THE DISTRIBUTION OF PRIME NUMBERS

K. Soundararajan

What follows is an expanded version of my lectures at the NATO School on Equidistribution. I have tried to keep the informal style of the lectures. In particular, I have sometimes oversimplified matters in order to convey the spirit of an argument.

Lecture 1: The Cramér model and gaps between consecutive primes

The prime number theorem tells us that $\pi(x)$, the number of primes below x, is $\sim x/\log x$. Equivalently, if p_n denotes the n-th smallest prime number then $p_n \sim n \log n$. What is the distribution of the gaps between consecutive primes, $p_{n+1} - p_n$?

We have just seen that $p_{n+1} - p_n$ is approximately $\log n$ "on average". How often do we get a gap of size $2 \log n$, say; or of size $\frac{1}{2} \log n$? One way to make this question precise is to fix an interval $[\alpha, \beta]$ (with $0 \le \alpha < \beta$) and ask for

$$(1.1) \qquad \lim_{N \to \infty} \frac{1}{N} \# \left\{ 2 \le n \le N : \quad \frac{p_{n+1} - p_n}{\log n} \in [\alpha, \beta] \right\}.$$

Does this limit exist, and if so what does it equal?

Here is another way to formulate this question. Consider intervals of the form $[n, n + \log n]$ as n ranges over integers up to N. On average such an interval contains one prime. But of course some intervals may not contain any prime, and others may contain several. Given a non-negative integer k, how often does such an interval contain exactly k primes? What is

$$(1.2) \qquad \lim_{N \to \infty} \frac{1}{N} \# \{ n \le N : \quad \pi(n + \log n) - \pi(n) = k \}?$$

Or more generally, for a fixed real number $\lambda > 0$ we may ask for

$$(1.3) \qquad \lim_{N \to \infty} \frac{1}{N} \# \{ n \le N : \quad \pi(n + \lambda \log n) - \pi(n) = k \}?$$

In this lecture we will describe the conjectured answers to these questions, but we confess at the outset that no one knows how to prove those conjectures. While conjecturing the prime number theorem, Gauss stated that the 'density of primes' around x should be

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

$1/\log x$. He based his conjecture on extensive numerical investigations. In particular he divides the numbers up to three million into intervals of length 100 (a "centad") and meticulously tabulates the number of centads with no primes, exactly one prime etc.[1] While he does not seem to make a synthesis of his results (except to conjecture the prime number theorem) it seems clear that he was seeking to understand a question like (1.3). It was left to Harald Cramér [5] to set Gauss's work on a probabilistic footing.

Cramér's model. *The primes behave like independent random variables $X(n)$ ($n \geq 3$) with $X(n) = 1$ (the number n is 'prime') with probability $1/\log n$, and $X(n) = 0$ (the number n is 'composite') with probability $1 - 1/\log n$.*

Let us suppose that the primes behave like a typical sequence in this random model, and answer questions (1.1) and (1.3). We want the 'probability' that $p_{n+1} - p_n$ lies between $\alpha \log n$ and $\beta \log n$. Thus, given the prime p_n, we want $p_n + 1$, ..., $p_n + h - 1$ to be composite, and $p_n + h$ to be prime, where $\alpha \log n \leq h \leq \beta \log n$. According to Cramér's model, this occurs with probability

$$\sum_{\alpha \log n \leq h \leq \beta \log n} \prod_{j=1}^{h-1} \left(1 - \frac{1}{\log(p_n + j)}\right) \frac{1}{\log(p_n + h)} \sim \sum_{\alpha \log n \leq h \leq \beta \log n} \left(1 - \frac{1}{\log n}\right)^{h-1} \frac{1}{\log n}$$

since $\log(p_n + j) \sim \log n$ as $p_n \sim n \log n$ and $j \ll \log n$. This is

$$\sim \sum_{\alpha \leq h/\log n \leq \beta} e^{-h/\log n} \frac{1}{\log n} \sim \int_\alpha^\beta e^{-t} dt,$$

for large n, because the LHS looks like a Riemann sum approximation to the integral in the RHS. This is the conjectured answer to question (1.1): the probability "density" of finding $p_{n+1} - p_n$ close to $t \log n$ is e^{-t}. This is an example of what is known as a "Poisson process" in the probability literature, see Feller [7].

Exercise 1. *Show similarly that the Cramér model predicts that the answer to question (1.3) is $\frac{\lambda^k}{k!} e^{-\lambda}$. This is the Poisson distribution with parameter λ.*

The reader may well object that these predictions are dubious: clearly the probability that n and $n + 1$ are both primes must be zero, but the Cramér model assigns this event a probability $1/(\log n \log(n + 1))$. More generally, suppose we are given a set $\mathcal{H} = \{h_1, \ldots, h_k\}$ of k distinct integers, and we ask for the number of integers $n \leq x$ with $n + h_1$, $n + h_2$, ..., $n + h_k$ all being prime. The Cramér model would predict an answer of $\sim x/(\log x)^k$, but clearly we must take into account arithmetic properties of the set $\mathcal{H}$. For example, if there were a prime p such that the integers h_1, ..., h_k occupied all the residue classes (mod p) then the integers $n + h_1$, ..., $n + h_k$ would also occupy all the residue classes (mod p). In particular one of these numbers would be a multiple of p, and so there can only be finitely many values of n with $n + h_1$, ..., $n + h_k$ all being prime.

---

[1]We refer the reader to www.math.princeton.edu/~ytschink/.gauss for scans of Gauss's manuscripts showing these calculations.

In [17] Hardy and Littlewood proposed the prime k-tuple conjecture that

$$(1.4) \qquad \#\{n \le x : \ n+h_1, n+h_2, \ldots, n+h_k \text{ prime}\} \sim \mathfrak{S}(\mathcal{H})\frac{x}{(\log x)^k},$$

for a certain constant $\mathfrak{S}(\mathcal{H})$ called the 'singular series.' The constant $\mathfrak{S}(\mathcal{H})$ equals 0 if the elements $h_1, \ldots, h_k$ occupy a complete set of residue classes $\pmod p$ for some prime p, and $\mathfrak{S}(\mathcal{H})$ is positive otherwise. We will describe this conjecture in more detail below. The aim of this lecture is to describe a beautiful calculation of Gallagher [9] which shows that the Hardy-Littlewood conjecture (1.4) implies the same distribution of gaps between primes predicted by the Cramér random model. The crux of his proof is that although $\mathfrak{S}(\mathcal{H})$ is not always 1 (as the Cramér model would have), it is ~ 1 on average over all k-element sets $\mathcal{H}$ with the $h_j \le h$.

The Hardy-Littlewood Conjecture. We now motivate the Hardy-Littlewood conjecture (1.4) and describe the singular series $\mathfrak{S}(\mathcal{H})$ that arises there. As a toy model for prime numbers let us fix an integer q and consider the reduced residue classes $\pmod q$. Out of the q total residue classes, there are $\phi(q)$ reduced classes, and we may think of $\phi(q)/q$ as the 'probability' of a class being reduced. Now suppose we are given the set $\mathcal{H} = \{h_1, \ldots, h_k\}$ and we ask for the number of $n \pmod q$ such that $n+h_1, \ldots, n+h_k$ are all coprime to q. For convenience, let us just think of square-free q. If these k events were independent then the answer would be $q(\phi(q)/q)^k$. The correct answer is a little different: for each prime p that divides q we need n to avoid the residue classes $-h_1, \ldots, -h_k \pmod p$. Let $\nu_{\mathcal{H}}(p)$ denote the number of distinct residue classes occupied by $\mathcal{H} \pmod p$. Thus n must lie in one of $p - \nu_{\mathcal{H}}(p)$ residue classes $\pmod p$. Using the chinese remainder theorem we see easily that the correct answer is

$$\prod_{p|q}(p - \nu_{\mathcal{H}}(p)) = q\prod_{p|q}\left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right) = q\left(\frac{\phi(q)}{q}\right)^k \prod_{p|q}\left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

Let us write $\mathfrak{S}(\mathcal{H};q) = \prod_{p|q}(1 - \frac{\nu_{\mathcal{H}}(p)}{p})(1 - \frac{1}{p})^{-k}$. We have seen that the answer for the number of $n \pmod q$ with $n+h_1, \ldots, n+h_k$ all being coprime to q involves correcting the guess $q(\phi(q)/q))^k$ by the factor $\mathfrak{S}(\mathcal{H};q)$ which keeps track of the arithmetic properties of the set $\mathcal{H}$. Now let us consider what happens when we take $q = q_\ell = \prod_{p \le \ell} p$ and let ℓ go to infinity. As $\ell \to \infty$ we see that

$$\mathfrak{S}(\mathcal{H};q_\ell) \to \prod_{p}\left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

The infinite product above converges because if p is larger than all the h_j's then $\nu_{\mathcal{H}}(p) = k$ and so $(1 - \nu_{\mathcal{H}}(p)/p)(1 - 1/p)^{-k} = (1 - k/p)(1 - 1/p)^{-k} = 1 + O(p^{-2})$. This infinite product is the singular series[2]:

$$(1.5) \qquad\qquad \mathfrak{S}(\mathcal{H}) := \prod_{p}\left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

---

[2]The term arises from Hardy and Littlewood's original derivation of their conjecture using the circle method. Here $\mathfrak{S}(\mathcal{H})$ arose as a series rather than the product given above.

Further if $n+h_1, \ldots, n+h_k$ are coprime to q_ℓ with ℓ large, then they have no small prime divisors, and may reasonably be viewed as a kind of approximation to primes. Thus in formulating a conjecture on the number of $n \leq x$ with $n + h_1, \ldots, n + h_k$ being prime, a natural guess is to take the random answer $x/(\log x)^k$, and multiply it by the arithmetical correction factor $\mathfrak{S}(\mathcal{H})$. This is precisely the Hardy-Littlewood conjecture (1.4). It is immediate from (1.5) that $\mathfrak{S}(\mathcal{H}) = 0$ if and only if $\mathcal{H}$ exhausts a compete set of residue classes $\pmod{p}$ for some p.

Gallagher's calculation. We will now describe Gallagher's argument, using the Hardy-Littlewood conjecture (1.4) to justify the distribution of gaps between primes predicted by the Cramér model. The precise problem we consider is a close variant of question (1.3). Let λ be a positive real number, and let N be large. We set $h = \lambda \log N$ and seek to understand the distribution of $\pi(n + h) - \pi(n)$ as n varies over the natural numbers below N. To understand this quantity, we consider the moments

$$(1.6) \qquad \frac{1}{N} \sum_{n \leq N} (\pi(n+h) - \pi(n))^r = \frac{1}{N} \sum_{n \leq N} \Big(\sum_{\substack{\ell=1 \\ n+\ell \text{ prime}}}^{h} 1 \Big)^r,$$

where r is a natural number. If the Cramér prediction is right, then we may expect these moments to be approximately

$$(1.7) \qquad \frac{1}{N} \mathbb{E}\Big(\sum_{2 \leq n \leq N} \Big(\sum_{\ell=1}^{h} X(n+\ell) \Big)^r \Big),$$

where $\mathbb{E}$ denotes expectation, and the $X(n)$'s are independent random variables as in Cramér's model. If these moments are roughly equal for $r \leq R$ for any $R = R(N)$ tending to infinity, then we would know that $\pi(n + h) - \pi(n)$ has a Poisson distribution with parameter λ. This is because of a well-known principle from probability, that nice distributions including the Poisson distribution are determined by their moments.

Let us expand out the r-th powers in (1.6) and (1.7). We then get numbers $\ell_1, \ldots, \ell_r$ below h not necessarily distinct and would like to understand how often $n + \ell_1, \ldots, n + \ell_r$ are all prime (for (1.6)), or to understand $\mathbb{E}(X(n + \ell_1) \cdots X(n + \ell_r))$ (for (1.7)). Let us suppose that there are exactly k distinct numbers among the $\ell_1, \ldots, \ell_r$ and write these distinct numbers as $(1 \leq) h_1 < h_2 < \ldots < h_k (\leq h)$. The number of choices for $\ell_1, \ldots, \ell_r$ that lead to the same ordered set of distinct numbers $h_1, \ldots, h_k$ is the number of different ways of mapping $\{1, 2, \ldots, r\}$ onto $\{1, \ldots, k\}$; let us denote this[3] by $\sigma(r, k)$. Thus we see that (1.6) may be written as

$$(1.8) \qquad \sum_{k=1}^{r} \sigma(r, k) \sum_{1 \leq h_1 < h_2 < \ldots < h_k \leq h} \Big(\frac{1}{N} \sum_{\substack{n \leq N \\ n+h_1, \ldots, n+h_k \text{ prime}}} 1 \Big),$$

---

[3]This is a 'Stirling number of the second kind.'

while (1.7) may be written as

$$(1.9) \qquad \sum_{k=1}^{r} \sigma(r,k) \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} \left(\frac{1}{N} \sum_{2 \le n \le N} \mathbb{E}(X(n+h_1) \cdots X(n+h_k)) \right).$$

Since the same quantity appears in both (1.8) and (1.9) and is non-negative, we don't need to worry about what $\sigma(r,k)$ is.

Invoking the Hardy-Littlewood conjecture (1.4)[4] we get that (1.8) is

$$\sim \sum_{k=1}^{r} \frac{\sigma(r,k)}{(\log N)^k} \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} \mathfrak{S}(\{h_1, \ldots, h_k\}).$$

Clearly the quantity in (1.9) is

$$\sim \sum_{k=1}^{r} \frac{\sigma(r,k)}{(\log N)^k} \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} 1.$$

Thus, to show that (1.6) and (1.7) are approximately equal, we need only show that

$$(1.10) \qquad \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} \mathfrak{S}(\{h_1, \ldots, h_k\}) \sim \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} 1.$$

This is Gallagher's crucial result in [9]. It shows that although the Hardy-Littlewood probabilities are different from the Cramér probabilities, on average they are roughly equal. This explains why the Cramér model makes accurate predictions for the distribution of primes in such short intervals.

Exercise 2. *For a prime p put $\mathfrak{S}(\mathcal{H}; p) = (1 - \nu_{\mathcal{H}}(p)/p)(1 - 1/p)^{-k}$. Prove that as $h \to \infty$*

$$\sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} \mathfrak{S}(\mathcal{H}; p) \sim \sum_{1 \le h_1 < h_2 < \ldots < h_k \le h} 1.$$

Explain why this morally implies (1.10); better still prove (1.10) rigorously (or read Gallagher's argument [9]).

Exercise 3. *We have sketched how the Hardy-Littlewood conjecture implies that for a given positive real number λ, and a fixed non-negative integer k,*

$$\frac{1}{N} \#\{n \le N : \quad \pi(n + \lambda \log N) - \pi(n) = k\} \sim \frac{\lambda^k}{k!} e^{-\lambda}.$$

Deduce that

$$\frac{1}{N} \#\left\{ 2 \le n \le N : \quad \frac{p_{n+1} - p_n}{\log n} \in [\alpha, \beta] \right\} \sim \int_{\alpha}^{\beta} e^{-t} dt.$$

---

[4]Precisely, we need this conjecture uniformly for all $h_1, \ldots, h_k$ below h, and for all $k \le R$ with $R = R(N)$ tending slowly to infinity.

Proof of (1.10) when $k = 2$. From the definition (1.5) note that $\mathfrak{S}(\{h_1, h_2\}) = \mathfrak{S}(\{0, h_2 - h_1\})$ and so, letting $\ell = h_2 - h_1$ we see that the LHS of (1.10) is (in the case $k = 2$)

$$\sum_{\ell \le h} \mathfrak{S}(\{0, \ell\}) \Big(\sum_{\substack{1 \le h_1 < h_2 \le h \\ h_2 - h_1 = \ell}} 1 \Big) = \sum_{\ell \le h} \mathfrak{S}(\{0, \ell\})(h - \ell).$$

To evaluate the above asymptotically, it is useful to study the generating Dirichlet series

$$F(s) := \sum_{\ell=1}^{\infty} \frac{\mathfrak{S}(\{0, \ell\})}{\ell^s}.$$

The definition (1.5) gives that $\mathfrak{S}(\{0, \ell\}) = \prod_{p|\ell}(1 - 1/p)^{-1} \prod_{p \nmid \ell}(1 - 2/p)(1 - 1/p)^{-2}$. From this, we may see that $F(s)$ converges absolutely in the half-plane $\mathrm{Re}(s) > 1$, and moreover in that region has the Euler product

$$F(s) = \prod_p \left(\Big(1 - \frac{2}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-2} + \frac{1}{p^s}\Big(1 - \frac{1}{p}\Big)^{-1} + \frac{1}{p^{2s}}\Big(1 - \frac{1}{p}\Big)^{-1} + \frac{1}{p^{3s}}\Big(1 - \frac{1}{p}\Big)^{-1} + \dots \right).$$

Mutliplying and dividing by $\zeta(s) = \prod_p (1 - 1/p^s)^{-1}$ we see (with a little calculation) that in $\mathrm{Re}(s) > 1$,

$$F(s) = \zeta(s) \prod_p \Big(1 - \frac{1}{(p-1)^2} + \frac{1}{p^{s-1}(p-1)^2} \Big) = \zeta(s)G(s),$$

say. The Euler product for $G(s)$ converges absolutely in $\mathrm{Re}(s) > 0$ and so in that region we have obtained a meromorphic continuation of $F(s)$ with a simple pole at $s = 1$ coming from the simple pole of $\zeta(s)$ there. We now make use of the formula that for any $c > 0$

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{y^s}{s(s+1)} ds = \begin{cases} (1 - 1/y) & \text{if } y > 1 \\ 0 & \text{if } 0 < y \le 1. \end{cases}$$

This is easily proved by moving the line of integration to the left if $y > 1$ and to the right if $y \ge 1$; the term $1 - 1/y$ when $y > 1$ arises from the residues of the poles at $s = 0$ and $s = -1$. Therefore, if $c > 1$, we see that

$$h \sum_{\ell \le h} \mathfrak{S}(\{0, \ell\}) \Big(1 - \frac{\ell}{h} \Big) = h \sum_{\ell=1}^{\infty} \mathfrak{S}(\{0, \ell\}) \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Big(\frac{h}{\ell} \Big)^s \frac{ds}{s(s+1)}$$

$$= \frac{h}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(s)G(s) \frac{h^s}{s(s+1)} ds,$$

where the interchange of summation and integration is justified by the absolute convergence of $F(s)$ in the region $\mathrm{Re}(s) > 1$. To evaluate the contour integral, we shift the line of

integration to $\mathrm{Re}(s) = \epsilon > 0$. In the region traversed we encounter only a simple pole at $s = 1$ (because of $\zeta(s)$) and so our integral is

$$h\operatorname*{Res}_{s=1}\left(\frac{\zeta(s)G(s)h^s}{s(s+1)}\right) + \frac{h}{2\pi i}\int_{\epsilon-i\infty}^{\epsilon+i\infty}\zeta(s)G(s)\frac{ds}{s(s+1)}.$$

Since $G(1)$ is easily seen to be 1, the residue above equals $G(1)h^2/2 = h^2/2$. By bounding $\zeta(s)$ and $G(s)$ on the line $\mathrm{Re}(s) = \epsilon$ we may estimate the remaining integral on that line; we omit the standard, but technical, details and merely note that this term is $O(h^{1+\epsilon})$. We conclude that

$$\sum_{\ell \le h}\mathfrak{S}(\{0, \ell\})(h - \ell) = \frac{h^2}{2} + O(h^{1+\epsilon}) = \sum_{h_1 < h_2 \le h} 1 + O(h^{1+\epsilon}).$$

This proves (1.10) in the case $k = 2$.

Exercise 4. *Analyze $G(s)$ further by writing it as $\zeta(s+1)H(s)$, where $H(s)$ is now analytic in $\mathrm{Re}(s) > -\frac{1}{2}$. Evaluating residues, as above, prove that*

$$\sum_{\ell \le h}\mathfrak{S}(\{0, \ell\})(h - \ell) = \frac{h^2}{2} - \frac{h\log h}{2} + \frac{Bh}{2} + O(h^{\frac{1}{2}+\epsilon}),$$

with $B = 1 - \gamma - \log 2\pi$; here γ is Euler's constant.

Concluding remarks. Two important consequences of our predictions for the spacings between primes are that

$$\limsup_{n \to \infty}\frac{p_{n+1} - p_n}{\log n} = \infty, \qquad \text{and} \qquad \liminf_{n \to \infty}\frac{p_{n+1} - p_n}{\log n} = 0.$$

Happily both these results have now been proved. The first involves constructing long strings of composite numbers, and was first proved by Westzynthius with important refinements due to Erdős and Rankin. The second is a recent breakthrough of Goldston, Pintz and Yıldırım , see [12]. The reader may consult the survey by Heath-Brown [18] for the lim sup result and much else besides, and my survey [31] for the lim inf result.

LECTURE 2: THE DISTRIBUTION OF PRIMES IN LONGER INTERVALS

Cramér's prediction. In the first lecture we considered the distribution of primes in intervals of length a constant times the average spacing. We now discuss what happens in longer intervals. Precisely, we consider $\pi(n + h) - \pi(n)$ for $n \le N$ and where $h/\log N$ is large, but h/N is small.

Exercise 5. *Using Stirling's formula, show that as λ gets large, a Poisson distribution with parameter λ begins to look like a normal distribution with mean λ and variance λ.*

Thus Cramér's model would suggest that, if $h/\log N$ is large but h/N is small, then for $n \le N$, $\pi(n + h) - \pi(n)$ has an approximately normal distribution with mean $\sim h/\log N$

and variance $\sim h/\log N$. Another way to arrive at this prediction is to calculate the moments (note that for most $n \le N$, $\sum_{\ell=1}^{h} 1/\log(n+\ell) \sim h/\log N$)

$$(2.1\text{a}) \qquad \frac{1}{N}\mathbb{E}\Big(\sum_{2 \le n \le N} \Big(\sum_{\ell=1}^{h} X(n+\ell) - \sum_{\ell=1}^{h} \frac{1}{\log(n+\ell)} \Big)^{k} \Big),$$

which we claim is

$$(2.1\text{b}) \qquad = \frac{k!}{2^{k/2}(k/2)!}\Big(\frac{h}{\log N}\Big)^{k/2}\Big(1 + O_k\Big(\frac{\log N}{h}\Big)\Big)$$

if k is even, and if k is odd it is

$$(2.1\text{c}) \qquad \ll \Big(\frac{h}{\log N} \Big)^{(k-1)/2}.$$

Exercise 6. *Justify (2.1a)-(2.1c) by arguing as follows. For $n \ge 3$, set $X_0(n) = 1 - 1/\log n$ with probability $1/\log n$ and $X_0(n) = -1/\log n$ with probability $1 - 1/\log n$: that is, $X_0(n) = X(n) - 1/\log n$. Note that $\mathbb{E}(X_0(n)) = 0$. Expand*

$$\frac{1}{N}\mathbb{E}\Big(\sum_{2 \le n \le N} \Big(\sum_{1 \le \ell \le h} X_0(n+\ell) \Big)^{k} \Big) = \frac{1}{N} \sum_{\ell_1,\dots,\ell_k \le h} \sum_{2 \le n \le N} \mathbb{E}(X_0(n+\ell_1) \cdots X_0(n+\ell_k)).$$

The expectation above is zero if any of the ℓ_i's occurs only once among $\ell_1, \dots, \ell_k$. When k is even there is a leading contribution from terms where the $\ell_1, \dots, \ell_k$ contain $k/2$ distinct numbers each occurring twice.

Calculating the variance via Hardy-Littlewood. However, we do not believe that this prediction, given by the Cramér model, is accurate. At this juncture, it is more convenient to deal with $\psi(n+h) - \psi(n)$, where $\psi(x) = \sum_{n \le x} \Lambda(n)$ with $\Lambda(n)$ denoting the von Mangoldt function. Note that the prime number theorem is equivalent to $\psi(x) \sim x$, and that the Hardy-Littlewood conjecture (1.4) may be recast as ($\mathcal{H} = \{h_1, \dots, h_k\}$ is a set of k distinct numbers)

$$(2.2) \qquad \sum_{n \le x} \Lambda(n+h_1) \cdots \Lambda(n+h_k) \sim \mathfrak{S}(\mathcal{H})x.$$

The Cramér model predicts that $\psi(n+h) - \psi(n)$ is approximately normal with mean $\sim h$ and variance $\sim h \log N$.

To see the flaw in this prediction, let us now calculate the variance using the Hardy-Littlewood conjectures. Note that

$$\frac{1}{N} \sum_{n \le N} (\psi(n+h) - \psi(n) - h)^2 = \frac{1}{N} \sum_{n \le N} \Big(\sum_{\ell \le h} \Lambda(n+\ell) \Big)^2 - 2\frac{h}{N} \sum_{n \le N} \sum_{\ell \le h} \Lambda(n+\ell) + h^2.$$

The middle term in the RHS above is $-2h^2(\psi(N) + O(h \log N))/N \sim -2h^2$. As for the first term in the RHS we may square it out, and invoke the Hardy-Littlewood conjecture (2.2). If we forget all the error terms, then the above is

$$\frac{1}{N} \sum_{n \leq N} \sum_{\ell \leq h} \Lambda(n + \ell)^2 + 2 \sum_{\ell \leq h} \mathfrak{S}(\{0, \ell\})(h - \ell) - h^2.$$

The prime number theorem and partial summation gives that the first term above is $\sim h(\log N - 1)$, while from Exercise 4 we see that the second term above is $\sim h^2 - h \log h + Bh$. So, ignoring all error terms, we conclude that the variance satisfies

(2.3)
$$\frac{1}{N} \sum_{n \leq N} (\psi(n + h) - \psi(n) - h)^2 \sim h\left(\log \frac{N}{h} + B - 1\right),$$

which is different from the $\sim h \log N$ predicted by Cramér's model.

Exercise 7. *Assume that the Hardy-Littlewood conjecture (2.2) holds in the quantitative form*

$$\sum_{n \leq x} \Lambda(n + h_1) \cdots \Lambda(n + h_k) = \mathfrak{S}(\mathcal{H})x + O(x^{\frac{1}{2} + \epsilon}),$$

uniformly for $k \leq K$, and distinct h_j satisfying $1 \leq h_j \leq x$. Using this, obtain (2.3) with an error term of $O(h^{\frac{1}{2} + \epsilon} + h^2 N^{-\frac{1}{2} + \epsilon} + h^3 N^{-1})$. Thus, even assuming the quantitative Hardy-Littlewood conjectures, one knows (2.3) only for $h \leq N^{\frac{1}{2} - \epsilon}$.

So although the Hardy-Littlewood probabilities and the Cramér probabilities are roughly equal on average, significant deviations show up when we consider h to be a small power of N. We believe that (2.3) is the right asymptotic for the variance and the Cramér model predicts the wrong answer.

The variance and zeros of the zeta function. Here is the sketch of a very different calculation which leads to the same answer as (2.3). Riemann's explicit formula (see [6]) says that

$$\psi(x) = x - \sum_{\rho} \frac{x^\rho}{\rho} + \text{negligible terms.}$$

Here ρ runs over the non-trivial zeros of the Riemann zeta-function. We assume the Riemann hypothesis and write $\rho = \frac{1}{2} + i\gamma$. The sum over zeros is only conditionally convergent, but we will argue loosely omitting such considerations. It follows that

$$\psi(x + h) - \psi(x) - h = - \sum_{\rho} \frac{(x + h)^\rho - x^\rho}{\rho} + \text{negligible terms.}$$

The sum over zeros above is weighted down with a factor $1/\rho$, and so we may expect that large zeros make a minor contribution. It turns out that zeros with $|\rho| \geq x/h$ are not so

important. For the small zeros, we replace $(x + h)^\rho - x^\rho$ by the Taylor approximation $\rho h x^{\rho-1}$. Therefore we may expect that

$$\frac{1}{X} \int_X^{2X} (\psi(x + h) - \psi(x) - h)^2 dx \approx \frac{h^2}{X^2} \int_X^{2X} \Big| \sum_{|\gamma| \leq X/h} x^{i\gamma} \Big|^2 dx$$

(2.4)
$$= \frac{h^2}{X} \sum_{|\gamma_1|, |\gamma_2| \leq X/h} X^{i(\gamma_1 - \gamma_2)} \frac{2^{1+i(\gamma_1 - \gamma_2)} - 1}{1 + i(\gamma_1 - \gamma_2)}.$$

There are $\ll \log T$ zeros of the zeta-function with ordinates lying between T and $T + 1$. Using this observation in (2.4), and estimating the magnitude of the sums over zeros there, we "deduce" that, assuming RH,[5]

(2.5)
$$\frac{1}{X} \int_X^{2X} (\psi(x + h) - \psi(x) - h)^2 dx \ll h(1 + \log X/h)^2.$$

A result like this was established by Selberg [30]. If we want an asymptotic in (2.4), then we need some understanding of the spacings $\gamma_1 - \gamma_2$ between zeros of the Riemann zeta-function. Such an understanding is provided by the pair correlation conjecture of Montgomery [24], which predicts that these ordinates are distributed like eigenvalues of large random matrices. Using such information Mueller obtained an asymptotic formula much like (2.3), and Goldston and Montgomery [11] showed conversely that a formula like (2.3) also conveys information about the zeros of $\zeta(s)$. For more discussion on this set of ideas consult Goldston's recent survey [10].

Higher moments. Recently, Montgomery and I (see [25] and [26]) used a quantitative form of the Hardy-Littlewood conjecture (see Exercise 7) to study higher moments of $\psi(n + h) - \psi(n) - h$. We now describe these results briefly. They support the conjecture that if $(\log N)^{1+\delta} \leq h \leq N^{1-\delta}$ then for $n \leq N$ the distribution of $\psi(n + h) - \psi(n)$ is approximately normal with mean h and variance $h \log(N/h)$.

We assume that $(\log N)^{1+\delta} \leq h \leq N^{1-\delta}$ and wish to evaluate

(2.6)
$$\frac{1}{N} \sum_{n \leq N} (\psi(n + h) - \psi(n) - h)^r.$$

For even r we expect that this is $\sim \frac{r!}{2^{r/2}(r/2)!} (h \log N/h)^{r/2}$, while for odd r we expect it to be $o((h \log N/h)^{r/2})$. If we simply expanded $(\psi(n + h) - \psi(n) - h)^r$ in powers of $(\psi(n + h) - \psi(n))$ and h (as we did in the case $r = 2$) then we would get many terms all of size h^r, and a careful cancellation of these and lower order terms is needed before we get to the actual delicate main term of size essentially $h^{r/2}$. To circumvent this, we define $\Lambda_0(n) = \Lambda(n) - 1$, in analogy with Exercise 6. This eliminates the unnecessary higher order terms at the outset, and simplifies calculations considerably. For other situations where this trick helps, see my paper with Granville [15] in this volume. Using this notation, and expanding (2.6) we want to understand

(2.7)
$$\sum_{\ell_1, \dots, \ell_r \leq h} \frac{1}{N} \sum_{n \leq N} \Lambda_0(n + \ell_1) \cdots \Lambda_0(n + \ell_r).$$

---

[5]In fact we would only deduce $\ll h(1 + \log X/h)^3$ but the extra "log" may be removed by smoothing.

Exercise 8. *Define the modified singular series $\mathfrak{S}_0(\mathcal{H})$ by*

$$\mathfrak{S}_0(\mathcal{H}) = \sum_{\mathcal{J} \subset \mathcal{H}} (-1)^{|\mathcal{H}| - |\mathcal{J}|} \mathfrak{S}(\mathcal{J}), \qquad so \ that \quad \mathfrak{S}(\mathcal{H}) = \sum_{\mathcal{J} \subset \mathcal{H}} \mathfrak{S}_0(\mathcal{J}).$$

Here we understand that $\mathfrak{S}(\emptyset) = \mathfrak{S}_0(\emptyset) = 1$. Show that the quantitative Hardy-Littlewood conjecture of Exercise 7 is the same as

$$\sum_{n \leq x} \Lambda_0(n + h_1) \cdots \Lambda_0(n + h_k) = \mathfrak{S}_0(\mathcal{H}) x + O(x^{\frac{1}{2} + \epsilon}),$$

keeping the hypotheses there.

For simplicity, consider first the terms in (2.7) when the ℓ_i are distinct. If we use the asymptotic of Exercise 8 we are led to the problem of evaluating

$$\sum_{\substack{h_1, \ldots, h_k \leq h \\ h_i \ \text{distinct}}} \mathfrak{S}_0(\mathcal{H}),$$

which is a problem analogous to, but more delicate than, Gallagher's calculation (1.10). The main result in [26] is the asymptotic

$$(2.8) \qquad \sum_{\substack{h_1, \ldots, h_k \leq h \\ h_i \ \text{distinct}}} \mathfrak{S}_0(\mathcal{H}) \ = \ \begin{cases} \{1 + o(1)\} \frac{k!}{2^{k/2}(k/2)!} (-h \log h + B + 1)^{k/2} & \text{if } k \text{ is even} \\ o((h \log h)^{k/2}) & \text{if } k \text{ is odd}. \end{cases}$$

Exercise 9. *Show the following refinement of Gallagher's (1.10):*

$$\sum_{\substack{h_1, \ldots, h_k \leq h \\ h_i \ \text{distinct}}} \mathfrak{S}(\mathcal{H}) = h^k - \binom{k}{2} h^{k-1} \log h + \binom{k}{2} B h^{k-1} + O(h^{k-3/2+\epsilon}).$$

Returning to (2.7), we must analyze the terms when the ℓ_i are not necessarily distinct. Suppose that h_1, ..., h_k are the distinct elements among ℓ_1, ..., ℓ_r and that each h_i appears $m_i (\geq 1)$ times among the ℓ_i. After a little combinatorics, we may write (2.7) as

$$(2.9) \qquad \sum_{k=1}^{r} \sum_{\substack{m_1, \ldots, m_k \geq 1 \\ \sum m_i = r}} \binom{r}{m_1, \ldots, m_k} \frac{1}{k!} \sum_{\substack{h_1, \ldots, h_k \leq h \\ h_j \ \text{distinct}}} \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{k} \Lambda_0(n + h_i)^{m_i}.$$

We must distinguish the indices where $m_i = 1$ and the remaining indices where $m_i > 1$. Let $\mathcal{I}$ denote the subset of $\{1, \ldots, k\}$ such that $m_i = 1$ for $i \in \mathcal{I}$. For $i \notin \mathcal{I}$ (so $m_i \geq 2$) we think of $\Lambda_0(n + h_i)^{m_i}$ as being essentially $(\log N)^{m_i - 1} \Lambda(n + h_i)$: the point is that both quantities have about the same expected value $(\log N)^{m_i - 1}$, unlike the case when $m_i = 1$

where the expected value of $\Lambda(n + h_i)$ and $\Lambda_0(n + h_i)$ are 1 and 0 respectively. Therefore the inner sum over n in (2.9) is essentially

$$\frac{(\log N)^{r-k}}{N} \sum_{n=1}^{N} \prod_{i \in \mathcal{I}} \Lambda_0(n + h_i) \prod_{\substack{1 \leq i \leq k \\ i \notin \mathcal{I}}} (\Lambda_0(n + h_i) + 1)$$

$$= \frac{(\log N)^{r-k}}{N} \sum_{\mathcal{I} \subset \mathcal{J} \subset \{1,\dots,k\}} \sum_{n=1}^{N} \prod_{j \in \mathcal{J}} \Lambda_0(n + h_j).$$

Now we invoke the Hardy-Littlewood conjecture of Exercise 8, and use (2.8).

Exercise 10. *Complete the details in evaluating (2.7). Show that when r is odd or any of the m_i's is ≥ 3 we get a contribution of $o(h \log N)^{r/2}$. In the case r is even, the main term $\frac{r!}{2^{r/2}(r/2)!} (h \log N/h)^{r/2}$ arises from contributions to (2.9) where the m_i are all 1 or 2.*

The proof of (2.8) is quite complicated, and we do not go into it here. Let us however point out one important ingredient. While motivating the Hardy-Littlewood conjecture in Lecture 1, we considered the toy problem of reduced residues (mod q). If $1 = a_1 < a_2 < \dots < a_{\phi(q)} < q$ are the reduced residues below q, then we may ask for the distribution of $(a_{i+1} - a_i)(\phi(q)/q)$; we have multiplied by $\phi(q)/q$ so that this is 1 'on average.' If, for example, q is the product of the first ℓ primes, then as in Lecture 1 we may think of these a_i as being like primes, and expect that, for $0 < \alpha < \beta$,

$$\#\left\{1 \leq i \leq \phi(q) - 1 : \quad (a_{i+1} - a_i)\frac{\phi(q)}{q} \in [\alpha, \beta]\right\} \sim \phi(q) \int_{\alpha}^{\beta} e^{-x} dx.$$

A beautiful result of Hooley [20] shows that this holds provided $\phi(q)/q$ is small. Obviously, some restriction on $\phi(q)/q$ is necessary; for example, if q is prime then clearly $a_{i+1} - a_i = 1$ for $1 \leq i \leq \phi(q) - 1$. Moreover, Montgomery and Vaughan [27] have even estimated the moments:

$$M_k(q; h) = \sum_{n=1}^{q} \left(\sum_{\substack{\ell \leq h \\ (n+\ell, q)=1}} 1 - h\frac{\phi(q)}{q} \right)^k.$$

The proof of (2.8) builds on the techniques developed there.

In our discussion above we have ignored error terms altogether. If one argues carefully using the quantitative Hardy-Littlewood conjectures of Exercises 7 and 8, we can evaluate the r-th moment (2.6) provided that $h \leq N^{1/r-\epsilon}$. We expect that the same asymptotics hold even when h is larger with $h \leq N^{1-\epsilon}$. Thus these arguments suggest that for $(\log N)^{1+\delta} \leq h \leq N^{1-\delta}$, the distribution of $\psi(n + h) - \psi(n)$ (for $n \leq N$ is approximately normal with mean h and variance $h \log N/h$. For numerical support for this conjecture, see [3] and [25]. For other work related to this circle of ideas, see [3] and [4].

Connections with zeros of $\zeta(s)$? We mentioned earlier the work of Goldston and Montgomery relating the variance of primes in short intervals to the pair correlation of

zeros of $\zeta(s)$. Our calculations on the higher moments of primes in short intervals suggest that if $X \geq T^{1+\epsilon}$ then[6]

$$\int_X^{2X} \Big(\sum_{|\gamma| \leq T} x^{i\gamma} \Big)^k dx = \int_X^{2X} \Big(\sum_{0 \leq \gamma \leq T} 2\cos(\gamma \log x) \Big)^k dx$$

is $\sim \frac{k!}{2^{k/2}(k/2)!} X(2N(T))^{k/2}$ if k is even, and is $o(XN(T)^{k/2})$ if k is odd. Here $N(T) \sim \frac{T}{2\pi} \log T$ denotes the number of zeros of $\zeta(s)$ with $0 \leq \gamma \leq T$. Viewed this way, Montgomery's pair correlation conjecture may be thought of as saying that for $x \geq T^{1+\epsilon}$ the sum $\sum_{0 \leq \gamma \leq T} \cos(\gamma \log x)$ behaves like a sum of uncorrelated random variables. The higher moments suggest that it behaves in fact like a sum of independent random variables.[7] These statements are quite vague, and it would be nice to flesh out the precise connection between these higher moments and zeros of $\zeta(s)$. For other connections between zeros of $\zeta(s)$ and Hardy-Littlewood type conjectures see [2].

Chebyshev's bias. We have considered above the distribution of primes in short intervals. What happens to the distribution in long intervals $[1, x]$? That is what can be said about the distribution of $\psi(x) - x$. Assuming RH, we get from Riemann's explicit formula that this is essentially $-2x^{\frac{1}{2}}\text{Re} \sum_{0 < \gamma} x^{i\gamma}/(1/2+i\gamma)$. It is expected[8] that the zeros of $\zeta(s)$ are all simple, and have no non-trivial $\mathbb{Q}$-linear relations among them. In that case the sum over zeros above may be modeled by $\text{Re} \sum_{0 < \gamma} X(\gamma)/(1/2+i\gamma)$ where the $X(\gamma)$ are independent random variables, taking uniformly distributed values on the unit circle. Precisely, as t varies from 1 to T, the distribution of $(\psi(e^t) - e^t)/(2e^{t/2})$ is like the distribution of our random sum above.[9] This is a certain non-universal distribution, which has been investigated in, for example, [23] and [29]. To gain a flavor of this distribution the reader may contemplate $\sum_{n=1}^{\infty} X_n/n$ where the X_n are independent random variables taking the values ± 1 with equal probability.

The distribution above is symmetric about the origin, and so $\psi(x)$ is as likely to be larger than x as it is to be smaller than x. However, $\psi(x) = \theta(x) + \theta(x^{\frac{1}{2}}) + \theta(x^{\frac{1}{3}}) + \dots$ where $\theta(x) = \sum_{p \leq x} \log p$. Thus it is much more likely for $\theta(x)$ to be smaller than x than for it to be larger than x. By partial summation one gets that $\pi(x) < \text{li}(x)$ much more often than $\pi(x) > \text{li}(x)$. In fact, in a certain sense the probability that $\pi(x)$ 'beats' $\text{li}(x)$ is only $0.00000026\dots$! We stop here, referring the reader to Rubinstein and Sarnak [29], and the delightful recent survey [14] for more information.

To summarize, we found three distinct behaviors for the distribution of primes in intervals. At the "microscopic" scale ($h \asymp \log N$) there is Poisson behavior, at the "mesoscopic" scale ($h/\log N \to \infty$, $h = o(N)$) there is Gaussian (normal) behavior, and at the "macroscopic" scale ($h \gg N$) there is a specific non-universal distribution law. Such division into three regimes occurs in many other problems as well; for example, in the distribution of

---

[6]We take this opportunity to point out that the important constraint $X \geq T^{1+\epsilon}$ has been erroneously omitted in a similar discussion (on page 594) in [26].

[7]This is analogous to a result of E. Rains [28] in random matrix theory.

[8]There is perhaps no good reason for this belief, except that the contrary situation is harder to imagine!

[9]The change of variable $x = e^t$ means that $x^{i\gamma} = e^{it\gamma}$ now takes values uniformly on the unit circle as t varies.

lattice points in the plane. As a starting point, we refer the reader to the recent paper of Hughes and Rudnick [21] and to the references therein.

LECTURE 3: MAIER'S METHOD AND AN "UNCERTAINTY PRINCIPLE"

If the Riemann Hypothesis is true, then from Selberg's result (2.5) we easily deduce that (for $h \leq N$) the number of $n \leq N$ with $|\psi(n+h) - \psi(n) - h| \geq \sqrt{h}(\log N)^{1+\delta}$ is $\ll N/(\log N)^{2\delta}$. It follows that if $N \geq h \geq (\log N)^{2+\delta}$ then "almost all" intervals $(n, n+h]$ with $n \leq N$ contain about the correct number of primes, $\sim h/\log N$. If (2.3) holds then we can even conclude that if $h \geq (\log N)^{1+\delta}$ then "almost all" intervals $(n, n+h]$ with $n \leq N$ contain approximately the correct number of primes. In Cramér's model, one can show that almost surely $\sum_{x \leq n \leq x+h} X(n) \sim h/\log x$ if $(\log x)^{2+\delta} \leq h \leq x$. Thus it seems quite plausible that if x is large and $x \geq h \geq (\log x)^{2+\delta}$ then $\psi(x+h) - \psi(x) \sim h$.

The classical prime number theorem with error term $x\exp(-C\sqrt{\log x})$ tells us that such a result holds if $h \geq x\exp(-C\sqrt{\log x})$. An important advance was made by Hoheisel who showed that the asymptotic $\psi(x+h) - \psi(x) \sim h$ holds if $x \geq h \geq x^\theta$ for some number $\theta < 1$. He was able to take $\theta = 1 - 1/33000$, but this has been improved subsequently, with the best result known, due to Huxley, being $\theta = \frac{7}{12} + \epsilon$. If the Riemann hypothesis is true then θ may be taken as $\frac{1}{2} + \epsilon$. The arguments pioneered by Hoheisel depend on the fact that while we don't know RH, we do know that most zeros of $\zeta(s)$ lie close to the $\frac{1}{2}$ line. For a nice account of these results see Heath-Brown [18]. If the asymptotics given for (2.6) are true then we may take θ to be any positive number.

Thus it seems the conjecture that $\psi(x+h) - \psi(x) \sim h$ for $x \geq h \geq (\log x)^{2+\delta}$, if true, lies quite deep. This conjecture was widely believed until the mid 1980's when Maier [22] shattered this belief by showing that for any $A > 1$ there are arbitrarily large x such that the interval $(x, x + (\log x)^A]$ contains significantly more primes than usual (that is, $\geq (1 + \delta_A)(\log x)^{A-1}$ primes for some $\delta_A > 0$) and also intervals $(x, x + (\log x)^A]$ containing significantly fewer primes than usual. In this lecture we will sketch Maier's ingenious method, and describe some extensions of his idea. The reader may also consult [13] for another exposition of related ideas.

Maier's "matrix" method. Let x be large, and h be on the scale of a power of $\log x$. Let P be an integer which we will eventually take to be the product of many small (below $\log x$) primes. Consider the $[x/P]$-by-h "matrix" whose (i,j)-th entry is the number $([x/P] + i)P + j$. Thus the entries of this matrix are numbers lying between x and $2x + h$. Note that each row of the matrix consists of an interval $([x/P] + i)P$ to $([x/P] + i)P + h$. Each column of the matrix consists of an arithmetic progression with common difference P: namely, $x \leq n \leq 2x + h$ with $n \equiv j \pmod{P}$. The idea is to count the number of primes in this matrix in two ways: by counting the primes row by row, and by counting the primes column by column, and then comparing the two answers. If we assume that the asymptotic formula for primes in short intervals holds then we get an answer for the row by row calculation. The prime number theorem for arithmetic progressions allows us to do the column by column calculation. Of course the two answers should match. However when h is very small, like a power of $\log x$, there are choices of P for which the answers don't match! This leads to Maier's result.

Consider the row by row calculation. The number of primes is

$$(3.1a) \qquad \sum_{x/P \leq n \leq 2x/P} (\pi(nP + h) - \pi(nP)),$$

and if we assume that intervals of length h contain the right number of primes, this is

$$(3.1b) \qquad \sim \frac{x}{P} \frac{h}{\log x}.$$

Consider next the column by column calculation. If the progression $n \equiv j \pmod{P}$ is to contain primes, we must have $(j, P) = 1$. In that case the prime number theorem in arithmetic progressions would say that such a progression contains a proportion $1/\phi(P)$ of all primes. Of course, in order to use the prime number theorem in arithmetic progressions rigorously we must pay attention to the size of the modulus P compared with x. Assuming that this is not an issue, we find that the column by column contribution is

$$(3.2) \qquad \sum_{\substack{j \leq h \\ (j,P)=1}} (\pi(2x + h; P, j) - \pi(x; P, j)) \sim \frac{x}{\phi(P) \log x} \sum_{\substack{j \leq h \\ (j,P)=1}} 1.$$

If we compare (3.2) and (3.1b) we find the relation

$$(3.3) \qquad \sum_{\substack{j \leq h \\ (j,P)=1}} 1 \sim h \frac{\phi(P)}{P}$$

should hold. At first glance, (3.3) is eminently reasonable: the probability that j is coprime to P is $\phi(P)/P$. It is even easy to make this precise: write the condition $(j, P) = 1$ as $\sum_{\ell|(j,P)} \mu(\ell)$ and we easily get

$$(3.4) \qquad \sum_{\substack{j \leq h \\ (j,P)=1}} 1 = h \frac{\phi(P)}{P} + O(d(P)),$$

where $d(P)$ is the number of divisors of P. Thus, if h is just a bit larger than $d(P)$ (which is always quite small, that is $\ll P^\epsilon$) then (3.3) will hold. So where is the contradiction? The point is that in Maier's application h is very small compared with P, and so (3.4) is useless.

For the purpose of illustration suppose that P is the product of all primes between $(\log x)^{\frac{9}{10}}$ and $(\log x)/100$. Then, by the prime number theorem, P is about size $x^{\frac{1}{100} + o(1)}$. For such moduli P we don't know the prime number theorem in arithmetic progressions used in (3.2), but such a result does hold if the Riemann hypothesis for Dirichlet L-functions is true; let us postpone a discussion of this point. Suppose now that h is a number of size $(\log x)^\theta$ with $2 < \theta < 2.7$. By inclusion-exclusion, the LHS of (3.3) is

$$\sim h - \sum_{(\log x)^{0.9} \leq p \leq (\log x)/100} \frac{h}{p} + \sum_{(\log x)^{0.9} \leq p < q \leq (\log x)/100} \frac{h}{pq}$$

$$\sim h \left(1 - \log \frac{10}{9} + \frac{1}{2} \left(\log \frac{10}{9} \right)^2 \right),$$

where we have used the prime number theorem to evaluate $\sum 1/p$ for p between $(\log x)^{\frac{9}{10}}$ and $(\log x)/100$. On the other hand, by Mertens' theorem, the RHS of (3.3) is

$$\sim h \prod_{(\log x)^{\frac{9}{10}} \leq p \leq (\log x)/100} \left(1 - \frac{1}{p}\right)^{-1} \sim \frac{9}{10} h.$$

The formula for the LHS has the first three terms in the usual expansion of $\frac{9}{10} = e^{-\log(10/9)}$, so the two answers are certainly close, but obviously they are not equal! Indeed the LHS is a little bit larger.

Exercise 11. *Conclude from the above that for any $2 < \theta < 2.7$ there exist arbitrarily large x such that the interval $[x, x + (\log x)^\theta]$ contains significantly more primes than expected. Taking such an interval and cutting it up into smaller intervals, deduce that the same conclusion holds for all $1 < \theta < 2.7$. Using the same P as above, and taking four terms in the inclusion-exclusion formula, show that if $\theta < 3.6$ there exist intervals $[x, x + (\log x)^\theta]$ with significantly fewer primes than expected. In this manner one can proceed for $\theta < 8.1$, just using inclusion-exclusion and easy calculations. Now replace $(\log x)^{0.9}$ in the definition of P with $(\log x)^{1-\delta}$ and prove Maier's theorem.*

More on the contradiction. Now let us describe a different way of seeing a contradiction to (3.3). This method is very flexible, and works for many choices of P, and also generalizes readily. Let y be a large parameter; in the application we may think of y as being some power of $\log x$. From each dyadic block $[2^{-j}y, 2^{-j+1}y]$ with $j \leq [\log y/(2\log 2)]$ select about half the primes. Take P to be the product of these selected primes. Thus P is composed of about half the primes in $[\sqrt{y}, y]$, and there are plenty of choices for P. Let $u \geq 1$ be a real number, set $h = y^u$ and consider whether (3.3) can hold. We will show that for arbitrarily large u the LHS is appreciably larger than the RHS, and for arbitrarily large u it is smaller.

To see this we consider the Dirichlet series $\zeta_P(s) = \sum_{(n,P)=1} n^{-s}$. Plainly we have

$$(3.5) \qquad\qquad \zeta_P(s) = \zeta(s) \prod_{p|P} \left(1 - \frac{1}{p^s}\right),$$

so that $\zeta_P(s)$ extends to a meromorphic function in all of $\mathbb{C}$ with a simple pole at $s = 1$. The point is that if something like (3.3) holds then $\zeta_P(s)$ must approximately look like $\zeta(s)\phi(P)/P$, and by choosing s appropriately we can obtain a contradiction to (3.5). More precisely, set

$$E(u) = \frac{1}{y^u} \Big(\sum_{\substack{n \leq y^u \\ (n,P)=1}} 1 - [y^u] \frac{\phi(P)}{P} \Big).$$

Then, for $\mathrm{Re}(s) > 1$,

$$\zeta_P(s) - \zeta(s) \frac{\phi(P)}{P} = \int_{1^-}^\infty \frac{1}{z^s} d\Big(\sum_{\substack{n \leq z \\ (n,P)=1}} 1 - \sum_{n \leq z} \frac{\phi(P)}{P} \Big) = \int_1^\infty \frac{s}{z^s} E\Big(\frac{\log z}{\log y}\Big) dz,$$

upon integrating by parts. Changing variables $u = \log z / \log y$ we obtain that

$$(3.6) \qquad \zeta_P(s) = \zeta(s) \frac{\phi(P)}{P} + s \log y \int_0^\infty E(u) y^{-u(s-1)} du.$$

To start with, (3.6) is valid for $\mathrm{Re}(s) > 1$, but since $E(u) \ll d(P) y^{-u}$ by (3.4), we see that (3.6) makes sense for $\mathrm{Re}\, s > 0$.

Exercise 12. *Let $(\log y)/2 \geq \xi \geq 1$ be a real number, and take $s = 1 - \xi / \log y + i\pi / \log y$. Using (3.5) prove that*

$$|\zeta_P(s)| \gg \frac{\log y}{\xi} \exp\left(\frac{e^\xi}{2\xi} + O\left(\frac{e^\xi}{\xi^2}\right) \right).$$

Then using (3.6) deduce that

$$\int_0^\infty |E(u)| e^{\xi u} du \gg \frac{1}{\xi} \exp\left(\frac{e^\xi}{2\xi} + O\left(\frac{e^\xi}{\xi^2}\right) \right).$$

Show that $\int_0^\infty E(u) e^{\xi u} du \ll 1/\xi$, so that in the LHS above both positive and negative values of $E(u)$ make roughly equal contributions.

Morally, Exercise 12 shows that $E(u)$ cannot be too small for large u. To make this precise, one also needs an upper bound for $E(u)$ so as to be able to bound the tail of the integral in Exercise 12. Developing this argument carefully, one may show that there is a positive constant A such that every interval $[u(1 - A/\log u), u(1 + A/\log u)]$ contains points $u_\pm$ satisfying

$$E(u_+) \geq \exp(-u_+(\log u_+ + \log \log u_+ + O(1))),$$

and

$$E(u_-) \leq -\exp(-u_-(\log u_- + \log \log u_- + O(1))).$$

For more details, see section 3 of [16], especially Corollary 3.3.

Earlier, we postponed discussion of the prime number theorem in arithmetic progressions. We refer the reader to Davenport [6] for an account of this. In Chapter 20 there one finds Page's result that $\psi(x; q, a) \sim x/\phi(q)$ for all $q \leq \exp(C\sqrt{\log x})$ with the possible exception of multiples of a particular modulus q_1 which may depend on x. If we choose y a little less than $\sqrt{\log x}$ then our moduli P above are below $\exp(C\sqrt{\log x})$ and certainly we can find P that are not multiples of the exceptional modulus q_1. Thus our appeal to the prime number theorem in arithmetic progressions can be made rigorous.

The flexibility in choosing P is quite useful. Exploiting this, Granville and I (see [16]) showed that the asymptotic

$$(3.7) \qquad \psi(x + h) - \psi(x) = h + O(h^{\frac{1}{2} + \epsilon}),$$

suggested by Cramér's model, sometimes fails to hold if $h \leq \exp((\log x)^{\frac{1}{2} - \epsilon})$. This improves work of Hildebrand and Maier [19] who had obtained this result assuming the Generalized

Riemann Hypothesis, and a weaker result unconditionally. It seems safe to conjecture that (3.7) holds if $h \geq x^{\delta}$, and perhaps it holds when $h \geq \exp((\log x)^{\frac{1}{2}+\delta})$.

An uncertainty principle. Maier's method can be adapted to establish limitations to the equidistribution of primes in arithmetic progressions. For example, Friedlander and Granville [8] proved that for every $A \geq 1$ there exist large x and an arithmetic progression $a \pmod q$ with $(a, q) = 1$ and $q \leq x/(\log x)^A$ such that

$$\left| \pi(x; q, a) - \frac{\pi(x)}{\phi(q)} \right| \gg_A \frac{\pi(x)}{\phi(q)}.$$

More recently, Balog and Wooley [1] showed that the sequence of integers which may be written as the sum of two squares also exhibits "Maier type" irregularities in intervals $(x, x + (\log x)^A)$ for any fixed, positive A. Previously Maier's work had seemed inextricably linked to the mysteries of primes, but Balog and Wooley's result suggests that such results should be part of a more general phenomenon. This has been formalized by Granville and me as an "uncertainty principle" for arithmetic sequences. What Maier's argument shows is that the primes cannot be simultaneously well distributed in short intervals, and in arithmetic progressions. Then a suitable version of the prime number theorem in arithmetic progressions is used to remove the second possibility, leaving us with the irregularities of distribution in short intervals. The first conclusion of irregularities in short intervals or progressions turns out to be a general feature of many interesting arithmetical sequences.

A rough description of this result is as follows: Let $\mathcal{A}$ denote a sequence $a(n)$ of non-negative real numbers, and let $\mathcal{A}(x) = \sum_{n \leq x} a(n)$. If $\mathcal{A}$ is well-distributed in short intervals, then we may expect that

$$(3.8) \qquad\qquad \mathcal{A}(x + y) - \mathcal{A}(x) \approx y \frac{\mathcal{A}(x)}{x}.$$

To understand the distribution of $a(n)$ in arithmetic progressions we begin with n that are multiples of a given number d. We suppose that there is a non-negative multiplicative function h such that

$$(3.9a) \qquad\qquad \mathcal{A}_d(x) = \sum_{\substack{n \leq x \\ d|n}} a(n) \approx \frac{h(d)}{d} \mathcal{A}(x).$$

We assume that the asymptotic behavior of

$$\mathcal{A}(x; q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} a(n)$$

depends only on the g.c.d. of a and q. Then (3.9a) leads to the prediction that

$$(3.9b) \qquad\qquad \mathcal{A}(x; q, a) \approx \frac{f_q(a)}{q \gamma_q} \mathcal{A}(x),$$

with $\gamma_q = \prod_{p|q}(p-1)/(p-h(p))$ and $f_q(a)$ is a certain non-negative multiplicative function of a, defined in terms of h, such that $f_q(a) = f_q((a,q))$ so that $f_q(a)$ is periodic (mod q). We can be flexible in how we want to assume (3.9b); for example, sometimes it is convenient to assume it only for q that are coprime to a certain fixed set of primes.

To illustrate the framework consider the following examples.

Example 1. Take $a(n) = 1$ for all n. It is natural to take $h(d) = 1$ for all d, $\gamma_q = 1$, and $f_q(a) = 1$. Then (3.9a) and (3.9b) are both good approximations with errors at most 1.

Example 2. Take $a(n)$ to be the indicator function of the primes. Then $h(1) = 1$ and $h(d) = 0$ for $d > 1$. One has $\gamma_q = \phi(q)/q$ and $f_q(a) = 1$ if $(a,q) = 1$ and 0 otherwise. The prime number theorem in arithmetic progressions gives (3.9b) for small values of q. The result of Friedlander and Granville places restrictions on the approximation (3.9b) when q is large. Maier's results place restrictions on (3.9a) for small y.

Example 3. Take $a(n)$ to be the indicator function of the sums of two squares. The multiplicative function h is defined by $h(p^k) = 1$ if $p^k \equiv 1 \pmod 4$ and $h(p^k) = 1/p$ if $p^k \equiv 3 \pmod 4$. Here Balog and Wooley's result places restrictions on (3.9a).

The main results of [16] give that if $h(p)$ is not always close to 1 (as in the regular example 1) then there will be moduli q for which (3.9b) cannot hold. Typically these moduli will be large as in the Friedlander-Granville result for primes in progressions. Furthermore, either there exist values y larger than an arbitrary power of $\log x$ for which (3.9a) is false, or there exist small moduli q (below $\exp((\log x)^\delta)$) for which (3.9b) is false. These results include the previous results on primes and sums of two squares, and also cover many other examples.

Consider sets containing roughly half of the prime numbers. There are uncountably many such sets, and so maybe we can find a set which is very well distributed in arithmetic progressions. One amusing example from [16] shows that this cannot be done, and the Friedlander-Granville limitations persist for any such set.

We content ourselves with this vague description of the uncertainty principle, referring the reader to [16] for more examples and a precise description of the results.

References

[1] A. Balog and T.D. Wooley, *Sums of two squares in short intervals*, Canad. J. Math. **52** (2000), 673–694.

[2] E.B. Bogomolny and J.P. Keating, *Random matrix theory and the Riemann zeros. II. n-point correlations*, Nonlinearity **9** (1996), 911–935.

[3] T.H. Chan, *Pair correlation and distribution of prime numbers*, Ph. D. Thesis, University of Michigan (2002), 101pp.

[4] T.H. Chan, *A note on primes in short intervals*, Int. J. Number Theory **2** (2006), 105–110.

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

[6] H. Davenport, *Multiplicative number theory*, Springer Graduate Texts in Math. 74, 2000.

[7] W. Feller, *An introduction to probability theory and its applications*, Wiley, 1966.

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

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

[10] D. Goldston, *Notes on pair correlation of zeros and prime numbers*, Recent perspectives in random matrix theory and number theory, London Math. Soc. Lecture Notes Ser. 322, Cambridge U. Press, 2005, pp. 79–110.

[11] D. Goldston and H.L. Montgomery, *On pair correlations of zeros and primes in short intervals*, Analytic Number Theory and Diophantine Problems, vol. 70, Prog. in Math. Birkhäuser, 1987, pp. 183–203.

[12] D. Goldston, J. Pintz and C. Yıldırım, *Primes in tuples, I*, Ann. of Math. (to appear), preprint available at www.arxiv.org.

[13] A. Granville, *Unexpected irregularities in the distribution of prime numbers*, Proc. of the Int. Congr. of Math., Vol. 1, 2 (Zürich, 1994) (1995), Birkhäuser, Basel, 388-399.

[14] A. Granville and G. Martin, *Prime number races*, Amer. Math. Monthly **113** (2006), 1–33.

[15] A. Granville and K. Soundararajan, *Sieving and the Erdős-Kac theorem*, these proceedings.

[16] A. Granville and K. Soundararajan, *An uncertainty principle for arithmetic sequences*, Ann. of Math. (to appear), preprint available at www.arxiv.org.

[17] G.H. Hardy and J.E. Littlewood, *Some problems of Paritio Numerorum (III): On the expression of a number as a sum of primes*, Acta Math. **44** (1922), 1–70.

[18] D.R. Heath-Brown, *Differences between consecutive primes*, Jahresber. Deutsch. Math.-Verein. **90** (1998), 71–89.

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

[20] C. Hooley, *On the difference between consecutive numbers prime to n: II*, Publ. Math. Debrecen **12** (1965), 39–49.

[21] C.P. Hughes and Z. Rudnick, *On the distribution of lattice points in thin annuli*, Int. Math. Res. Not. **13** (2004), 637–658.

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

[23] W.R. Monach, *Numerical investigation of several problems in number theory*, Ph. D. Thesis, University of Michigan (1980), 171 pp.

[24] H.L. Montgomery, *The pair corelation of zeros of the zeta function*, Analytic Number Theory (St. Louis Univ. 1972), vol. 24, Proc. Sympos. Pure Math. (Amer. Math. Soc.), 1973, pp. 181-193.

[25] H.L. Montgomery and K. Soundararajan, *Beyond pair correlation*, Paul Erdős and his mathematics, I, Bolyai Soc. Math. Stud., 11, Budapest, 2002, pp. 507-514.

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

[27] H.L. Montgomery and R.C. Vaughan, *On the distribution of reduced residues*, Annals of Math. **123** (1986), 311–333.

[28] E. M. Rains, *High powers of random elements of compact Lie groups*, Probab. Theory Related Fields **107** (1997), 219–241.

[29] M. Rubinstein and P. Sarnak, *Chebyshev's Bias*, Experimental Math. **3** (1994), 173–197.

[30] A. Selberg, *On the normal density of primes in short intervals, and the difference between consecutive primes*, Collected papers (Volume I), Springer, 1989, pp. 160–178.

[31] K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım*, Bull. Amer. Math. Soc. (to appear), preprint available at www.arxiv.org.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
E-mail address: ksound@umich.edu