The distribution of intermediate prime factors

Nathan McNew
Towson University

Joint work with Paul Pollack and Akash Singha Roy (UGA)

33èmes Journées Arithmétiques
University of Luxembourg
July 3rd, 2025

Extreme prime factors

The study of the largest and smallest prime factors of integers has a rich history, motivated by the analysis of factoring algorithm run times.

Extreme prime factors

The study of the largest and smallest prime factors of integers has a rich history, motivated by the analysis of factoring algorithm run times.

Let $P^{+}(n)$ and $P^{-}(n)$ denote the largest and smallest prime factors of n.

Extreme prime factors

The study of the largest and smallest prime factors of integers has a rich history, motivated by the analysis of factoring algorithm run times.

Let $P^+(n)$ and $P^-(n)$ denote the largest and smallest prime factors of n.

What can we say about the distribution of these functions for $n \in [2, x]$?

Extreme prime factors

The study of the largest and smallest prime factors of integers has a rich history, motivated by the analysis of factoring algorithm run times.

Let $P^+(n)$ and $P^-(n)$ denote the largest and smallest prime factors of n.

What can we say about the distribution of these functions for $n \in [2, x]$?

These questions are closely related to smooth/rough numbers: An integer is y-smooth (rough) if all of its prime factors are no larger (smaller) than y.

The distribution of the largest prime factor

Mean value of $P^+(n), n \in [2, x]$

The distribution of the largest prime factor

Mean value of $P^+(n), n \in [2, x]$

$$\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \frac{\pi^2 x}{12\log x}$$ (Knuth, Pardo, 1976,

The distribution of the largest prime factor

Mean value of $P^+(n), n \in [2, x]$

$$\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \frac{\pi^2 x}{12\log x}$$ (Knuth, Pardo, 1976, Erdős, Alladi, 1977, Kemeny, 1993)

The distribution of the largest prime factor

Mean value of $P^+(n), n \in [2, x]$

$$\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \frac{\pi^2 x}{12\log x}$$ (Knuth, Pardo, 1976, Erdős, Alladi, 1977, Kemeny, 1993)

In fact

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

Largest Prime Factor

Mean value:

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

Largest Prime Factor

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:

Largest Prime Factor

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: $$x^{\frac{1}{\sqrt{e}}+o(1)}$$ (Selfridge and Wunderlich, 1974)

Largest Prime Factor

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:

$$x^{\frac{1}{\sqrt{e}}+o(1)}$$ (Selfridge and Wunderlich, 1974)

$$e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}}\left(1+O\left(\frac{1}{\log x}\right)\right)$$ (Naslund, 2013)

Largest Prime Factor

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:
$$x^{\frac{1}{\sqrt{e}}+o(1)}$$
(Selfridge and Wunderlich, 1974)

$$e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}}\left(1 + O\left(\frac{1}{\log x}\right)\right)$$
(Naslund, 2013)

Mode:

Largest Prime Factor

Mean value:

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

Median value:

$$x^{\frac{1}{\sqrt{e}}+o(1)}$$ (Selfridge and Wunderlich, 1974)

$$e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}}\left(1+O\left(\frac{1}{\log x}\right)\right)$$ (Naslund, 2013)

Mode:

$$e^{\sqrt{\frac{1}{2}\log x(\log_2 x+\log_3 x)}+O\left(\sqrt{\frac{\log x}{\log_2 x}}\right)}$$ (De Koninck, 1994)

Largest Prime Factor

Mean value:

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

Median value:

$$x^{\frac{1}{\sqrt{e}}+o(1)}$$ (Selfridge and Wunderlich, 1974)

$$e^{\frac{\gamma-1}{\sqrt{e}}} x^{\frac{1}{\sqrt{e}}}\left(1+O\left(\frac{1}{\log x}\right)\right)$$ (Naslund, 2013)

Mode:

$$e^{\sqrt{\frac{1}{2}\log x(\log_2 x+\log_3 x)}+O\left(\sqrt{\frac{\log x}{\log_2 x}}\right)}$$ (De Koninck, 1994)

$$\exp\left(\sqrt{\nu(x)\log x}+\frac{1}{4}-\frac{\nu(x)-3}{8\nu(x)^2-12\nu(x)+4}\right)\left(1+O\left(\left(\frac{\log\log x}{\log x}\right)^{1/4}\right)\right)$$ (M, 2015)

where $\nu(x)$ is defined implicitly as the solution to $e^{\nu(x)} = 1+\sqrt{\nu(x)\log x}-\nu(x)$.

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\leq n\leq x} P^{-}(n) \sim \frac{x}{2\log x}$$

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\leq n\leq x} P^-(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\leq x} p$$

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\le x} p$$

Median:

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\le x} p$$

Median: 2

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\leq n\leq x} P^{-}(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\leq x} p$$

Median: 2 (or 2.5)

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\le x} p$$

Median: 2 (or 2.5)

Mode:

Smallest Prime Factor

Mean value:

$$\frac{1}{x}\sum_{2\leq n\leq x} P^-(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{p\leq x} p$$

Median: 2 (or 2.5)

Mode: 2

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

$$\frac{1}{x}\sum_{2\le n\le x}\log P^+(n)\sim\lambda\log x \qquad \lambda=\int_0^1 e^{\mathrm{Li}(t)}\mathrm{d}t=0.624329\ldots$$ (Golomb-Dickman constant)

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

$$\frac{1}{x}\sum_{2\le n\le x}\log P^+(n)\sim\lambda\log x \qquad \lambda=\int_0^1 e^{\mathrm{Li}(t)}\mathrm{d}t=0.624329\ldots \text{ (Golomb-Dickman constant)}$$

- On average $P^+(n)$ has about $5/8{=}0.625$ as many digits as n.

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

$$\frac{1}{x}\sum_{2\le n\le x}\log P^+(n)\sim\lambda\log x \qquad \lambda=\int_0^1 e^{\mathrm{Li}(t)}\mathrm{d}t=0.624329\ldots$$ (Golomb-Dickman constant)

- On average $P^+(n)$ has about $5/8{=}0.625$ as many digits as n.
- The geometric mean of the largest prime factors is $\left(\prod_{2\le n\le x}P^+(n)\right)^{\frac{1}{x}}=x^{\lambda+o(1)}$.

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

$$\frac{1}{x}\sum_{2\le n\le x}\log P^+(n) \sim \lambda \log x \qquad \lambda = \int_0^1 e^{\mathrm{Li}(t)}\mathrm{d}t = 0.624329\ldots \text{ (Golomb-Dickman constant)}$$

- On average $P^+(n)$ has about $5/8{=}0.625$ as many digits as n.
- The geometric mean of the largest prime factors is $\left(\prod_{2\le n\le x} P^+(n)\right)^{\frac{1}{x}} = x^{\lambda+o(1)}$.

For smallest prime factors Tennenbaum shows: $$\frac{1}{x}\sum_{2\le n\le x}\log P^-(n) \sim e^{-\gamma}\log_2 x$$

Average size

Average size: Knuth and Trabb Pardo (1976) consider the average size of the largest prime factors as part of an analysis of factoring algorithm runtimes. (Size of n is $\log n$.)

$$\frac{1}{x}\sum_{2\le n\le x}\log P^+(n)\sim\lambda\log x \qquad \lambda=\int_0^1 e^{\mathrm{Li}(t)}\mathrm{d}t=0.624329\ldots \text{ (Golomb-Dickman constant)}$$

- On average $P^+(n)$ has about $5/8{=}0.625$ as many digits as n.
- The geometric mean of the largest prime factors is $\left(\prod_{2\le n\le x}P^+(n)\right)^{\frac{1}{x}}=x^{\lambda+o(1)}$.

For smallest prime factors Tennenbaum shows: $\displaystyle\frac{1}{x}\sum_{2\le n\le x}\log P^-(n)\sim e^{-\gamma}\log_2 x$

Connection to permutations: On average the longest cycle in a permutation of length n is λn (Golomb, 1964) and the average shortest cycle is $e^{-\gamma}\log n$ (Shepp, Lloyd, 1966).

Other distributional properties of $P^+(n)$

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod{q}\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod{q}\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Benford's leading digit law: Fix a base b and an initial string of digits D.

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a,q)=1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod{q}\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Benford's leading digit law: Fix a base b and an initial string of digits D.

$$\#\{n \le x : P^+(n) \text{ starts with } D \text{ in base } b\} \sim \frac{x \log(1+D^{-1})}{\log b}$$ (Pollack, Singha Roy, 2022)

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod q\} \sim \frac{x}{\varphi(q)} \qquad \text{(Banks, Harmon, Shparlinski, 2005)}$$

Benford's leading digit law: Fix a base b and an initial string of digits D.

$$\#\{n \le x : P^+(n) \text{ starts with } D \text{ in base } b\} \sim \frac{x \log(1+D^{-1})}{\log b} \qquad \text{(Pollack, Singha Roy, 2022)}$$

Both also hold for the kth largest prime factor $\left(q \le \frac{\log x}{(\log_2 x)^{k-1+\epsilon}}\right)$.

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod{q}\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Benford's leading digit law: Fix a base b and an initial string of digits D.

$$\#\{n \le x : P^+(n) \text{ starts with } D \text{ in base } b\} \sim \frac{x \log(1+D^{-1})}{\log b}$$ (Pollack, Singha Roy, 2022)

Both also hold for the kth largest prime factor $\left(q \le \frac{\log x}{(\log_2 x)^{k-1+\epsilon}}\right)$. Neither holds for $P^-(n)$.

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod q\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Benford's leading digit law: Fix a base b and an initial string of digits D.

$$\#\{n \le x : P^+(n) \text{ starts with } D \text{ in base } b\} \sim \frac{x \log(1+D^{-1})}{\log b}$$ (Pollack, Singha Roy, 2022)

Both also hold for the kth largest prime factor $\left(q \le \frac{\log x}{(\log_2 x)^{k-1+\epsilon}}\right)$. Neither holds for $P^-(n)$.

What about "in between" these extremes?

Other distributional properties of $P^+(n)$

Equidistribution in residue classes: If $q \le \log^A x$ (for fixed A) and $(a, q) = 1$ then

$$\#\{n \le x : P^+(n) \equiv a \pmod{q}\} \sim \frac{x}{\varphi(q)}$$ (Banks, Harmon, Shparlinski, 2005)

Benford's leading digit law: Fix a base b and an initial string of digits D.

$$\#\{n \le x : P^+(n) \text{ starts with } D \text{ in base } b\} \sim \frac{x \log(1+D^{-1})}{\log b}$$ (Pollack, Singha Roy, 2022)

Both also hold for the kth largest prime factor $\left(q \le \frac{\log x}{(\log_2 x)^{k-1+\epsilon}}\right)$. Neither holds for $P^-(n)$.

What about "in between" these extremes? ...the middle prime factor?

The middle prime factor

De Koninck and Luca (2013), initiated the study of “middle” prime factors.

The middle prime factor

De Koninck and Luca (2013), initiated the study of "middle" prime factors. Suppose

$$n = p_1 p_2 \cdots p_{\Omega(n)} \qquad p_1 \le p_2 \le \cdots \le p_{\Omega(n)}$$

The middle prime factor

De Koninck and Luca (2013), initiated the study of "middle" prime factors. Suppose

$$n = p_1 p_2 \cdots p_{\Omega(n)} \qquad p_1 \le p_2 \le \cdots \le p_{\Omega(n)}$$

Define the middle prime factor $P^{(\frac{1}{2})}(n) := p_{\lceil \Omega(n)/2 \rceil}$.

The middle prime factor

De Koninck and Luca (2013), initiated the study of "middle" prime factors. Suppose

$$n = p_1 p_2 \cdots p_{\Omega(n)} \qquad p_1 \le p_2 \le \cdots \le p_{\Omega(n)}$$

Define the middle prime factor $P^{(\frac{1}{2})}(n) := p_{\lceil \Omega(n)/2 \rceil}$.

Note that when an integer has an even number of prime factors we take the smaller of the two "middle" primes.

The middle prime factor

De Koninck and Luca (2013), initiated the study of "middle" prime factors. Suppose

$$n = p_1 p_2 \cdots p_{\Omega(n)} \qquad p_1 \le p_2 \le \cdots \le p_{\Omega(n)}$$

Define the middle prime factor $P^{(\frac{1}{2})}(n) := p_{\lceil \Omega(n)/2 \rceil}$.

Note that when an integer has an even number of prime factors we take the smaller of the two "middle" primes. $P^{(\frac{1}{2})}(140) = P^{(\frac{1}{2})}(2 \cdot 2 \cdot 5 \cdot 7) = 2$.

The middle prime factor

De Koninck and Luca (2013), initiated the study of "middle" prime factors. Suppose

$$n = p_1 p_2 \cdots p_{\Omega(n)} \qquad p_1 \le p_2 \le \cdots \le p_{\Omega(n)}$$

Define the middle prime factor $P^{(\frac{1}{2})}(n) := p_{\lceil \Omega(n)/2 \rceil}$.

Note that when an integer has an even number of prime factors we take the smaller of the two "middle" primes. $P^{(\frac{1}{2})}(140) = P^{(\frac{1}{2})}(2 \cdot 2 \cdot 5 \cdot 7) = 2$.

We can generalize this to any percentile. For $\alpha \in (0,1)$ define $P^{(\alpha)} := p_{\lceil \alpha \Omega(n) \rceil}$

The distribution of the intermediate prime factor

For any $\alpha \in (0,1)$,

- Mean: $\frac{1}{x}\sum_{2\le n\le x} P^{(\alpha)}(n) \sim \frac{x}{2\log x}$

The distribution of the intermediate prime factor

For any $\alpha \in (0,1)$,

- Mean: $\frac{1}{x}\sum_{2\le n\le x} P^{(\alpha)}(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{2\le n\le x} P^{-}(n) \sim \frac{1}{x}\sum_{p\le x} p$

The distribution of the intermediate prime factor

For any $\alpha \in (0,1)$,

- Mean: $\frac{1}{x} \sum_{2 \le n \le x} P^{(\alpha)}(n) \sim \frac{x}{2\log x} \sim \frac{1}{x} \sum_{2 \le n \le x} P^{-}(n) \sim \frac{1}{x} \sum_{p \le x} p$
- Mode: 2

The distribution of the intermediate prime factor

For any $\alpha \in (0,1)$,

- Mean: $\frac{1}{x}\sum_{2\le n\le x} P^{(\alpha)}(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{2\le n\le x} P^{-}(n) \sim \frac{1}{x}\sum_{p\le x} p$
- Mode: 2
- $\log\log P^{(\alpha)}(n)$ has normal order $\alpha \log\log n$. (De Koninck, Katai, 2014) $P^{(\frac{1}{2})}(n) \approx e^{\sqrt{\log n}}$

The distribution of the intermediate prime factor

For any $\alpha \in (0,1)$,

- Mean: $\frac{1}{x}\sum_{2\le n\le x} P^{(\alpha)}(n) \sim \frac{x}{2\log x} \sim \frac{1}{x}\sum_{2\le n\le x} P^{-}(n) \sim \frac{1}{x}\sum_{p\le x} p$
- Mode: 2
- $\log\log P^{(\alpha)}(n)$ has normal order $\alpha \log\log n$. (De Koninck, Katai, 2014) $P^{(\frac{1}{2})}(n) \approx e^{\sqrt{\log n}}$
- $\log_2 P^{(\alpha)}(n)$ follows a normal distribution. (De Koninck, Doyon, Ouellet, 2019)

$$\frac{1}{x}\#\left\{n \le x : \frac{\log_2 P^{(\alpha)}(n) - \alpha \log_2 x}{\sqrt{\log_2 x}} < t\right\} = \Phi(2t) + O_\alpha\left(\frac{1}{\sqrt{\log_3 x}}\right)$$

for $|t| \ll (\log_2 x)^{1/8-\epsilon}$. $\Phi(t) = \int_{-\infty}^{t} e^{-u^2/2} du$ is the Gaussian distribution function.

The distribution of the middle prime factor

Does $P^{(\alpha)}(n)$ follow Benford's law? Is it equidistributed in coprime residue classes?

The distribution of the middle prime factor

Does $P^{(\alpha)}(n)$ follow Benford's law? Is it equidistributed in coprime residue classes? **Yes!**

The distribution of the middle prime factor

Does $P^{(\alpha)}(n)$ follow Benford's law? Is it equidistributed in coprime residue classes? **Yes!**

Theorem (M.,P.,S., $P^{(\alpha)}(n)$ follows Benford's law)

Fix an integer $b \geq 2$, and fix a positive integer D. The set of n for which $P^{(\alpha)}(n)$ begins with D in base b has asymptotic density $\log(1 + D^{-1})/\log b$.

The distribution of the middle prime factor

Does $P^{(\alpha)}(n)$ follow Benford's law? Is it equidistributed in coprime residue classes? **Yes!**

Theorem (M.,P.,S., $P^{(\alpha)}(n)$ follows Benford's law)

Fix an integer $b \geq 2$, and fix a positive integer D. The set of n for which $P^{(\alpha)}(n)$ begins with D in base b has asymptotic density $\log(1 + D^{-1})/\log b$.

Theorem (M.,P.,S., $P^{(\alpha)}(n)$ is equidistributed in residue classes)

The number of $n \leq x$ for which $P^{(\frac{1}{2})}(n) \equiv a \pmod q$ is $(1 + o(1))x/\phi(q)$ as $x \to \infty$, uniformly in moduli $q \leq (\log x)^{\frac{1}{2}-\epsilon}$ and in coprime residue classes a mod q.

The distribution of the middle prime factor

Does $P^{(\alpha)}(n)$ follow Benford's law? Is it equidistributed in coprime residue classes? **Yes!**

Theorem (M.,P.,S., $P^{(\alpha)}(n)$ follows Benford's law)

Fix an integer $b \geq 2$, and fix a positive integer D. The set of n for which $P^{(\alpha)}(n)$ begins with D in base b has asymptotic density $\log(1 + D^{-1})/\log b$.

Theorem (M.,P.,S., $P^{(\alpha)}(n)$ is equidistributed in residue classes)

The number of $n \leq x$ for which $P^{(\frac{1}{2})}(n) \equiv a \pmod{q}$ is $(1 + o(1))x/\phi(q)$ as $x \to \infty$, uniformly in moduli $q \leq (\log x)^{\frac{1}{2}-\epsilon}$ and in coprime residue classes a mod q.

The uniformity is essentially optimal: we can show it fails for $q > (\log x)^{\frac{1}{2}+\epsilon}$.

Proof Methods

We found two different methods of proving each of these results.

Proof Methods

We found two different methods of proving each of these results.

Method 1: General result about $P^{(\alpha)}(n)$ in sets of primes with logarithmic weights.

Proof Methods

We found two different methods of proving each of these results.

Method 1: General result about $P^{(\alpha)}(n)$ in sets of primes with logarithmic weights.

Theorem

Let $\mathcal{P}$ be any subset of the primes such that whenever X, Y, and $Y/X \to \infty$,

$$\sum_{\substack{p \in (X,Y] \\ p \in \mathcal{P}}} \frac{1}{p} \sim \nu \sum_{p \in (X,Y]} \frac{1}{p}$$

for a fixed $\nu \in (0,1]$. Then the set of n with $P^{(\alpha)}(n) \in \mathcal{P}$ has asymptotic density ν.

Proof Methods

We found two different methods of proving each of these results.

Method 1: General result about $P^{(\alpha)}(n)$ in sets of primes with logarithmic weights.

Theorem

Let $\mathcal{P}$ be any subset of the primes such that whenever X, Y, and $Y/X \to \infty$,

$$\sum_{\substack{p\in(X,Y]\\ p\in\mathcal{P}}} \frac{1}{p} \sim \nu \sum_{p\in(X,Y]} \frac{1}{p}$$

for a fixed $\nu \in (0,1]$. Then the set of n with $P^{(\alpha)}(n) \in \mathcal{P}$ has asymptotic density ν.

Primes starting with D in base b and those $p \equiv a \pmod q$ for fixed q have this property.

Proof Idea

For most integers n, $P^{(\frac{1}{2})}(n)$ isn't too close to its neighbors in the factorization of n.

Proof Idea

For most integers n, $P^{(\frac{1}{2})}(n)$ isn't too close to its neighbors in the factorization of n.

Suppose $p = P^{(\frac{1}{2})}(n)$, $n = ApB$ with $P^+(A) \leq p \leq P^-(B)$.

Proof Idea

For most integers n, $P^{(\frac{1}{2})}(n)$ isn't too close to its neighbors in the factorization of n.

Suppose $p = P^{(\frac{1}{2})}(n)$, $n = ApB$ with $P^+(A) \le p \le P^-(B)$. "Typically" $\frac{P^-(B)}{P^+(A)} > \log_3 x$.

Proof Idea

For most integers n, $P^{(\frac{1}{2})}(n)$ isn't too close to its neighbors in the factorization of n.

Suppose $p = P^{(\frac{1}{2})}(n)$, $n = ApB$ with $P^+(A) \le p \le P^-(B)$. "Typically" $\frac{P^-(B)}{P^+(A)} > \log_3 x$.

This gives us wiggle room to average over all of the primes between $P^+(A)$ and $P^-(B)$.

Proof Idea

For most integers n, $P^{(\frac{1}{2})}(n)$ isn't too close to its neighbors in the factorization of n.

Suppose $p = P^{(\frac{1}{2})}(n)$, $n = ApB$ with $P^+(A) \leq p \leq P^-(B)$. "Typically" $\frac{P^-(B)}{P^+(A)} > \log_3 x$.

This gives us wiggle room to average over all of the primes between $P^+(A)$ and $P^-(B)$.

For fixed A and B, we can multiply AB by any prime $p \in [P^+(A), P^-(B)]$ and that prime will be the middle prime factor.

The number of integers whose middle prime factor is p

Method 2: Count the number $M_p^{(\frac{1}{2})}(x)$ of integers up to x with middle prime factor p.

The number of integers whose middle prime factor is p

Method 2: Count the number $M_p^{(\frac{1}{2})}(x)$ of integers up to x with middle prime factor p. It will be useful to define $\beta := \frac{\log_2 p}{\log_2 x}$.

The number of integers whose middle prime factor is p

Method 2: Count the number $M_p^{\left(\frac{1}{2}\right)}(x)$ of integers up to x with middle prime factor p. It will be useful to define $\beta := \frac{\log_2 p}{\log_2 x}$. So $p = e^{(\log x)^\beta}$.

The number of integers whose middle prime factor is p

Method 2: Count the number $M_p^{(\frac{1}{2})}(x)$ of integers up to x with middle prime factor p. It will be useful to define $\beta := \frac{\log_2 p}{\log_2 x}$. So $p = e^{(\log x)^\beta}$.

Theorem

Suppose $p \to \infty$ is prime and either $\beta < \frac{1}{5} - \epsilon$ or $\frac{1}{5} + \epsilon < \beta < 1 - \epsilon$. Then we have

$$M_p^{(\frac{1}{2})}(x) \sim \begin{cases} \dfrac{C_\beta x}{p(\log x)^{1-2\sqrt{\beta(1-\beta)}}\sqrt{\log_2 x}} & \frac{1}{5} + \epsilon < \beta < 1 - \epsilon, \\[2ex] \dfrac{Cx}{p(\log x)^{\frac{1}{2}-\frac{3}{2}\beta}} & 0 < \beta < \frac{1}{5} - \epsilon, \end{cases}$$

The constants

The constants

The constants in the previous theorem are given by

$$C_\beta := \frac{\exp\left(\frac{\gamma(1-2\beta)}{\sqrt{\beta(1-\beta)}}\right)}{\Gamma\left(1+\sqrt{\frac{\beta}{1-\beta}}\right)} \frac{\sqrt{\beta}+\sqrt{1-\beta}}{2\sqrt{\pi}\beta^{1/4}(1-\beta)^{3/4}} \prod_{q \text{ prime}} \left(1-\frac{1}{q}\right)^{\sqrt{\frac{1-\beta}{\beta}}} \left(1-\frac{\sqrt{\frac{1-\beta}{\beta}}}{q}\right)^{-1}$$

$$C := \frac{3e^{\frac{3\gamma}{2}}}{4\sqrt{\pi}} \prod_{q>2 \text{ prime}} \left(1+\frac{1}{q(q-2)}\right) = 1.523555\ldots.$$

The constants

The constants in the previous theorem are given by

$$C_\beta := \frac{\exp\left(\frac{\gamma(1-2\beta)}{\sqrt{\beta(1-\beta)}}\right)}{\Gamma\left(1+\sqrt{\frac{\beta}{1-\beta}}\right)} \frac{\sqrt{\beta}+\sqrt{1-\beta}}{2\sqrt{\pi}\beta^{1/4}(1-\beta)^{3/4}} \prod_{q \text{ prime}} \left(1-\frac{1}{q}\right)^{\sqrt{\frac{1-\beta}{\beta}}} \left(1-\frac{\sqrt{\frac{1-\beta}{\beta}}}{q}\right)^{-1}$$

$$C := \frac{3e^{\frac{3\gamma}{2}}}{4\sqrt{\pi}} \prod_{q>2 \text{ prime}} \left(1+\frac{1}{q(q-2)}\right) = 1.523555\ldots.$$

. . . Note that when $\beta = \frac{1}{2}$ we have $C_{\frac{1}{2}} = \sqrt{\frac{2}{\pi}}$.

Applications

- $P^{(\frac{1}{2})}(n)$ follows Benford's law and is equidistributed in coprime residue classes.

Applications

- $P^{(\frac{1}{2})}(n)$ follows Benford's law and is equidistributed in coprime residue classes.

Theorem (Improved convergence to normal distribution)

For all t we have

$$\frac{1}{x}\#\left\{n \le x : \frac{\log_2 P^{(\frac{1}{2})}(n) - \frac{1}{2}\log_2 x}{\sqrt{\log_2 x}} < t\right\} = \Phi(2t) + O\left(\frac{(\log_3)^{3/2}}{\sqrt{\log_2 x}}\right).$$

Average size of the middle prime factor

Recall that $\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \lambda \log x, \quad \frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim e^{-\gamma}\log_2 x.$

Average size of the middle prime factor

Recall that $\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \lambda \log x$, $\frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim e^{-\gamma} \log_2 x$.

Theorem

Let $\varphi = \frac{1+\sqrt{5}}{2}$ *the golden ratio, and* $\varphi' = \frac{1}{\varphi} = \varphi - 1 = \frac{\sqrt{5}-1}{2} = 0.6180\ldots$.

$$\frac{1}{x}\sum_{2\le n\le x} \log P^{\left(\frac{1}{2}\right)}(n) \ \sim \ A(\log x)^{\varphi'}$$

Average size of the middle prime factor

Recall that $\frac{1}{x}\sum_{2\le n\le x} P^+(n) \sim \lambda \log x$, $\quad \frac{1}{x}\sum_{2\le n\le x} P^-(n) \sim e^{-\gamma}\log_2 x$.

Theorem

Let $\varphi = \frac{1+\sqrt{5}}{2}$ *the golden ratio, and* $\varphi' = \frac{1}{\varphi} = \varphi - 1 = \frac{\sqrt{5}-1}{2} = 0.6180\ldots$.

$$\frac{1}{x}\sum_{2\le n\le x} \log P^{\left(\frac{1}{2}\right)}(n) \;\sim\; A(\log x)^{\varphi'}$$

where

$$A := \frac{e^{-\gamma}}{\varphi!}\frac{\varphi+1}{\sqrt{5}}\prod_p \left(1-\tfrac{1}{p}\right)^{\varphi'}\left(1-\tfrac{\varphi'}{p}\right)^{-1} = 1.313314\ldots$$

Generalizations to other α

These results generalize to $P^{(\alpha)}(n)$ for any $\alpha \in (0,1)$.

Generalizations to other α

These results generalize to $P^{(\alpha)}(n)$ for any $\alpha\in(0,1)$. Consider $M_p^{(\alpha)}(x)$, the count of $n \leq x$ with $P^{(\alpha)}(n) = p$.

Generalizations to other α

These results generalize to $P^{(\alpha)}(n)$ for any $\alpha \in (0,1)$. Consider $M_p^{(\alpha)}(x)$, the count of $n \le x$ with $P^{(\alpha)}(n) = p$. Find a phase transition at $\beta_\alpha := \left(1 + 2^{\frac{1}{1-\alpha}}(\alpha^{-1} - 1)\right)^{-1}$

Generalizations to other α

These results generalize to $P^{(\alpha)}(n)$ for any $\alpha\in(0,1)$. Consider $M_p^{(\alpha)}(x)$, the count of $n \le x$ with $P^{(\alpha)}(n) = p$. Find a phase transition at $\beta_\alpha := \left(1+2^{\frac{1}{1-\alpha}}(\alpha^{-1} - 1)\right)^{-1}$

Theorem

Fix $\alpha \in (0,1)$ and suppose $\beta_\alpha + \epsilon < \beta < 1 - \epsilon$. Then,

$$M_p^{(\alpha)}(x) \sim \frac{C_{\beta,\alpha} x}{p(\log x)^{1-\left(\frac{\beta}{\alpha}\right)^{\alpha}\left(\frac{1-\beta}{1-\alpha}\right)^{1-\alpha}} \sqrt{\log_2 x}}$$

$C_{\alpha,\beta}$

Set $\chi := \frac{(1-\alpha)\beta}{(1-\beta)\alpha}$,

$C_{\alpha,\beta}$

Set $\chi := \frac{(1-\alpha)\beta}{(1-\beta)\alpha}$, then

$$C_{\beta,\alpha} := \frac{e^{\gamma(\chi^{\alpha-1}-\chi^{\alpha})}}{\Gamma\left(1+\chi^{\alpha}\right)} \frac{\chi^{\frac{3}{2}\alpha-1}\rho_{\chi,\alpha}}{\sqrt{2\pi\alpha(1-\beta)}} \prod_{q} \left(1-\frac{1}{q}\right)^{\chi^{\alpha-1}} \left(1-\frac{\chi^{\alpha-1}}{q}\right)^{-1}$$

$C_{\alpha,\beta}$

Set $\chi := \frac{(1-\alpha)\beta}{(1-\beta)\alpha}$, then

$$C_{\beta,\alpha} := \frac{e^{\gamma(\chi^{\alpha-1}-\chi^{\alpha})}}{\Gamma\left(1+\chi^{\alpha}\right)} \frac{\chi^{\frac{3}{2}\alpha-1}\rho_{\chi,\alpha}}{\sqrt{2\pi\alpha(1-\beta)}} \prod_{q}\left(1-\frac{1}{q}\right)^{\chi^{\alpha-1}}\left(1-\frac{\chi^{\alpha-1}}{q}\right)^{-1}$$

$$\rho_{\chi,\alpha} := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi^{\{n(1-\alpha)\}}$$

$C_{\alpha,\beta}$

Set $\chi := \frac{(1-\alpha)\beta}{(1-\beta)\alpha}$, then

$$C_{\beta,\alpha} := \frac{e^{\gamma(\chi^{\alpha-1}-\chi^{\alpha})}}{\Gamma\left(1+\chi^{\alpha}\right)} \frac{\chi^{\frac{3}{2}\alpha-1}\rho_{\chi,\alpha}}{\sqrt{2\pi\alpha(1-\beta)}} \prod_{q}\left(1-\frac{1}{q}\right)^{\chi^{\alpha-1}}\left(1-\frac{\chi^{\alpha-1}}{q}\right)^{-1}$$

$$\rho_{\chi,\alpha} := \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\chi^{\{n(1-\alpha)\}} = \begin{cases} \sum_{i=0}^{b-1}\chi^{i/b} & \alpha=\frac{a}{b} \text{ rational} \\ \int_0^1 \chi^t dt = \frac{\chi-1}{\log\chi} & \alpha \text{ irrational} \end{cases}.$$

$C_{\alpha,\beta}$

Set $\chi := \frac{(1-\alpha)\beta}{(1-\beta)\alpha}$, then

$$C_{\beta,\alpha} := \frac{e^{\gamma(\chi^{\alpha-1}-\chi^{\alpha})}}{\Gamma(1+\chi^{\alpha})} \frac{\chi^{\frac{3}{2}\alpha-1}\rho_{\chi,\alpha}}{\sqrt{2\pi\alpha(1-\beta)}} \prod_q \left(1-\frac{1}{q}\right)^{\chi^{\alpha-1}} \left(1-\frac{\chi^{\alpha-1}}{q}\right)^{-1}$$

$$\rho_{\chi,\alpha} := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi^{\{n(1-\alpha)\}} = \begin{cases} \sum_{i=0}^{b-1} \chi^{i/b} & \alpha = \frac{a}{b} \text{ rational} \\ \int_0^1 \chi^t dt = \frac{\chi-1}{\log \chi} & \alpha \text{ irrational} \end{cases}.$$

As a function of α (for a fixed β) $C_{\beta,\alpha}$ is continuous at every irrational, but discontinuous at every rational value of α, besides $\alpha = \beta$.

New improvements by Rotgé, (2025)

Focuses on the middle prime factor, $P^{\frac{1}{2}}(n)$.

New improvements by Rotgé, (2025)

Focuses on the middle prime factor, $P^{\frac{1}{2}}(n)$.

- Works out the details of the "phase transition" at $\beta = \frac{1}{5}$, showing that there is a Gaussian transition between the two regimes.

New improvements by Rotgé, (2025)

Focuses on the middle prime factor, $P^{\frac{1}{2}}(n)$.

- Works out the details of the "phase transition" at $\beta = \frac{1}{5}$, showing that there is a Gaussian transition between the two regimes.
- Refines the computation of

$$\frac{1}{x}\sum_{n<x} \log P^{(\frac{1}{2})}(n) = A(\log x)^{\varphi'}\left\{1 + \sum_{1\le j\le J} \frac{Q_j(\theta_x)}{(\log_2 x)^j} + O\left(\frac{1}{(\log_2 x)^{J+1}}\right)\right\}.$$

For any fixed J where the Q_j are polynomials, $\theta_x = \{\log_2 x/\sqrt{5}\}$.

New improvements by Rotgé, (2025)

Focuses on the middle prime factor, $P^{\frac{1}{2}}(n)$.

- Works out the details of the "phase transition" at $\beta = \frac{1}{5}$, showing that there is a Gaussian transition between the two regimes.
- Refines the computation of

$$\frac{1}{x}\sum_{n<x} \log P^{(\frac{1}{2})}(n) = A(\log x)^{\varphi'}\left\{1 + \sum_{1\le j\le J} \frac{Q_j(\theta_x)}{(\log_2 x)^j} + O\left(\frac{1}{(\log_2 x)^{J+1}}\right)\right\}.$$

  For any fixed J where the Q_j are polynomials, $\theta_x = \{\log_2 x/\sqrt{5}\}$.
- Obtains an improved (and optimal) error term for the convergence the typical values,

$$\frac{1}{x}\#\left\{n \le x : \frac{\log_2 P^{(\frac{1}{2})}(n) - \frac{1}{2}\log_2 x}{\sqrt{\log_2 x}} < t\right\} = \Phi\left(2t\right) + O\left(\frac{1}{\sqrt{\log_2 x}}\right).$$

Open Questions

Open Questions

The median of the middle prime factor is still pretty open.

Open Questions

The median of the middle prime factor is still pretty open.
After Rotgé's improvments we now know this median is $\exp\left(O\left(\sqrt{\log x}\right)\right)$.

Open Questions

The median of the middle prime factor is still pretty open.
After Rotgé's improvments we now know this median is $\exp\left(O\left(\sqrt{\log x}\right)\right)$.

We also don't know the behavior of $M_p^{\left(\frac{1}{2}\right)}(x)$ as when $\beta = \frac{\log_2 p}{\log_2 x} \to 1$.

Open Questions

The median of the middle prime factor is still pretty open.
After Rotgé's improvments we now know this median is $\exp\left(O\left(\sqrt{\log x}\right)\right)$.

We also don't know the behavior of $M_p^{\left(\frac{1}{2}\right)}(x)$ as when $\beta = \frac{\log_2 p}{\log_2 x} \to 1$.

THANK YOU!