Results in multiplicative combinatorial number theory:

Popular values of the largest prime divisor function

Nathan McNew
Dartmouth College
Hanover, New Hampshire

University of Maine
Math Colloquium
October 1st, 2014
Denote by $P(n)$ the largest prime divisor of $n$.

What is the distribution of $P(n)$ for $n \in [2, x]$?

**Mean value**

\[
(1+o(1)) \frac{\pi^2 x}{12 \log x} + \frac{\pi^2 x}{12 \log x} \frac{d_2 x}{\log^2 x} + \ldots + \frac{d_m x}{\log^m x} + O \left( \frac{x}{\log^{m+1} x} \right)
\]

(De Koninck and Ivić, 1984)

Uniformly for all $m$.

\[
d_m = \frac{1}{2^{m+1}} \sum_{j=0}^{m} \frac{(-2)^j \zeta(j)(2)}{j!}.
\]

(Naslund, 2013)
Histogram of $P(n)$ for $n \leq 1,000,000$

Mean: 64937.45
Denote by $P(n)$ the largest prime divisor of $n$.

What is the distribution of $P(n)$ for $n \in [2, x]$?

**Mean value**

$$\frac{\pi^2 x}{12 \log x} + \frac{d_2 x}{\log^2 x} + \ldots + \frac{d_m x}{\log^m x} + O\left(\frac{x}{\log^{m+1} x}\right)$$

**Median value**

$$e^{\frac{\gamma - 1}{\sqrt{e}}} \frac{1}{x^{\sqrt{e}}} \left(1 + O\left(\frac{1}{\log x}\right)\right)$$

(Selfridge and Wunderlich, 1974)

(Naslund, 2013)
Histogram of $P_1(n)$ for $n \leq 1,000,000$

Mean: 64937.45  
Median: 3259
Histogram of $P_1(n)$ for $n \leq 1,000,000$

Mean: 64937.45  Median: 3259
Histogram of $P(n)$ for $n \leq 1,000,000$

Mean: 64937.45  
Median: 3259  
Mode: 73
Distribution of the largest prime divisor

Denote by $P(n)$ the largest prime divisor of $n$.

What is the distribution of $P(n)$ for $n \in [2, x]$?

**Mean value**

\[
\frac{\pi^2 x}{12 \log x} + \frac{d_2 x}{\log^2 x} + \ldots + \frac{d_m x}{\log^m x} + O\left(\frac{x}{\log^{m+1} x}\right)
\]

**Median value**

\[
e^{\frac{\gamma - 1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}} \left(1 + O\left(\frac{1}{\log x}\right)\right)
\]

**Mode**

\[
e^{\sqrt{\frac{1}{2} \log x (\log \log x + \log \log \log x)}} + O\left(\sqrt{\frac{\log x}{\log \log x}}\right)
\]

(de Koninck, 1994)
Counting integers by their largest prime factor

To study the distribution of $P(n)$ we need to count integers up to $x$ whose largest prime divisor is $p$.

Suppose $n \leq x$ and $P(n) = p$, then $\frac{n}{p} \leq \frac{x}{p}$ and $\frac{n}{p}$ is $p$-smooth.

An integer is $y$-smooth (or $y$-friable) if all its prime factors are at most $y$.

Every $p$-smooth number up to $\frac{x}{p}$ can be multiplied by $p$ to produce a unique integer whose largest prime divisor is $p$.

$$\# \left\{ n \leq x : P(n) = p \right\} = \# \left\{ n \leq \frac{x}{p} : P(n) \leq p \right\}$$
Denote by $\Psi(x, y)$ the number of $y$-smooth numbers up to $x$.

Actuary Karl Dickman (1930) showed that for any fixed $u$

$$\lim_{x \to \infty} \frac{1}{x} \Psi(x, x^{1/u}) = \rho(u).$$

$\rho(u)$ satisfies a differential delay equation:

$$u\rho'(u) + \rho(u - 1) = 0$$

$$\rho(u) = 1 \quad 0 \leq u \leq 1$$

$$\rho(u) = 1 - \log u \quad 1 \leq u \leq 2$$

$$\vdots$$
Dickman’s Rho Function

\[ \rho(u) \approx u^{-u} \]
When $1 \leq u \leq 2$, $\rho(u) = 1 - \log u$ and Dickman’s result follows from Mertens’ theorem

$$\sum_{p < x} \frac{1}{p} = \log \log x + B + o(1)$$

**Proof:** If $u < 2$ and $n < x$ has a prime divisor greater than $x^{1/u}$ it is unique. For each $p \in (x^{1/u}, x)$, the number of $n \leq x$ with $p \mid n$ is $\left\lfloor \frac{x}{p} \right\rfloor$.

$$\sum_{x^{1/u} < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor = x \sum_{x^{1/u} < p \leq x} \frac{1}{p} + O \left( \frac{x}{\log x} \right)$$

$$= x \left( \log \log x + B - \log \log x^{1/u} - B + o(1) \right)$$

$$= x \log u + o(x)$$

So the density of integers without such a factor is $1 - \log u$. 
The median largest prime factor \( m \) of integers in \([1, x]\) satisfies:

\[
\frac{1}{x} \psi(x, m) = \frac{1}{2}
\]

\[
\lim_{x \to \infty} \frac{1}{x} \psi(x, x^{1/u}) = \rho(u) \quad \text{(Dickman)}
\]

\[
\rho(u) = 1 - \log u \quad 1 \leq u \leq 2
\]

\[
1 - \log u = 1/2 \quad \Rightarrow u = \sqrt{e}
\]

\[
m = x^{\frac{1}{\sqrt{e}}} + o(1)
\]
De Bruijn (1951) shows that if \( u = \frac{\log x}{\log y} \)

\[
\Psi(x, y) = x \rho(u) \left( 1 + O \left( \frac{\log(u + 1)}{\log y} \right) \right)
\]

as long as \( 1 \leq u \leq (\log y)^{3/5-\epsilon} \) or equivalently \( y > \exp((\log x)^{5/8+\epsilon}) \).

He also gives an estimate for \( \rho(u) \):

\[
\rho(u) = \left( 1 + o(1) \right) \frac{1}{\sqrt{2\pi u}} \exp \left\{ \gamma + u \xi(u) + \int_0^{\xi(u)} \frac{e^s - 1}{s} \, ds \right\}
\]

where \( \xi(u) \) is the positive solution to

\[
e^{\xi(u)} = 1 + u \xi(u).
\]
Hildebrand (1986) shows that De Bruijn’s approximation

\[ \Psi(x, y) = x \rho(u) \left( 1 + O \left( \frac{\log(u + 1)}{\log y} \right) \right) \]

holds for \( \exp((\log \log x)^{5/3+\epsilon}) < y \leq x \).

Alladi (1984) shows that

\[ \rho(u) = \left( 1 + O \left( \frac{1}{u} \right) \right) \sqrt{\frac{\xi'(u)}{2\pi}} \exp \left\{ \gamma + u \xi(u) + \int_0^{\xi(u)} \frac{e^s - 1}{s} \, ds \right\}. \]
Mode of \( \{ P(n) | n \leq X \} \)

Use Hildebrand and Alladi’s results to approximate \( \Psi \left( \frac{x}{p}, p \right) \).

Let \( Q(x) \) be the prime \( p \) which maximizes \( \Psi \left( \frac{x}{p}, p \right) \).

\( Q(x) \) is the most popular large prime divisor on the interval \([2, x]\).

**Theorem**

The mode, \( Q(x) \), satisfies

\[
Q(x) = e^{\sqrt{v(x) \log x} + O((\log \log x)^{1/4})}
\]

where \( v(x) \) is the unique solution to

\[
e^{v(x)} = 1 + \sqrt{v(x) \log x - v(x)^2}.
\]

\[
v(x) = \frac{1}{2} \log \log x + \frac{1}{2} \log \log \log x - \frac{1}{2} \log 2 + o(1)
\]
Consider the largest prime divisor of the first few integers:

|   | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
|   | 2 | 3 | 2 | 5 | 3 | 7 | 2 | 3 | 5  | 11 | 3  | 13 | 7  | 5  | 2  | 17 | 3  |

| Prime | First popular | Last popular |
|-------|---------------|--------------|
| 2     | 2             | 17           |
| 3     | 3             | 119          |
| 5     | 45            | 279          |
| 7     | 70            | 1858         |
| 13    | 1456          | 5471         |

Call a prime $p$ a **Popular Prime** if there exists an $N$ such that

$$p = \text{mode}\{P(n) : n \in [2, N]\}.$$
Histogram of $P(n)$ for $n \leq 1,000,000$

Mean: 64937.45  Median: 3259  Mode: 73
Histogram of $P(n)$ for $n \leq 1,000,000$

Zoomed: 1000x

Mean: 64937.45
Median: 3259
Mode: 73
Most Popular Largest Prime Divisor up to 90 Billion

Nathan McNew (Dartmouth College)

Popular Primes

October 1st, 2014
Most Popular Largest Prime Divisor up to 90 Billion

Nathan McNew (Dartmouth College)
The primes popular for some $x \leq 7 \times 10^{13}$

| Popular prime | First Popular | Count |
|---------------|---------------|-------|
| 2             | 2             | 1     |
| 3             | 12            | 4     |
| 5             | 80            | 12    |
| 7             | 196           | 21    |
| 13            | 1638          | 72    |
| 19            | 4864          | 141   |
| 23            | 22425         | 365   |
| 31            | 46500         | 565   |
| 43            | 109779        | 965   |
| 47            | 158625        | 1224  |
| 73            | 603564        | 2880  |
| 83            | 2552416       | 7490  |
| 109           | 2620142       | 7622  |
| 113           | 2627250       | 7636  |
| 199           | 41163747      | 48358 |
| 283           | 237398795     | 161542|
| 467           | 1966466950    | 697876|
| 661           | 13690729828   | 2760914|
| 773           | 64322158656   | 8354318|
| 887           | 79838739611   | 9754754|
| 1109          | 220355987735  | 20284681|
| 1129          | 232268774850  | 21082413|
| 1327          | 618745972214  | 43030538|
| 1627          | 1882062476406 | 96835109|
| 2143          | 9607713772982 | 318536261|
| 2399          | 19364051829855| 534252391|
| 2477          | 26393150937218| 672026919|
| 2803          | 37636607806688| 873944931|
Which primes are popular?

Write: \( p_n = \text{nth prime} \)
\( d_n = p_{n+1} - p_n \) \( (\text{nth prime gap}) \)

**Theorem**

*For sufficiently large \( n \), if \( d_{n-1} > d_n \), then \( p_n \) is not popular.*

**Theorem**

*If \( p_n \) and \( p_{n+k} \) are both popular primes then they satisfy*

\[
\frac{p_{n+k} - p_n}{k} = \log p_{n+k} + O(\log \log p_{n+k}).
\]
How many popular primes are there?

Very average spacing + Brun’s Sieve:

**Theorem**

The number of popular primes up to $x$ is at most $\frac{x}{(\log x)^{4/3+o(1)}}$.

By the prime number theorem there are about $\frac{x}{\log x}$ primes up to $x$. However unlike the primes...

**Corollary**

The sum of the reciprocals of the popular primes converges.

Recall $Q(x) = e^{\sqrt{\nu(x) \log x} + O((\log \log x)^{1/4})}$.

**Theorem**

The number of popular primes up to $x$ is asymptotically at least $\frac{\log x}{(\log \log x)^{1/4}}$.
An application: factoring

A critical step in many factoring algorithms (Dixon’s random squares, quadratic sieve, number field sieve...) is to generate integers $a_1, a_2 \ldots$ such that $b_i \equiv a_i^2 \pmod{n}$ is $y$-smooth until some product of the $a_i$ is a square mod $n$. (Square dependence)

How to pick the smoothness bound $y$?

- Expected number of trials to pick a $y$-smooth integer is $\frac{x}{\Psi(x,y)}$.
- Having $\pi(y) + 1$ integers forces a square dependence.

Optimal value of $y$ minimizes the expression $\frac{x\pi(y)}{\Psi(x,y)} \approx \frac{xy}{\Psi(x,y)}$. (Maximizes $\frac{\Psi(x,y)}{y}$.)
Distribution of $\Psi(x, y)/y$ for $x = 1,000,000$
\( \psi(x, y)/y \) for \( x = 1,000,000 \)

Peak: 113
Croot, Granville, Pemantle and Tetali showed (2008, Annals of Math) that the optimal smoothness bound $y_0$, which maximizes $\Psi(x, y)/y$, satisfies

$$y_0 = e^{\sqrt{\frac{1}{2} \log x (\log \log x + \log \log \log x - \log 2 + o(1))}}.$$ 

**Future work:** Compare the set of “Popular Primes” to the set of those “Fast Primes” $p$ which minimize $\Psi(x, p)/p$ for some value of $x$. 
Consider the collection of points \((n, p_n)\) in \(\mathbb{R}^2\). (The prime number graph) Generally (but somewhat erratically) curve upward, \(p_n \sim n \log n\).

Take the convex hull of these points. Say that a prime \(p_n\) is a convex prime if \((n, p_n)\) is a vertex point of this convex hull.
The prime number graph
The prime number graph
\( p_n \) is a convex prime if and only if for every \( j < n \) and \( k > n \) the line segment connecting \((j, p_j)\) to \((k, p_k)\) passes above the point \((n, p_n)\).

A weaker requirement would ask that for each \( i < n \) the line segment from \((n-i, p_{n-i})\) to \((n+i, p_{n+i})\) pass through or above the point \((p_n, n)\). Call such primes **midpoint convex**.

\[
\forall i < n \quad p_{n-i} + p_{n+i} \geq 2p_n
\]

Every convex prime is midpoint convex. **Pomerance** uses this and the geometry of the prime number graph to prove there are infinitely many of each.

Discussed in problem **A14** in **Guy**’s Unsolved Problems in Number Theory
The first few popular primes are:
   2, 3, 5, 7, 13, 19, 23, 31, 43, 47, 73, 83, 109, 113, 199, 283, 467, 661, 773, 887, 1109, 1129, 1327, 1627, 2143, 2399, 2477, 2803

The first few convex primes are:
   2, 3, 7, 19, 47, 73, 113, 199, 283, 467, 661, 887, 1129, 1327, 1627, 2803

Thus far every convex prime is a popular prime.
(Also 5, 13, 23, 31, 43, lie on the convex hull but are not vertex points.)

Every known popular prime is midpoint convex besides 773.
Popular primes must satisfy $p_{n-1} + p_{n+1} > 2p_n$. 
The count of the convex primes up to $x$ is at least $\exp\left(c\left(\log x\right)^{3/5}\right)$ for some constant $c$. (Pomerance)

Assuming the Riemann hypothesis this can be improved to $x^{1/4}/\log^{3/2} x$.

A result of Erdős and Prachar implies that the count is $o(\pi(x))$.

**Theorem**

The count of the convex primes up to $x$ is $O\left(\frac{x^{2/3}}{\log^{2/3} x}\right)$. 

Nathan McNew (Dartmouth College)

Popular Primes

October 1st, 2014
Theorem

The count of the convex primes up to $x$ is $O\left(\frac{x^{2/3}}{\log^{2/3} x}\right)$.

Proof:

Using the prime number theorem we can show that the slope of the convex hull following a vertex point $(n, p_n)$ is $\log n + \log \log n - 1 + o(1))$.

Let $p_{i_1}, p_{i_2}, \ldots$ be the consecutive convex primes in the interval $(\frac{1}{2}x, x]$.

The slopes between consecutive convex primes are increasing rational numbers

$$s_j = \frac{p_{i_{j+1}} - p_{i_j}}{i_{j+1} - i_j}.$$

Because $p_{i_j} \in (\frac{1}{2}x, x]$, we have $\frac{1}{2} \pi(x) < \pi \left(\frac{1}{2}x\right) < i_j \leq \pi(x)$ so each $p_{i_j}$ is contained in some interval of length $\log 2 + o(1)$. 
The count of the convex primes up to $x$ is $O \left( \frac{x^{2/3}}{\log^{2/3} x} \right)$.

Proof Continued:

$$s_j = \frac{p_{i_{j+1}} - p_{i_j}}{i_{j+1} - i_j}$$

$s_j < s_{j+1} < \ldots$ and all are contained in an interval of length $\log 2 + o(1)$.

$i_j - i_{j-1} = k \implies$ at most $k(\log 2 + o(1))$ possible values of $p_{i_j} - p_{i_{j-1}}$.

For each $k$, there are $O(k)$ convex primes $p_{i_j}$ with $i_j - i_{j-1} = k$, or $O(K^2)$ convex primes which follow a gap of at most $K$ convex primes.

The number of consecutive convex primes with $i_j - i_{j-1} > K$ is $O \left( \frac{x}{K \log x} \right)$.

Optimizing $K$ we find $K = (x/\log x)^{1/3}$. So the count of convex primes in $(\frac{1}{2}x, x]$ is $O \left( \frac{x^{2/3}}{\log^{2/3} x} \right)$. Sum dyadically to complete the proof.
The most popular $k$th-largest prime divisor ($k \geq 2$) is always 3. (de Koninck)

The most common largest part of factorizations in more general settings. (Rings of integers in a number field, polynomials over a finite field...)

When factoring polynomials of degree $n$ over $\mathbb{F}_q$, which polynomials (degrees) show up most often as the irreducible factor with highest degree?

All polynomials with the same degree occur with the same frequency: No analogue of a popular polynomial.
A partition of a positive integer $n$ is a way of writing $n$ as a sum of positive integers. (Order doesn’t matter)

For example: $7 = 3 + 2 + 1 + 1$

Over all partitions of $n$, what is the most popular largest component?

Let $K(n)$ denote the most popular largest component of the partitions of $n$.

Theorem (Auluck, Chowla, Gupta, 1942)

$$n^{1/2} < K(n) < \frac{\sqrt{6}}{\pi} n^{1/2} \log n^{1/2}$$
Additive factorizations: partitions

Let $K(n)$ denote the most popular largest component among the partitions of $n$.

**Theorem (Erdős, 1946)**

$$K(n) = \frac{\sqrt{6}}{\pi} n^{1/2} \log \left( \frac{\sqrt{6}}{\pi} n^{1/2} \right) + o(n^{1/2})$$

**Theorem (Szekeres, 1953)**

$K(n)$ is the closest integer to

$$\frac{\sqrt{6}}{2\pi} n^{1/2} \log \left( \frac{6}{\pi^2} n \right) - \frac{6}{\pi^2} \left( \frac{1}{16} \log^2 \left( \frac{6}{\pi^2} n \right) - \frac{3}{4} \log \left( \frac{6}{\pi^2} n \right) - \frac{3}{2} \right) - \frac{1}{2} + O \left( \frac{\log^4 n}{n^4} \right)$$
Thank you