Small sieves: a consumer's introduction

Paul Pollack

University of British Columbia

June 9, 2011

Variations on a theme of Eratosthenes

List the integers from 2 to x. Take the first uncrossed number and cross out all its proper multiples. Iterate.

$$
\begin{array}{cccccccccc}
 & 2 & 3 & \cancel{4} & 5 & \cancel{6} & 7 & \cancel{8} & \cancel{9} & \cancel{10} \\
11 & \cancel{12} & 13 & \cancel{14} & \cancel{15} & \cancel{16} & 17 & \cancel{18} & 19 & \cancel{20} \\
\cancel{21} & \cancel{22} & 23 & \cancel{24} & \cancel{25} & \cancel{26} & \cancel{27} & \cancel{28} & 29 & \cancel{30}
\end{array}
$$

Variations on a theme of Eratosthenes

List the integers from 2 to x. Take the first uncrossed number and cross out all its proper multiples. Iterate.

$$\begin{array}{cccccccccc} & 2 & 3 & \cancel{4} & 5 & \cancel{6} & 7 & \cancel{8} & \cancel{9} & \cancel{10} \\ 11 & \cancel{12} & 13 & \cancel{14} & \cancel{15} & \cancel{16} & 17 & \cancel{18} & 19 & \cancel{20} \\ \cancel{21} & \cancel{22} & 23 & \cancel{24} & \cancel{25} & \cancel{26} & \cancel{27} & \cancel{28} & 29 & \cancel{30} \end{array}$$

Remaining integers are the primes in $[1, x]$.

Question: How many integers remain? Is this a useful way to get a handle on $\pi(x)$?

How not to prove the prime number theorem

Let $z := \sqrt{x}$. Let $f(p) = 0$ if $p \le z$, and let $f(p) = 1$ if $p > z$. Extend f to a completely multiplicative function. Then for $n > z$,

$$f(n) = \begin{cases} 1 & \text{if } \gcd(n, \prod_{p \le z} p) = 1, \\ 0 & \text{if } \gcd(n, \prod_{p \le z} p) > 1. \end{cases}$$

Hence,

$$\pi(x) = O(\sqrt{x}) + \sum_{n \le x} f(n).$$

How can we estimate the remaining sum?

How not to prove the prime number theorem, ctd.

Rewrite $f(n)$. Let $P = \prod_{p \leq z} p$. Then $f(n)$ is the characteristic function of those integers coprime to p.

Lemma (Fundamental property of the Möbius function)

For $n > 1$,

$$\sum_{d|n} \mu(d) = 0.$$

Proof.

This is equivalent to the Dirichlet series identity $\sum_{n=1}^{\infty} \frac{1}{n^s} \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$, which is in turn equivalent to

$$\left(\sum_{n=1}^{\infty} \frac{1}{n^s} \right) \left(\prod_p (1 - 1/p^s) \right) = 1.$$

How not to prove the prime number theorem, ctd.

Now $f(n) = 1 \iff \gcd(n, P) = 1$. So

$$f(n) = \sum_{d \mid \gcd(n,P)} \mu(d) = \sum_{\substack{d \mid n \\ d \mid P}} \mu(d).$$

Hence,

$$\sum_{n \le x} f(n) = \sum_{n \le x} \sum_{\substack{d \mid n \\ d \mid P}} \mu(d) = \sum_{d \mid P} \mu(d) \sum_{\substack{n \le x \\ d \mid n}} 1 = \sum_{d \mid P} \mu(d) \lfloor x/d \rfloor.$$

We might guess that

$$\begin{aligned} \sum_{d \mid P} \mu(d) \lfloor x/d \rfloor &\approx x \sum_{d \mid P} \mu(d)/d \\ &= x \prod_{p \le z} (1 - 1/p). \end{aligned}$$

Theorem (Mertens)

For $z \geq 2$, we have

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

where $\gamma = 0.57721566\ldots$ is the Euler–Mascheroni constant.

Theorem (Mertens)

For $z \geq 2$, we have

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

where $\gamma = 0.57721566\ldots$ is the Euler–Mascheroni constant.

So our guess is that

$$\pi(x) \approx x \prod_{p \leq z} (1 - 1/p) \sim \frac{x}{e^{\gamma} \log(x^{1/2})} \sim 2e^{-\gamma} \frac{x}{\log x},$$

and $2e^{-\gamma} = 1.12291\ldots$.

Theorem (Mertens)

For $z \geq 2$, we have

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

where $\gamma = 0.57721566\ldots$ is the Euler–Mascheroni constant.

So our guess is that

$$\pi(x) \approx x \prod_{p \leq z} (1 - 1/p) \sim \frac{x}{e^{\gamma} \log(x^{1/2})} \sim 2e^{-\gamma} \frac{x}{\log x},$$

and $2e^{-\gamma} = 1.12291\ldots$.

This is wrong. We know $\pi(x) \sim x/\log x$, as $x \to \infty$.

Theorem (Mertens)

For $z \geq 2$, we have

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

where $\gamma = 0.57721566\ldots$ is the Euler–Mascheroni constant.

So our guess is that

$$\pi(x) \approx x \prod_{p \leq z} (1 - 1/p) \sim \frac{x}{e^{\gamma} \log(x^{1/2})} \sim 2e^{-\gamma} \frac{x}{\log x},$$

and $2e^{-\gamma} = 1.12291\ldots$.

This is wrong. We know $\pi(x) \sim x/\log x$, as $x \to \infty$.
But it's not *that* wrong. We're off by a constant.

What did we really prove?

Let $\Phi(x,z) = \#\{n \le x : p \mid n \Rightarrow p > z\}$. Then with $P := \prod_{p \le z} p$,

$$\begin{aligned}\Phi(x,z) = \sum_{d \mid P} \mu(d) \lfloor x/d \rfloor &= x \sum_{d \mid P} \mu(d)/d + O(\tau(P)) \\ &= x \prod_{p \le z} (1 - 1/p) + O(2^{\pi(z)}).\end{aligned}$$

What did we really prove?

Let $\Phi(x,z) = \#\{n \leq x : p \mid n \Rightarrow p > z\}$. Then with $P := \prod_{p \leq z} p$,

$$\Phi(x,z) = \sum_{d|P} \mu(d)\lfloor x/d \rfloor = x \sum_{d|P} \mu(d)/d + O(\tau(P))$$

$$= x \prod_{p \leq z} (1 - 1/p) + O(2^{\pi(z)}).$$

Assuming $z \to \infty$, "main term" is

$$\sim e^{-\gamma} x / \log z.$$

Error term is essentially exponential in z. So this method gives an asymptotic for $\pi(x,z)$ only if z is very small, e.g., $z \leq \frac{1}{2} \log x$.

What is a small sieve problem?

For all primes $p \leq z$, suppose we are given a set of $\omega(p)$ residue classes modulo p.

Suppose also that $\omega(p) \ll 1$ (the *small sieve* condition).

Problem: Let $\mathscr{S}$ be the largest subset of $[1, x]$ which does not contain any integer from any of the $\omega(p)$ excluded residue classes, for any prime $p \leq z$. Estimate $\#\mathscr{S}$.

More specific problem: Compare $\#\mathscr{A}$ with the naive guess, namely

$$x \prod_{p \leq z} (1 - \omega(p)/p).$$

Estimating $\pi(x, z)$ is an example, where $\omega(p) = 1$.

Why do we care?

Conjecture (Twin prime conjecture)

There are infinitely many pairs of primes $p, p+2$.

Example

Let $\pi_2(x)$ be the number of $p \leq x$ for which $p+2$ is also prime.
Let $\Phi_2(x,z)$ be the number of $n \leq x$ for which n and $n+2$ have no prime factors $\leq z$.

The function $\Phi_2(x,z)$ arises in a small sieve problem:
For all primes $p \leq z$, exclude the residue classes $0 \pmod{p}$ and $-2 \pmod{p}$. So $\omega(2) = 1$ and $\omega(p) = 2$ if $p > 2$. Then

$$\Phi_2(x,z) = \#\mathscr{S}.$$

Observation: For any $z < x$, we have

$$\pi_2(x) \leq z + \Phi_2(x, z).$$

On the other hand, if $z = \sqrt{x+2}$, then

$$\Phi_2(x, z) \leq \pi_2(x).$$

So:

- upper bounds for $\Phi_2(x, z) \Rightarrow$ upper bounds for $\pi_2(x)$
- lower bounds for $\Phi_2(x, \sqrt{x+2}) \Rightarrow$ lower bounds for $\pi_2(x)$.

Observation: For any $z < x$, we have

$$\pi_2(x) \leq z + \Phi_2(x, z).$$

On the other hand, if $z = \sqrt{x+2}$, then

$$\Phi_2(x, z) \leq \pi_2(x).$$

So:

- upper bounds for $\Phi_2(x, z) \Rightarrow$ upper bounds for $\pi_2(x)$
- lower bounds for $\Phi_2(x, \sqrt{x+2}) \Rightarrow$ lower bounds for $\pi_2(x)$.

Almost-prime results: Suppose $z = (x+2)^{1/u}$, and we have a lower bound on $\Phi_2(x, z)$. If n is counted by $\Phi_2(x, z)$, then n and $n+2$ both have at most u prime factors.

Brun's twin primes theorem

Theorem

We have $\sum \frac{1}{p} < \infty$, where p ranges over all primes for which $p+2$ is also prime.

By partial summation, it's enough to show that

$$\pi_2(x) \ll x(\log\log x)^2/(\log x)^2.$$

Brun showed that if $z = x^{\frac{1}{20\log\log x}}$, then

$$\begin{aligned}\Phi_2(x,z) &\sim x\prod_{p\le z}(1-\omega(p)/p)\\ &\sim C\frac{x(\log\log x)^2}{(\log x)^2}.\end{aligned}$$

N.B.: The first asymptotic presumably fails if $z = \sqrt{x+2}$.

Some ideas of the proof.

Let $P = \prod_{p \le z} p$. Clearly,

$$\Phi_2(x, z) = \sum_{n \le x} \chi(n),$$

where $\chi(n)$ is the characteristic function of $n(n+2)$ being prime to P. We can write

$$\chi(n) = \sum_{\substack{d | n(n+2) \\ d | P}} \mu(d).$$

Insert this above and reverse order of summation. Get:

$$\Phi_2(x, z) = \sum_{d | P} \mu(d) \sum_{\substack{n \le x \\ d | n(n+2)}} 1.$$

Let $\omega(d)$ denote the number of solutions n mod d to $n(n+2) \equiv 0 \pmod{d}$. (Multiplicative!) For inner sum,

$$\sum_{\substack{n \le x \\ d | n(n+2)}} 1 = \omega(d)\frac{x}{d} + r_d, \quad |r_d| \le \omega(d).$$

So

$$\begin{aligned} \Phi_2(x, z) &= x \sum_{d|P} \frac{\mu(d)\omega(d)}{d} + \sum_{d|P} \mu(d) r_d \\ &= x \prod_{p \le z} (1 - \omega(p)/p) + O(\sum_{d|P} |r_d|). \end{aligned}$$

Problem: The error sum has too many terms!

Key idea: Replace the Möbius function with an imitator! Suppose we have a sequence λ_d^+ with

$$|\lambda_d^+| \leq 1 \quad \text{and supported on} \quad d \leq y \text{ (say)},$$

with $\lambda_1^+ = 1$ and

$$0 = \sum_{d|n} \mu(d) \leq \sum_{d|n} \lambda_d^+ \quad (\text{for } n > 1).$$

Then

$$\Phi_2(x,z) \leq \sum_{n \leq x} \sum_{\substack{d|n \\ d|P}} \lambda_d^+.$$

Proceeding as before gives

$$\Phi_2(x,z) \leq x \sum_{d|P} \frac{\lambda_d^+ \omega(d)}{d} + O(\sum_{\substack{d|P \\ d \leq y}} |r_d|).$$

Key idea: Replace the Möbius function with an imitator! Suppose we have a sequence λ_d^+ with

$$|\lambda_d^+| \leq 1 \quad \text{and supported on} \quad d \leq y \text{ (say)},$$

with $\lambda_1^+ = 1$ and

$$0 = \sum_{d|n} \mu(d) \leq \sum_{d|n} \lambda_d^+ \quad (\text{for } n > 1).$$

Then

$$\Phi_2(x, z) \leq \sum_{n \leq x} \sum_{\substack{d|n \\ d|P}} \lambda_d^+.$$

Proceeding as before gives

$$\Phi_2(x, z) \leq x \sum_{d|P} \frac{\lambda_d^+ \omega(d)}{d} + O\Big(\sum_{\substack{d|P \\ d \leq y}} |r_d|\Big).$$

Error term has at most y terms now! Trivially $\ll y^{1+\epsilon}$.

The fundamental lemma (twin prime case)

Let $y > 1$. There is a sequence of real numbers $\{\lambda_d^+\}$ supported on $d \leq y$ with the properties that $\lambda_1^+ = 1$ and

$$0 = \sum_{d|n} \mu(d) \leq \sum_{d|n} \lambda_d^+ \quad (\text{for } n > 1),$$

and so that if $P = \prod_{p \leq z} p$,

$$\sum_{d|P} \frac{\lambda_d^+ \omega(d)}{d} = \left(\prod_{p \leq z} \left(1 - \frac{\omega(p)}{p} \right) \right) (1 + O(e^{-s})),$$

where $s = \log y / \log z$.

The fundamental lemma (twin prime case)

Let $y > 1$. There is a sequence of real numbers $\{\lambda_d^+\}$ supported on $d \leq y$ with the properties that $\lambda_1^+ = 1$ and

$$0 = \sum_{d|n} \mu(d) \leq \sum_{d|n} \lambda_d^+ \quad (\text{for } n > 1),$$

and so that if $P = \prod_{p \leq z} p$,

$$\sum_{d|P} \frac{\lambda_d^+ \omega(d)}{d} = \left(\prod_{p \leq z} \left(1 - \frac{\omega(p)}{p}\right)\right) (1 + O(e^{-s})),$$

where $s = \log y / \log z$.

In our application: Take $y = \sqrt{x}$ (say). Second term is $1 + o(1)$ as $s = \frac{1}{2} \frac{\log x}{\log z} \to \infty$.

The fundamental lemma (twin prime case), ctd.

Hence,

$$\Phi_2(x,z) \leq \left(x \prod_{p \leq z} \left(1 - \frac{\omega(p)}{p}\right)\right)(1 + O(e^{-s})) + O(x^{1/2+\epsilon}),$$

where $s = \frac{1}{2}\frac{\log x}{\log z}$. Introducing λ_d^-, we can change "$\leq$" to "$=$".

The fundamental lemma (twin prime case), ctd.

Hence,

$$\Phi_2(x,z) \leq \left(x \prod_{p \leq z} \left(1 - \frac{\omega(p)}{p}\right)\right)(1 + O(e^{-s})) + O(x^{1/2+\epsilon}),$$

where $s = \frac{1}{2}\frac{\log x}{\log z}$. Introducing λ_d^-, we can change "$\leq$" to "$=$".

Write $z = x^{1/u}$. Whenever $z, u \to \infty$ (e.g., $u = 20 \log\log x$),

$$\Phi_2(x,z) \sim x \prod_{p \leq z} (1 - \omega(p)/p) \sim C \frac{x}{(\log z)^2} = u^2 C \frac{x}{(\log x)^2}.$$

Whenever u is large enough, get

$$\Phi_2(x,z) \asymp \prod_{p \leq z} (1 - \omega(p)/p) \asymp \frac{x}{(\log z)^2}.$$

Theorem (Brun)

There are infinitely many n for which both n and $n+2$ have at most 9 prime factors.

Theorem (Brun)

If N is a large even number, then there is a decomposition $N = n_1 + n_2$, where n_1 and n_2 have at most 9 prime factors.

Theorem (Brun)

There are infinitely many n for which both n and $n+2$ have at most 9 prime factors.

Theorem (Brun)

If N is a large even number, then there is a decomposition $N = n_1 + n_2$, where n_1 and n_2 have at most 9 prime factors.

Theorem (Chen)

In both cases, you can make one of the expressions prime and the other a product of at most two primes.

The fundamental lemma (general case)

Let $\kappa > 0$ and let $y > 1$. There are sequences $\{\lambda_d^{\pm}\}$ supported on $d \leq y$ with the properties that $\lambda_1^{\pm} = 1$ and

$$\sum_{d|n} \lambda_d^{-} \leq 0 \leq \sum_{d|n} \lambda_d^{+} \quad (\text{for } n > 1),$$

and so that if g is any multiplicative function with $0 \leq g(p) < 1$ for all primes p satisfying

$$\prod_{w \leq p < z} (1 - g(p))^{-1} \leq (\log z / \log w)^{\kappa} (1 + K / \log w)$$

for all $2 \leq w < z \leq y$, then with $P = \prod_{p \leq z} p$ and $s = \log y / \log z$,

$$\sum_{d|P} \frac{\lambda_d^{+} \omega(d)}{d} = \left(\prod_{p \leq z} \left(1 - \frac{\omega(p)}{p} \right) \right) (1 + O(e^{-s}(1 + K/\log z)^{10})).$$

Brun's sieve (corollary of the fundamental lemma)

In general, we approach asymptotics as $\frac{\log X}{\log z}$ gets larger.

Brun's sieve (corollary of the fundamental lemma)

In general, we approach asymptotics as $\frac{\log X}{\log z}$ gets larger.

Theorem

Suppose we are given $\omega(p)$ residue classes modulo each prime $p \leq z$. Suppose also that

$$\omega(p) \leq C.$$

Let $\mathscr{S} \subset [1, x]$ be the unsieved integers. For $z \leq x$,

$$\#\mathscr{S} \ll_{A,C} x \prod_{p \leq z} (1 - \omega(p)/p).$$

Also, if $z \leq x^A$ and $A > 0$ is sufficiently small (depending on C), then

$$\#S \gg_{A,C} x \prod_{p \leq z} (1 - \omega(p)/p).$$

Application: the Brun–Titchmarsh theorem

Theorem (Prime number theorem for arithmetic progressions)

Fix $A > 0$. If $q \leq (\log x)^A$ and $\gcd(a, q) = 1$, then

$$\begin{aligned}\pi(x; q, a) :&= \#\{p \leq x : p \equiv a \pmod q\} \\ &\sim \frac{1}{\phi(q)} \frac{x}{\log x},\end{aligned}$$

as $x \to \infty$.

We expect this holds in a much wider range of q, e.g. $q \leq x^{1-\epsilon}$.

Application: the Brun–Titchmarsh theorem

Theorem (Prime number theorem for arithmetic progressions)

Fix $A > 0$. If $q \leq (\log x)^A$ and $\gcd(a, q) = 1$, then

$$\begin{aligned}\pi(x; q, a) &:= \#\{p \leq x : p \equiv a \pmod{q}\} \\ &\sim \frac{1}{\phi(q)} \frac{x}{\log x},\end{aligned}$$

as $x \to \infty$.

We expect this holds in a much wider range of q, e.g. $q \leq x^{1-\epsilon}$.

Theorem (Brun–Titchmarsh)

If $q \leq x/2$, then $\pi(x; q, a) \ll \frac{x}{\phi(q) \log \frac{x}{q}}$.

This gives an upper bound of the correct order of magnitude whenever $q \leq x^{1-\epsilon}$.

To prove this, we sieve the integers m from 1 to x/q, for each prime $p \le z = \sqrt{x/q}$ removing those m for which

$$a + mq \equiv 0 \pmod{p}.$$

- If $p \mid q$, then this congruence has no solution, and so we are removing $\omega(p) = 0$ residue classes.
- If $p \nmid q$, then we are removing $\omega(p) = 1$ residue class, namely $-aq^{-1} \pmod{p}$.

By Brun's sieve, the set $\mathscr{S}$ of remaining numbers m is

$$\ll \frac{x}{q} \prod_{p \le \sqrt{x/q}} (1 - \omega(p)/p) = \frac{x}{q} \prod_{p \le \sqrt{x/q}} (1 - 1/p) \prod_{\substack{p \le \sqrt{x/q} \\ p \mid q}} (1 - 1/p)^{-1}$$

$$\ll \frac{1}{q \prod_{p \mid q}(1 - 1/p)} \frac{x}{\log(x/q)} = \frac{x}{\phi(q) \log(x/q)}.$$

If $a + mq \le x$ is prime, then either $a + mq \le \sqrt{x/q}$, or m belongs to the set $\mathscr{S}$ described above. Hence,

$$\pi(x; q, a) \le \#\mathscr{S} + \sqrt{x/q} \ll \frac{x}{\phi(q)\log(x/q)} + \sqrt{x/q}.$$

But

$$\frac{x}{\phi(q)\log(x/q)} \gg \frac{x/q}{\log(x/q)} \gg \sqrt{x/q},$$

so the first term above dominates and we get the theorem.

If $a + mq \le x$ is prime, then either $a + mq \le \sqrt{x/q}$, or m belongs to the set $\mathscr{S}$ described above. Hence,

$$\pi(x; q, a) \le \#\mathscr{S} + \sqrt{x/q} \ll \frac{x}{\phi(q)\log(x/q)} + \sqrt{x/q}.$$

But

$$\frac{x}{\phi(q)\log(x/q)} \gg \frac{x/q}{\log(x/q)} \gg \sqrt{x/q},$$

so the first term above dominates and we get the theorem.

Theorem (Montgomery–Vaughan)

If $\gcd(a, q) = 1$ *and* $x > q$, *then*

$$\pi(x; q, a) \le 2\frac{x}{\phi(q)\log(x/q)}.$$

See Chapter 7 of the book draft for a different proof (when x/q is large enough).

Application: Schnirelmann's theorem

Let $R(N)$ be the number of representations of N as a sum of two primes. In other words,

$$R(N) := \sum_{\substack{p_1, p_2 \\ p_1 + p_2 = N}} 1.$$

Goldbach's conjecture states that $R(N) > 0$ if $N > 4$ is even.

Theorem (Schnirelmann)

The set of N for which $R(N) > 0$ has positive lower density.

Application: Schnirelmann's theorem

Let $R(N)$ be the number of representations of N as a sum of two primes. In other words,

$$R(N) := \sum_{\substack{p_1, p_2 \\ p_1 + p_2 = N}} 1.$$

Goldbach's conjecture states that $R(N) > 0$ if $N > 4$ is even.

Theorem (Schnirelmann)

The set of N for which $R(N) > 0$ has positive lower density.

> *Landau left Brun's manuscript untouched in a drawer for six years, until he saw Schnirelmann's skillful use of it. – Halberstam and Richert, Sieve methods*

Brief sketch of the proof of Schnirelmann's theorem.

We use second moments. By Cauchy–Schwarz,

$$\left(\sum_{\substack{n\leq x\\ R(n)>0}} R(n)\right)^2 \leq \sum_{\substack{n\leq x\\ R(n)>0}} R(n)^2 \sum_{\substack{n\leq x\\ R(n)>0}} 1^2,$$

so that

$$\#\{n \leq x : R(n) > 0\} \geq \frac{\left(\sum_{n\leq x} R(n)\right)^2}{\sum_{n\leq x} R(n)^2}.$$

We need a lower bound on the numerator and an upper bound on the denominator.

Revisiting $\Phi(x, y)$: Applications to mean values of multiplicative functions

If we sieve out the multiples of the the primes $\leq y$ from an interval of length x, the number of remaining integers is

$$x \prod_{p \leq y} (1 - 1/p)(1 + O(1/u^u)),$$

where $y = x^{1/u}$. (Follows from version of fund. lemma.)

In other words, if f is the multiplicative function for which $f(p^k) = 0$ if $p \leq y$ and $f(p^k) = 1$ if $p > y$, then

$$\sum_{n \leq x} f(n) = \left(x \prod_{p \leq y} (1 - 1/p) \right) (1 + O(1/u^u)).$$

Recall the notation

$$\mathcal{P}(f;x) = \prod_{p \le x} \left(1 - \frac{1}{p}\right) \left(1 + \frac{f(p)}{p} + \frac{f(p)}{p^2} + \dots\right).$$

Then the last estimate reads

$$\frac{1}{x} \sum_{n \le x} f(n) = \mathcal{P}(f;x)(1 + O(1/u^u)).$$

Recall the notation

$$\mathcal{P}(f;x) = \prod_{p\le x}\left(1-\frac{1}{p}\right)\left(1+\frac{f(p)}{p}+\frac{f(p)}{p^2}+\dots\right).$$

Then the last estimate reads

$$\frac{1}{x}\sum_{n\le x} f(n) = \mathcal{P}(f;x)(1+O(1/u^u)).$$

Suppose that f is multiplicative with $|f|\le 1$. And suppose f only varies at small primes, in that $f(p^k)=1$ for all $p>y$.

Conjecture (?)

We have

$$\frac{1}{x}\sum_{n\le x} f(n) = \mathcal{P}(f;x)(1+O(1/u^u)).$$

Theorem

If $|f| \leq 1$ and $f(p^k) = 1$ for all $p > y$, then

$$\frac{1}{x} \sum_{n \leq x} f(n) = \mathcal{P}(f; x) + O(1/u^{u/3}).$$

Lemma

If $\Psi(x, y)$ denotes the number of $n \leq x$ all of whose prime factors are $\leq y$ (so-called y-smooth numbers), then in a wide range of x and y,

$$\Psi(x, y) \approx x/u^u, \quad \textit{where} \quad u = \frac{\log x}{\log y}.$$

In fact (Erdős–Canfield–Pomerance),

$$\Psi(x, y) = x/u^{u+o(u)},$$

if $u \to \infty$ and $y \geq (\log x)^{1+\epsilon}$.

Theorem

If $|f| \leq 1$ and $f(p^k) = 1$ for all $p > y$, then

$$\frac{1}{x}\sum_{n\leq x} f(n) = \mathcal{P}(f;x) + O(1/u^{u/3}).$$

Proof.

Define $g(p^k) = 0$ if $p < y$ and $g(p^k) = 1$ if $p > y$. Define $h(p^k) = f(p^k)$ if $p \leq y$ and $h(p^k) = 0$ if $p > y$. Then

$$f = g * h; \quad \text{i.e.} \quad f(n) = \sum_{ab=n} g(a)h(b).$$

Then (cf. Dirichlet's hyperbola method)

$$\sum_{n\leq x} f(n) = \sum_{a\leq\sqrt{x}} h(a) \sum_{b\leq x/a} g(b) + \sum_{b\leq\sqrt{x}} g(b) \sum_{\sqrt{x}<a\leq x/b} h(a).$$

Now using

$$\sum_{n\le x} f(n) = \sum_{a\le\sqrt{x}} h(a) \sum_{b\le x/a} g(b) + \sum_{b\le\sqrt{x}} g(b) \sum_{\sqrt{x}<a\le x/b} h(a),$$

We already have estimated the first inner sum; this gives

$$\sum_{a\le\sqrt{x}} h(a) \sum_{b\le\sqrt{x}} g(b) = \kappa_y x \sum_{a\le\sqrt{x}} \frac{h(a)}{a}(1+O((u/2)^{-u/2})),$$

where $\kappa_y := \prod_{p\le y}(1-1/p)$. Extending the sum on a,

$$\kappa_y \sum_{a} \frac{h(a)}{a} = \mathcal{P}(h;x) = \mathcal{P}(f;x).$$

So the first double sum gives the main term of $x\mathcal{P}(f;x)$.

We have an error which is

$$\ll \kappa_y \sum_{a>\sqrt{x}} \frac{|h(a)|}{a} + (u/2)^{-u/2}\kappa_y x \sum_{a} \frac{|h(a)|}{a}.$$

Notice h is supported on y-smooth numbers; estimates for smooths gives error which is $\ll x(u/2)^{-u/2}$.

It remains to estimate the second double sum:

$$\sum_{b\leq\sqrt{x}} g(b) \sum_{\sqrt{x}<a\leq x/b} h(a).$$

Remaining inner sum is $\leq \#$ of y-smooths up to x/b. Get

$$\ll \sum_{b\leq\sqrt{x}} g(b)\frac{x}{b}(u/2)^{-u/2} \ll x(u/2)^{-u/2} \sum_{b\leq\sqrt{x}} \frac{g(b)}{b} \ll x(u/2)^{1-u/2}$$

Thank you!