Introduction to Sieve Methods

Frank Thorne

March 30, 2006

Overview:

- Introduce notation and discuss the Sieve of Erasthosthenes-Legendre, use to give a simple upper bound on $\pi(x)$.

- Discuss the idea of the *density* or *dimension* of a sieve.

- Discuss *Selberg's sieve* $-$ an effective and simple upper bound sieve method.

- Application: Use this sieve to derive the 'correct' asymptotic for the number of twin primes.

- Give an introduction to recent results on small gaps between primes - work of Goldston, Graham, Pintz, and Yildirim.

The Sieve of Erasthosthenes-Legendre

Classical Problem: Give an upper bound on $\pi(X)$, the number of primes between 1 and X.

Prime Number Theorem :

$$\pi(X) \sim \frac{X}{\log X}.$$

We will do a lot worse than this. But let's have some fun anyway.

Estimating $\pi(X)$

Technique: Start with all numbers between 1 and X. Then, for all small primes p, exclude all of the numbers divisible by p.

The probability that a prime p divides some number X is roughly $1/p$, and these probabilities are roughly independent for different primes p and q. Accordingly, we expect an estimate that looks like

$$\pi(X) \leq X \prod_{p} (1 - 1/p) \, .$$

We now formalize the above, introduce some notation, and get an estimate for the error.

Notation : A will denote a set of numbers to be sifted, here $[1, X]$.

P will denote some set of primes, **or** (by abuse of notation) the product of these same primes.

$S(A, P)$ will denote the elements of A which are divisible by no element of P.

The general sieve problem: Estimate $|S(A, P)|$, for arbitrary A and P.

Our problem : Choose $P = \{p|p < z\}$ for some z to be chosen later. If we do this then

$$\pi(X) \leq S(A, P) + \pi(z).$$

Estimation of $S(A, P)$

Use the inclusion-exclusion principle of combinatorics.

Notation : For an integer d, let

$$A_d = \{a \in A : d | A\}$$

Then, by inclusion-exclusion, we have

$$|S(A, P)| = X - \sum_{p \in P} |A_p| + \sum_{p_1, p_2 \in P} |A_{p_1 p_2}|$$

$$- \sum_{p_1, p_2, p_3 \in P} |A_{p_1 p_2 p_3}| + \dots$$

or

$$|S(A, P)| = \sum_{d | P} \mu(d) |A_d|.$$

Estimation of $|A_d|$

We have the equality

$$|A_d| = \left[\frac{X}{d}\right]$$

and it is convenient to write this

$$|A_d| = \frac{X}{d} + r_A(d)$$

where $|r_A(d)| < 1$. Then the main term X/d is *multiplicative* in d, so our sum will turn into a product.

In general - $\rho(d)$: For general sieve problems, we hope to find a multiplicative function $\rho(d)$ so that $\rho(d)/d$ approximates the proportion of elements of A divisible by d. Then we define $r_A(d)$ by

$$|A_d| = \frac{\rho(d)}{d} X + r_A(d).$$

Approximating $|S(A,P)|$ Returning to the problem of approximating $|S(A,P)|$, we have

$$|S(A,P)| = \sum_{d|P} \mu(d)|A_d|$$

$$|S(A,P)| = \sum_{d|P} \mu(d)(X/d + r_A(d))$$

$$|S(A,P)| = X \sum_{d|P} \mu(d)/d + \sum_{d|P} \mu(d) r_A(d)$$

$$|S(A,P)| = X \prod_{p \in P} \left(1 + \frac{\mu(p)}{p}\right) + \sum_{d|P} \mu(d) r_A(d)$$

and therefore

$$\pi(X) \le X \prod_{p<z} \left(1 - \frac{1}{p}\right) + \sum_{d|P} \mu(d) r_A(d) + \pi(z).$$

We can estimate this stuff.

Approximating $|S(A, P)|$

Mertens' Product:

$$\prod_{p<z}\left(1-\frac{1}{p}\right)^{-1} = e^{\gamma}\log z + O(1).$$

$\gamma = 0.577\ldots$ is Euler's constant.

An alternative elementary estimate:

$$\prod_{p<z}\left(1-\frac{1}{p}\right)^{-1} = \prod_{p<z}\sum_{r=0}^{\infty}\frac{1}{p^r} > \sum_{n<z}\frac{1}{n} > \log z.$$

In either case we have

$$\pi(X) < \frac{X}{\log z} + \pi(z) + \sum_{d|P}\mu(d)r_A(d)$$

Give the trivial estimate $\pi(z) \leq z$ and so

$$\pi(X) < \frac{X}{\log z} + z + \sum_{d|P}\mu(d)r_A(d)$$

and if we can estimate this last error term then we're done.

Estimation of the Error Term We want to estimate

$$\sum_{d|P} \mu(d) r_A(d).$$

What we hope for: The $r_A(d)$ are all positive, and the Möbius function should cause some cancellation.

The sad reality: This method is ill-equipped to give a nontrivial estimate. We are obliged to estimate

$$\mu(d) \leq 1$$

$$r_A(d) \leq 1$$

and therefore

$$\sum_{d|P} \mu(d) r_A(d) \leq \sum_{d|P} 1 = 2^{\pi(z)}.$$

This stinks.

Putting it All Together

We have the estimate

$$\pi(X) < X/\log z + z + 2^{\pi(z)}$$

and we *choose* $z = \log X$ to derive

Theorem : $\pi(X) << \frac{X}{\log\log X}$.

This is a mediocre estimate, but we can immediately generalize to other sets A. For example, the same result is immediate for sets $[N, N + X - 1]$. Still, we want to try to do better.

How To Do Better?

- We *can* be more clever about the error - use "Rankin's Trick", etc. For example,

$$\pi(x) << \frac{x \log \log x}{\log x}.$$

See Ch. 5 of Cojocaru.

- *Brun's Combinatorial Sieve (1915, 1920)* : Came up with a vastly improved version of this sieve. Brun proved, for example, that there are infinitely many P_9's. Unfortunately, Brun's notation was horrible and no one tried to read his papers.

A General Theorem

We apply the same methods to (easily) prove the following

Theorem (Greaves, p. 20): Assume the notation given previously. Suppose that we are given a function $\lambda_D(d)$ with the property that

$$\sum_{d|A} \mu(d) \leq \sum_{d|A} \lambda(d)$$

for any $A|P$. Then, we have

$$S(A, P) \leq XV(D, P) + R(D, P),$$

where

$$V(D, P) = \sum_{d|P} \frac{\lambda_D(d)\rho(d)}{d}$$

$$R(D, P) = \sum_{d|P} \lambda_D(d) r_A(d).$$

How to choose $\lambda_D(d)$? If we set $\lambda_D(d) = \mu(d)$, then we recover the following...

Theorem: (Greaves, p. 23) Set

$$V(D,P)=\sum_{d|P}\frac{\mu(d)\rho(d)}{d}=\prod_{p|P}\left(1-\frac{\rho(p)}{p}\right).$$

Then for some θ with $|\theta|\leq 1$,

$$S(A,P)=XV(P)+\theta\sum_{d|P}|r_A(d)|.$$

Set $\rho(p)=1$, $|r_A(d)|\leq 1$ to recover the previous theorem.

How to prove something better? Find an interesting function λ_D. In particular find λ_D with *level of support* D; i.e., $\lambda_D(d)=0$ for $d>D$. (This explains the D in the notation.)

If we can do this, then we may make the estimate

$$R(D,P)\leq \sum_{d|P,d\leq D}|r_A(d)|$$

and this will allow us to sieve out more primes while maintaining control over the error. The *Selberg sieve* will give us such a λ_D.

Sifting Density

Idea: Give an 'average value' for the function $\rho(p)$. If $\rho(p)$ is frequently high, more elements will be sieved out – and so we can expect better estimates.

Definition 1: (Greaves, p. 28) κ is a **sifting density** for the function ρ if there exists an $L > 1$ (depending on κ) such that

$$\prod_{w \le p \le z} \left(1 - \frac{\rho(p)}{p}\right)^{-1} \le K_w \left(\frac{\log z}{\log w}\right)^{\kappa}$$

with

$$K_w \le 1 + \frac{L}{\log w},$$

for any w and z with $2 \le w < z$.

Should look sort of like Mertens' product.

An Alternative Definition

Definition 2: (Greaves, p. 33) We say κ is a **sifting density** for ρ in the strong sense if there is a constant A such that

$$\sum_{w \le p < z} g(p) \log p \le \kappa \log \frac{z}{w} + A$$

where $g(p)$ is defined by

$$g(p) = \frac{\rho(p)}{p - \rho(p)}.$$

In other words, a weighted average of the $\rho(p)$ is sufficiently small. $g(p)$ is 'almost' like $\rho(p)/p$ if $\rho(p)$ remains bounded.

This function g will appear again, and its definition will be extended to all squarefree d by multiplicativity.

Basic Facts about Sifting Densities

Theorem :

(0) If ρ has sifting density κ then it has sifting density κ' for any $\kappa' > \kappa$.

(1) If $\rho(p) < p$ always, and $\rho(p) \leq \kappa$ for all p, then ρ has sifting density κ in both senses. Moreover, if $\rho(p) = \kappa$ almost always, then ρ does not have a sifting density less than κ.

(2) If κ is a sifting density for ρ in the strong sense, then it is in the weak sense as well.

None of the proofs are terribly deep.

The Selberg Sieve

Developed by Atle Selberg in the 1940's. An excellent small sieve method. 'Small' means, it will provide good results for sieve problems of finite density.

References: See books by Greaves, Halberstam and Richert, Iwaniec and Kowalski, Cojocaru and Murty.

What can you use it to prove?

Selberg Sieve Applications

Theorem :

$$\pi(X) << \frac{X}{\log X}.$$

Theorem : Let $TP(X)$ denote the number of twin primes less than X. Then

$$TP(X) << \frac{X}{\log^2 X}.$$

Theorem (Brun-Titchmarsh): Let $\pi(y; l, k)$ denote the number of primes less than y congruent to l mod k. Then

$$\pi(y; l, k) - \pi(y - x; l, k) \leq$$

$$\frac{2x}{\phi(k)\log(x/k)} + O(\frac{x}{k\log^2(x/k)}),$$

the implied constant being absolute.

Spectacular Selberg Sieve Applications

Theorem (Goldston, Graham, Pintz, Yildirim): Let p_n denote the n-th prime, and let q_n denote the n-th number that is the product of exactly two distinct primes. Then:

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0;$$

$$\liminf_{n\to\infty} q_{n+1} - q_n \leq 26.$$

Moreover, these results are vastly improved under the hypothesis of a stronger version of the Bombieri-Vinogradov Theorem.

The first proof of the first result is by complex analytic methods, due to Goldston, Pintz, and Yildirim. The paper with Graham uses the Selberg sieve to prove both results.

The Selberg Upper Bound Method

Lemma : Suppose that for each d a real number λ_d is defined, with $\lambda_1 = 1$. Then

$$|S(A,P)| \leq \sum_{a \in A} \left(\sum_{d|(a,P)} \lambda_d \right)^2 .$$

The proof is trivial: Every element $a \in S(A,P)$ is coprime to P, and therefore contributes $\lambda_1^2 = 1$ to the sum at right. Every $a \in A - S(A,P)$ contributes nothing to the left, and something squared to the right.

We now analyze the quantity at right. We have

$$\sum_{a \in A} \left(\sum_{d|(a,P)} \lambda_d \right)^2 = \sum_{a \in A} \sum_{d,e|(a,P)} \lambda_d \lambda_e$$

$$= \sum_{d,e|P} \lambda_d \lambda_e \sum_{a \in A; [d,e]|a} = \sum_{d,e|P} \lambda_d \lambda_e A_{[d,e]}.$$

Selberg's Method

$$|S(A,P)| \le \sum_{d,e|P} \lambda_d \lambda_e A_{[d,e]}.$$

We now split the quantity $A_{[d,e]}$ into our usual estimate and error and write

$$|S(A,P)| \le XG + R(A,P)$$

where

$$G = \sum_{d,e|P} \lambda_d \lambda_e \frac{\rho([d,e])}{[d,e]}$$

$$R(A,P) = \sum_{d,e|P} \lambda_d \lambda_e r_A([d,e]).$$

Now the λ_d and λ_e are still undetermined, so we regard G as a bilinear form and choose λ_d and λ_e to minimize this form. Moreover, we choose a **level of support** R for the λ_d; that is, we stipulate that $\lambda_d = 0$ for $d > R$. This will allow us to control the error $R(A,P)$.

Estimating G

$$G = \sum_{d,e|P} \lambda_d \lambda_e \frac{\rho([d,e])}{[d,e]}.$$

By multiplicativity we can write

$$G = \sum_{d,e|P} \left(\frac{\lambda_d \rho(d)}{d}\right)\left(\frac{\lambda_e \rho(e)}{e}\right)\frac{(d,e)}{\rho(d,e)}.$$

We want to convert the latter quantity into a sum, so we can switch the order of summation:

$$\frac{(d,e)}{\rho(d,e)} = \prod_{p|(d,e)} \frac{p}{\rho(p)}$$

$$= \prod_{p|(d,e)} \left(1 + \frac{p-\rho(p)}{\rho(p)}\right).$$

Tonghai Yang's dictum: 'If you don't understand something, give it a name.'

The Function $g(n)$

Notation : Define a function $g(n)$ as follows: For prime p define

$$g(p) = \frac{\rho(p)}{p - \rho(p)}$$

and then extend to all squarefree n by multiplicativity.

Notice that this coincides with the definition given during the discussion of sifting density.

Remark: We can express this as a Dirichlet convolution. If we first define the multiplicative function

$$f(n) = \frac{n}{\rho(n)}$$

then

$$\frac{1}{g(n)} = (f * \mu)(n).$$

Estimating G (cont.)

With this notation we have

$$\frac{(d,e)}{\rho(d,e)} = \prod_{p|(d,e)} \left(1 + \frac{1}{g(p)}\right)$$

$$= \sum_{h|(d,e)} \frac{1}{g(h)}$$

and thus

$$G = \sum_{d,e|P} \left(\frac{\lambda_d \rho(d)}{d}\right) \left(\frac{\lambda_e \rho(e)}{e}\right) \sum_{h|(d,e)} \frac{1}{g(h)}.$$

Interchanging the order of summation,

$$G = \sum_{h|P} \frac{1}{g(h)} \sum_{d,e|P;h|d;h|e} \left(\frac{\lambda_d \rho(d)}{d}\right) \left(\frac{\lambda_e \rho(e)}{e}\right)$$

$$= \sum_{h|P} \frac{1}{g(h)} \left(\sum_{d|P;h|d} \left(\frac{\lambda_d \rho(d)}{d}\right)\right)^2.$$

A Bilinear Form

$$G = \sum_{h|P} \frac{1}{g(h)} \left(\sum_{d|P;h|d} \left(\frac{\lambda_d \rho(d)}{d} \right) \right)^2 .$$

Notation / Proposition : Denote

$$x(h) = \sum_{d|P;h|d} \frac{\lambda_d \rho(d)}{d}.$$

Then, if $x(h)$ is given for each h, the λ_d are given as

$$\lambda_d = \frac{d}{\rho(d)} \sum_k \mu(k) x(kd).$$

This is 'dual' Möbius inversion - see p. 45 of Greaves or p. 4 of Cojocaru.

Notice: $x(h)$ has the same level of support as λ_d; $x(h) = 0$ for $h > R$.

A Bilinear Form (cont.)

With the change of variables

$$x(h) = \sum_{d|P;h|d} \frac{\lambda_d \rho(d)}{d}.$$

we may write

$$G = \sum_{h|P} \frac{1}{g(h)} x(h)^2$$

and our Möbius inversion formula gave the constraint

$$\sum_k \mu(k) x(k) = \lambda_1 = 1$$

and subject to this constraint we may choose $x(h)$ freely to minimize the form

$$\sum_{h|P} \frac{1}{g(h)} x(h)^2.$$

Minimizing the Form

Proposition : This bilinear form attains its minimum at

$$x(k) = C\mu(k)g(k),$$

or equivalently,

$$\lambda_d = C\frac{\mu(d)d}{\rho(d)} \sum_{h;d|h,h|P,h\leq R} g(h)$$

with $\lambda_d = 0$ whenever $\rho(d) = 0$. The constant C is chosen to normalize $\lambda_d = 1$; it is determined by

$$C = \frac{1}{G(R)},$$

where $G(R)$ is defined to be

$$G(R) = \sum_{n<R;n|P} g(n).$$

We have, for each d, that $|\lambda_d| \leq 1$. And, finally, the **minimum of this form is** given by $1/G(R)$.

Selberg's Main Theorem

The proof is elementary, and follows by Lagrange multipliers or Cauchy's inequality. (See p. 46 of Greaves.)

Collecting all of this, we have the following

Theorem:

$$|S(A, P)| \leq \frac{X}{G(R)} + R(A, P)$$

where

$$G(R) = \sum_{n \leq R; n|P} \prod_{p|n} \frac{\rho(n)}{p - \rho(n)}$$

$$R(A, P) = \sum_{d,e \leq R; d,e|P} \lambda_d \lambda_e r_A([d, e]).$$

Main Theorem - Remarks

- The cutoff for the support of the λ_d makes the error term a lot better. In the previous problem we had to take $z = \log X$. Using this sieve, we may take R much larger, and sieve out many more primes simultaneously.

For example, choices such as $R = X^{1/2}$ will frequently be practical.

- As $|\lambda_d| \leq 1$ for all d, the trivial estimate of the error term will often be acceptable in applications.

These quantities are a bit messy to define but are often simple to estimate. For example...

A Sample Theorem

Recall that one of our definitions of *sifting density* involved the quantity $g(d)$.

Theorem : (Greaves, p. 55) Suppose a set A satisfies a certain sifting density hypothesis, with sifting density κ.

Then,

$$G(R) \geq \left(\mathfrak{S}(\rho, R) + O\left(\frac{1}{\log R}\right)\right) \log^{\kappa} R$$

where

$$\mathfrak{S}(\rho, R) = \prod_{p \leq R} \left(1 - \frac{1}{p}\right)^{\kappa} \left(1 - \frac{\rho(p)}{p}\right)^{-1}.$$

The upshot: Often, we can often expect that

$$S(A, P) << \frac{|A|}{\log^{\kappa} |A|}.$$

The Quantity $\mathfrak{S}(\rho, R)$

We remark that the reciprocal of the infinite version of

$$\mathfrak{S}(\rho, R) = \prod_{p \le R} \left(1 - \frac{1}{p}\right)^{\kappa} \left(1 - \frac{\rho(p)}{p}\right)^{-1}$$

is the **singular series**

$$\mathfrak{S}(\rho) = \prod_{p} \left(1 - \frac{1}{p}\right)^{-\kappa} \left(1 - \frac{\rho(p)}{p}\right).$$

In general, such series will converge, and can be estimated without too much trouble. Singular series crop up commonly in analytic number theory – in sieve methods, in the circle method, etc.

A Simple Application to Twin Primes

As a more concrete illustration of the preceding theory, we prove the following

Proposition : Let $TP(x)$ denote the number of twin primes less than x. Then

$$TP(x) << \frac{x}{\log^2 x}.$$

In particular, the sum of the reciprocals of the twin primes converges.

It is easy to see that the first statement implies the second: the number of twin primes in the interval $[2^n, 2^{n+1}]$ is $O(2^n/n^2)$, and therefore the sum of the reciprocals in this interval is $O(1/n^2)$. Summing over all positive n we obtain the result.

The Right Asymptotic Before proving this result, we indicate why we believe that it is the best possible. In particular we have the

Conjecture :

$$TP(X) = \frac{X}{\log^2 X}(\mathfrak{S}(H) + o(1)),$$

where

$$H = \{0, 2\}$$

$$\mathfrak{S}(H) = \frac{1}{2} \prod_{p \neq 2} \left(1 - \frac{2}{p}\right) \left(1 - \frac{1}{p}\right)^{-2}.$$

This conjecture is easily explained using heuristic reasoning: The probability that a large number x is prime is roughly $1/\log x$, so very roughly $TP(x) \sim X/\log^2 X$. However, the probability, for each p, that p divides x and that p divides $x+2$ are not independent, and so we multiply by a correction term for each p.

The Selberg Sieve Setup We will use the Selberg sieve to prove the first assertion. We choose

$$A = \{n(n+2) | n \le x\}$$

and we will sift A by the set of primes P that are less than some number R. In doing so, we will give an upper bound on $TP(X)$ with an error less than $R/\log R$.

We recall that A_d denotes the subset of elements divisible by d, and we write

$$A_d = \frac{\rho(d)}{d} + r_A(d).$$

For primes $p > 2$, we will have $\rho(p) = 2$: p will divide $n(n+2)$ for n congruent to either 0 or -2 mod p, exactly twice in each residue class. We will also have $\rho(2) = 1$.

Estimation of $G(R)$

The bulk of the proof is to estimate the quantity

$$G(R) = \sum_{d \le R; d|P} \prod_{p|d} \frac{\rho(d)}{p - \rho(d)}.$$

We can do this using standard techniques as follows:

$$G(R) = \sum_{d \le R; d|P} \prod_{p|d; p \ge 3} \left(\frac{2}{p} + \left(\frac{2}{p}\right)^2 + \left(\frac{2}{p}\right)^3 + \dots \right)$$

$$G(R) \ge \sum_{d \le R; (d,2)=1} \frac{2^{\nu(d)}}{d}.$$

We denote by $\nu(d)$ the number of divisors of d, and note that the condition $d|P$ in the sum is redundant by our choice of $P = \prod_{p \le R} p$.

Estimation of $G(R)$

$$G(R) \geq \sum_{d \leq R; (d,2)=1} \frac{2^{\nu(d)}}{d}.$$

We write the second sum as a sum over all divisors and interchange the order of summation:

$$G(R) \geq \sum_{d \leq R; (d,2)=1} \frac{1}{d} \sum_{k|d} 1$$

$$= \sum_{k \leq R; (k,2)=1} \sum_{d \leq R; k|d; (d,2)=1} \frac{1}{d}$$

$$= \sum_{k \leq R; (k,2)=1} \frac{1}{k} \sum_{a \leq R/k; (a,2)=1} \frac{1}{a}$$

$$= \sum_{k \leq R; (k,2)=1} \frac{1}{k} \left(\frac{1}{2} \log \frac{R}{k} + O(1) \right)$$

$$= \frac{1}{2} \sum_{k \leq R} \left(\frac{\log R}{k} - \frac{\log k}{k} \right) + O(\log R)$$

$$= \frac{1}{2}\log^2 R - \frac{1}{2}\sum_{k\le R}\frac{\log k}{k} + O(\log R).$$

We can use partial summation to show that

$$\sum_{k\le R}\frac{\log k}{k} = \frac{1}{2}\log^2 R + O(\log R)$$

and therefore

$$G(R) = \frac{1}{4}\log^2 R + O(\log R),$$

which is of the desired order of magnitude.

Estimation of $R(A,P)$

We now turn to the cumulative error

$$R(A,P) = \sum_{d,e|P} \lambda_d \lambda_e r_A([d,e]).$$

We can give a simple estimate for this:

- The λ_d are only supported on $r < R$, so that the summand is only supported on $[d,e] < R^2$.

- $|\lambda_d| \leq 1$ for all d.

- All $[d,e]$ will be squarefree. So,

$$R(A,P) \leq \sum_{n<R^2} \mu^2(n)\alpha(n) r_A(n)$$

where we use $\alpha(n)$ to denote the number of ways to write $n = [d,e]$ for some d and e.

Estimation of $R(A,P)$

$$R(A,P) \leq \sum_{n<R^2} \mu^2(n)\alpha(n)r_A(n)$$

$\alpha(n)$: Multiplicative, with $\alpha(p) = 3$ for any prime p, so $\alpha(n) = 3^{\omega(n)}$, where $\omega(n)$ counts the number of prime divisors of n.

$r_A(n)$: Again multiplicative. By the Chinese Remainder Theorem, there are $2^{\omega(d)}$ elements of A_d in each residue class mod d (for d odd, squarefree). Making no assumption of 'nice' distribution, $|r_A(n)| \leq 2^{\omega(n)}$.

Putting this together,

$$R(A,P) \leq \sum_{n<R^2} \mu^2(n)6^{\omega(n)}.$$

Estimation of $R(A,P)$

Claim:

$$\sum_{n<R^2} \mu^2(n) 6^{\omega(n)} \le R^2 (\log R^2 + 1)^6.$$

Proof ([GGPY], Lemma 1): Clearly the above will follow if we can prove that

$$\sum_{n<R^2} \frac{\mu^2(n) 6^{\omega(n)}}{n} \le (\log R^2 + 1)^6.$$

But the left hand side is

$$\sum_{d_1 d_2 \ldots d_6 < R^2} \frac{\mu^2(d_1 \ldots d_n)}{d_1 \ldots d_n} \le \left(\sum_{n \le R^2} \frac{1}{n} \right)^6$$

$$\le (\log R^2 + 1)^6.$$

The Exciting Finish

Recalling our main theorem

$$|S(A,P)| \le \frac{X}{G(R)} + R(A,P)$$

and also

$$TP(X) \le |S(A,P)| + R/\log R$$

we put all of this together to obtain the estimate

$$TP(X) \le \frac{X}{\frac{1}{4}\log^2 R + O(\log R)} + (\log R^2 + 1)^6 + \frac{R}{\log R}$$

$$= \frac{4X}{\log^2 R} + O\left(\frac{4X}{\log^3 R} + \log^6 R + \frac{R}{\log R}\right).$$

Choosing $R = X^{1-\epsilon}$, the result follows.

Introduction To Small Gaps Between Primes

We make some brief remarks on how the ideas considered figure in [GGPY].

Let $H = h_1, h_2, \ldots, h_k$ be a set of distinct primes. For each prime p denote by $\rho_p(H)$ the number of distinct residue classes mod p in H.

The **singular series** :

$$\mathfrak{S}(H) = \prod_p \left(1 - \frac{\rho_p(H)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}$$

Conjecture : The number of $n \le N$ such that $n + h_1, \ldots, n + h_k$ are all prime is

$$\frac{N}{\log^k N}\left(\mathfrak{S}(H) + o(1)\right).$$

The Hardy-Littlewood Conjecture

Conjecture : The number of $n \le N$ such that $n + h_1, \ldots, n + h_k$ are all prime is

$$\frac{N}{\log^k N} \left(\mathfrak{S}(H) + o(1) \right).$$

Note that the case $H = \{0, 2\}$ is the twin prime conjecture with the asymptotic discussed previously. The case $H = \{0\}$ is the prime number theorem, and is the only nontrivial case which has been proven (or disproven).

Note: It may be the case that the h_i cover all the residue classes mod p for some p. In this case we call the sequence H *inadmissable.* We will have $\mathfrak{S}(H) = 0$ and the conjecture is trivial in this case.

The Function $\varpi(n)$

We can rephrase the above by introducing a function $\varpi(n)$, defined as follows:

$$\varpi(n) = \begin{cases} \log n & \text{if } n \text{ prime} \\ 0 & \text{otherwise.} \end{cases}$$

Then, we expect that

$$\sum_{n \le N} \varpi(n+h_1)\dots\varpi(n+h_k) = N(\mathfrak{S}(N) + o(1)).$$

The Bombieri-Vinogradov Theorem

The prime number theorem for arithmetic progressions (see Davenport) implies that for any a and q with $(a, q) = 1$, and for any x, we have

$$\sum_{x<n\le 2x;\, n\equiv a \pmod q} \varpi(n) \sim \frac{x}{\phi(q)}.$$

Define $E(x; q, a)$ by

$$\sum_{x<n\le 2x;\, n\equiv a \pmod q} \varpi(n) = \frac{x}{\phi(q)} + E(x; q, a)$$

$$E(x, q) = \max_{a;(a,q)=1} |E(x; q, a)|$$

$$E^*(N, q) = \max_{x\le N} E(x, q)$$

Then...

The Bombieri-Vinogradov Theorem

Theorem : If $A > 0$ then there exists $B > 0$ such that for $Q \le N^{1/2} \log^{-B} N$, then

$$\sum_{q \le Q} E^*(N, q) <<_A N \log^{-A} N.$$

In other words, the worst case error in the prime number theorem, as you average over q, is small. The 'smallness' is approximately that predicted by GRH.

GGPY: The Basic Setup

The idea is to find short, closely packed sequences H so that for infinitely many n, at least two of the numbers $n+h_i$ are prime, almost prime, or what have you. To that end, write

$$P(n;H) = \prod_{h \in H} (n+h)$$

and consider a sum

$$S = \sum_{n=N}^{2N} \left(\sum_{h \in H} \varpi(n+h) - \log 3N \right) \left(\sum_{d \mid P(n,H)} \lambda_d \right)^2$$

with indeterminates λ_d. If the sum is positive, then at least two of the numbers $n+h$ will be prime for some n. Therefore, we want to prove that S is always positive.

GGPY - cont.

Goldston et al. only manage to prove this under an appropriate (conjectural) generalization of the Bombieri-Vinogradov Theorem. They do get close enough that they are able to prove that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0,$$

essentially by comparing results for a set of different H.

To apply the method to almost primes, replace ϖ with $\varpi * \varpi$, which picks out numbers with exactly two prime factors.

GGPY - The Basic Setup

The proof proceeds by analyzing the quantities

$$S = \sum_{N<n\leq 2N} \left(\sum_{d|P(n,H)} \lambda_d \right)^2$$

and

$$S = \sum_{N<n\leq 2N} \varpi(n) \left(\sum_{d|P(n,H)} \lambda_d \right)^2 .$$

One wishes to choose λ_d so that the first quantity is small and the last quantity is large. That is, the λ_d will be used to 'pick out primes'.

The problem of minimizing the first quantity alone is a standard sieve problem. This is a bit more sophisticated.

The Choice of λ_d

Recall in our discussion of the Selberg sieve we defined a quantity

$$x(h) = \sum_{d|P;h|d} \frac{\lambda_d \rho(d)}{d}.$$

and used Möbius inversion to recover the λ_d.

Goldston et al. use the notation $y_{r,\ell}$ in place of $x(h)$, and define

$$y_{r,\ell} = y_{r,\ell}(H) = \begin{cases} \frac{\mu^2(r)\mathfrak{S}(H)(\log R/r)^\ell}{\ell!} & \text{if } r < R \\ 0 & \text{otherwise.} \end{cases}$$

Note that ℓ is an indeterminate which will be chosen optimally later.

Then you get something explicit (and messy) for λ_d. In this case λ_1 is not necessarily normalized to be 1.

GGPY - The Details

The paper then devotes about twenty pages of the paper to finding precise asymptotics for the sums related to those above. Although the estimation is relatively straightforward, the sums involved are somewhat more complicated, as a few tricks are needed at the conclusion to get optimal estimates.

"Only experts worry about the details." -Ken Ono

References

A. C. Cojocaru and M. Ram Murty. *An Introduction to Sieve Methods and their Applications.* Cambridge University Press, 2005.

D.A. Goldston, S.W. Graham, J. Pintz, and C.Y. Yildirim, *Small gaps between primes and almost primes,* preprint.

G. Greaves. *Sieves in Number Theory.* Springer, Berlin, 2001.

H. Halberstam and H.-E. Richert. *Sieve Methods* , Academic Press, New York, 1974.