UNEXPECTED IRREGULARITIES IN THE DISTRIBUTION OF PRIME NUMBERS

Andrew Granville

In 1849 the Swiss mathematician Encke wrote to Gauss, asking whether he had ever considered trying to estimate $\pi(x)$, the number of primes up to x, by some sort of 'smooth' function. On Christmas Eve 1849, Gauss replied that "*he had pondered this problem as a boy*" and had come to the conclusion that "*at around x, the primes occur with density* $1/\log x$." Thus, he concluded, $\pi(x)$ could be approximated by

$$\mathrm{Li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t} = \frac{x}{\log x} + \frac{x}{\log^2 x} + O\left(\frac{x}{\log^3 x}\right).$$

Comparing Gauss's guess to the best data available today (due to Rivat), we have:

x	$\pi(x)$	$[\mathrm{Li}(x) - \pi(x)]$
10^8	5761455	754
10^9	50847534	1701
10^{10}	455052511	3104
10^{11}	4118054813	11588
10^{12}	37607912018	38263
10^{13}	346065536839	108971
10^{14}	3204941750802	314890
10^{15}	29844570422669	1052619
10^{16}	279238341033925	3214632
10^{17}	2623557157654233	7956589
10^{18}	24739954287740860	21949555

The number of primes, $\pi(x)$, up to x.

This data certainly seems to support Gauss's prediction, since $\mathrm{Li}(x)-\pi(x)$ appears to be no bigger than a small power of $\pi(x)$. In 1859, Riemann, in a now famous memoir, illustrated how the question of estimating $\pi(x)$ could be turned into a question in analysis: Define $\zeta(s) := \sum_{n\geq 1} n^{-s}$ for $\mathrm{Re}(s) > 1$, and then analytically continue $\zeta(s)$ to the rest of the complex plane. We have, for sufficiently large x,

$$x^{b-\varepsilon} \ll \max_{y\leq x} |\pi(y) - \mathrm{Li}(y)| \ll x^{b+\varepsilon} \quad \text{where } b := \sup_{\zeta(\beta+i\gamma)=0} \beta. \tag{1}$$

The author is an Alfred P. Sloan Research Fellow; and is also supported, in part, by the National Science Foundation.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

The (as yet unproven) *Riemann Hypothesis* (RH) asserts that $b = 1/2$ (in fact that $\beta = 1/2$ whenever $\zeta(\beta+i\gamma) = 0$ with $0 \leq \beta \leq 1$) leading to the sharp estimate

$$\pi(x) \ = \ \mathrm{Li}(x) + O(x^{1/2} \log x) \ . \tag{2a}$$

It was not until 1896 that HADAMARD and DE LA VALLÉE POUSSIN independently proved that $\beta < 1$ whenever $\zeta(\beta + i\gamma) = 0$, which implies *The Prime Number Theorem*: that is, GAUSS's prediction that

$$\pi(x) \ \sim \ \mathrm{Li}(x) \ \sim \ \frac{x}{\log x}.$$

In 1914, LITTLEWOOD showed, unconditionally, that

$$\pi(x) - \mathrm{Li}(x) = \Omega_{\pm}\left(x^{1/2} \ \frac{\log\log\log x}{\log x} \right), \tag{2b}$$

the first proven 'irregularities' in the distribution of primes[1].

Since GAUSS's vague 'density assertion' was so prescient, CRAMÉR [4] decided, in 1936, to interpret GAUSS's statement more formally in terms of probability theory, to try to make further predictions about the distribution of prime numbers: Let $Z_2, Z_3, \ldots$ be a sequence of independent random variables with

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

Let S be the space of sequences $T = z_2, z_3, \ldots$ and, for each $x \geq 2$ define

$$\pi_T(x) \ = \ \sum_{2 \leq n \leq x} z_n \ .$$

The sequence $P = \pi_2, \pi_3, \ldots$, where $\pi_n = 1$ if and only if n is prime, belongs to S. CRAMÉR wrote: "*In many cases it is possible to prove that,* with probability *1, a certain relation R holds for sequences in S ... Of course we cannot in general conclude that R holds for the particular sequence P, but results suggested in this way may sometimes afterwards be rigorously proved by other methods.*" For example CRAMÉR was able to show, with probability 1, that

$$\max_{y \leq x} |\pi_T(y) - \mathrm{Li}(y))| \sim \sqrt{2x \cdot \frac{\log\log x}{\log x}},$$

which corresponds well with the estimates in (2); and if true for $T = P$ implies RH, by (1).

GAUSS's assertion was really about primes in short intervals, and so is best applied to $\pi(x+y) - \pi(x)$, where y is "small" compared to x. The binomial random

[1] $f(x) = \Omega_{\pm}(g(x))$ means that there exists a constant $c > 0$ such that $f(x_+) > cg(x_+)$ and $f(x_-) < -cg(x_-)$ for certain arbitrarily large values of $x_{\pm}$.

variables Z_n are more-or-less the same for all integers n in such an interval. If we take $y = \lambda \log x$ so that the 'expected' number of primes, λ, in the interval is fixed then we would expect that the number of primes in such intervals should follow a Poisson distribution. Indeed, we can prove that for any fixed $\lambda > 0$ and integer $k \geq 0$, we have

$$\#\{\text{integers } x \leq X : \quad \pi_T(x + \lambda \log x) - \pi_T(x) = k\} \sim e^{-\lambda} \frac{\lambda^k}{k!} X \tag{3}$$

as $X \to \infty$, with probability 1 for $T \in S$. In 1976, Gallagher [11] showed that this holds for the sequence of primes (that is, for P) under the assumption of a reasonable "uniform" version of Hardy and Littlewood's *Prime k-tuplets conjecture* [14]. This conjecture is the case where we take each $f_j(x)$ to be a linear polynomial in Schinzel and Sierpiński's [22]

Hypothesis H. *Let $F = \{f_1(x), f_2(x), \ldots, f_k(x)\}$ be a set of irreducible polynomials with integer coefficients. Then the number of integers $n \leq x$ for which each $|f_j(n)|$ is prime is*

$$\pi_F(x) = \{C_F + o(1)\} \frac{x}{\log|f_1(x)| \log|f_2(x)| \ldots \log|f_k(x)|}$$
$$\text{where} \quad C_F = \prod_{p \text{ prime}} \left(1 - \frac{\omega_F(p)}{p}\right) \bigg/ \left(1 - \frac{1}{p}\right)^k,$$

and $\omega_F(p)$ counts the number of integers n, in the range $1 \leq n \leq p$, for which $f_1(n) f_2(n) \ldots f_k(n) \equiv 0 \pmod p$[2,3].

Estimates analogous to (2a) should hold for the number of primes in intervals of various lengths, if we believe that what almost always occurs in S, should also hold for P. Specifically, if $10 \log^2 x \leq y \leq x$ then

$$\pi_T(x + y) - \pi_T(x) \;=\; \mathrm{Li}(x + y) - \mathrm{Li}(x) + O(y^{1/2}) \tag{4}$$

with probability 1 for $T \in S$. In 1943 Selberg [23] showed that primes do, on the whole, behave like this by proving, under the assumption of RH, that

$$\pi(x + y) - \pi(x) \sim \frac{y}{\log x} \tag{5}$$

for 'almost all' integers x, provided $y / \log^2 x \to \infty$ as $x \to \infty$.

Montgomery [17] has shown that one can deduce estimates about primes in short intervals by understanding local distribution properties of the zeros of $\zeta(s)$:

[2] Elementary results on prime ideals guarantee that the product defining C_F converges if the primes are taken in ascending order.

[3] The asymptotic formula proposed here for $\pi_F(x)$ has a 'local part' C_F, which has a factor corresponding to each rational prime p, and an 'analytic part' $x / \prod_i \log|f_i(x)|$. This reminds one of formulae which arise when counting points on varieties.

Pair Correlation Conjecture (PC). *Assume* RH. *For any fixed* $\alpha > 0$, *the average number of zeros* $1/2 + i\gamma'$ *of* $\zeta(s)$ *'close' to a given zero* $1/2 + i\gamma$, *that is with* $\gamma < \gamma' \leq \gamma + 2\pi\alpha/\log(|\gamma| + 1)$, *is*

$$\sim \int_0^\alpha \left\{ 1 - \left(\frac{\sin \pi u}{\pi u} \right)^2 \right\} \mathrm{d}u. \tag{6a}$$

Inspired by the work of MONTGOMERY and others, GOLDSTON [12] showed that (6a) holds for any fixed $\alpha > 0$ *if and only if* for any fixed $\beta > 0$,

$$\int_T^{T^{1+\beta}} \left(\psi\left(x + \frac{x}{T} \right) - \psi(x) - \frac{x}{T} \right)^2 \frac{\mathrm{d}x}{x^2} \sim \beta \frac{\log^2 T}{T}, \tag{6b}$$

where $\psi(x) := \sum_{p^m \leq x} \log p$. Since (6b) is predicted by CRAMÉR's model, thus so is PC. DYSON predicted an analogous density function for the correlation of n-tuples of zeros of $\zeta(s)$,[4] which presumably may be shown to be equivalent to estimates for primes in short intervals, and thus be predicted by CRAMÉR's model.

CRAMÉR's model does seem to accurately predict what we already believe to be true about primes for more substantial reasons[5]. To be sure, one can find small discrepancies[6] but the probabilistic model usually gives one a strong indication of the truth. CRAMÉR made one conjecture, based on his model, which does not seem to be attackable by other methods: If $p_1 = 2 < p_2 = 3 < p_3 = 5 < \ldots$ is the sequence of prime numbers then

$$\max_{p_n \leq x} (p_{n+1} - p_n) \sim \log^2 x.$$

This statement (or the weaker $O(\log^2 x)$) is known as '*Cramér's Conjecture*'; there is some computational evidence to support it:

p_n	$p_{n+1} - p_n$	$(p_{n+1} - p_n)/\log^2 p_n$
31397	72	.6715
370261	112	.6812
2010733	148	.7025
20831323	210	.7394
25056082087	456	.7953
2614941710599	652	.7975
19581334192423	778	.8177

Record-breaking gaps between primes, up to 10^{14}

[4] Based somewhat surprisingly on the fact that (6a) also describes the distribution of eigenvalues of random Hermitian matrices, which arise in physics as models for various naturally occurring quanta. The 'Fourier transforms' of these pair correlation and n-tuple correlation conjectures for the zeros of $\zeta(s)$, are now known to hold in a natural restricted range, under the assumption of RH (see [17] and [21])

[5] Though HARDY AND LITTLEWOOD [14] remarked thus on probabilistic models: "*Probability is not a notion of pure mathematics, but of philosophy or physics*"

[6] As has been independently pointed out to me by SELBERG, MONTGOMERY and PINTZ: for example, PINTZ noted that the mean square of $|\psi(y) - y|$ for $y \leq x$, is $\gg x^{2b-\varepsilon}$ (with b as in (1)), in fact $\asymp x$ assuming RH, whereas the probabilistic model predicts $\asymp x \log x$.

In 1985 Maier [16] surprisingly proved that, despite Selberg showing (5) holds 'almost all' the time when $y = \log^B x$ (assuming RH) for fixed $B > 2$, it cannot hold all of the time for such y. This not only radically contradicts what is predicted by the probabilistic model, but also what most researchers in the field had believed to be true, whether or not they had faith in the probabilistic model. Specifically, Maier showed the existence of a constant $\delta_B > 0$ such that for occasional, but arbitrarily large, values of x_+ and x_-,

$$\begin{aligned} &\pi(x_+ + \log^B x_+) - \pi(x_+) \ > \ (1+\delta_B)\log^{B-1} x_+ \ , \\ \text{and} \quad &\pi(x_- + \log^B x_-) - \pi(x_-) \ < \ (1-\delta_B)\log^{B-1} x_- . \end{aligned} \tag{7}$$

Outline of the Proof. There are $\sim e^{-\gamma} x/\log z$ integers $\leq x$, all of whose prime factors are $> z$, provided z is not too large. Among these we have all but $\pi(z)$ of the primes $\leq x$, and so the probability that a randomly chosen such integer is prime is $\sim e^{\gamma} \log z/\log x$. Thus in a specific interval $(x, x+y]$ we should 'expect' $\sim e^{\gamma} \Phi \log z/\log x$ primes, where Φ is the number of integers in the interval that are free of prime factors $\leq z$. Now if we can select our interval so that $\Phi \not\sim e^{-\gamma} y/\log z$ then our new prediction is not the same as that in (5).

If x is divisible by $P = \prod_{p \leq z} p$ then

$$\Phi = \Phi(y,z) := \#\{1 \leq n \leq y : \ p|n \Rightarrow p > z\} \sim \omega(u)\frac{y}{\log z}$$

for $y = z^u$ with u fixed (see [3]), where $\omega(u) = 0$ if $0 < u < 1$ and satisfies the differential-delay equation $u\,\omega(u) = 1 + \int_1^{u-1} \omega(t)\mathrm{d}t$ if $u \geq 1$. Obviously $\lim_{u\to\infty} \omega(u) = \mathrm{e}^{-\gamma}$. Iwaniec showed that $\omega(u) - \mathrm{e}^{-\gamma}$ oscillates, crossing zero either once or twice in every interval of length 1. Thus if we fix $u > B$, chosen so that $\omega(u) > \mathrm{e}^{-\gamma}$ or $< \mathrm{e}^{-\gamma}$ (as befits the case of (7)), select $y = \log^B x$ and $z = y^{1/u}$, and 'adjust' x so that it is divisible by P, then we expect (5) to be false.

To convert this heuristic into a proof, Maier considered a progression of intervals of the form $(rP, rP+y]$, with $R \leq r < 2R$ for a suitable value of R. Visualizing this as a 'matrix', with each such interval represented by a different row, we see that the primes in the matrix are all contained in those columns j for which $(j, P) = 1$.

$$\begin{array}{|ccccc|} \hline RP+1 & RP+2 & RP+3 & \cdots & RP+y \\ (R+1)P+1 & (R+1)P+2 & (R+1)P+3 & \cdots & (R+1)P+y \\ (R+2)P+1 & (R+2)P+2 & . & . & \vdots \\ (R+3)P+1 & \vdots & \boxed{\begin{array}{c}(i,j)\text{th entry}: \\ (R+i)P+j\end{array}} & \vdots & \vdots \\ \vdots & \vdots & . & \vdots & \vdots \\ (2R-1)P+1 & (2R-1)P+2 & \cdots & \cdots & (2R-1)P+y \\ \hline \end{array}$$

The 'Maier Matrix' for $\pi(x+y) - \pi(x)$

Now, for any integer $q > 1$, the primes are roughly equi-distributed amongst those arithmetic progressions $a \pmod q$ with $(a, q) = 1$: in fact up to x we expect that the number of such primes

$$\pi(x; q, a) \sim \frac{\pi(x)}{\phi(q)}. \tag{8}$$

If so, then the number of primes in the jth column, when $(j, P) = 1$, is

$$\pi(2X; P, j) - \pi(X; P, j) \sim \frac{1}{\phi(P)} \frac{X}{\log X} \quad \text{where } X = RP.$$

To get the total number of primes in the matrix we sum over all such j, and then we can deduce that, on average, a row contains

$$\sim \frac{\Phi(y,z)}{R} \frac{1}{\phi(P)} \frac{RP}{\log RP} \sim \omega(u) \frac{y}{\log z} \frac{P}{\phi(P)} \frac{1}{\log RP} \sim e^{\gamma} \omega(u) \frac{y}{\log RP}$$

primes. MAIER's result follows provided we can prove a suitable estimate in (8)

In general, it is desirable to have an estimate like (8) when x is not too large compared to q. It has been proved that (8) holds uniformly for

i) All $q \le \log^B x$ and all $(a, q) = 1$, for any fixed $B > 0$ (SIEGEL-WALFISZ).
ii) All $q \le \sqrt{x}/\log^{2+\varepsilon} x$ and all $(a, q) = 1$, assuming GRH[7]. In fact (8) then holds with error term $O\left(\sqrt{x}\log^2(qx)\right)$.
iii) Almost all $q \le \sqrt{x}/\log^{2+\varepsilon} x$ and all $(a, q) = 1$ (BOMBIERI-VINOGRADOV)[8].
iv) Almost all $q \le x^{1/2+o(1)}$ with $(q, a) = 1$, for fixed $a \ne 0$ (BOMBIERI-FRIEDLANDER-IWANIEC, FOUVRY)
v) Almost all $q \le x/\log^{2+\varepsilon} x$ and almost all $(a, q) = 1$ (BARBAN-DAVENPORT-HALBERSTAM, MONTGOMERY, HOOLEY).

Thus, when GRH is true, we get a good enough estimate in (8) with $R = P^2$ to complete MAIER's proof. However MAIER, in the spirit of the BOMBIERI-VINOGRADOV Theorem, showed how to pick a '*good*' value for P (see [8, Prop. 2]), so that (8) is off by, at worst, an insignificant factor when R is a large, but fixed, power of P (thus proving his result unconditionally).

In [15], HILDEBRAND AND MAIER extended the range for y in the proof above, establishing that there are arbitrarily large values of x for which (4) fails to hold for some $y > \exp\left((\log x)^{1/3-\varepsilon}\right)$; and, assuming GRH, for some $y > \exp\left((\log x)^{1/2-\varepsilon}\right)$. Moreover they show that such intervals $(x, x+y]$ occur within every interval $[X, 2X]$.[9]

It is plausible that (5) holds uniformly if $\log y/\log\log x \to \infty$ as $x \to \infty$; and that (4) holds uniformly for $T = P$ if $y > \exp\left((\log x)^{1/2+\varepsilon}\right)$ (at least, we can't

[7] The *Generalized Riemann Hypothesis* (GRH) states that if $\beta + i\gamma$ is a zero of *any* Dirichlet L-function then $\beta \le 1/2$

[8] This result is often referred to as 'GRH on average'

[9] A far better localization than those obtained in any proof of (2b).

disprove these statements as yet). We conjecture, presumably safely, that (4) and (5) hold uniformly when $y > x^{\varepsilon}$.

One can show that there are more than $x/\exp((\log x)^{c_B})$ integers $x_\pm \leq x$ satisfying the unexpected inequalities in (7). Although this may not be enough to upset (6b), it surely guarantees that the error term there will not be as small might have been hoped. Thus we should not expect the pair correlation conjecture to hold with as much uniformity as might have been believed. Evidence that this is so may be seen in the computations represented by [19, Figures 2.3.1,2,3]: The pair correlation function for the nearest 10^6 zeros to the 10^{12}th zero fits well with (6a) for $\alpha < .8$ and then has larger amplitude for $.8 < \alpha < 3$; also, the pair correlation function for the nearest 8×10^6 zeros to the 10^{20}th zero fits well with (6a) for $\alpha < 3$ and then has larger amplitude for $3 < \alpha < 5$.

Maier's work suggests that Cramér's model should be adjusted to take into account divisibility of n by 'small' primes[10]. It is plausible to define 'small' to mean those primes up to a fixed power of $\log n$. Then we are led to conjecture that there are infinitely many primes p_n with $p_{n+1} - p_n > 2\mathrm{e}^{-\gamma} \log^2 p_n$, contradicting Cramér's conjecture![11]

If we analyze the distribution of primes in arithmetic progressions using a suitable analogue of Cramér's model, then we would expect (8), and even

$$(8') \qquad \pi(x; q, a) \;=\; \frac{\pi(x)}{\phi(q)} + O\left(\left(\frac{x}{q}\right)^{1/2} \log(qx)\right) \;,$$

to hold uniformly when $(a, q) = 1$ in the range

$$(9) \qquad q \leq Q = x/\log^B x \;,$$

for any fixed $B > 2$. However the method of Maier is easily adapted to show that neither (8) nor (8′) cannot hold in at least part of the range (9): For any fixed $B > 0$ there exists a constant $\delta_B > 0$ such that for any modulus q, with 'not too many small prime factors', there exist arithmetic progressions $a_\pm \pmod q$ and values $x_\pm \in [\phi(q)\log^B q, 2\phi(q)\log^B q]$ such that

$$(10) \quad \pi(x_+; q, a_+) \;>\; (1 + \delta_B)\frac{\pi(x_+)}{\phi(q)} \quad \text{and} \quad \pi(x_-; q, a_-) \;<\; (1 - \delta_B)\frac{\pi(x_-)}{\phi(q)} \;.$$

[10] One has to be careful about the meaning of 'small' here, since if we were to take into account the divisibility of n by all primes up to $\sqrt{n}$, then we would conclude that there are $\sim \mathrm{e}^{-\gamma} x/\log x$ primes up to x.

[11] It is unclear what the 'correct conjecture' here should be since, to get at it with this approach, we would need more precise information on 'sifting limits' than is currently available.

The proof is much as before, though now using a modified 'Maier matrix':

$$\boxed{\begin{array}{ccccc}
RP & RP+q & RP+2q & \cdots & RP+yq \\
(R+1)P & (R+1)P+q & (R+1)P+2q & \cdots & (R+1)P+yq \\
(R+2)P & (R+2)P+q & . & . & \vdots \\
(R+3)P & \vdots & \boxed{\begin{array}{l}(i,j)\text{th entry :}\\ (R+i)P+jq\end{array}} & \vdots & \vdots \\
\vdots & \vdots & . & \vdots & \vdots \\
(2R-1)P & (2R-1)P+q & \cdots & \cdots & (2R-1)P+yq
\end{array}}$$

The Maier Matrix for $\pi(yq;q,a)$

The BOMBIERI-VINOGRADOV Theorem is usually stated in a stronger form than above: For any given $A>0$, there exists a value $B=B(A)>0$ such that

$$\sum_{q\leq Q}\max_{(a,q)=1}\max_{y\leq x}\left|\pi(y;q,a)-\frac{\pi(y)}{\phi(q)}\right| \ll \frac{x}{\log^A x} \tag{11}$$

where $Q=\sqrt{x}/\log^B x$. It is possible [6] to take the same values of R and P in the Maier matrix above for many different values of q, and thus deduce that there exist arbitrarily large values of a and x for which

$$\left|\sum_{\substack{Q\leq q\leq 2Q\\(q,a)=1}}\left\{\pi(x;q,a)-\frac{\pi(x)}{\phi(q)}\right\}\right| \gg x; \tag{12}$$

thus refuting the conjecture that for any given $A>0$, (11) should hold in the range (9) for some $B=B(A)>0$. In [7] we showed that (11) even fails with

$$Q \;=\; x/\exp\left((A-\varepsilon)(\log\log x)^2/(\log\log\log x)\right).$$

We also showed that ($8'$) cannot hold for every integer a, prime to q, for

i) Any $q\geq x/\exp\left((\log x)^{1/5-\varepsilon}\right)$.

ii) Any $q\geq x/\exp\left((\log x)^{1/3-\varepsilon}\right)$ which has $<1.5\log\log\log q$ distinct prime factors $<\log q$.[12]

iii) Almost any $q\in(y,2y]$, for any $y\geq x/\exp\left((\log x)^{1/2-\varepsilon}\right)$.

Moreover, under the assumption of GRH we can improve the values $1/5$ and $1/3$ in (i) and (ii), respectively, to $1/3$ and $1/2$.

It seems plausible that (8) holds uniformly if $\log(x/q)/\log\log q\to\infty$ as $q\to\infty$; and that (11) holds uniformly for $Q<x/\exp\left((\log x)^{1/2+\varepsilon}\right)$. At least we can't disprove these statements as yet, though we might play it safe and conjecture only that they hold uniformly for $q,Q<x^{1-\varepsilon}$.

---

[12] Which, by the TURAN-KUBILIUS inequality, includes 'almost all' integers.

Notice that in the proof described above, the values of a increase with x, leaving open the possibility that (8) might hold uniformly for all $(a, q) = 1$ in the range (9) if we fix a.[13] However, in [8] we observed that when a is fixed one can suitably modify the Maier matrix, by forcing the elements of the second column to all be divisible by P:

$$\begin{array}{|ccccc|}
\hline
1 & 1+(RP-1) & 1+2(RP-1) & \cdots & 1+y(RP-1) \\
1 & 1+((R+1)P-1) & 1+2((R+1)P-1) & \cdots & 1+y((R+1)P-1) \\
1 & 1+((R+2)P-1) & \cdot & \cdot & \vdots \\
1 & \vdots & \boxed{\begin{array}{c}(i,j)\text{th entry :} \\ 1+j((R+i)P-1)\end{array}} & \vdots & \vdots \\
\vdots & \vdots & \cdot & \vdots & \vdots \\
1 & 1+((2R-1)P-1) & \cdots & \cdots & 1+y((2R-1)P-1) \\
\hline
\end{array}$$

The Maier Matrix for $\pi(yq; q, 1)$

Notice that the jth column here is now part of an arithmetic progression with a varying modulus, namely $1 - j \pmod{jP}$. With this type of Maier matrix we can deduce that, for almost all $0 < |a| < x/\log^B x$ (including all fixed $a \neq 0$), there exist $q \in (x/\log^B x, 2x/\log^B x]$, coprime to a, for which (8) does not hold. However (8) cannot be false too often (like in (12)), since this would contradict the Barban-Davenport-Halberstam Theorem. So for which a is (8) frequently false? It turns out that the answer depends on the number of prime factors of a: In [10], extending the results of [2], we show that for any given $A > 1$ there exists a value $B = B(A) > 0$ such that, for any $Q \le x/\log^B x$ and any integer a which satsfies $0 < |a| < x$ and has $\ll \log\log x$ distinct prime factors[14], we have

$$\left| \sum_{\substack{Q \le q \le 2Q \\ (q,a)=1}} \left\{ \pi(x; q, a) - \frac{\pi(x)}{\phi(q)} \right\} \right| \ll \frac{x}{\log^A x}. \tag{13}$$

On the other hand, for every given $A, B > 0$, there exists $Q \le x/\log^B x$ and an integer a which satisfies $0 < |a| < x$ and has $< (\log\log x)^{6/5+\varepsilon}$ distinct prime factors, for which (13) does not hold (and assuming GRH we may replace $6/5 + \varepsilon$ here by $1 + \varepsilon$).

Finding primes in $(x, x + y]$ is equivalent to finding integers $n \le y$ for which $f(n)$ is prime, where $f(t)$ is the polynomial $t + x$. Similarly, finding primes $\le x$ which belong to the arithmetic progression $a \pmod q$, is equivalent to finding integers $n \le y := x/q$ for which $f(n)$ is prime, where $f(t)$ is the polynomial $qt + a$. Define the *height*, $h(f)$, of a given polynomial $f(t) = \sum_i c_i t^i$ to be $h(f) := \sqrt{\sum_i c_i^2}$. In the

[13] Which would be consistent with the Barban-Davenport-Halberstam Theorem.

[14] which includes almost all integers a once the inexplicit constant here is > 1 (by a famous result of Hardy and Ramanujan).

cases above, in which the degree is always 1, we proved that we do not always get the asymptotically expected number of prime values $f(n)$ with $n \leq y = \log^B h(f)$, for any fixed $B > 0$. In [9] we showed that this is true for polynomials of arbitrary degree d, which is somewhat ironic since it is not known that *any* polynomial of degree ≥ 2 takes on infinitely many prime values, nor that the prime values are ever 'well-distributed'. NAIR AND PERELLI [18] showed that some of the polynomials $F_R(n) = n^d + RP$ attain more than, and others attain less than, the number of prime values expected in such a range, by considering the following Maier matrix:

$$\begin{array}{|ccccc|}
\hline
F_R(1) & F_R(2) & F_R(3) & \cdots & F_R(y) \\
F_{R+1}(1) & F_{R+1}(2) & F_{R+1}(3) & \cdots & F_{R+1}(y) \\
F_{R+2}(1) & F_{R+2}(2) & . & . & \vdots \\
F_{R+3}(1) & \vdots & \boxed{\begin{array}{c}(i,j)\text{th entry}: \\ F_{R+i}(j)\end{array}} & \vdots & \vdots \\
\vdots & \vdots & . & \vdots & \vdots \\
F_{2R-1}(1) & F_{2R-1}(2) & \cdots & \cdots & F_{2R-1}(y) \\
\hline
\end{array}$$

The Maier Matrix for $\pi_F(y)$

Notice that the jth column here is part of the arithmetic progression $j^d \pmod P$.

Using Maier matrices it is possible to prove 'bad equi-distribution' results for primes in other interesting sequences, such as the values of binary quadratic forms, and of prime pairs. For example, if we fix $B > 0$ then, once x is sufficiently large, there exists a positive integer $k \leq \log x$ such that there are at least $1 + \delta_B$ times as many prime pairs $p, p + 2k$, with $x < p \leq x + \log^B x$, as we would expect from assuming that the estimate in Hypothesis H holds uniformly for $n \ll \log^B h((t + x)(t + (x + 2k)))$.

We have now seen that the asymptotic formula in Hypothesis H fails when x is an arbitrary fixed power of $\log h(F)(:= \sum_i \log h(f_i))$, for many different non-trivial examples F. Presumably the asymptotic formula *does* hold uniformly as $\log x / \log\log h(F) \to \infty$. However, to be safe, we only make the following prediction:

Conjecture. *Fix $\varepsilon > 0$ and positive integer k. The asymptotic formula in Hypothesis H holds uniformly for $x > h(F)^\varepsilon$ as $h(F) \to \infty$.*

Our work here shows that the 'random-like' behaviour exhibited by primes in many situations does not carry over to *all* situations. It remains to discover a model that will always accurately predict how primes are distributed, since it seems that minor modifications of CRAMÉR's model will not do. We thus agree that:

> *"It is evident that the primes are randomly distributed but, unfortunately, we don't know what 'random' means."* — R.C. VAUGHAN (February 1990).

Final remarks

Armed with Maier's ideas it seems possible to construct incorrect conclusions from, more-or-less, any variant of Cramér's model. This flawed model may still be used to make conjectures about the distribution of primes, but one should be very cautious of such predictions!

There are no more than $O(x^2/\log^{3B} x)$ arithmetic progressions $a \pmod q$, with $1 \leq a < q < x/\log^B x$ and $(a,q) = 1$, for which (8) fails, by the Barban-Davenport-Halberstam Theorem. However our methods here may be used to show that (8) does fail for more than $x^2/\exp((\log x)^\varepsilon)$ such arithmetic progressions.

Maier's matrix has been used in other problems too: Konyagin recently used it to find unusually large gaps between consecutive primes. Maier used it to find long sequences of consecutive primes, in which there are longer than average gaps between each pair. Shiu has used it to show that every arithmetic progression $a \pmod q$ with $(a,q) = 1$ contains arbitrarily long strings of consecutive primes.

Balog [1] has recently shown that the prime k-tuplets conjecture holds 'on average' (in the sense of the Bombieri-Vinogradov Theorem)[15].

As we saw in the table above, $\mathrm{Li}(x) > \pi(x)$ for all $x \leq 10^{18}$. However (2b) implies that this inequality does not persist for ever; indeed, it is reversed for some $x < 10^{370}$ (te Riele). Recently, however, Rubinstein and Sarnak [20][16] showed that it does hold more often than not, in the sense that there exists a constant $\delta \approx 1/(4 \cdot 10^6)$ such that the (logarithmically scaled) proportion of x for which $\pi(x) > \mathrm{Li}(x)$ exists and equals δ. Such biases may also be observed in arithmetic progressions, in that there are 'more' primes belonging to arithmetic progressions that are quadratic non-residues than those that are quadratic residues. In particular they prove that $\pi(x;4,3) > \pi(x;4,1)$ for a (logarithmically scaled) proportion $0.9959\ldots$ of the time.

Delicate questions concerning the distribution of prime numbers still seem to be very mysterious. It may be that by taking into account divisibility by small primes we can obtain a very accurate picture; or it may be that there are other phenomena, disturbing the equi-distribution of primes, that await discovery ...

> *"Mathematicians have tried in vain to discover some order in the sequence of prime numbers but we have every reason to believe that there are some mysteries which the human mind will never penetrate."*
>
> — L. Euler (1770).

Acknowledgements: I'd like to thank Red Alford, Nigel Boston, John Friedlander, Dan Goldston, Ken Ono, Carl Pomerance and Trevor Wooley for their helpful comments on earlier drafts of this paper.

[15] Which was improved to the exact analogue of (11) by Mikawa for $k = 2$, and by Kawada for all $k \geq 1$.

[16] All of their results are proved assuming appropriate conjectures such as RH, GRH, and that the zeros of the relevant L-functions are linearly independent over $\mathbb{Q}$.

References

[1] A. Balog, *The prime k-tuplets conjecture on average*, Analytic Number Theory (B.C. Berndt, H.G. Diamond, H. Halberstam, A. Hildebrand, eds.), Birkhäuser, Boston, 1990, pp. 165-204.

[2] E. Bombieri, J.B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli III*, J. Amer. Math. Soc. **2** (1989), 215-224.

[3] A.A. Buchstab, *On an asymptotic estimate of the number of numbers of an arithmetic progression which are not divisible by relatively small prime numbers*, Mat. Sbornik **28/70** (1951), 165-184. (Russian)

[4] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 23-46.

[5] H. Davenport, *Multiplicative Number Theory* (2nd ed.), Springer-Verlag, New York, 1980.

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

[7] J.B. Friedlander, A. Granville, A. Hildebrand and H. Maier, *Oscillation theorems for primes in arithmetic progressions and for sifting functions*, J. Amer. Math. Soc. **4** (1991), 25-86.

[8] J.B. Friedlander and A. Granville, *Limitations to the equi–distribution of primes III*, Comp. Math. **81** (1992), 19-32.

[9] ———, *Limitations to the equi–distribution of primes IV*, Proc. Royal Soc. A **435** (1991), 197-204.

[10] ———, *Relevance of the residue class to the abundance of primes*, Proc. Amalfi Conf. on Analytic Number Theory (E. Bombieri, A. Perelli, S. Salerno, U. Zannier, eds.), Salerno, Italy, 1993, pp. 95-104.

[11] P.X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976), 4-9.

[12] D.A. Goldston, *On the pair correlation conjecture for zeros of the Riemann zeta-function*, J. Reine Angew. Math. **385** (1988), 24-40.

[13] A. Granville, *Harald Cramér and the distribution of prime numbers*, Act. J. Scand. (to appear).

[14] G.H. Hardy and J.E. Littlewood, *Some problems on partitio numerorum III On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1-70.

[15] A. Hildebrand and H. Maier, *Irregularities in the distribution of primes in short intervals*, J. Reine Angew. Math. **397** (1989), 162-193.

[16] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), 221-225.

[17] H.L. Montgomery, *The Pair Correlation of Zeros of the Zeta Function*, Proc. Symp. Pure Math., vol. 24, Amer. Math. Soc., Providence, 1973, pp. 181-193.

[18] M. Nair and A. Perelli, *On the prime ideal theorem and irregularities in the distribution of primes* (to appear).

[19] A.M. Odlyzko, *The 10^{20}th Zero of the Riemann Zeta function and 70 Million of its neighbors* (to appear).

[20] M. Rubinstein and P. Sarnak, *Chebyshev's Bias* (to appear).

[21] Z. Rudnick and P. Sarnak, *The n-level correlations of L-functions* (to appear).

[22] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185-208.

[23] A. Selberg, *On the normal density of primes in small intervals and the difference between consecutive primes*, Arch. Math. Naturvid. J. **47** (1943), 87-105.

Dept, of Mathematics, University of Georgia, Athens, Georgia 30602, USA
E-mail address: `andrew@math.uga.edu`