Lecture 2: Small Prime Gaps: Short Divisor Sums, Correlations, and Moments

Daniel Goldston

A few comments on the last lecture:

1. We replaced the usual major arcs approximation for $S(\alpha)$ with

$$J_Q(\alpha) = \sum_{q \le Q} \sum_{\substack{1 \le a \le q \\ (a,q)=1}} \frac{\mu(q)}{\phi(q)} I(\alpha - \frac{a}{q}) \chi_Q(\alpha, \frac{a}{q})$$

with

$$\begin{aligned} V_Q(\alpha) &= \sum_{q \le Q} \sum_{\substack{1 \le a \le q \\ (a,q)=1}} \frac{\mu(q)}{\phi(q)} I(\alpha - \frac{a}{q}) \\ &= \sum_{n \le N} \left(\sum_{q \le Q} \frac{\mu(q)}{\phi(q)} c_q(-n) \right) e(n\alpha) \\ &= \sum_{n \le N} \lambda_Q(n) e(n\alpha) \end{aligned}$$

and since

$$S(\alpha) = \sum_{n \le N} \Lambda(n) e(n\alpha).$$

this suggests that the content of the circle method is to approximate $\Lambda(n)$ by $\lambda_Q(n)$. In turn, an approximation of $\lambda_Q(n)$ is

$$\Lambda_Q(n) = \sum_{\substack{d|n \\ d \le Q}} \mu(d) \log(Q/d),$$

but from elementary number theory

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(1/d).$$

Thus the content of the circle method for primes is reduced to a short smoothed truncation of this elementary formula.

But is this reasonable? For approximating the twin prime problem both J_Q and V_Q give the same correct conjectured singular series formula if $Q \leq N^{1/2}$. However, as we saw,

$$\int_0^1 |S(\alpha)|^2 \, d\alpha = \sum_{n \leq N} \Lambda(n)^2 \sim N \log N$$

but

$$\int_0^1 |J_Q(\alpha)|^2 \, d\alpha \sim N \log \Big(\min(Q, \frac{N}{Q}) \Big).$$

Now Sid Graham (1976) proved, for $1 \leq Q \leq N$

$$\sum_{n \leq N} \Lambda_Q(n)^2 = N \log Q + O(N)$$

so our approximation really does approximate $\Lambda(n)$ if Q is large enough. But since $\Lambda_Q(n) = \Lambda(n)$ if $Q \geq n$, this isn't a total surprise. What about $\lambda_Q(n)$? Here it is easy to prove

$$\sum_{n \leq N} \lambda_Q(n)^2 = N \log Q + O(N) + O(Q^2)$$

which is only good for $Q \le N^{1/2}$. Hooley (1999 ?) proved, $Q \le N^{1-\epsilon}$

$$\sum_{n \le N} \lambda_Q(n)^2 \sim N \log Q$$

This says that for $Q \leq N^{1-\epsilon}$

$$\int_0^1 |V_Q(\alpha)|^2 \, d\alpha \sim N \log Q$$

The diagonal terms here contribute

$$\int_0^1 \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \frac{\mu(q)^2}{\phi(q)^2} |I(\alpha - \frac{a}{q})|^2 \, d\alpha$$

$$= \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \frac{\mu(q)^2}{\phi(q)^2} \int_0^1 |I(\alpha - \frac{a}{q})|^2 \, d\alpha$$

$$= N \sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \frac{\mu(q)^2}{\phi(q)^2}$$

$$= N \sum_{q \leq Q} \frac{\mu(q)^2}{\phi(q)} \sim N \log Q.$$

Thus this proves the overlapping spikes end up not contributing anything extra here.

Given the idea of seeking a truncated divisor sum approximation of $\Lambda(n)$,

Consider

$$\text{Minimize} \sum_{n \le N} \left(\Lambda(n) - \sum_{\substack{d|n \\ d \le Q}} a(d,Q) \right)^2$$

It is easy to see that this problem is equivalent to the one Selberg solved in his 5 page 1950 sieve paper, the solution is that (with mild conditions on $a(d,Q)$) the minimum is attained with the choice $\lambda_Q(n)$, and this and $\Lambda_Q(n)$ give the minimum

$$\sim N \log(N/Q)$$

1. DIVISOR SUMS

We will use the von Mangoldt function

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^m,\ p \text{ prime},\ m \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

The prime number theorem (PNT) implies

$$\psi(x) \sim x, \qquad \text{as } x \to \infty.$$

Now

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(1/d),$$

which is easily proved directly, or seen analytically by noting

$$\frac{\zeta'}{\zeta} = \frac{1}{\zeta} \times \zeta'.$$

Is this a worthless formula or not?

Let's try to prove the PNT:

$$\psi(N) = \sum_{n \le N} \Lambda(n) = \sum_{n \le N} \sum_{d|n} \mu(d) \log(1/d)$$
$$= \sum_{d \le N} \mu(d) \log(1/d) \sum_{\substack{n \le N \\ d|n}} 1.$$

Since

$$\sum_{\substack{n \le N \\ d|n}} 1 = \left[\frac{N}{d}\right],$$

we have

$$\psi(N) = \sum_{d \le N} \mu(d) \log(1/d) \left[\frac{N}{d}\right]$$
$$= N \sum_{d \le N} \frac{\mu(d) \log(1/d)}{d} + O(\sum_{d \le N} \log d)$$
$$= N \sum_{d \le N} \frac{\mu(d) \log(1/d)}{d} + O(N \log N).$$

We have two problems:

1) The error term is bigger than the main term.

2) The main term still needs to be evaluated. In fact, PNT is equivalent to an asymptotic formula for main term and implies

$$\sum_{d \le N} \frac{\mu(d) \log(1/d)}{d} = 1 + O(\frac{1}{\log^A N}).$$

But on the positive side, it was an easy argument.

2. SHORT SMOOTHED DIVISOR SUMS

We can handle the error term above by truncating the divisor sum. Following Selberg, we also smooth it. Smoothing is critical, and most of the results that we obtain would be false for a straight truncation. Thus, let

$$\Lambda_R(n) = \sum_{\substack{d|n \\ d\le R}} \mu(d) \log(R/d).$$

Since

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(R/d), \qquad \text{for } n \ge 2,$$

we see that $\Lambda_R(n) = \Lambda(n)$ for $R \ge n \ge 2$.

New viewpoint: $\Lambda_R(n)$ should retain to some extent the properties we know or conjecture that $\Lambda(n)$ has.

PNT for $\psi_R(N)$. We have as before

$$\begin{aligned}\psi_R(N) :&= \sum_{n\le N} \Lambda_R(n)\\ &= N\sum_{d\le R}\frac{\mu(d)\log(R/d)}{d} + O(\sum_{d\le R}\log R/d)\\ &= N + O(\frac{N}{\log^A R}) + O(R),\end{aligned}$$

Hence we obtain a PNT for ψ_R if $N^\epsilon \le R \le o(N)$.

Twin Prime Theorem for $\psi_R(N)$. Consider

$$\mathcal{S}_R(k) = \sum_{n\le N}\Lambda_R(n)\Lambda_R(n+k).$$

Then

$$\mathcal{S}_2(k) = \sum_{d,e\le R}\mu(d)\mu(e)\log(R/d)\log(R/e)\sum_{\substack{n\le N\\ d|n\\ e|n+k}} 1.$$

The two divisibility conditions imply that n will run through a residue class modulo $[d,e]$ provided $(d,e)|k$, and there will be no solution for

n otherwise. Therefore

$$\sum_{\substack{n \le N \\ d|n \\ e|n+k}} 1 = [(d,e)|k]\Big(\frac{N}{[d,e]} + O(1)\Big),$$

where we use the Iverson notation

$$[P(x)] = \begin{cases} 1, & \text{if } P(x) \text{ is true,} \\ 0, & \text{if } P(x) \text{ is false.} \end{cases}$$

Thus

$$\mathcal{S}_R(k) = N \sum_{\substack{d,e \le R \\ (d,e)|k}} \frac{\mu(d)\mu(e)\log(R/d)\log(R/e)}{[d,e]}$$
$$+ O\Big(\sum_{d,e \le R} \log(R/d)\log(R/e)\Big)$$

Using standard PNT type arguments we will discuss later in the talk we obtain, if $k \neq 0$,

$$\sum_{n \leq N} \Lambda_R(n)\Lambda_R(n+k) = N\mathfrak{S}(k) + O(\frac{N}{(\log R)^A}) + O(R^2),$$

where

$$\mathfrak{S}(n) = \begin{cases} 2C_2 \prod_{\substack{p|n \\ p>2}} \left(\frac{p-1}{p-2}\right), & \text{if } n \text{ is even, } n \neq 0; \\ 0, & \text{if } n \text{ is odd;} \end{cases}$$

and

$$C_2 = \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right),$$

while in the case $k = 0$ we have

$$\sum_{n \leq N} {\Lambda_R}^2(n) = N \log R + O(N) + O(R^2).$$

Thus we get a twin prime type result for $\Lambda_R(n)$ if $R \leq o(N^{1/2})$, but fail to get the binary form of the prime number theorem, since

$$\sum_{n \leq N} \Lambda^2(n) = N \log N + O(N).$$

3. MIXED CORRELATIONS

We can obtain information about primes through the previous correlations of $\Lambda_R(n)$ and with the mixed correlation with $\Lambda(n)$

$$\tilde{\mathcal{S}}_R(k) = \sum_{n\le N} \Lambda(n)\Lambda_R(n+k) \sim \log N \sum_{p\le N} \Lambda_R(p+k).$$

We have, for $k \neq 0$

$$\begin{aligned}\tilde{\mathcal{S}}_R(k) &= \sum_{d\le R} \mu(d)\log(R/d)\Big(\sum_{\substack{n\le N\\ d|n+k}} \Lambda(n)\Big)\\ &= \sum_{d\le R} \mu(d)\log(R/d)\Big(\sum_{\substack{n\le N\\ n\equiv -k(d)}} \Lambda(n)\Big)\end{aligned}$$

We let

$$\psi(x;q,a) = \sum_{\substack{n\le x\\ n\equiv a(q)}} \Lambda(n),$$

and

$$E(x;q,a) = \psi(x;q,a) - [(a,q)=1]\frac{x}{\phi(q)}.$$

We need the estimate, for a fixed $0 < \vartheta \le 1$,

$$\sum_{1\le q\le x^{\vartheta-\epsilon}} \max_{\substack{a \\ (a,q)=1}} |E(x;q,a)| \ll \frac{x}{\log^{\mathcal{A}} x},$$

for any $\epsilon > 0$, any $\mathcal{A} = \mathcal{A}(\epsilon) > 0$, and x sufficiently large. This is a weakened form of the Bombieri-Vinogradov theorem if $\vartheta = \frac{1}{2}$. Elliott and Halberstam conjectured that $\vartheta = 1$ holds. Hence

$$\begin{aligned}\tilde{\mathcal{S}}_R(k) = N &\sum_{\substack{d\le R \\ (d,k)=1}} \frac{\mu(d)}{\phi(d)} \log(R/d) \\ &+ O\Big(\sum_{d\le R} \log(R/d)|E(N;d,-k)|\Big) \\ &= N\mathfrak{S}(k) + O\Big(\frac{N}{(\log R)^A}\Big)\end{aligned}$$

provided $R \le N^{\theta}$, where $0 < \theta < \vartheta$. We also have immediately by PNT

$$\sum_{n\le N} \Lambda_R(n)\Lambda(n) = N\log R + O(N).$$

Thus the mixed correlations are the same asymptotically as the mixed correlations in the range $R = N^{\theta}$, $0 < \theta < 1/2$.

Detecting Prime Gaps Recall

$$Z(N;k) := \sum_{\substack{n \\ 1\le n, n+k\le N}} \Lambda(n)\Lambda(n+k)$$
$$= \int_0^1 |S(\alpha)|^2 e(-k\alpha)\, d\alpha.$$

The minor arcs here are hopeless, but Hardy-Littlewood had the following trick up their sleeve:

$$\int_0^1 |S(\alpha)|^2 |P(\alpha)|^2\, d\alpha \ge \int_{\mathcal{M}\subset[0,1]} |S(\alpha)|^2 |P(\alpha)|^2\, d\alpha$$
$$= \int_{\mathcal{M}} |J_Q(\alpha) + R_Q(\alpha)|^2 |P(\alpha)|^2\, d\alpha$$
$$= \int_{\mathcal{M}} |J_Q(\alpha)|^2 |P(\alpha)|^2\, d\alpha$$
$$+ 2\text{Re} \int_{\mathcal{M}} \overline{J_Q}(\alpha) R_Q(\alpha) |P(\alpha)|^2\, d\alpha$$
$$+ \int_{\mathcal{M}} |R_Q(\alpha)|^2 |P(\alpha)|^2\, d\alpha$$
$$\ge \int_{\mathcal{M}} |J_Q(\alpha)|^2 |P(\alpha)|^2\, d\alpha$$
$$+ 2\text{Re} \int_{\mathcal{M}} \overline{J_Q(\alpha)} R_Q(\alpha) |P(\alpha)|^2\, d\alpha$$

The first term gives the main term. In the second, one uses GRH or in Bombieri-Davenport's paper the Bombieri-Vinogradov Theorem to show this is an error.

Here

$$P(\alpha) = \sum_{k=-H}^{H} t(k)e(k\alpha),$$

but the optimal choice turns out to be

$$P(\alpha) = \sum_{k=1}^{H} e(k\alpha),$$

so

$$|P(\alpha)|^2 = \sum_{k=-H}^{H} (H - |k|)e(k\alpha).$$

Thus we get the lower bound: $Q = N^{1/2}$,

$$\sum_{k=-H}^{H} (H - |k|) Z(N,k) > KN \log Q + N \sum_{k=-H}^{H} (H - |k|) \mathfrak{S}(k)$$

so using a singular series average (ignore ϵ's)

$$HN \log N + 2 \sum_{k=1}^{H} (H - |k|) Z(N,k) > \frac{1}{2} HN \log N + NH^2$$

so we get

$$\begin{aligned} 2 \sum_{k=1}^{H} (H - |k|) Z(N,k) &> (H^2 - \frac{1}{2} H \log N) \\ &> H(H - \frac{1}{2} \log N) \\ &> 0 \end{aligned}$$

provided $H > \frac{1}{2} \log N$.

While $\mathcal{M}$ is made use of by Bombieri-Davenport, one can use the full Farey decomposition with no minor arcs in this argument. Thus throwing away the $|R|^2$term is a true loss due to the approximation.

Bombieri-Davenport Proof using $\Lambda_R(n)$ and Moments

We consider the second moment

$$M_2(N,\psi) = \sum_{n=1}^{N} \Big(\psi(n+h) - \psi(n)\Big)^2$$

and the corresponding moments

$$M_2(N,\psi_R) = \sum_{n=1}^{N} \Big(\psi_R(n+h) - \psi_R(n)\Big)^2$$

and

$$\tilde{M}_2(N,\psi_R) = \sum_{n=1}^{N} \Big(\psi(n+h)-\psi(n)\Big)\Big(\psi_R(n+h)-\psi(n)\Big).$$

All three moments can be resolved into correlations the result being, for $h \ll N^{1/2}$,

$$M_2(h,F) \sim \sum_{|k|\le h} (h-|k|)\bigg(\sum_{n\le N} f(n)f(n+k)\bigg).$$

Hence, letting $h = \lambda \log N$, $R = N^\theta$ we have for $0 < \theta < 1/2$,

$$\begin{aligned} M_2(N,\psi_R) &\sim Nh\log R + 2N \sum_{1\le k\le h} (h-|k|)\mathfrak{S}(k) \\ &\sim Nh\log R + Nh^2 \\ &\sim \Big(\theta\lambda + \lambda^2\Big) N\log^2 N, \end{aligned}$$

and similarly for $\tilde{M}_2(N,\psi_R)$.

Bombieri and Davenport's is equivalent to the inequality

$$\sum_{n=1}^{N} \bigg(\Big(\psi(n+h)-\psi(n)\Big) - \Big(\psi_R(n+h) - \psi_R(n)\Big)\bigg)^2 \ge 0.$$

Expanding gives the lower bound

$$M_2(N,h,\psi) \ge 2\tilde{M}_2(N,h,\psi_R) - M_2(N,h,\psi_R)$$

which implies on taking $\theta = 1/2 - \epsilon$

$$M_2(N,h,\psi) \ge (\frac{1}{2}\lambda + \lambda^2 - \epsilon) N\log^2 N.$$

If there are never two primes as close as $\lambda \log N$,

$$M_2(N,h,\psi) \sim (\log N) M_1(N,h,\psi) \sim \lambda N \log^2 N$$

so that

$$\lambda \geq \frac{1}{2}\lambda + \lambda^2 - \epsilon$$

so

$$0 \geq \lambda(\lambda - \frac{1}{2}) - \epsilon$$

which is false if $\lambda > \frac{1}{2}$. We conclude that

$$\liminf_{n\to\infty} \left(\frac{p_{n+1} - p_n}{\log p_n}\right) \leq \frac{1}{2}.$$

This argument has an additional advantage over Bombieri-Davenport's argument: assuming the Elliott-Halberstam for $\Lambda(n)$ AND $\Lambda_R(n)$, you do prove small gaps.

Gallagher (1976) Hardy-Littlewood Prime Tuple conjecture implies Poisson distribution

(OBTAINING MOMENTS FROM CORRELATIONS)

The Hardy-Littlewood prime-tuple conjecture is that for $\boldsymbol{j} = (j_1, j_2, \ldots, j_r)$ with the j_i's distinct integers,

$$\begin{aligned}\psi_{\boldsymbol{j}}(N) = & \sum_{n=1}^{N} \Lambda(n+j_1)\Lambda(n+j_2)\cdots\Lambda(n+j_r) \\ & \sim \mathfrak{S}(\boldsymbol{j})N\end{aligned}$$

when $\mathfrak{S}(\boldsymbol{j}) \neq 0$, where

$$\mathfrak{S}(\boldsymbol{j}) = \prod_p \left(1 - \frac{1}{p}\right)^{-r} \left(1 - \frac{\nu_p(\boldsymbol{j})}{p}\right)$$

and $\nu_p(\boldsymbol{j})$ is the number of distinct residue classes modulo p that the j_i's occupy.

Consider the k-th moment

$$M_k(N, h, \psi) = \sum_{n=1}^{N} (\psi(n+h) - \psi(n))^k$$

We have

$$M_k(N,h,\psi) = \sum_{n=1}^{N} \left(\sum_{1\le m\le h} \Lambda(n+m) \right)^k$$

$$= \sum_{\substack{1\le m_i\le h\\ 1\le i\le k}} \sum_{n=1}^{N} \Lambda(n+m_1)\Lambda(n+m_2)\cdots\Lambda(n+m_k).$$

Now suppose that the k numbers $m_1, m_2, \ldots, m_k$ take on r distinct values $j_1, j_2, \ldots, j_r$ with j_i having multiplicity a_i, so that $\sum_{1\le i\le r} a_i = k$. Grouping the terms above, we have that

$$M_k(N,h,\psi) = \sum_{r=1}^{k} \sum_{\substack{a_1,a_2,\ldots,a_r\\ a_i\ge 1, \sum a_i = k}} \binom{k}{a_1, a_2, \ldots, a_r}$$

$$\times \sum_{\substack{1\le j_1<j_2<\cdots<j_r\le h\\ \text{distinct}}} \psi_k(N, \boldsymbol{j}, \boldsymbol{a}),$$

where

$$\psi_k(N,\boldsymbol{j},\boldsymbol{a}) = \sum_{n=1}^{N} \Lambda(n+j_1)^{a_1}\Lambda(n+j_2)^{a_2}\cdots\Lambda(n+j_r)^{a_r}$$

and the multinomial coefficient counts the number of different innermost sums that occur. If $n+j_i$ is a prime then

$$\Lambda(n+j_i)^{a_i} = \Lambda(n+j_i)(\log(n+j_i))^{a_i-1},$$

and thus

$$\psi_k(N,\boldsymbol{j},\boldsymbol{a}) \sim (\log N)^{k-r}\psi_{\boldsymbol{j}}(N).$$

Since

By the Hardy-Littlewood conjecture assuming it is valid uniformly for $\max|j_i| \le h$,

$$M_k(N,h,\psi) \sim N\sum_{r=1}^{k}(\log N)^{k-r}\sum_{\substack{a_1,a_2,\dots,a_r\\ a_i\ge 1,\sum a_i=k}}\binom{k}{a_1,a_2,\dots,a_r} \times \sum_{\substack{1\le j_1<j_2<\dots<j_r\le h\\ \text{distinct}}}\mathfrak{S}(\boldsymbol{j}).$$

Gallagher proved that, as $h\to\infty$,

$$\sum_{\substack{1\le j_1,j_2,\cdots,j_r\le h\\ \text{distinct}}}\mathfrak{S}(\boldsymbol{j}) \sim h^r,$$

and since this sum includes $r!$ permutations of the specified vector $\boldsymbol{j}$ when the components are ordered, we have

$$M_k(N,h,\psi) \sim$$
$$N(\log N)^k \sum_{r=1}^{k} \frac{1}{r!}\left(\frac{h}{\log N}\right)^r \sum_{\substack{a_1,a_2,\ldots,a_r \\ a_i \geq 1, \sum a_i = k}} \binom{k}{a_1,a_2,\ldots,a_r}.$$

Letting $\left\{ \begin{matrix} k \\ r \end{matrix} \right\}$ denote the Stirling numbers of the second type, then

$$\sum_{\substack{a_1,a_2,\ldots,a_r \\ a_i \geq 1, \sum a_i = k}} \binom{k}{a_1,a_2,\ldots,a_r} = r! \left\{ \begin{matrix} k \\ r \end{matrix} \right\}.$$

We conclude that for $h \sim \lambda \log N$,

$$M_k(N,h,\psi) \sim N(\log N)^k \sum_{r=1}^{k} \left\{ \begin{matrix} k \\ r \end{matrix} \right\} \lambda^r,$$

which are the moments of a Poisson distribution with mean λ. The first 4 moments are

$$\begin{aligned}
M_1(N,h,\psi) &\sim \lambda N \log N, \\
M_2(N,h,\psi) &\sim (\lambda+\lambda^2) N \log^2 N, \\
M_3(N,h,\psi) &\sim (\lambda+3\lambda^2+\lambda^3) N \log^3 N, \\
M_4(N,h,\psi) &\sim (\lambda+7\lambda^2+6\lambda^3+\lambda^4) N \log^4 N.
\end{aligned}$$

These Poisson moments determine the distribution function for small gaps:

$$\sum_{\substack{p_n \le N \\ p_{n+1}-p_n \le \lambda \log p_n}} 1 \sim (1-e^{-\lambda})\pi(N),$$

This quantitative result suggests small gaps occur fairly frequently.

5. HIGHER CORRELATIONS

In 1999 Yildirim and I were at MSRI. After working for two months we figured out: For

$R \leq N^{\frac{1}{3}}$

$$\sum_{n \leq N} \Lambda_R(n)^3 \sim \frac{3}{4} N \log^2 R,$$

Over the next year we added

$$\sum_{n \leq N} \Lambda_R(n)^2 \Lambda_R(n+k) \sim \mathfrak{S}(k) N \log R,$$

$$\sum_{n \leq N} \Lambda_R(n) \Lambda_R(n+k_1) \Lambda_R(n+k_2) \sim \mathfrak{S}(k_1, k_2) N,$$

Why did it take so long?

8. Calculating Correlations

For j a non-negative integer, define

$$\phi_j(p) = p - j,$$

$\phi_j(1) = 1$, and extend the definition to square-free integers by multiplicativity. Let

$$p(j) = \begin{cases} j, & \text{if } j \text{ is a prime,} \\ 1, & \text{otherwise,} \end{cases},$$

and define

$$H_j(n) = \prod_{\substack{p \mid n \\ p \neq j-1,\ p \neq j}} \left(1 + \frac{1}{p-j}\right)$$

We define the singular series for $j \geq 1$ and $n \neq 0$ by

$$\mathfrak{S}_j(n) = \begin{cases} C_j G_j(n) H_j(n), & \text{if } p(j) \mid n, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$G_j(n) = \prod_{\substack{p \mid n \\ p = j-1 \text{ or } p = j}} \left(\frac{p}{p-1}\right),$$

and

$$C_j = \prod_{\substack{p \\ p\neq j-1,\ p\neq j}} \left(1 - \frac{j-1}{(p-1)(p-j+1)}\right).$$

Theorem 1 *For* $R \geq 1$, $j \geq 0$, $p(j)|k$, *and* $0 \leq \log|k| \ll \log R$, *we have*

$$\sum_{\substack{d\leq R \\ (d,k)=1}} \frac{\mu(d)}{\phi_j(d)} \log\frac{R}{d} = \mathfrak{S}_{j+1}(k) + r_j(R,k),$$

where

$$r_j(R,k) \ll_j e^{-c_1\sqrt{\log R}},$$

and c_1 *is an absolute positive constant.*

The next lemma is a generalization of a result of Hildebrand.

Theorem 2 *For* $R \geq 1$, $j \geq 1$ *and* $p(j)|k$, *we have*

$$\sum_{\substack{d \leq R \\ (d,k)=1}} \frac{\mu^2(d)}{\phi_j(d)} = \begin{cases} \frac{1}{\mathfrak{S}_j(k)} \Big(\log R + D_j + h_j(k) \Big) + O(\frac{m(k)}{\sqrt{R}} \\ O(\frac{m(k)}{\sqrt{R}}), \end{cases}$$

where

$$D_j = \gamma + \sum_{p \neq j-1} \frac{(2-j)\log p}{(p-j+1)(p-1)},$$

$$h_j(k) = \sum_{p|k} \frac{\log p}{p-1} - \sum_{\substack{p|k \\ p \neq j-1}} \frac{(2-j)\log p}{(p-j+1)(p-1)},$$

and

$$m(k) = \sum_{d|k} \frac{\mu^2(d)}{\sqrt{d}} = \prod_{p|k} \left(1 + \frac{1}{\sqrt{p}} \right).$$

Let

$$\mathcal{S}_3(k_1,k_2,k_3) = \sum_{n=1}^{N} \Lambda_R(n+k_1)\Lambda_R(n+k_2)\Lambda_R(n+k_3)$$

Expanding, we have

$$\mathcal{S}_3(k_1,k_2,k_3) = \sum_{d_1,d_2,d_3 \le R} \prod_{i=1}^{3} \mu(d_i)\log(R/d_i) \sum_{\substack{n\le N \\ d_1|n+k_1 \\ d_2|n+k_2 \\ d_3|n+k_3}} 1.$$

The sum over n is zero unless $(d_1,d_2)|k_2-k_1$, $(d_1,d_3)|k_3-k_1$, and $(d_2,d_3)|k_3-k_2$, in which case the sum runs through a residue class modulo $[d_1,d_2,d_3]$, and we have

$$\sum_{\substack{n\le N \\ d_1|n+k_1 \\ d_2|n+k_2 \\ d_3|n+k_3}} 1 = \frac{N}{[d_1,d_2,d_3]} + O(1).$$

We conclude

$$\mathcal{S}_3(k_1, k_2, k_3)$$
$$= N \sum_{\substack{d_1,d_2,d_3 \le R \\ (d_1,d_2)|k_2-k_1 \\ (d_1,d_3)|k_3-k_1 \\ (d_2,d_3)|k_3-k_2}} \frac{\prod_{i=1}^{3} \mu(d_i) \log(R/d_i)}{[d_1, d_2, d_3]} + O(R^3).$$

We now decompose d_1, d_2, and d_3 into relatively prime factors

$$\begin{aligned} d_1 &= a_1 b_{12} b_{13} a_{123} \\ d_2 &= a_2 b_{12} b_{23} a_{123} \\ d_3 &= a_3 b_{13} b_{23} a_{123} \end{aligned}$$

where a_χ or b_χ is a divisor of the d_i's where i occurs in χ. Since the d_i's are squarefree, these new variables are pairwise relatively prime. The letters a and b reflect the parity of the number of d_i's that the new variable divides. We will let $\mathcal{D}$ denote the set of a_χ's and b_χ's which satisfy the conditions

$$\begin{aligned}
&a_1 b_{12} b_{13} a_{123} \le R\\
&a_2 b_{12} b_{23} a_{123} \le R\\
&a_3 b_{13} b_{23} a_{123} \le R\\
&b_{12} a_{123} | k_2 - k_1\\
&b_{13} a_{123} | k_3 - k_1\\
&b_{23} a_{123} | k_3 - k_2.
\end{aligned}$$

Letting

$$L_i(R) = \log \frac{R}{d_i}, \quad L = L_1(R) L_2(R) L_3(R)$$

we have

$$T_3(k_1, k_2, k_3) =$$

$$\sum_{\mathcal{D}}{}' \frac{\mu(a_1)\mu(a_2)\mu(a_3)\mu^2(b_{12})\mu^2(b_{13})\mu^2(b_{23})\mu(a_{123})}{a_1 a_2 a_3 b_{12} b_{13} b_{23} a_{123}} L$$

We first sum over a_1, a_2, and a_3. This eventually leads to

$$-\sum_{\substack{\mathcal{D}(D_3)\\ b_{12}b_{13}a_{123}\le R_1\\ b_{12}b_{23}a_{123}\le R_2\\ b_{13}b_{23}a_{123}\le R_3}}{}^{\prime} \frac{\mu^2(D_3)\mu(a_{123})\mu((D_3,2))}{\phi(D_3)} \times \mathfrak{S}_2(2D_3)\mathfrak{S}_3(2D_3)$$

where $D_3 = b_{12}b_{13}b_{23}a_{123}$, and R_1, R_2, R_3 will be chosen to either in terms of the $k_i's$ or to be about $Re^{-c(\log\log R)^2}$.

Application to primes.

Using 4th moment approximations, we knew that we would probably prove

$$\liminf_{n\to\infty} \left(\frac{p_{n+1} - p_n}{\log p_n}\right) \le .57$$

but we didn't care.

In mid-2000 we had obtained from our triple correlations:

Theorem 3 *For* $h \sim \lambda \log N$, $\lambda \ll R^{\epsilon}$, *and* $R = N^{\theta_k}$, *where* θ_k *is fixed and* $0 < \theta_k < \frac{1}{k}$ *for* $1 \le k \le 3$, *we have*

$$\begin{aligned}
M_1(N,h,\psi_R) &\sim \lambda N \log N, \\
M_2(N,h,\psi_R) &\sim (\theta_2\lambda + \lambda^2) N \log^2 N, \\
M_3(N,h,\psi_R) &\sim (\frac{3}{4}{\theta_3}^2\lambda + 3\theta_3\lambda^2 + \lambda^3) N \log^3 N.
\end{aligned}$$

Theorem 4 *For $h \sim \lambda \log N$, $\lambda \ll R^{\epsilon}$, and $R = N^{\theta_k}$, where θ_k is fixed, $0 < \theta_1 \leq 1$, and $0 < \theta_k < \frac{\vartheta}{k-1}$ for $2 \leq k \leq 3$, we have,*

$$\begin{aligned}
\tilde{M}_1(N,h,\psi_R) &\sim \lambda N \log N, \\
\tilde{M}_2(N,h,\psi_R) &\sim (\theta_2\lambda + \lambda^2) N \log^2 N, \\
\tilde{M}_3(N,h,\psi_R) &\sim ({\theta_3}^2\lambda + 3\theta_3\lambda^2 + \lambda^3) N \log^3 N.
\end{aligned}$$

Luckily, we stumbled on a better way to detect primes:

$$\tilde{M}_3(N,h,\psi_R,C) = \sum_{n=N+1}^{2N} \psi_h(n)\Big(\psi_R(n+h) - \psi_R(n) - C\log N\Big)^2,$$

where $\psi_h(n) = \psi(n+h) - \psi(n)$, and C may be chosen as a function of h and R to optimize the argument. Our moment results can now be applied, and on choosing C optimally we obtain

Theorem 5 *For $r \geq 1$, we have*

$$\Xi_r := \liminf_{n\to\infty} \left(\frac{p_{n+r} - p_n}{\log p_n}\right) \leq r - \frac{1}{2}\sqrt{r}.$$

$r = 1$ gives $1/2$, thus providing a 50 page proof of Bombieri-Davenport result.

We decided we could ''In Principle'' obtain asymptotic formulas for

$$\mathcal{S}_k(N, \boldsymbol{k}, \boldsymbol{a}) = \sum_{n=1}^{N} \Lambda_R(n+k_1)^{a_1}\Lambda_R(n+k_2)^{a_2}\cdots\Lambda_R(n+k_r)^{a_r}$$

and

$$\tilde{\mathcal{S}}_k(N, \boldsymbol{k}, \boldsymbol{a}) = \sum_{n=1}^{N} \Lambda_R(n+k_1)^{a_1}\cdots\Lambda_R(n+k_{r-1})^{a_{r-1}}\Lambda(n+k_r)$$

where $\boldsymbol{k} = (k_1, k_2, \ldots, k_r)$ and $\boldsymbol{a} = (a_1, a_2, \ldots a_r)$, the k_i's are distinct integers, $a_i \geq 1$ and $\sum_{i=1}^{r} a_i = k$. In the mixed correlation we assume that $r \geq 2$ and take $a_r = 1$.

Another 6 months, its 2001. Did: For $R \leq N^{\frac{1}{4}}$

$$\sum_{n \leq N} \Lambda_R(n)^3 \Lambda_R(n+k) \sim \frac{3}{4} \mathfrak{S}(k) N \log^2 R,$$

$$\sum_{n \leq N} \Lambda_R(n)^2 \Lambda_R(n+k_1) \Lambda_R(n+k_2) \sim \mathfrak{S}(k_1, k_2) N \log R,$$

and still needed to do

$$\sum_{n \leq N} \Lambda_R(n)^4, \qquad \sum_{n \leq N} \Lambda_R(n)^2 \Lambda_R(n+k)^2,$$

$$\sum_{\leq N} \Lambda_R(n) \Lambda_R(n+k_1) \Lambda_R(n+k_2) \Lambda_R(n+k_3)$$

At 2001 AIM meeting a shouted suggestion from the audience let us actually do all the correlations in a few weeks. Thanks to Sarnak, Friedlander, Conrey, Farmer.

Theorem 6 *Given* $k \geq 1$, $\max_i |j_i| \leq R$ *and* $R \geq 2$. *Then*

$$\mathcal{S}_k(N, \boldsymbol{j}, \boldsymbol{a}) = \Big(\mathcal{C}_k(\boldsymbol{a}) \mathfrak{S}(\boldsymbol{j}) + o_k(1)\Big) N (\log R)^{k-r} + O(R^k).$$

For $N^\epsilon \ll R \ll N^{\frac{1}{2(k-1)}}$,

$$\tilde{\mathcal{S}}_k(N, \boldsymbol{j}, \boldsymbol{a}) = \Big(\mathcal{C}_k(\boldsymbol{a})\mathfrak{S}(\boldsymbol{j}) + o(1)\Big)N(\log R)^{k-r}.$$

The $\mathcal{C}_k(\boldsymbol{a})$ are rational numbers, and Denoting $\mathcal{C}_k(k)$ as $\mathcal{C}_k$. With $\boldsymbol{a} = (a_1, a_2, \ldots, a_r)$

$$\mathcal{C}_k(\boldsymbol{a}) = \prod_{i=1}^{r} \mathcal{C}_{a_i}.$$

Here

$$\sum_{n\le N} \Lambda_R(n)^k \sim \mathcal{C}_k N(\log R)^{k-1}$$

The correlation constants $\mathcal{C}_k$ are defined by the absolutely convergent integrals

$$\mathcal{C}_1 = \frac{1}{2\pi i} \int_{(c_1)} \frac{e^{s_1}}{s_1}\, ds_1,$$

$$\mathcal{C}_2 = \frac{1}{(2\pi i)^2} \int_{(c_2)} \int_{(c_1)} \frac{e^{s_1+s_2}}{s_1 s_2 (s_1+s_2)}\, ds_1\, ds_2,$$

$$\mathcal{C}_3 = \frac{1}{(2\pi i)^3} \int_{(\sigma_3)} \int_{(\sigma_2)} \int_{(\sigma_1)} \frac{(s_1+s_2+s_3)e^{s_1+s_2+s_3}}{s_1 s_2 s_3 (s_1+s_2)(s_1+s_3)(s_2+s_3)} \, ds_1 \, ds_2 \, ds_3.$$

Currently (2006) we only know the values of the first six correlation constants:

$$\mathcal{C}_1 = 1, \quad \mathcal{C}_2 = 1, \quad \mathcal{C}_3 = \frac{3}{4}, \quad \mathcal{C}_4 = \frac{3}{4},$$
$$\mathcal{C}_5 = \frac{11065}{2^{14}} = .675\ldots,$$
$$\mathcal{C}_6 = \frac{11460578803}{2^{34}} = .667\ldots \ .$$

These values are obtained by residue calculations, the last three were found by David Farmer using a Mathematica program he wrote. It is routine to compute these moments from correlations. We have used Mathematica for the calculations. On taking $\mathcal{C}_1 = 1$ and

$$h = \lambda \log N, \quad R = N^\theta,$$

$$M_1 \sim \lambda$$
$$M_2 \sim \mathcal{C}_2\theta\lambda + \lambda^2,$$
$$M_3 \sim \mathcal{C}_3\theta^2\lambda + 3\mathcal{C}_2\theta\lambda^2 + \lambda^3,$$
$$M_4 \sim \mathcal{C}_4\theta^3\lambda + \Big(4\mathcal{C}_3 + 3C_2{}^2\Big)\theta^2\lambda^2 + 6\mathcal{C}_2\theta\lambda^3 + \lambda^4,$$
$$M_5 \sim \mathcal{C}_5\theta^4\lambda + \Big(5\mathcal{C}_4 + 10\mathcal{C}_3\mathcal{C}_2\Big)\theta^3\lambda^2 + \Big(10\mathcal{C}_3 + 15\mathcal{C}_2{}^2\Big)\theta^2\lambda^3 + 10\mathcal{C}_2\theta\lambda^4 + \lambda^5.$$

For the mixed moments we find

$$\tilde{M}_1 \sim \lambda,$$
$$\tilde{M}_2 \sim \theta\lambda + \lambda^2,$$
$$\tilde{M}_3 \sim \theta^2\lambda + \Big(2 + \mathcal{C}_2\Big)\theta\lambda^2 + \lambda^3,$$
$$\tilde{M}_4 \sim \theta^3\lambda + \Big(3 + \mathcal{C}_3 + 3C_2\Big)\theta^2\lambda^2 + \Big(3 + 3\mathcal{C}_2\Big)\theta\lambda^3 + \lambda^4,$$
$$\tilde{M}_5 \sim \theta^4\lambda + \Big(4 + \mathcal{C}_4 + 4\mathcal{C}_3 + 6\mathcal{C}_2\Big)\theta^3\lambda^2 + \Big(6 + 4\mathcal{C}_3 + 12\mathcal{C}_2 + 3\mathcal{C}_2{}^2\Big)\theta^2\lambda^3 + \Big(4 + 6\mathcal{C}_2\Big)\theta\lambda^4 + \lambda^5.$$

If all of the correlation constants are one, as we expect in the case when all the Λ_R's are replaced by Λ, then we obtain truncation Poisson moments

$$PM_1 \sim \lambda$$
$$PM_2 \sim \theta\lambda + \lambda^2$$
$$PM_3 \sim \theta^2\lambda + 3\theta\lambda^2 + \lambda^3$$
$$PM_4 \sim \theta^3\lambda + 7\theta^2\lambda^2 + 6\theta\lambda^3 + \lambda^4$$
$$PM_5 \sim \theta^4\lambda + 15\theta^3\lambda^2 + 25\theta^2\lambda^3 + 10\theta\lambda^4 + \lambda^5.$$

or in general

$$PM_k \sim \theta^k \mathcal{PM}_k(\frac{\lambda}{\theta})$$

where the Poisson moments are given by

$$\mathcal{PM}_k(\lambda) = \sum_{r=1}^{k} \left\{ \begin{array}{c} k \\ r \end{array} \right\} \lambda^r,$$

and the coefficients are the Stirling numbers of the second type.

OBTAINING GAPS FROM MOMENTS

To detect small gaps between primes, let $L = \log N$ and

$$\mathcal{S}_{2k+1} = \frac{1}{NL^{2k+1}} \sum_{n=N+1}^{2N} (\psi(n,h) - \rho L)\Big(P_k(\psi_R(n,h))\Big)^2$$

where $\rho \geq 0$ is a number we will eventually take to approach 1^+, and

$$P_k(\psi_R(n,h)) = \sum_{j=0}^{k} a_j (\psi_R(n,h))^j L^{k-j},$$

where the a_j's are arbitrary functions of N, R, λ, and ρ which are to be chosen to optimize the argument.

If we can prove $\mathcal{S}_{2k+1} > 0$ for a $\rho > 1$ then it follows that some of the intervals $(n, n+h]$ must contain two primes, which produces a gap between consecutive primes of size less than h.

We need to take

$$\theta < \frac{1}{4k}$$

in order to evaluate $\mathcal{S}_{2k+1}$. Assuming the Elliott-Halberstam conjecture we can take $\theta < \frac{1}{2k}$.

We found the following results for consecutive gaps:

n	Ξ	Ξ_2	EH Ξ	EH Ξ_2
3	.5	1.29289	.29289	1
5	.42580	1.16934	.22466	.85161
7	.38767	1.10435	.19068	.77529

We have also worked out a multiple truncation method: For the fifth moment we use the weight

$$a_3(\psi_{R_2}(n,h))^2 + a_2 L\psi_{R_2}(n,h) + a_1 L\psi_{R_1}(n,h) + a_0 L^2$$

Now we can take

$$R_1 = N^{1/4}, R_2 = N^{1/8}.$$

Using constants computed by David Farmer we found

n	Ξ	Ξ_2
5	.401220	1.12665
7	.355892	.970955

compared to single truncation result

n	Ξ	Ξ_2
3	.5	1.29289
5	.42580	1.16934
7	.38767	1.10435

A HINT : $a_2 = 0$ in

$$a_3(\psi_{R_2}(n,h))^2 + a_2 L\psi_{R_2}(n,h) + a_1 L\psi_{R_1}(n,h) + a_0 L^2$$

ruins the result completely.

Speculation in an NSF proposal

The following table compares the results for Ξ for truncated Poisson moments with moments where we take the first 6 values of $\mathcal{C}_k$ to be the correct values and set all other $\mathcal{C}_k = 2/3$:

n	Ξ Poisson	$\mathcal{C}_k$ correct, $k \leq 6$, $\mathcal{C}_k = 2/3$, $k \geq 7$
3	.5	.5
5	.43454	.42580
7	.40092	.38764
9	.37976	.36364
11	.36494	.34686
13	.35387	.33433
15	.34522	.32454
17	.33823	.31664

n	Ξ Poisson	Ξ EH Poisson
21	.327	.143
31	.311	.130
41	.301	.123
51	.294	.118
61	.290	.114
71	.285	.111
81	.284	.110
91	.281	.108
101	.279	.106
201	.268	.099
401	.262	.094
801	.257	.091
2001	.254	.088
4001	.252	.087
5001	.252	.085

Michael Rubinstein determined that the polynomials here are associated Laguerre polynomials, the smallest zero asymptotics are known, and that the above values approach

$$1/4, \qquad 1/12 \text{ (EH)}$$

He also apologized for finding this out.

Soundararajan (Oct 15, 2002) proved

$$\mathcal{C}_k \to \infty$$

actually

$$\mathcal{C}_k \gg k^k$$

He apologized about repercussions on the argument proposed in the NSF proposal.

Soundararajan's result shows that

$$\Lambda_R(n) = \sum_{\substack{d|n \\ d\leq R}} \mu(d) \log(R/d)$$

fails to detect primes correctly in higher correlations.

Thus perhaps the factor $\log(R/d)$ is the wrong weight when trying to approximate k-tuples of primes, and something else is needed.

In November 2003 Heath-Brown suggested using

$$\sum_{1\le i,j\le K} a_{ij}\psi_{R_i}(n,h)\psi_{R_j}(n,h)$$

with $R_iR_j = N^{1/4-\epsilon}$ in the moment argument, determine optimal a_{ij}, and then rewriting above as

$$\sum_{\substack{d,d' \\ dd'\le N^{1/4-\epsilon}}} h(d,d')$$

one can get an approximation for h(d,d') from which one can maybe guess the right weight.

THE NEW IDEA (BUT TOTALLY WRONG)

Up to now we used the truncated divisor sum

$$\Lambda_R(n) = \sum_{\substack{d|n \\ d\le R}} \mu(d)\log(R/d)$$

which is a truncated and smoothed approximation for the von Mangoldt function $\Lambda(n)$ because

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(1/d).$$

$\Lambda_R(n)$ gives asymptotically the best least square fit to $\Lambda(n)$.

But what should be used to approximate a 2-tuples of primes:

$$\Lambda(n)\Lambda(n+k)?$$

Up to now we have used

$$\Lambda(n)\Lambda(n+k) \approx \Lambda_{R_1}(n)\Lambda_{R_2}(n+k), \quad R_1R_2 = R.$$

The problem with this: We can not take different choices of R_1 and R_2 within the same sum. However:

$$d_1 \le R_1 \text{ and } d_2 \le R_2 \Rightarrow d_1d_2 \le R_1R_2 = R,$$

and the latter inequality includes all possible divisors that can occur for every value of R_1 and R_2 with $R_1 R_2 \leq R$.

Therefore consider divisor sums with this new truncation, and then smooth in an appropriate fashion. AND: This also allows us to eliminate the correlation constants encountered before.

NEW IDEA- BUT WRONG Approximate $\Lambda(n)\Lambda(n+k)$ by the truncated double divisor sum

$$\Lambda_R(n, n+k) = \sum_{\substack{d_1|n,\ d_2|n+k \\ [d_1,d_2]\le R}} \left(1 - \frac{[d_1, d_2]}{R}\right) \mu(d_1)\mu(d_2) \log \frac{R}{d_1} \log \frac{R}{\frac{d_2}{(d_1,d_2)}}.$$

The weight $1 - \frac{[d_1,d_2]}{R}$ was only inserted to simplify the proofs, and may be deleted.

Generalizing for r-tuples, we let $D_0 = 1$ and for $r \geq 1$, let

$$D_r = [d_1, d_2, \ldots, d_r],$$

and define for $\boldsymbol{m} = (m_1, m_2, \ldots, m_r)$

$$\Lambda_R(\boldsymbol{m}) = \sum_{\substack{d_j | m_j,\ 1 \leq j \leq r \\ D_r \leq R}} \left(1 - \frac{D_r}{R}\right) \prod_{j=1}^{r} \mu(d_j) \log \frac{R}{\frac{d_j}{(D_{j-1}, d_j)}},$$

which is an approximation of $\Lambda(m_1)\Lambda(m_2)\cdots\Lambda(m_r)$

Granville and Soundararajan found this doesn't work, but also set the stage so that all the problems above - Correlation coefficients and detecting primes correctly were solved. Only a new approximation idea was needed.