jz^A \\ 0 \le j \le z^{B-A}-1}} \#\{n \in (X, X + z^A] : (n, P(z)) = 1\} \\ &\ge \frac{1}{z^{B-A}}\#\{n \in (x, x + z^B] : (n, P(z)) = 1\} = \frac{S^+(z^B, z)}{z^{B-A}}, \end{aligned}$$

so that $\sigma_+(A) \ge \sigma_+(B)$. The analogous proof, with the inequalities reversed, yields the result for σ_-. That these tend to the limit 1 is a well-known result from the theory of the small sieve. □

[7] That is, putative counterexamples to the Generalized Riemann Hypothesis, the most egregious that cannot be ruled out by current methods.

3.1. **Best bounds known.** In Maier's paper he used the well-known fact that for all $u \geq 1$,

$$\#\{n \leq z^u : \ (n, P(z)) = 1\} \sim \omega(u) \frac{z^u}{\log z}$$

where $\omega(u)$ is the *Buchstab function*, defined by $\omega(u) = \frac{1}{u}$ for $1 \leq u \leq 2$, and $(u\omega(u))' = \omega(u-1)$ for all $u \geq 2$. By Lemma 1 we have

$$\sigma_+(A) = \max_{B \geq A} \sigma_+(B) \geq e^\gamma \max_{B \geq A} \omega(B),$$

and, similarly, $\sigma_-(A) \leq e^\gamma \min_{B \geq A} \omega(B)$. For all we know, it could be that

$$\sigma_+(A) = e^\gamma \max_{B \geq A} \omega(B).$$

That is, it could be that the most extreme example of sieving an interval occurs at the beginning, but there is little evidence that there are no other intervals with even more extreme behaviour.

In [14], Maier and Stewart noted one could obtain smaller upper bounds for $\sigma_-(A)$ for small A. Their idea was to construct a sieve based on the ideas used to prove that there are long gaps between primes: Fix $2 > u > 1$. One first sieves the interval $[1, x]$ where $x = z^u$ with the primes in $(z^{1/v}, z]$ where $1 \leq v \leq \frac{1}{u-1}$. The integers left are the $z^{1/v}$-smooth integers up to x, and the integers of the form $mp \leq x$ for some prime $p \in (z, x]$ (note that $m \leq x/p < x/z = z^{u-1} \leq z^{1/v}$). The number of these is

$$\psi(z^u, z^{1/v}) + \sum_{z < p \leq x} \left[\frac{x}{p}\right] \lesssim x\rho(uv) + x \sum_{z < p \leq z^u} \frac{1}{p} \sim x(\rho(uv) + \log u).$$

Next we sieve "greedily" with the primes $\leq z^{1/v}$ so that the number of integers left is

$$\lesssim \prod_{p \leq z^{1/v}} \left(1 - \frac{1}{p}\right) \cdot x(\rho(uv) + \log u) \sim v(\rho(uv) + \log u) \frac{e^{-\gamma} x}{\log z}$$

We now select $v = v_u \in [1, \frac{1}{u-1}]$ to minimize $r_u(v) := v(\rho(uv) + \log u)$. Now

$$r_u(v)' = \rho(uv) + \log u + uv\rho'(uv) = \rho(uv) + \log u - \rho(uv - 1)$$

and so we select v_u so that $r_u'(v_u) = 0$. If $u = 1 + 1/\Delta$ with $1/\Delta = o(1)$ then $v_u \sim \frac{\log \Delta}{\log \log \Delta}$ and so $r_u(v_u) \sim \frac{\log \Delta}{\Delta \log \log \Delta}$. On the other hand if we use the Buchtab function then we cannot obtain a constant smaller than $e^\gamma/2$. Thus for $1 \leq A \leq 2$, we have

$$\sigma_-(A) \leq \min\{e^\gamma/2, r_A(v_A)\}$$

In [14] this argument is extended to show that $r_A(v_A)$ is the minimum exactly when $1 \leq A \leq 1.50046\ldots$. Unfortunately we are only really interested in $\sigma_-(A)$ for $A \geq 2$ in this article.

Now $\omega'(u)$ changes sign in every interval of length 1, so it has lots of minima and maxima, which occur where $\omega(u) = \omega(u-1)$. Nonetheless its global minimum occurs at $u = 2$ so that $\sigma_-(2) \leq e^\gamma \omega(2) = \frac{e^\gamma}{2}$ (and we saw earlier that the linear sieve bounds give $\sigma_-(2) \geq 0$). We are most interested in $\sigma_+(2)$, which is bounded below by $e^\gamma \max_{B \geq 2} \omega(B)$. This maximum occurs at $B \approx 2.75$ with $\omega(B) \approx 0.57$, so that

$\sigma_+(2) \geq 1.015\ldots$ (and we saw earlier that the linear sieve bounds give $\sigma_+(2) \leq e^\gamma = 1.78107\ldots$)

In section 1.3 we have $c_+ = \sigma_+(2)$ and took this equal to $1.015\ldots$ in our computations as this is the best lower bound known on $\sigma_+(2)$. Similarly in section 1.4 we have $c_- = \sigma_-(2)$ and took this equal to $\frac{e^\gamma}{2}$ in our model as this is the best upper bound known on $\sigma_-(2)$. It could be that these are equalities, but there is little evidence either way.

4. Very short intervals ($y \leq \log x$)

A set of integers A is *inadmissible* if there exists a prime p such that for every integer n at least one of $n+a, a \in A$ is divisible by p, and so cannot be prime if n is sufficiently large. On the other hand, Hardy and Littlewood's prime k-tuplets conjecture [10] states that if A is an *admissible set* then there are infinitely many integers n for which $n+a$ is prime for every $a \in A$, and this seems to be supported by an accumulation of evidence.

We are interested in $\pi(n, n+y]$, the number of primes in intervals $(n, n+y]$ of length y (with y small compared to n), particularly the minimum, $m(x,y)$, and the maximum, $M(x,y)$, as n varies between x and $2x$. If the primes in $(n, n+y]$ are $\{n+a : a \in A\}$ with $n > y$, then A is an admissible set, say of size k, and therefore

$$\pi(n, n+y] := \pi(n+y) - \pi(n) = k \leq S(y),$$

where $S(y)$ is the maximum size of an admissible set A of length y. Moreover this implies that if the prime k-tuplets conjecture holds then

$$\max_{n \geq y} \pi(n, n+y] = S(y).$$

How large is $S(y)$? One can show that the primes in $(y, 2y]$ yield an admissible set and so $S(y) \gtrsim \frac{y}{\log y}$ (by the prime number theorem). It is believed that

$$S(y) \sim \frac{y}{\log y}$$

but the best upper bound known is $S(y) \lesssim \frac{2y}{\log y}$ (by the upper bound in (7)), and this upper bound seems unlikely to be significantly improved in the foreseable future (as we again run into the Siegel zero obstruction). Calculations support the believed size of $S(y)$. One interesting theorem, due to Hensley and Richards, is that if y is sufficiently large then $S(y) > \pi(y)$ and so, if the prime k-tuplets conjecture is true then for all sufficiently large y there exist infinitely many intervals of length y that have more primes than the initial interval $[1, y]$. The known values of $S(y)$ and bounds, can be found on `http://math.mit.edu/~primegaps/` and from there we see that $S(3432) \geq 481 > \pi(3432) = 480$. Therefore we believe that there are infinitely many intervals of length 3432 with more primes than at the start, though finding such an interval (via methods based on this discussion) involves finding a prime 481-tuple, which would be an extraordinary challenge unless one is very lucky.

So assuming the prime k-tuplets conjecture we know that $\max_{n \geq y} \pi(n, n+y] = S(y)$ for fixed y, but we might expect that $M(x,y) = S(y)$ for y which (slowly) grows with

x. In sections 4.1 and 8.1 we present two quite different heuristics to suggest that

$$M(x,y) = S(y) \text{ for all } y \le \{1+o(1)\}\log x; \tag{9}$$

and we saw, in section 2.1, that this is well supported by the data that we have.

By a simple sieving argument Westzynthius showed in the 1930s that for any constant $C > 0$ there exist intervals $[x, x + C\log x]$ which do not contain any primes. This argument is easily modified to show that for any $c > 0$

$$m(x, c\log x) := \min_{X\in(x,2x]} \pi(X + c\log x) - \pi(X) = 0 \text{ if } x \text{ is sufficiently large.}$$

We will give two theoretical justifications for our prediction (9), supporting the conclusions we have drawn from the data represented in the graphs above. The first relies on guessing at what point a given admissible set yields roughly as many prime k-tuplets as conjectured; the second a more traditional approach, developing the Gauss-Cramér heuristic by taking account of divisibility by small primes.

4.1. **An explicit prime k-tuplets conjecture.** For a given admissible set of linear forms $b_jn + a_j,\ j = 1,\dots,k$, Hardy and Littlewood [10] conjectured that

$$\#\{x < n \le 2x : \text{ Each } b_jn+a_j \text{ is prime}\} \sim \prod_p \left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\omega(p)}{p}\right)\cdot\frac{x}{(\log x)^k}, \tag{10}$$

where $\omega(p)$ is the number of $n \pmod p$ for which p divides $\prod_{j=1}^k(b_jn + a_j)$. (Here *admissible* can be defined to be those k-tuples for which every $\omega(p) < p$.) We wish to know for what x are the two sides of (10) equal up to a small factor, and for what x can we obtain a good lower bound on the right-hand side.

This conjecture is known to be true as $x \to \infty$ for $k = 1$ (where we may assume that $1 \le a \le b-1$). There is a lot of data on primes in arithmetic progression and these all suggest that (10) holds uniformly for all $x \ge b^\epsilon$ for any fixed $\epsilon > 0$.[8]

Let A be an admissible set of size $k = S(y) \sim \frac{\alpha y}{\log y}$ (where we believe $\alpha = 1$), a subset of the positive integers $\le y$. Since there are $\ll \frac{y}{(\log y)^2}$ integers in $S(y)$ that are $< \frac{y}{\log y}$ (by the sieve), we deduce that $Q := \prod_{a\in A} a = e^{(\alpha+o(1))y} = k^{(1+o(1))k}$. Now $\omega(p) = k$ for all $p \ge y$ (since no two elements of A can be in the same congruence class mod p), so that

$$\prod_{p>y}\left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\omega(p)}{p}\right) = \prod_{p>y}\left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{k}{p}\right) = e^{-o(k^2/y)}.$$

Otherwise $1 \le \omega(p) \le \min\{k, p-1\}$ so that

$$e^{o(k)} = \left(\frac{\log 2y}{\log k}\right)^k \gg \prod_{y\ge p>k}\left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\omega(p)}{p}\right) \ge e^{-o(k)}.$$

[8]Surprisingly there is no way known to try to prove this. The best we know how to obtain, assuming the Generalized Riemann Hypothesis, is that if $k = 1$ then (10) holds for all $x \ge b^{1+\epsilon}$, though this can be obtained "on average" unconditionally. Linnik's Theorem implies that there exists a constant L such that one can obtain a lower bound on the left-hand side of (10) once $x \gg b^L$ ($L = 4$ is the smallest L known in this theorem at the moment).

For the primes $\le k$ we have $p-1 \ge \omega(p) \ge 1$ and so

$$1 \ge \prod_{p\le k}\left(1-\frac{\omega(p)}{p}\right) \ge 1\Big/\prod_{p\le k} p = e^{-k+o(k)}.$$

Therefore, by Mertens' theorem, we expect that

$$\prod_p \left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\omega(p)}{p}\right) = (e^{O(1)}\log k)^k.$$

So the right-hand side of (10) is ≥ 1 when $(C\log k)^k x > (\log x)^k$. This certainly happens when $x = k^{ck}$ for any fixed $c > 1$; that is, $x > Q^{1+\epsilon}$. One might guess that there is an error term in (10) of size $x^{1/2+o(1)}$, in which case we must take $c > 2$, that is $x > Q^{2+\epsilon}$, to guarantee that the left-hand side of (10) is positive.

Now if $\#\{x < n \le 2x : \text{ Each } n+a \text{ is prime, for each } a \in A\} \ge 1$ then $M(x,y) = S(y)$. From the above we might guess this holds when $x > Q^{1+\epsilon}$ where $Q = e^{(1+o(1))y}$; indeed we only need the above heuristic discussion to be roughly correct "on average" over all such admissible sets, which supports the conjecture in (9).[9]

4.2. **A problem to think about.** We would guess that if $p \le \frac{y}{(\log y)^{2+\epsilon}}$ then there is a prime $q \in (y, 2y]$ in each non-zero arithmetic progression mod p. Therefore if $A = \{q - y : q \text{ prime } \in (y, 2y]\}$ then $\omega(p) = p-1$. Perhaps something like this is true for any admissible set A of size $\sim \frac{y}{\log y}$ and length y? We conjecture that for such sets A, and $Y := \frac{y}{(\log y)^2}$ we have

$$\sum_{p\le Y}\frac{p-\omega(p)}{p} \ll \log\log y.$$

5. Cramér's heuristic

Gauss noted from calculations of the primes up to 3 million, that the density of primes at around x is about $\frac{1}{\log x}$. Cramér used this as his basis for a heuristic to make predictions about the distribution of primes: Consider an infinite sequence of independent random variables $(\mathbb{X}_n)_{n\ge 3}$ for which

$$\text{Prob}(X_n = 1) = \frac{1}{\log n} \text{ and } \text{Prob}(X_n = 0) = 1 - \frac{1}{\log n}.$$

By determining what properties are true with probability $1 + o(1)$ for the sequence of 0's and 1's given by $X_3, X_4, \ldots$, Cramér suggested that such properties must also be true of the sequence $1, 0, 1, 0, 1, 0, 0, 0, 1, \ldots$ of 0's and 1's which is characteristic of the odd prime numbers. For example, if N is sufficiently large then

$$S_N := \sum_{n=3}^{N} X_n$$

[9]This reasoning suggests that even if we are pessimistic then we would simply change the range in (9) to $y \le (c + o(1))\log x$ for some constant $c \in (0, 1)$.

has mean $\int_2^N \frac{dt}{\log t} + O(1)$ and roughly the same variance, which suggests the conjecture that $\pi(N) = \int_2^N \frac{dt}{\log t} + O(N^{1/2+o(1)})$; it is known that this conjecture is equivalent to the Riemann Hypothesis. So for this particular statistic, Cramér's heuristic makes an important prediction and it can be applied to many other problems to make equally suggestive predictions.

However Cramér's heuristic does have an obvious flaw: Since it treats all the random variables as independent, we have $\text{Prob}(X_n = X_{n+1} = 1) \approx \frac{1}{(\log n)^2}$, so that

$$\mathbb{E}\Big(\sum_{n=3}^{N-1} X_n X_{n+1}\Big) = \int_2^N \frac{dt}{(\log t)^2} + O(N^{1/2+o(1)})$$

with probability $1 + o(1)$, which, Cramér's heuristic suggests, implies that there are infinitely many prime pairs $n, n+1$. But we have seen this is not so as $\{0, 1\}$ is an inadmissible set. More dramatically this heuristic would even suggest that $M(x, y) = y$ for values of $y \le \{1 + o(1)\} \log x$. From the previous section we know that this is false because $M(x, y) \le S(y)$, as every $\pi(n, n + y]$ is restricted by those integers that are divisible by "small" primes, that is primes $\le y^{1+o(1)}$. This heuristic also suggests that the primes are equi-distributed amongst all of the reduced residue classes modulo a given integer q, rather than just the reduced classes.

It therefore makes sense to modify Cramér's probabilistic model for the primes to take account of divisibility by "small" primes. The obvious way to proceed is to begin by sieving out the integers n that are divisible by a prime $p \le z$ (perhaps with $z = y$), and then apply an appropriate modification of Cramér's model to the remaining integers, that is the integers that have no prime factor $\le z$. The number of such integers up to x is

$$\sim \kappa x \text{ where } \kappa = \kappa(z) := \prod_{p \le z} \Big(1 - \frac{1}{p}\Big)$$

if $z = x^{o(1)}$, and so the density of primes amongst such integers is $\frac{1}{\kappa \log x}$. We therefore proceed as follows:

Define $P = P(z) := \prod_{p \le z} p$ so that $\kappa(z) = \frac{\phi(P)}{P}$, and then an infinite sequence of independent random variables $(\mathbb{X}_n)_{n \ge 3}$ for which $X_n = 0$ if $(n, P) > 1$; and

$$\text{Prob}(X_n = 1) = \frac{1}{\kappa \log n} \text{ and } \text{Prob}(X_n = 0) = 1 - \frac{1}{\kappa \log n} \text{ if } (n, P) = 1.$$

With this model we can again accurately predict the prime number theorem (and the Riemann Hypothesis), as well as asymptotics for primes in arithmetic progressions, for prime pairs, and even for admissible prime k-tuplets (with $k \le z$). Moreover, as we shall see, this will allow us to obtain our predictions for maximal and minimal values of $\pi(x, x + y]$ (including the prediction for $y \ll \log x$ which we already deduced from assuming enough uniformity in the prime k-tuplets conjecture).

If $n \in (x, 2x]$ with $(n, P) = 1$ then $\text{Prob}(X_n = 1) = \frac{1}{L} + O(\frac{1}{L \log x})$ where $L := \kappa \log x$, so for convenience we will work with a model where each $\text{Prob}(X_n = 1) = \frac{1}{L}$. There are, say, N integers in $(X, X + y]$ that are coprime to P where, a priori, N could be

any number between 0 and y (though we can refine that to $0 \le N \le S^+(y,z) \ll \frac{y}{\log z}$ by the sieve). We now develop a model where L and N are fixed:

6. The maxima and minima of a binomial distribution

Suppose that we have a sequence of independent, identically distributed random variables $X_1, \dots, X_N$ with

$$\mathbb{P}(X_n = 1) = \frac{1}{L} \text{ and } \mathbb{P}(X_n = 0) = 1 - \frac{1}{L},$$

where L is large. Let

$$\mathbb{Y} := \sum_{n \le N} X_n.$$

Thus $\mathbb{Y}$ is a *binomially distributed random variable*, which is often denoted $B(N, \frac{1}{L})$.

Proposition 1. *Suppose that $N \ll L \log x$, and that $L \to \infty$ as $x \to \infty$. If $k_- = k_-(N, L, x)$ is the largest integer for which*

$$\mathbb{P}(\mathbb{Y} < k_-) \le \frac{1}{x}$$

then

$$k_- = \begin{cases} 0 & \text{if } N \le \{1 + o(1)\} L \log x; \\ \{\delta_-(\lambda) + o(1)\} \frac{N}{L} & \text{if } N = \{\lambda + o(1)\} L \log x \text{ with } \lambda > 1; \end{cases}$$

where $\delta_- = \delta_-(t)$ is the smallest positive solution to $\delta(\log \delta - 1) + 1 = 1/t$.

If $k_+ = k_+(N, L, x)$ be the smallest integer for which

$$\mathbb{P}(\mathbb{Y} \ge k_+) \le \frac{1}{x}.$$

then

$$k_+ = \begin{cases} N & \text{if } N \le \frac{\log x}{\log L}; \\ \{1 + o(1)\} \frac{\log x}{\log\left(\frac{L \log x}{N}\right)} & \text{if } \frac{\log x}{\log L} \le N = o(L \log x); \\ \{\delta_+(\lambda) + o(1)\} \frac{N}{L} & \text{if } N = \{\lambda + o(1)\} L \log x \text{ with } \lambda > 0; \end{cases}$$

where $\delta_+ = \delta_+(t)$ is the largest positive solution to $\delta(\log \delta - 1) + 1 = 1/t$. We observe that $k_- \le k_+ \ll \log x$ if $N \ll L \log x$.

Proof. From the independent binomial distributions we deduce that

$$\mathbb{P}(\mathbb{Y} = k) = \mathbb{P}\left(\sum_{n \le N} X_n = k\right) = \binom{N}{k} \left(\frac{1}{L}\right)^k \left(1 - \frac{1}{L}\right)^{N-k}.$$

Therefore $\mathbb{P}(\mathbb{Y} = N) = 1/L^N$ and this is $> 1/x$ provided $N \le \frac{\log x}{\log L}$.

Also $\mathbb{P}(\mathbb{Y} = 0) = (1 - \frac{1}{L})^N = e^{-N/L + O(N/L^2)}$ which is $> \frac{1}{x}$ for $N \le \{L + O(1)\} \log x$.[10]

[10] To be more precise we obtain $N \le \frac{\log x}{-\log(1 - \frac{1}{L})} = (L - \frac{1}{2} - \frac{1}{12L} + O(\frac{1}{L^2})) \log x$.

We now estimate the terms in our formula for $\mathbb{P}(\mathbb{Y}=k)$:

$$\binom{N}{k} = \frac{N^k}{k!}\prod_{i=0}^{k-1}\left(1-\frac{i}{N}\right) = \frac{N^k}{(k/e)^k}k^{O(1)}\exp\left(\sum_{i=0}^{k-1}O\left(\frac{i}{N}\right)\right)$$
$$= \frac{N^k}{(k/e)^k}\exp\left(O\left(\frac{k^2}{N}+\log k\right)\right).$$

by Stirling's formula. We also have $(1-\frac{1}{L})^{N-k} = \exp(-\frac{N}{L}+O(\frac{k}{L}+\frac{N}{L^2})$, and so

$$\mathbb{P}(\mathbb{Y}=k) = \left(\frac{eN}{kL}\right)^k\exp\left(-\frac{N}{L}+O\left(\frac{k^2}{N}+\log k+\frac{k}{L}+\frac{N}{L^2}\right)\right)$$

Therefore if $N=o(L\log x)$ and $k=o(\log x)$ then

$$\mathbb{P}(\mathbb{Y}=k) = \left(\frac{eN}{kL}\right)^k x^{o(1)}$$

and this equals $x^{-1+o(1)}$ if and only if

$$k \sim \frac{\log x}{\log(\frac{L\log x}{N})}$$

Finally we deal with the range $N=\lambda L\log x$ with $\lambda>0$. If $k=\delta\lambda\log x$ with $\delta>0$ then, by the above estimate,

$$\mathbb{P}(\mathbb{Y}=k) = \left(\frac{e\lambda\log x}{k}\right)^k\exp\left(-\lambda\log x+O\left(\frac{\log x}{L}\right)\right) = 1/x^{\lambda(1-\delta\log(e/\delta))+o(1)},$$

which equals $1/x^{1+o(1)}$ if $\delta=\delta_\pm(\lambda)$ so that $\lambda(1-\delta\ \log(e/\delta))=1$. □

Remark. *There are well-known bounds on the tail of the binomial distribution (see, e.g., wikipedia) which can be used to obtain this last result:*

$$\frac{1}{\sqrt{8k(1-\frac{k}{N})}}\exp\left(-N\,\mathbb{D}\left(\frac{k}{N}\middle|\frac{1}{L}\right)\right) \le \begin{cases}\mathbb{P}(\mathbb{Y}\le k) & \text{if } k\le\frac{N}{L}\\ \mathbb{P}(\mathbb{Y}\ge k) & \text{if } k\ge\frac{N}{L}\end{cases} \le \exp\left(-N\,\mathbb{D}\left(\frac{k}{N}\middle|\frac{1}{L}\right)\right)$$

where

$$\mathbb{D}(a|p) := a\log\frac{a}{p}+(1-a)\log\frac{1-a}{1-p}$$

which is called the relative entropy *in some circles (this clean upper bound can be obtained by an application of Hoeffding's inequality); the two cases are equivalent since if $k\ge\frac{N}{L}$ then $\mathbb{D}(1-a|1-p)=\mathbb{D}(a|p)$. Using these inequalities we would determine $\delta=\delta(t,L)$ from the functional equation*

$$L\,\mathbb{D}\left(\frac{\delta}{L}\middle|\frac{1}{L}\right) = \frac{1}{t}\left(1+O\left(\frac{\log\log x}{\log x}\right)\right),$$

which is slightly different, but yields $\delta(t,L)=\delta(t)+O(\frac{1}{\log\delta(t)}(\frac{1}{L}+\frac{\log\log x}{\log x}))$, a negligible difference in the ranges we are concerned about.

7. Asymptotics

In section 1.3 we used the solutions $u = u_- \in (0,t)$ and $u = u_+ \in (t,\infty)$ to

$$u(\log u - \log t - 1) + t = 1$$

where $u(t) = t\delta(t)$, and $\delta = \delta_- \in (0,1)$ and $\delta = \delta_+ \in (1,\infty)$ are the solutions to

$$f(\delta) := 1 - \delta\log(e/\delta) = \frac{1}{t}.$$

To verify these claims, we note that $f(0) = 1, f(1) = 0$ and $f(\infty) = \infty$ We have $\frac{df}{d\delta} = \log\delta$ so f (as a function of δ) has its minimum $f(1) = 0$ with $f''(\delta) > 0$ for all $\delta > 0$. Therefore there exists a unique $\delta_- \in (0,1)$ with $f(\delta_-) = 1/t$ for all $t > 1$ and no such δ_- otherwise. Moreover $\delta_-(t)$ is an increasing function with limit 1. Also, there exists a unique $\delta_+ > 1$ with $f(\delta_+) = 1/t$ for all $t > 0$. Moreover $\delta_+(t)$ is a decreasing function with limit 1.

We will now show that $u_+(t)$ is increasing in $t > 0$ and $u_-(t)$ is increasing in $t \geq 1$ Differentiating $f(\delta) = \frac{1}{t}$ we obtain $\log\delta \cdot \frac{d\delta}{dt} = -\frac{1}{t^2}$. Therefore

$$\frac{d}{dt}\log u(t) = \frac{d}{dt}\log t\delta = \frac{1}{\delta}\frac{d\delta}{dt} + \frac{1}{t} = \frac{1}{t} - \frac{1}{t^2\delta\log\delta} = \frac{t\delta\log\delta - 1}{t^2\delta\log\delta} = \frac{\delta - 1}{t\delta\log\delta} > 0$$

for all $\delta > 0$.

We can be more precise about the limits:

7.1. Estimates as $t \to \infty$.

Write $\delta = 1 + \theta$ so that

$$1 - 1/t = (1+\theta)(1 - \log(1+\theta)) = 1 - \frac{\theta^2}{2} + \frac{\theta^3}{6} - \frac{\theta^4}{12} + \dots$$

Therefore $\theta = \pm\frac{2^{1/2}}{t^{1/2}} + \frac{1}{3t} \pm \frac{1}{9(2t)^{3/2}} + O(\frac{1}{t^2})$ as $t \to \infty$, so that

$$u_+(t) = t\delta_+(t) = t + (2t)^{1/2} + \frac{1}{3} + \frac{1}{9\cdot 2^{3/2}t^{1/2}} + O(\frac{1}{t})$$
$$u_-(t) = t\delta_-(t) = t - (2t)^{1/2} + \frac{1}{3} - \frac{1}{9\cdot 2^{3/2}t^{1/2}} + O(\frac{1}{t}),$$

for large t. So if t is large and $N = tL\log x$ then, in Proposition 1,

$$k_\pm = \left(t \pm (2t)^{1/2} + \frac{1}{3} - O\left(\frac{1}{t^{1/2}}\right)\right)\log x \text{ as } t \to \infty.$$

7.2. Approximating the normal distribution.

A random variable given as the sum of enough independent binomial distributions tends to look like the normal distribution, at least at the center of the distribution. However since we are looking here at tail probabilities, the explicit meaning of "enough" is larger than we are used to. To be specific. $\mathbb{Y}$ has mean $\mu := \frac{N}{L}$ and variance $\sigma^2 = \frac{N}{L}(1 - \frac{1}{L})$, and we expect $\mathbb{Y}$ will eventually be normally distributed with these parameters. If so, then

$$\mathbb{P}(\mathbb{Y} < \mu - \tau\sigma), \mathbb{P}(\mathbb{Y} > \mu + \tau\sigma) \approx \frac{1}{\sqrt{2\pi}}\int_\tau^\infty e^{-t^2/2}dt \sim \frac{e^{-\tau^2/2}}{\tau\sqrt{2\pi}}$$

and if this is $\approx 1/x$ then $\tau \sim \sqrt{2\log x}$. Therefore $\tau\sigma \sim (2\frac{N}{L}\log x)^{1/2}$. Writing $N = \lambda L\log x$ we have $\tau\sigma \sim (2\lambda)^{1/2}\log x$. Therefore we might expect the maximum and

minimum values of $\mathbb{Y}$ to be $(\lambda \pm (2\lambda)^{1/2} + o(1)) \log x$. We see from section 7.1 that this is correct as $\lambda \to \infty$ (but not for fixed λ).

We can see this issue more simply: If $k = \kappa N/L$ with $\kappa > 1$ then the binomial distribution gives

$$\mathrm{Prob}(\mathbb{Y} \geq k) \asymp \left(1 - \frac{1}{L}\right)^N \binom{N}{k} \frac{1}{(L-1)^k} = \exp\left(- \frac{N}{L}(\kappa(\log \kappa - 1) + 1 + o(1))\right)$$

and the normal distribution (with the same mean and variance) gives

$$\mathrm{Prob}(\mathbb{Y} \geq k) = \exp\left(- \frac{N}{L}(\tfrac{1}{2}(\kappa - 1)^2 + o(1))\right)$$

and the main terms here are only the same when $\kappa \to 1^+$.

7.3. **Estimates as $t \to 0^+$.** In the other direction we obtain estimates for $\delta_\pm(t)$ as t gets smaller.

If $t \to 0^+$ then we deduce from $\delta_+(\log \delta_+ - 1) + 1 = 1/t$ that

$$\delta_+(t) = \frac{1/t}{\log\left(\frac{1/t}{e \log 1/t}\right)} \left(1 + O\left(\frac{\log\log 1/t}{(\log 1/t)^2}\right)\right) \tag{11}$$

so that

$$u_+(t) = t\delta_+(t) = \frac{1}{\log(1/t)} \left(1 + O\left(\frac{\log\log 1/t}{\log 1/t}\right)\right)$$

and therefore

$$k_+ \sim u_+(t) \log x \sim \frac{\log x}{\log(1/t)} \text{ as } t \to 0^+.$$

Combining this with the second estimate for k_+ in Proposition 1, we deduce that $k_+(N)$ is a continuous function in N in the range of Proposition 1.

If $t \to 1^+$ then writing $t = 1 + \eta$ with $\eta > 0$ small and $\delta_- = 1/B$, we deduce from $\delta_-(1 - \log \delta_-) + 1 = 1 - 1/t$ that $\frac{1 + \log B}{B} = \eta + O(\eta^2)$ and so

$$1/\delta_- = B = (1/\eta)\log(1/\eta)\left(1 + O\left(\frac{\log\log 1/\eta}{\log 1/\eta}\right)\right).$$

This implies that

$$u_-(t) = t\delta_-(t) = \frac{\eta}{\log(1/\eta)} \left(1 + O\left(\frac{\log\log 1/\eta}{\log 1/\eta}\right)\right)$$

and therefore

$$k_- \sim u_-(t) \log x \sim \frac{(t-1)\log x}{\log(\frac{1}{t-1})} \text{ as } t \to 1^+,$$

which $\to 0$ as $t \to 1^+$. This suggests that $k_- = 0$ for $N < \{1 - o(1)\} L \log x$ but grows like

$$\frac{N - L\log x}{L \log \frac{N}{N - L\log x}}$$

for a small range near $L \log x$ which we denote by $L \log x < N < \{1 + o(1)\} L \log x$.

8. Applying the modified Cramér heuristic

Here is the general set-up. For some $z \le y$ define $P = P(z) := \prod_{p \le z} p$ so that $P(z) = e^{(1+o(1))z}$ by the prime number theorem. We define

$$I(N) = \{X \in (x, 2x] : S(X, y, z) = N\}.$$

for each integer N in the range $0 \le N \le S^+(y, z)$. Our heuristic is that the values

$$\pi(X, X + y] \text{ for } X \in I(N),$$

are distributed like the binomially distributed random variable

$$B(N, \tfrac{1}{L}) \text{ where } L = \frac{\phi(P)}{P} \log x.$$

We therefore use Proposition 1 (with x there equal to $\#I(N)$) to predict the value of

$$M_N(x, y) := \max_{X \in I(N)} \pi(X, X + y]$$

for each N with $I(N)$ non-empty. From here we obtain our predictions for

$$M(x, y) = \max_N M_N(x, y).$$

One can work out the details of this heuristic to make precise conjectures provided we can get a good estimate for $\log \#I(N)$. This is not difficult when $z \le \epsilon \log x$: For each $m, 0 \le m \le P - 1$ we have

$$S(X, y, z) = S(m, y, z) \text{ whenever } X \equiv m \pmod{P(z)},$$

since $(X + j, P) = (m + j, P)$ for all j. Moreover these intervals $(X, X + y]$ with $X \equiv m \pmod{P(z)}$ are all disjoint so can be considered to be independent. Therefore if $N = S(m, y, z)$ then $P = P(z) \le x^{\epsilon + o(1)}$ and so

$$\#I(N) \ge \#\{X \in (x, 2x] : X \equiv m \pmod{P(z)}\} = x/P + O(1) \ge x^{1-\epsilon+o(1)}.$$

Hence, when z is this small, the answer given by our heuristic depends only on the extreme values, $S^-(y, z)$ and $S^+(y, z)$.

Getting a good estimate for $\log \#I(N)$ is not straightforward if z (and therefore y) is significantly larger than $\log x$. However one expects our heuristic to be more accurate the larger z is, so we have to find the right balance in our selection of z.

8.1. **Very short intervals ($y \ll \log x$).** If $y \le \eta \log x$ with η small, then the above discussion suggests taking $z = y$. Hence $S^+(y, z) = S^+(y, y) = S(y)$. For each $m \pmod P$ we apply Proposition 1 with

$$N = S(m, y, y),\ L = \frac{\phi(P)}{P} \log x, \text{ and } x \text{ replaced by } x^{1-\eta}.$$

For fixed L and x, one obtains the largest value of k_+ in Proposition 1, when N is as large as possible. This happens here when $N = S(y)$, which we believe is $\sim \frac{y}{\log y}$ and know is no more than twice this. Now $\log L \sim \log y$ and Proposition 1 then implies that $k_+ = N = S(y)$ as long as $S(y) \lesssim (1 - \eta)\frac{\log x}{\log y}$, which should be true for $\eta < \frac{1}{2}$ (and at worst for $\eta < \frac{1}{3}$).

This supports the conjecture (9) in a range like $y \le (\frac{1}{2} - o(1)) \log x$. What about for larger y?

8.2. **Larger y with a different choice of intervals.** For larger y, say $\log x \ll y < (\log x)^A$ with $A > 2$, we need to decide how to select our value for z. One might guess that the right way to do so is to take $z = y$.[11] That is, to sieve the intervals of length y with all of the primes $\le z = y$, and then apply the modified Cramèr model. In this case the sets $\{j \in [1, y] : (X + j, P) = 1\}$ are probably different for every $X \in (x, 2x]$ (certainly they do not repeat periodically as in the earlier subsection), which seems difficult to cope with. However we do not need to understand these sets so precisely, we only need to understand their size, that is, to have good estimates for $\log \#I(N)$ for each N, but even this seems to be out of reach. Therefore this is the less desirable option (though we work through some of the details in Appendix C). In general, we do not know how to get good estimates for $\log \#I(N)$ whenever z is substantially larger than $\log x$.

These (for now insurmountable) issues, suggest that we should proceed as before, with a smallish value of z, so as to recover the sieved sets repeating predictably. Therefore we pre-sieve the intervals of length y with all of the primes $\le z := \epsilon \log x$, and then apply the modified Cramèr model. There might be a substantial difference when sieving with the primes $\le z$, as opposed to y, though we hope not. If there is a substantial difference then this needs further investigation.

8.3. **Larger y; Predictions by pre-sieving up to $z = o(\log x)$.** We pre-sieve with the primes up to $z = \epsilon \log x$ where $\epsilon \to 0$ very slowly as $x \to \infty$. In this case we have seen that we may cut to the chase by taking

$$N_+ = S^+(y, z) =: e^{-\gamma} \frac{y}{\log\ z} c_+ \text{ and } L = \frac{\phi(P)}{P} \log x \sim e^{-\gamma} \frac{\log x}{\log \log x}$$

Prediction: Pre-sieving up to $z = \epsilon \log x$: *If $\log x \ll y = o((\log x)^2)$ then*

$$M(x, y) = \min \left\{ S^+(y, z), \{1 + o(1)\} \frac{\log x}{\log\left(\frac{(\log x)^2}{y}\right)} \right\}$$

If $y = \lambda(\log x)^2$ with $\lambda > 0$ then

$$M(x, y) \sim u_+(\lambda c_+) \log x \text{ and } m(x, y) \sim u_-(\lambda c_-) \log x$$

If $y \asymp \log x$ then this might predict that $M(x, y) = S^+(y, z) > S(y)$ which is obviously false (though not by much) – in this range it therefore makes sense to sieve up to $z = y$, which will assure the feasible prediction $M(x, y) = S(y)$ (as we work out in Appendix C).

If λ is large and $y = \lambda(\log x)^2$ then

$$u_+(\lambda c_+) = \lambda c_+ + \sqrt{2\lambda c_+} + O(1),$$

[11] We do not wish to sieve with primes larger than the length of the interval, since any larger primes cannot divide more than one element in an interval of length y, so cannot be helpful in a sieve argument.

and so $M(x, \lambda(\log x)^2) \sim c_+ \frac{y}{\log x}$ as $\lambda \to \infty$; and analogously $m(x, \lambda(\log x)^2) \sim c_- \frac{y}{\log x}$.

Deduction from the predictions of Proposition 1. We apply Proposition 1 to predict, for each $0 \le j \le P-1$ where $P = P(z)$,

$$M_j(x,y) := \max_{\substack{X \in (x,2x] \\ X \equiv j \pmod P}} \pi(X+y) - \pi(X)$$

and then we guess that $M(x,y) = \max_j M_j(x,y)$. We observe that $\#\{X \in (x,2x] : X \equiv j \pmod P\} = \frac{x}{P} + O(1) = x^{1-o(1)}$ for each j, so we apply Proposition 1 to a set of this size, and the result follows directly. The analogous proof works for $m(x,y)$. □

9. Which choices should we make?

We will now distill these discussions, which each yield slightly different predictions.

9.1. Very short intervals ($y \ll \log x$).

In section 5.1, we predicted that if $y \ll \log x$ then $M(x,y) = S(y)$, and this was confirmed in sections 5.3 and 5.4. We also obtained the same prediction in section 2, with a very different approach.

From all three discussions it is not obvious what explicit constant one should take in place of the inexplicit "$\ll$". Our guess is that

$$M(x,y) = S(y) \text{ for } y \le (1-o(1)) \log x,$$

and

$$M(x,y) \sim \frac{\log x}{\log\log x} \text{ for } y \ge (1-o(1)) \log x.$$

The "$o(1)$" is inexplicit but our methods do not seem able to pinpoint the transition more accurately. The data represented in the graphs at the end of section 2 appear to more-or-less conform these predictions; indeed the correct constant for the transition in all of these examples is > 1.[12]

9.2. Intermediate length intervals ($\log x \le y = o((\log x)^2)$).

In the range $\log x \le y = o((\log x)^2)$ we predict (1) no matter whether we presieve up to z or up to y.

One can revisit the heuristic arguments above to try to get a more accurate approximation: By (11) we believe that if $y = \lambda(\log x)^2$ with $\lambda \to 0$ then

$$M(x,y) \text{ is better approximated by } \frac{\log x}{\log\left(\frac{1/\lambda}{e \log 1/\lambda}\right)}.$$

However the data for this prediction is no more compelling then for the less precise prediction in this range, presumably because x is so small.

[12] Though this cannot happen asymptotically without $M(x,y)$ looking significantly different from our predictions, as $M(x,y)$ is non-decreasing in y by definition.

9.3. **Comparatively long intervals ($y/(\log x)^2 \to \infty$ with $y \le x$).** Here we write $y = (\log x)^A$ with $A \ge 2$ and understanding that if $A = 2$ then $y/(\log x)^2 \to \infty$. If (6) holds then Proposition 1 suggests that

$$M(x,y) \sim \sigma_+(A)\frac{y}{\log x} \text{ and } m(x,y) \sim \sigma_-(A)\frac{y}{\log x}$$

which is what we believe.

If we were to pre-sieve up to y then Proposition 1 suggests that one should make a similar prediction but with $\sigma_+(A)$ replaced by

$$\max_{x<X\le 2x} \#\{j \le y : \ (X+j, P(y)) = 1\} \Big/ \frac{\phi(P(y))}{P(y)} y.$$

(and $\sigma_-(A)$ by the analogous expression with the min). However we have no idea how to study this ratio in this restricted range for X.

9.4. **Longish intervals ($y \asymp (\log x)^2$).** In section 5.3 we saw that if $y = \lambda(\log x)^2$ then we should expect that

$$M(x,y) \sim u_+(c_+\lambda) \cdot \log x$$

Now $u_+(c_+\lambda) \sim c_+\lambda$ as $\lambda \to \infty$ and so $M(x,y) \sim c_+ \frac{y}{\log x}$. This implies, letting $\lambda \to \infty$ and comparing this prediction to that in the last subsection, that $c_+ = \sigma_+(2)$.

Following the same heuristic but now focusing on the minimum we see that if $y = \lambda(\log x)^2$ then we should expect that

$$m(x,y) \sim u_-(c_-\lambda) \cdot \log x$$

for some constant $c_- > 0$. This analogously yields that $c_- = \sigma_-(2)$.

9.5. **More precise guesses for the maximal gap between primes.** We can be more precise about our prediction for gaps between primes using the footnote in the proof of Proposition 1. The estimate there $N \le (L - \frac{1}{2} + o(1)) \log x$ with $L = \frac{\phi(P)}{P} \log x$ which would suggest that

$$\max_{x<p_n\le 2x} p_{n+1} - p_n \approx c_-^{-1} \log x \left(\log x - \frac{1}{2}\frac{P}{\phi(P)} \right) \approx c_-^{-1} \log x \left(\log x - \tfrac{1}{2} \log\log x \right).$$

Here $P = P(z)$ and c_- depend on z.

Cadwell [2] presented a variant of Cramér's model. He took the viewpoint that the certain aspects of the distribution of $H := \pi(2x) - \pi(x)$ primes in $(x, 2x]$ can be assumed to be like the distribution of H randomly selected integers in $(x, 2x]$. He very elegantly proved that the expected largest gap has length $\frac{x}{H+1}(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{H+1})$. This can be used to predict that[13]

$$\max_{x<p_n\le 2x} p_{n+1} - p_n \approx \log(4x/e)(\log x - \log\log x + \gamma).$$

[13]Cadwell's conjecture $\log x(\log x - \log\log x)$ for the largest prime gap $\le x$ was briefly mentioned in section 1.4. However since $x/\pi(x)$ is more accurately approximated by $\log x - 1$, a famous correction of Legendre's prediction by Gauss, he should have deduced $(\log x - 1)(\log x - \log\log x)$ from his model!

It is not clear how to incorporate divisibility by small primes into this argument, particularly working only with those intervals with an unexpectedly small number of integers left unsieved.

There are some similarities in these two conjectural formulas but it is not clear which to choose and on what basis. We did see in Figure 5 that the data suggests that one should subtract a larger multiple of $\log\log x$ in the formulas above but we have not found a believable heuristic to do so, though finding a way to combine the two heuristics would be a good start.

10. Short arithmetic progressions

We can proceed similarly with the distribution of $\pi(qy; q, a)$, the number of primes among the smallest y positive integers in the arithmetic progression $\equiv a \pmod q$, as we vary over reduced residue classes $a \pmod q$ and where y is small compared to q. As before we sieve out with the primes $\le z$ (that do not divide q) before trying to find primes. If $P_q(z) := \prod_{p\le z,\ p\nmid q} p$ then the probability that a random such integer of size $q^{1+o(1)}$ is prime is

$$\sim \frac{qP_q(z)}{\phi(qP_q(z))}\frac{1}{\log q}$$

Now the number of unsieved integers in such an interval of length y is expected to be

$$\frac{\phi(P_q(z))}{P_q(z)}y,$$

and so the "expected" number of primes is

$$\sim \frac{q}{\phi(q)}\frac{y}{\log q}$$

(which is what suggested by the prime number theorem for arithmetic progressions). This set up allows us to proceed much as in the questions about primes on short intervals. We shall explore this in detail, with copious calculations, in a subsequent article.

Appendix A. The largest prime gap conjecture in computing range

In section 1.4, particularly in figure 5, we saw that our predictions for $G(x)$ appear to be significantly too large. The technique we used to make our prediction involves several asymptotic predictions for the distribution of primes and for the sieve and so any of these may be sufficiently far out for small integers that it might have led to the shortfall that we have seen. Our belief is that the main issue is the sieving and not the probabilistic argument and so we test that in this section. We take an example near to the end of what is currently computable:

We take $\log x = 40$: The largest prime gap up to x is 1248 immediately following 218034721194214273. The Cramér prediction is 1600 and ours is 1797. We follow the argument in this paper:

We want to determine the maximal gap y which should be (at a first guess) around $(\log x)^2 = 1600$ (at least according to Cramér), so we will now study sieving all intervals of length 1600 with the primes $\leq z = \frac{1}{2}\log x = 20$. Define

$$R(n) := \#\{X \pmod{P(20)} : S(X, y, z) = n\} \text{ where } n =: c_n \prod_{p \leq 20} \left(1 - \frac{1}{p}\right) y.$$

The arguments herein suggest that the maximal gap between primes is then

$$y \approx \max_n \frac{1}{c_n} \log(xR(n)/P(20)) \log x = \max_n \frac{20}{c_n}(23.91 + \log R(n)).$$

We can easily determine this function for each n on a computer, and from this we obtain a prediction of $y = 1536$, significantly smaller than either previous prediction, but still unaccountably larger than the truth. The data for each n is given in the following:

n	$R(n)$	$\frac{20}{c_n}(23.91 + \log R(n))$
254	44	1145.7
255	832	1262.4
256	6492	1341.8
257	36084	1406.7
258	137772	1455.8
259	386656	1492.0
260	823184	1516.8
261	1357900	1531.2
262	1772334	1536.0
263	1831984	1531.5
264	1513080	1518.1
265	992804	1495.6
266	516116	1464.2
267	221324	1425.4
268	75612	1377.9
269	21232	1323.2
270	4776	1260.2
271	1132	1199.8
272	232	1134.1
273	80	1089.0

FIGURE 10. Data when $y = 1536$

This shows that there are about 1.77 million intervals mod $P(20)$ of length 1536 which contain exactly 262 integers that are coprime to $P(20)$. The probabilistic argument then suggests that some of the corresponding intervals in $(x, 2x]$ contain no primes at all. If instead we work with $P(25)$ then our prediction reduces to 1531, not a big difference, and indeed we tried all the obvious possibilities and could not reduce the number significantly.

Appendix B. Is the model valid?

B.1. **A first example,** $x = 10^8, y = 340, z = 11$**.** For $x = 10^8$ we are going to study the distribution of primes in intervals of length $y = 340 \approx (\log x)^2$, which lie between x and $2x$, grouping them according to the value of $S(X, y, z)$ where $z = 11$.

A quick calculation reveals that $S(X, 340, 11)$ takes each value between 68 and 73. Let $C(N) := \#\{m \pmod P : S(m, y, z) = N\}$. As discussed in section 8, we have $S(X, y, z) = S(m, y, z)$ whenever $X \equiv m \pmod{P(z)}$, so that

$$I(N) = \bigcup_{m \in C(N)} \{X \in (x, 2x] : X \equiv m \pmod P\},$$

and therefore $\#I(N) = \frac{x}{P}\#C(N) + O(P)$. A simple calculation yields that $P(11) = 2310$ with

$$\#C(68) = 28, \#C(69) = 228, \#C(70) = 784, \#C(71) = 820, \#C(72) = 386, \#C(73) = 64$$

For each $N \in [68, 73]$ we define, for each integer h,

$$I(N, h) := \{X \in I(N) : \pi(X + y) - \pi(X) = h\}.$$

Then we create the bar graph where the column rooted at h on the vertical axis has height $\#I(N, h)$.

We wish to compare this to our assumptions, and to the binomial distribution. The first thing we might want to look at is how the sieving effects the probability of being prime. Thus if $\mu(N)$ is the calculated mean number of primes in an interval in $I(N)$, then we are interested in the probability of an unsieved integer being prime, namely $1/L(N)$ where $L(N) = N/\mu(N)$. In our model we would take $L = \frac{\phi(P)}{P} \log x = 3.82767\ldots$; but to compare this to small data we need to be more precise, noting that a better approximation to

$$\frac{x}{\pi(2x) - \pi(x)} \text{ is given by } \log 4x/e,$$

and using this we have $L = \frac{\phi(P)}{P} \log 4x/e = 3.90794\ldots$ Our data yields

$$L(68) = 3.8665\ldots, L(69) = 3.8847\ldots, L(70) = 3.8977\ldots,$$
$$L(71) = 3.9133\ldots, L(72) = 3.9265\ldots, L(73) = 3.9418\ldots,$$

which are all reasonably close to L (no more than about 1% out). The L-values here appear to be growing, more or less linearly, which deserves an explanation. A 'best fit' approximation yields that $L(N) \approx L + .0.01478(N - 70.69)$.

Next we compare what the binomial distribution predicts to the actual counts for primes when $S(X, y, z) = N$. Here N runs from 68 to 73 and we graph $I(N, h)$ compared to the prediction

$$\binom{N}{h} \frac{1}{L^h} \left(1 - \frac{1}{L}\right)^{N-h}$$

from the binomial distribution. We also mark the mean $\mu(N)$ number of primes in these intervals, as well as $m_N(x, y), M_N(x, y)$, the minimum and maximum number of primes in such intervals, and $m(x, y), M(x, y)$, the global minimum and maximum.

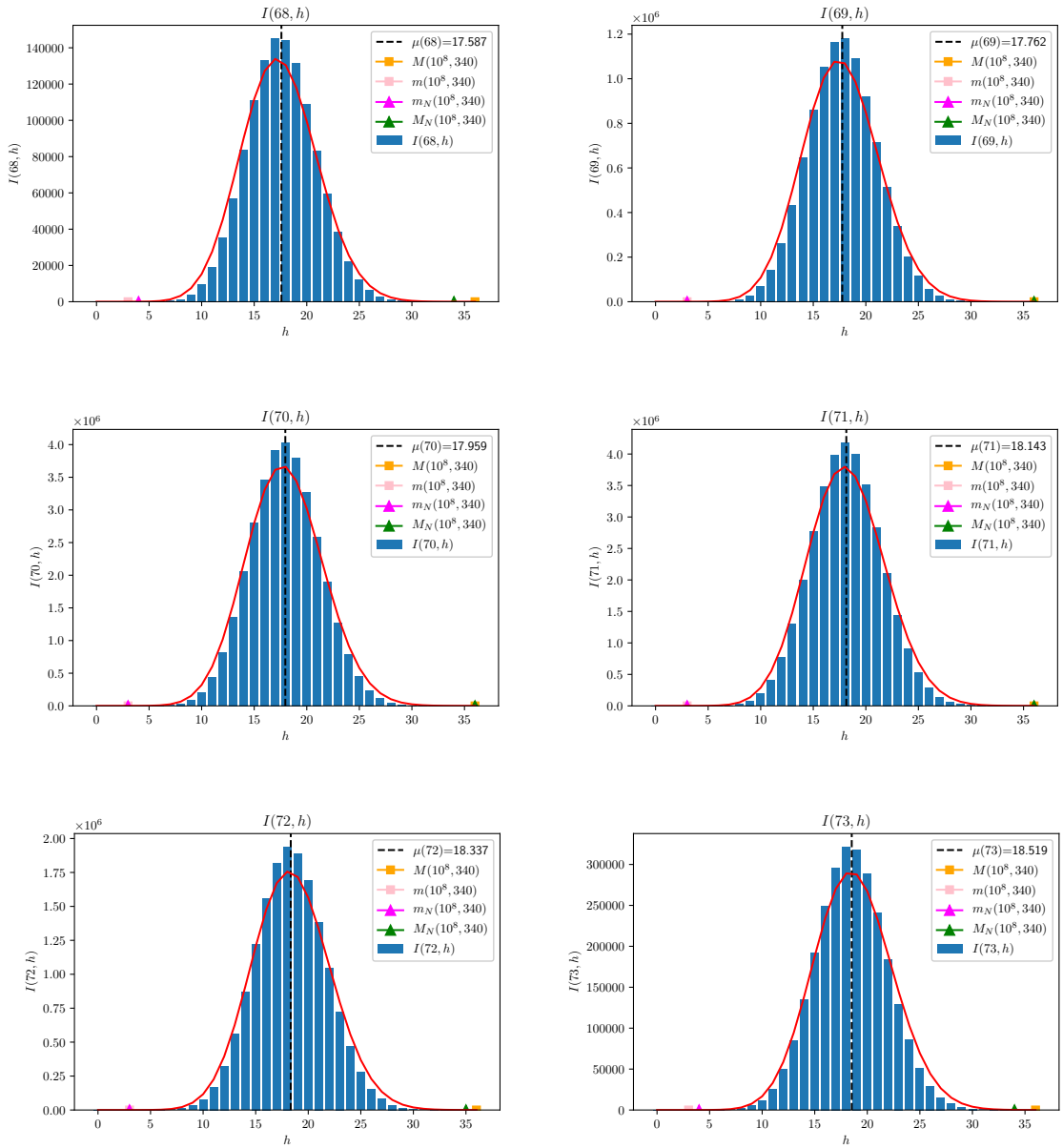

FIGURE 11. Testing the distributions, h vs $I(N,h)$, for each N in our range.

In each case we see that our prediction has the same basic shape as the data (a Bell curve) but is wider than the data, with less density around the mean. We can analyze this by simply looking at the mean and variance compared to what is expected from our model.

N :	68	69	70	71	72	73
Expected mean:	17.40	17.66	17.91	18.17	18.42	18.68
Actual mean:	17.59	17.76	17.96	18.14	18.34	18.52
Expected variance:	12.95	13.14	13.33	13.52	13.71	13.90
Actual variance:	10.82	10.93	11.06	11.17	11.25	11.34

Although both the actual and expected means increase with N we see that the actual mean increases more slowly than the expected. More striking is that the actual variance, that is the variance given by the data, is far smaller than in our prediction.

According to Montgomery and Sound [18] we should have

$$\sum_{X=x}^{2x}(\psi(X+y)-\psi(X)-y)^{2k} \sim y^k \cdot \int_{t=x}^{2x}\left(\log\frac{e^{-\gamma}t}{2\pi y}+1\right)^k dt$$

for $\log x \le y \le x^{1/2k}$. Therefore the variance here (for the primes) is, more-or-less

$$\frac{y}{x(\log x)^2}\cdot\int_{t=x}^{2x}\left(\log\frac{e^{-\gamma}t}{2\pi y}+1\right)dt = \frac{y}{\log x}\cdot\frac{\log\frac{2e^{-\gamma}x}{\pi y}}{\log x}$$

Thus a first approximation gives mean $\frac{y}{\log x} \approx 18.46$ and variance ≈ 11.586. If we replace $\log x$ by $\log 4x/e$ (since this gives a more accurate description of the density of primes in $[x, 2x]$) then we get ≈ 18.08 and ≈ 11.11, respectively. This corresponds very well to the data.

B.2. **A second example,** $x = 10^8, y = 500, z = 17$**.** Here $S(X, 500, 17)$ takes each value between 84 and 97. Now $P(17) = 510510$ and the C-values are given by

h	84	85	86	87	88	89	90
$\#C(h)$	52	576	3764	15836	47186	91432	125688

h	91	92	93	94	95	96	97
$\#C(h)$	115800	70096	29428	8050	1520	212	28

We see that there are very few such intervals for the outlying h-values, and indeed the data for these h-values does not conform to the patterns that we observe.

We have that $L = \frac{\phi(P(z))}{P(z)}\log(4x/e) = 3.39513\ldots$ and our data yields the following L-values to four decimal places

h	84	85	86	87	88	89	90
$L(h)$	3.3853	3.3805	3.3845	3.3843	3.3873	3.3906	3.3938

h	91	92	93	94	95	96	97
$L(h)$	3.3974	3.4011	3.4043	3.4062	3.4082	3.4156	3.4450

Again it is usually with 1-2% of the true L-value, but is slightly increasing. Our best linear approximation is $L(N) \approx L + .003054(N - 90.09)$. The corresponding graphs are given by

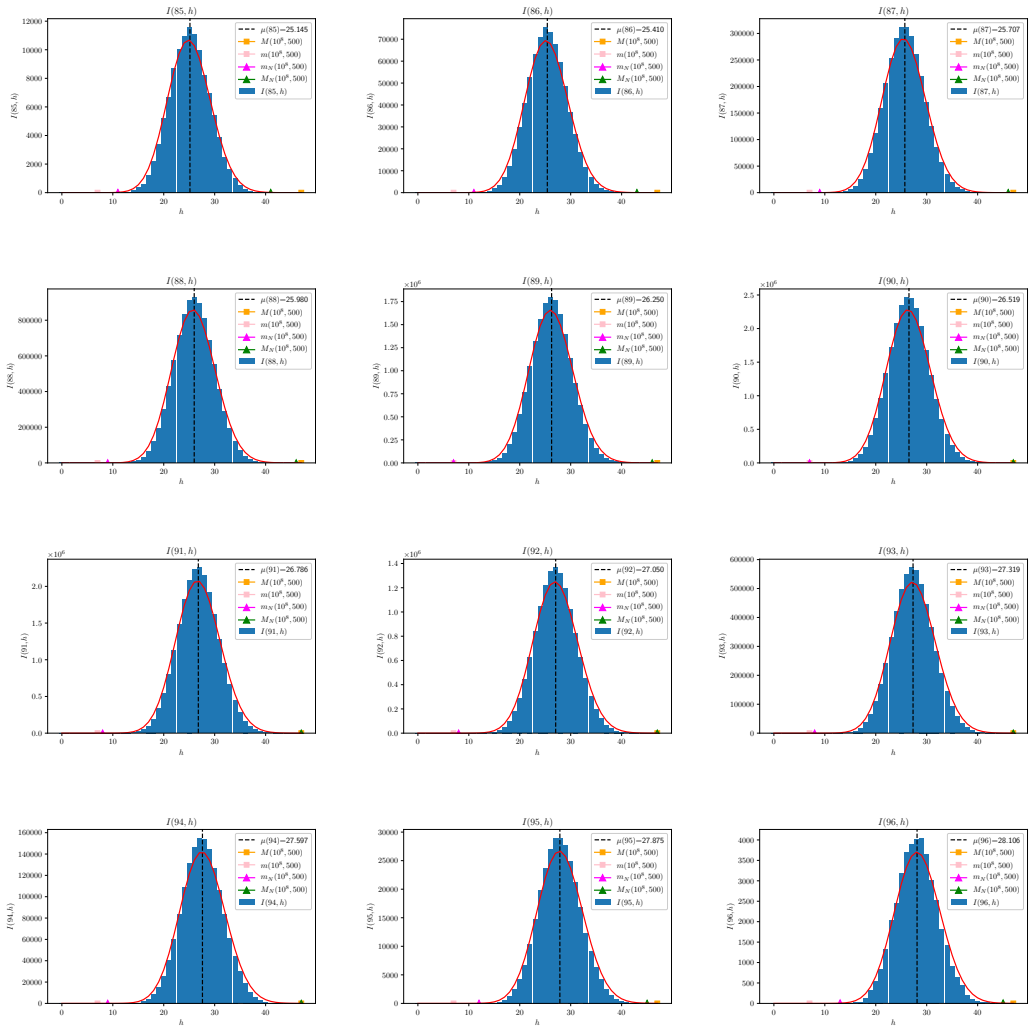

FIGURE 12. Testing the distributions, h vs $I(N, h)$, for $85 \leq N \leq 96$.

h	85	86	87	88	89	90	91	92	93	94	95	96
Data Mean	25.15	25.42	25.71	25.99	26.25	26.52	26.79	27.06	27.32	27.60	27.88	28.11
Exp Mean	25.04	25.33	25.62	25.92	26.21	26.51	26.80	27.1	27.39	27.69	27.98	28.28
Data Var	15.26	15.21	15.29	15.44	15.56	15.67	15.80	15.94	16.02	16.18	16.32	16.20
Exp Var	17.71	17.91	18.12	18.32	18.51	18.71	18.91	19.10	19.30	19.50	19.70	19.88

Replacing $\log x$ by $\log 4x/e$ as i the first example the overall expected mean is $26.5858\ldots$ and the new expected variance is $15.8003\ldots$, which again is a pretty good fit with this data.

The data in this appendix makes a compelling case that one should develop a different model, stemming from the binomial distribution, but in which the X_n are not independent. Instead, their dependence must imply that the number of primes in short intervals of length y between x and $2x$ satisfies the normal distribution with the variance predicted by Montgomery and Soundararajan, and then perhaps we might see what this new model might give for tail probabilities. We would thus revise our predictions for $M(x,y), m(x,y)$ and the largest gaps between consecutive primes.[14] We hope to return to this key topic in a further paper.

Appendix C. Pre-sieving intervals of length y by the primes up to y

In this case the intervals in $I(N)$ are not necessarily independent. Therefore we replace $I(N)$ by $I'(N)$, the largest subset of $I(N)$ of disjoint intervals. Evidently $\#I(N) \ge \#I'(N) \ge \#I(N)/y$ so that $\log \#I'(N) = \log \#I(N) + O(\log y)$, and this error term is irrelevant in applying Proposition 1 (with $x = \#I'(N)$) since $I(N)$ will typically contain x^θ elements for some constant $\theta > 0$.

Fix x and y. Recall that $I(N) = \{X \in (x, 2x] : \ S(X, y, y) = N\}$, where $0 \le N \le S(y)$. By Proposition 1 with $L = \frac{P}{\phi(P)} \log x$ we predict

$$M_N(x,y) := \max\{\pi(X, X+y] : \ x < X \le 2x \text{ and } S(X,y,y) = N\}$$

for each N with $I(N)$ non-empty. In section 5.1, the independence hypothesis of Proposition 1 was satisfied as the intervals were disjoint, Here the intervals in $I(N)$ might overlap, so replace $I(N)$ by $I'(N)$, the largest subset of $I(N)$ of disjoint intervals. Evidently $\#I(N) \ge \#I'(N) \ge \#I(N)/y$; in applying Proposition 1 (with $x = \#I'(N)$) the factor y here will be irrelevant, since $I(N)$ will typically contain x^θ elements for some constant $\theta > 0$. From here we obtain our predictions for

$$M(x,y) = \max_N M_N(x,y).$$

We write each $N = c_N \frac{\phi(P)}{P} y$ and $\#I(N) = x^{\theta_N}$, so that

$$\max_N \#I(N) \ge x/(S(y)+1) \ge x/y \ge x^{1-o(1)}.$$

Let $N_* = N_*(x,y)$ be that integer N which maximizes c_N over those N with $\theta_N = 1 + o(1)$, and let $c_* := c_{N_*}$.

Predictions, by pre-sieving up to y: *Assume that* $S(y) \sim \frac{y}{\log y}$.
If $\log x \ll y \le (e^\gamma/c_*) \log x$ *then*

$$M(x,y) = \max\left\{N : \ N \le \frac{\log \#I(N)}{\log\log x}\right\} \sim \max_N\{c_N : \ c_N y \le e^\gamma \theta_N \log x\} \cdot e^{-\gamma} \frac{y}{\log y}.$$

[14] Though hopefully only in the secondary terms, so as not to invalidate the conjectures in this paper!

If $(e^\gamma/c_*)\log x \le y = o((\log x)^2)$ *then*

$$M(x,y) \sim \frac{\log x}{\log(\frac{(\log x)^2}{y})}.$$

Finally if $y = \lambda(\log x)^2$ *with* $\lambda > 0$ *then*

$$M(x,y) \sim \max_N c_N \delta_+(c_N\lambda/\theta_N)\cdot\frac{y}{\log x}.$$

If λ is large and $y = \lambda(\log x)^2$ then

$$c_N\delta_+(c_N\lambda/\theta_N) = c_N + \sqrt{\frac{2\theta_N c_N}{\lambda}} + O\left(\frac{1}{\lambda}\right),$$

and so $M(x,\lambda(\log x)^2) \sim c_\dagger \frac{y}{\log x}$ as $\lambda\to\infty$ where $c_\dagger = \max_N c_N$ where the maximum is taken over all those c_N with $\theta_N \gg 1$.

These predictions are substantially more complicated than those obtained when presieving up to $\epsilon\log x$. By Occam's razor, we choose to follow the other path though it is feasible that both will yield the same prediction if only we could at least partly resolve the relevant sieve questions (that is, determine the values of $c_\dagger, c_*$ and $\max_N\{c_N:\ c_N \le u\theta_N\}$ for each $u>0$).

Deduction from the predictions of Proposition 1. We take the largest subset of the intervals in $I(N)$ that begin at least y apart (so there are $\#I(N)y^{O(1)}$ such intervals). We can employ Proposition 1 with $L \sim e^{-\gamma}\frac{\log x}{\log y}$, so that $\log L \sim \log\log x$. This yields that

$$M_N(x,y) \sim \begin{cases} N & \text{if } N \le \frac{\log \#I(N)}{\log\log x};\\ \frac{\log\#I(N)}{\log\left(\frac{L\log\#I(N)}{N}\right)} & \text{if } \frac{\log\#I(N)}{\log\log x} \le N = o(L\log\#I(N));\\ \delta_+(\lambda)\frac{N}{L} & \text{if } N = \lambda L\log\#I(N) \text{ with } \lambda>0;\end{cases}$$

Evidently $\sum_N \#I(N) = x$, each $N \le S(y)$ and

$$\sum_N N\#I(N) = y\#\{n\in(x,2x]:\ (n,P)=1\} + O(y^2) \sim \frac{\phi(P)}{P}xy,$$

so that $\#\{n\in(X,X+y]:\ (n,P)=1\}$ averages $\sim \frac{\phi(P)}{P}y$ over all $X\in(x,2x]$. From this we deduce that $c_*\ge 1$ which yields that

$$M(x,y) \ge M_{N_*}(x,y) \sim \begin{cases} c_* e^{-\gamma}\frac{y}{\log y} & \text{if } \frac{1}{3}\log x \le y \le c_*^{-1}e^\gamma\log x;\\ \frac{\log x}{\log(\frac{(\log x)^2}{y})} & \text{if } c_*^{-1}e^\gamma\log x \le y = o((\log x)^2);\\ \delta_+(c_*\lambda)c_*\frac{y}{\log x} & \text{if } y = \lambda(\log x)^2 \text{ with } \lambda>0.\end{cases}$$

One can also use Proposition 1 to bound $M(x,y)$ from above by a constant multiple of these lower bounds. We now show that if one obtains a larger prediction from some N (by a factor $>1+\epsilon$) then $N \asymp \frac{y}{\log y}$ and $\log\#I(N) \asymp \log x$ (that is, $c_N,\theta_N \asymp 1$):

In the first range for N we have $\frac{\log x}{\log\log x} \ge \frac{\log\#I(N)}{\log\log x} \ge N > N_* \gg \frac{y}{\log y}$ with $y \ll \log x$. We deduce that we are in the first range for N_*, and $\log\#I(N) \gg \log x$.

Now $N_* \gg \frac{y}{\log y} \gg N$ and so in the second range, $\frac{\log \#I(N)}{\log\left(\frac{L\log \#I(N)}{N}\right)} \lesssim \frac{\log x}{\log\left(\frac{L\log x}{N_*}\right)}$, and so we must be in the first range for N_*. Now $y \ll \log x$, in which case $N \ll L$, and so $M_N \sim \frac{\log \#I(N)}{\log\log \#I(N)}$, which implies that $\log \#I(N) \gg \log x$.

In the third range, $\frac{y}{\log x} \gg \frac{N}{L} \asymp M_N(x,y) > M_{N_*}(x,y) \gg \frac{y}{\log x}$, so that $N \asymp \frac{y}{\log y}$. Moreover $M_{N_*}(x,y) \ll \frac{y}{\log x}$ which implies $y \gg (\log x)^2$; and therefore $\log \#I(N) \asymp \frac{N}{L} \gg \log x$.

We can now construct a (completely analogous) table of values for $M_N(x,y)$ for each y, and compare to what we had for $M_{N_*}(x,y)$. From this we deduce our result. In the middle range this cannot be significantly bigger than $M_{N_*}(x,y)$ (and only equal if $\theta = 1 + o(1)$). □

References

[1] William Banks, Kevin Ford and Terence Tao, *Large prime gaps and probabilistic models*, (preprint).

[2] J.H. Cadwell, *Large intervals between consecutive primes*, Math. Comp. **25** (1971), 909–913.

[3] Harald Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arithmetica. **2** (1936), 23–46.

[4] Kevin Ford, Ben Green, Sergei Konyagin and Terence Tao, *Large gaps between consecutive prime numbers*, Ann. of Math. **183** (2016), 935–974.

[5] Kevin Ford, Ben Green, Sergei Konyagin, James Maynard, and Terence Tao, *Long gaps between primes*, J. Amer. Math. Soc. **31** (2018), 65–105.

[6] John Friedlander and Andrew Granville, *Limitations to the equi-distribution of primes. I*, Ann. of Math. **129** (1989), 363–382.

[7] Andrew Granville, *Harald Cramér and the distribution of prime numbers*, Harald Cramér Symposium (Stockholm, 1993). Scand. Actuar. J. **1** (1995), 12–28.

[8] Andrew Granville, *Primes in intervals of bounded length*, Bull. Amer. Math. Soc. **52** (2015), 171–222.

[9] Andrew Granville, *Sieving intervals and Siegel zeros*, (preprint)

[10] G. H. Hardy and J. E. Littlewood, *Some problems of "Partitio Numerorum", III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[11] Henryk Iwaniec, *On the problem of Jacobsthal*, Demonstratio Math. **11** (1978), 225–231.

[12] W.B. Jurkat and H.-E. Richert, *An improvement of Selberg's sieve method. I.* Acta Arith. **11** (1965), 217–240.

[13] Helmut Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), 221–225.

[14] Helmut Maier, and Cam L. Stewart, *On intervals with few prime numbers*, J. Reine Angew. Math. **608** (2007), 183–199.

[15] James Maynard, *Small gaps between primes*, Ann. of Math. **181** (2015), 383–413.

[16] James Maynard, *Large gaps between primes*, Ann. of Math. **183** (2016), 915–933.

[17] James Maynard, *Sums of two squares in short intervals*, arXiv:1910.13384

[18] Hugh L. Montgomery and K. Soundararajan *Primes in short intervals*, Comm. Math. Phys. **252** (2004), 589–617.

[19] A. Selberg, *Sieve methods*, ch. 36 of Collected Works, Vol I, Springer-Verlag, New York 1989. *published originally as:* Proc. Sympos. Pure Math **20** (1971), 311–351.

[20] Terence Tao, *Polymath8b: Bounded intervals with many primes, after Maynard*, Blog note. `https://terrytao.wordpress.com/2013/11/19/polymath8b-bounded-intervals-with-many-primes-after-maynard/`

[21] Yitang Zhang, *Bounded gaps between primes*, Ann. of Math. **179** (2014), 1121–1174.

Départment de Mathématiques et Statistique, Université de Montréal, CP 6128 succ Centre-Ville, Montréal, QC H3C 3J7, Canada.
E-mail address: andrew@dms.umontreal.ca

Centre recherche mathématiques, Université de Montréal, CP 6128 succ Centre-Ville, Montréal, QC H3C 3J7, Canada.
E-mail address: lumley@crm.umontreal.ca