Gaps between Consecutive Primes

$\mathbb{P} := \{\text{primes}\} = \{p_n\}_{n=1}^{\infty} \nearrow$

$\pi(x) := \mathbb{P} \cap [1, x]$.

PNT : $\pi(x) \sim \frac{x}{\log x} \sim \int_2^x \frac{dt}{\log t} \Rightarrow$ density function for $\mathbb{P} = \frac{1}{\log x}$.

$\Updownarrow$

$p_n \sim n \log n \Rightarrow$ average gap $p_{n+1} - p_n \approx \log n \sim \log p_n$

Q : What is the distribution of $p_{n+1} - p_n$?

$\limsup_{n\to\infty} (p_{n+1} - p_n) = \infty$.

Bertrand's postulate $\Rightarrow p_{n+1} < 2p_n$.

RH $\Rightarrow p_{n+1} - p_n = O(\sqrt{p_n} \log p_n)$.

$\liminf_{n\to\infty} (p_{n+1} - p_n) \stackrel{?}{=} 2$

More precisely, what is

(1) $$\lim_{x\to\infty} \frac{1}{\pi(x)} \# \left\{ p \le x : \frac{p_{next} - p}{\log p} \in [a, b] \right\} ?$$

Twin prime pair

$2996863034895 \cdot 2^{1290000} \pm 1$

Here $0 \le a < b$ are fixed.

Correspondingly, one may asks what is

$$(2) \quad \lim_{x\to\infty} \frac{1}{x} \# \{ n \leqslant x : \pi(n+\lambda\log n) - \pi(n) = k \} \ ?$$

Here $\lambda > 0$ and $k \geqslant 0$ are given.

Answer: We don't know. But we have a few reasonable conjectures.

Cramér's Model

Let $1_{\mathbb{P}}$ be the indicator function of $\mathbb{P}$, i.e.,

$$1_{\mathbb{P}}(n) := \begin{cases} 1 & \text{if } n \in \mathbb{P}, \\ 0 & \text{otherwise}. \end{cases}$$

Cramér's model: $1_{\mathbb{P}}$ behaves roughly like a sequence $\{X(n)\}_{n\in\mathbb{N}}$ of independent Bernoulli random variables $X(n)$ defined by $X(1)=0$, $X(2)=1$, and $\forall n \geq 3$,

$$X(n) = \begin{cases} 1 & \text{with probability } \frac{1}{\log n}, \\ 0 & \text{with probability } 1-\frac{1}{\log n}. \end{cases}$$

Accepting Cramér's model, let us provide an answer to (1). Let $p \in \mathbb{P}$ be a large prime. So $X(p)=1$. For every $h \in \mathbb{N}$, the probability that $X(p+1) = \cdots = X(p+h-1) = 0$ and $X(p+h)=1$ (i.e., $p_{next} = p+h$) is

$$\left(1-\frac{1}{\log(p+1)}\right)\left(1-\frac{1}{\log(p+2)}\right)\cdots\left(1-\frac{1}{\log(p+h-1)}\right)\frac{1}{\log(p+h)}.$$

If h is small compared to p, say $h \asymp \log p$, the above is

$$\sim \left(1-\frac{1}{\log p}\right)^{h-1}\frac{1}{\log p} = \exp\left((h-1)\log\left(1-\frac{1}{\log p}\right)\right)\frac{1}{\log p}$$

$$\sim e^{-\frac{h-1}{\log p}}\frac{1}{\log p}.$$

Given $0 \leq a < b$, the probability that $p_{next} \in [p+a\log p, p+b\log p]$ is

$$\approx \sum_{h\in[a\log p,\, b\log p]} e^{-\frac{h-1}{\log p}}\frac{1}{\log p} \approx \sum_{\frac{h}{\log p}\in[a,b]} e^{-\frac{h}{\log p}}\frac{1}{\log p} \approx \int_a^b e^{-t}\,dt.$$

Conjecture 1. Given $0 \leq a < b$, we have

$$\lim_{x\to\infty}\frac{1}{\pi(x)}\#\{p\leq x : p_{next}\in[p+a\log p,\, p+b\log p]\} = \int_a^b e^{-t}\,dt.$$

In other words, $\dfrac{p_{next}-p}{\log p}$ behaves like a random variable for exponential distribution with parameter 1.

Conjecture 2. Given $\lambda > 0$ and $k \geq 0$, we have

$$\lim_{x\to\infty} \frac{1}{x} \cdot \#\{2 \leq n \leq x : \pi(n+\lambda\log n) - \pi(n) = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In other words, the limit distribution of the number of primes in intervals of length $\asymp \log n$ as $n \to \infty$ is a Poisson distribution.

This is because $\mathbb{P}(X(j) = 1$ for precisely k of the j's in $[n+1, n+\lambda\log n])$ is

$$\sim \binom{\lfloor \lambda\log n \rfloor}{k} \left(\frac{1}{\log n}\right)^k \left(1 - \frac{1}{\log n}\right)^{\lfloor \lambda \log n\rfloor - k} \sim \frac{(\lambda\log n)^k}{k!} \left(\frac{1}{\log n}\right)^k e^{-\frac{\lambda\log n - k}{\log n}} \sim \frac{\lambda^k}{k!} e^{-\lambda}.$$

The Hardy–Littlewood Prime k-tuples Conjecture

Conjecture 3 (Hardy–Littlewood) Let $\mathcal{H} = \{f_1, \dots, f_k\} \subseteq \mathbb{Z}[x]$ be a set of $k \geq 1$ monic polynomials of degree 1. Then

$$\#\{n \leq x : f_1(n), \dots, f_k(n) \in \mathbb{P}\} \sim \mathfrak{S}(\mathcal{H}) \int_2^x \frac{dt}{(\log t)^k} \sim \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}.$$

Here
$$\mathfrak{S}(\mathcal{H}) := \prod_{p \in \mathbb{P}} \left(1-\frac{1}{p}\right)^{-k}\left(1-\frac{\nu_{\mathcal{H}}(p)}{p}\right),$$

where $\nu_{\mathcal{H}}(p) := \#\{ f_i(0) \in \mathbb{F}_p : 1 \leq i \leq k \}$.

$k=1 \Leftarrow$ PNT.

$k=2 \Leftarrow$ The Twin Prime Conjecture !

$$\pi_2(x) := \#\{ p \leq x : p+2 \in \mathbb{P} \} \sim 2C_2 \cdot \frac{x}{(\log x)^2}, \text{ where}$$

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

Cramér's model would yield a different value for C_2, mainly because n and $n+2$ are clearly dependent.

Nevertheless, one can show Conjecture 3 $\Rightarrow$ Conjectures 1 & 2.

Heuristics for the Hardy–Littlewood Conjecture based on Cramér's Model

Let $p \geq 2$ be a prime. Cramér's model predicts

$$\#\{n \leq x : f_1(n), \ldots, f_k(n) \in \mathbb{P}\} \sim \frac{x}{(\log x)^k} \qquad (*)$$

by assuming the events $p \nmid f_1(n), \ldots, p \nmid f_k(n)$ are mutually indenpent for a randomly chosen n. But this assumption is clearly faulty. In fact, $\mathbb{P}(p \nmid f_1(n), \ldots, p \nmid f_k(n))$ is $1 - \frac{\nu_H(p)}{p}$ rather than $(1 - \frac{1}{p})^k$. To adjust our prediction, we need to multiply the RHS of $(*)$ by $\prod_{p \in \mathbb{P}} (1 - \frac{\nu_H(p)}{p})(1 - \frac{1}{p})^{-k} = \mathfrak{S}(H)$ and assume that local condition $p \nmid f_1(n), \ldots, p \nmid f_k(n)$ is sufficient for the existence of infinitely many prime tuples.

Large Gaps between Primes

If $n \geq 2$, then $n!+2, n!+3, \cdots, n!+n$ are all composite. Hence

$$\limsup_{n\to\infty} (p_{n+1} - p_n) = \infty.$$

Cramér's model suggests

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} = \infty, \quad \limsup_{n\to\infty} \frac{p_{n+1}-p_n}{(\log p_n)^2} = c.$$

This was proved by Westzynthius (1931).

Westzynthius (1931):
$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_3 p_n}{\log_4 p_n}} \geq c_1$$

Erdös (1935):
$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_2 p_n}{(\log_3 p_n)^2}} \geq c_2.$$

Rankin (1938) :

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_2 p_n \cdot \log_4 p_n}{(\log_3 p_n)^2}} \geq C_3 .$$

Ford - Green - Konyagin - Tao (2014) and Maynard (2014) :

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_2 p_n \cdot \log_4 p_n}{(\log_3 p_n)^2}} = \infty .$$

Ford - Green - Konyagin - Maynard - Tao (2018) :

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_2 p_n \cdot \log_4 p_n}{\log_3 p_n}} \geq C_4 .$$

Tao conjectures

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\frac{\log p_n \cdot \log_2 p_n \cdot \log_4 p_n}{\log_3 p_n}} = \infty .$$

Here $\log_k x := \underbrace{\log \log \cdots \log}_{k\text{-fold}} x$.

Small Gaps between Primes

Cramér's model suggests

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

This was confirmed by Goldston, Pintz and Yildirim (2005).

Zhang (2013): $\liminf_{n\to\infty} (p_{n+1}-p_n) < 7 \cdot 10^7$.

Polymath (2013): $\liminf_{n\to\infty} (p_{n+1}-p_n) < 4680$.

Maynard (2013): $\liminf_{n\to\infty} (p_{n+1}-p_n) < 600$.

Polymath (2014): $\liminf_{n\to\infty} (p_{n+1}-p_n) < 246$.

Twin Prime Conjecture: $\liminf_{n\to\infty} (p_{n+1}-p_n) = 2$.

Heuristics for Conjecture 2 $\Rightarrow$ Conjecture 1.

$$A_k(x) := \#\{2 \leq n \leq x : \pi(n+\lambda \log n) - \pi(n) = k\} \sim \frac{\lambda^k}{k!} e^{-\lambda} \cdot x$$

$$\Rightarrow \quad \#\{p \leq x : \pi(p+\lambda \log p) - \pi(p) = k\}$$

$$= \sum_{n \leq x-1} A_k(n)\left(1_{\mathbb{P}}(n) - 1_{\mathbb{P}}(n+1)\right) + A_k(x) \cdot 1_{\mathbb{P}}(\lfloor x \rfloor)$$

$$\approx \frac{\lambda^k}{k!} e^{-\lambda}\left(\sum_{n \leq x-1} n\left(1_{\mathbb{P}}(n) - 1_{\mathbb{P}}(n+1)\right) + \lfloor x \rfloor \cdot 1_{\mathbb{P}}(\lfloor x \rfloor)\right) = \frac{\lambda^k}{k!} e^{-\lambda} \cdot \pi(x).$$

$$\Rightarrow \lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leq x : \pi(p+\lambda \log p) - \pi(p) = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

$$\Rightarrow \lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leq x : p_{next} \in (p+a\log p, p+b\log p]\}$$

$$= \lim_{x\to\infty} \frac{1}{\pi(x)} \left(\#\{p \leq x : \pi(p+a\log p) - \pi(p) = 0\} - \#\{p \leq x : \pi(p+b\log p) - \pi(p) = 0\}\right)$$

$$= e^{-a} - e^{-b} = \int_a^b e^{-t}\, dt.$$

Hardy-Littlewood $\Rightarrow$ Conjecture 2

$$\overline{\sum_{1\le h_1,\dots,h_k\le h}} \pi(x;\{h_1,\dots,h_k\})$$

Thm (Gallagher) $\overline{\sum_{1\le h_1<\dots<h_k\le h}} \mathfrak{S}(\mathcal{H}_k) \sim \overline{\sum_{1\le h_1<\dots<h_k\le h}} 1$ as $h\to\infty$.

Let $M_k(x) := \overline{\sum_{n\le x}} (\pi(n+h)-\pi(n))^k$. Then

$$M_k(x) = \overline{\sum_{n\le x}}\ \overline{\sum_{p_1,\dots,p_k\in(n,n+h]}} 1 = \overline{\sum_{1\le h_1,\dots,h_k\le h}} \pi(x;\mathcal{H}_k) = \sum_{r=1}^{k} \left\{ {k \atop r} \right\} r! \overline{\sum_{1\le h_1<\dots<h_r\le h}} \pi(x;\mathcal{H}_r)$$

where $\left\{ {k \atop r} \right\}$ is Stirling's number of the 2^{nd} kind, $\mathcal{H}_r := \{h_1,\dots,h_r\}$ and

$$\pi(x;\mathcal{H}_r) := \#\{n\le x : n+h_1,\dots,n+h_r\in\mathbb{P}\} \sim \mathfrak{S}(\mathcal{H}_r)\frac{x}{(\log x)^r}.$$

Thus $M_k(x) \sim \sum_{r=1}^{k}\left\{ {k \atop r} \right\} r!\binom{h}{r}\frac{x}{(\log x)^r} \sim \sum_{r=1}^{k}\left\{ {k \atop r} \right\}\left(\frac{h}{\log x}\right)^r x.$

For $h\sim\lambda\log x$, $M_k(x)\sim\sum_{r=1}^{k}\left\{ {k \atop r} \right\}\lambda^r x$. But $\sum_{r=1}^{k}\left\{ {k \atop r} \right\}\lambda^r$ is the

kth moment of a Poisson random variable with parameter λ.