Large gaps in sets of primes and other sequences
I. Heuristics and basic constructions

Kevin Ford

University of Illinois at Urbana-Champaign

May, 2018

Large gaps between primes

Def: $G(x) = \max_{p_n \leqslant x} (p_n - p_{n-1})$, p_n is the n^{th} prime.

$$2, 3, 5, 7, \ldots, 109, 113, 127, 131, \ldots, 9547, 9551, 9587, 9601, \ldots$$

Upper bound: $G(x) \ll x^{0.525}$ (Baker-Harman-Pintz, 2001).
Improve to $O(x^{1/2+\varepsilon})$ on RH.

Lower bound: $G(x) \gg (\log x)\frac{\log_2 x \log_4 x}{\log_3 x}$ (F,Green,Konyagin,Maynard,Tao,2018)

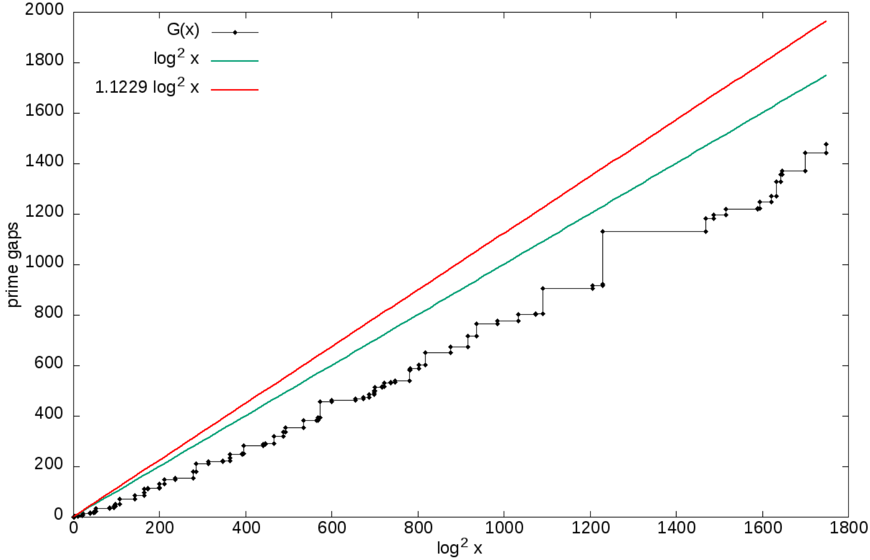
Conjectures

Cramér (1936): $\limsup_{x\to\infty} \frac{G(x)}{\log^2 x} = 1.$

Shanks (1964): $G(x) \sim \log^2 x.$

Granville (1995): $\limsup_{x\to\infty} \frac{G(x)}{\log^2 x} \geqslant 2e^{-\gamma} = 1.1229\ldots$

Computational evidence, up to 10^{18}



Cramér's model of large prime gaps

Let $X_3, X_4, X_5, \ldots$ be indep. random vars. s.t.

$$\mathbb{P}(X_n = 1) = \frac{1}{\log n}, \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{\log n}.$$

Let

$$\mathscr{P} = \{n : X_n = 1\} = \{P_1, P_2, \ldots\},$$

the set of "probabilistic primes".

Theorem (Cramér, 1936)

With probability 1,

$$\limsup_{N\to\infty} \frac{P_{N+1} - P_N}{\log^2 N} = 1.$$

Cramér: "*for the ordinary sequence of prime numbers p_n, some similar relation may hold*".

Sketch of the proof of Cramér's theorem

$$\mathbb{P}(X_n = 1) = \frac{1}{\log n}, \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{\log n}.$$

Suppose that $N^{1-o(1)} \leqslant k \leqslant N$. Then

$$\mathbb{P}(X_{k+1} = \cdots = X_{k+g} = 0) \approx \left(1 - \frac{1}{\log k}\right)^g \approx e^{-g/\log N}.$$

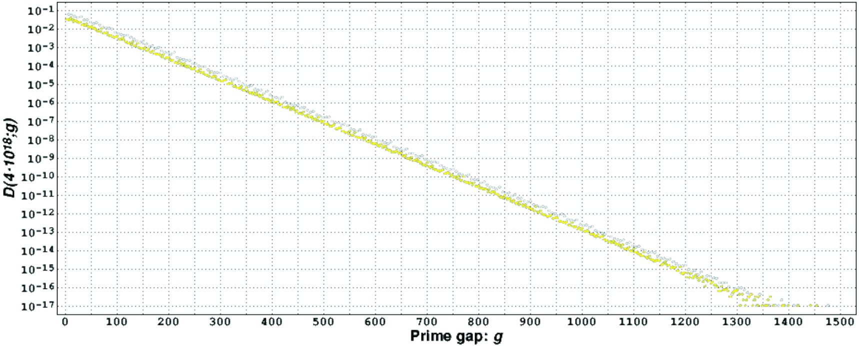
Hence (summing on k)

$$\mathbb{E}\#\{\text{gaps of length} \geqslant g \text{ below } N\} \approx N e^{-g/\log N}.$$

If $g > (1+\varepsilon)\log^2 N$, this is $o(1)$.
If $g < (1-\varepsilon)\log^2 N$, then this is very large.

More predictions of Cramér's model: distribtion of gaps

$$(1/N)\mathbb{E}\#\{\text{gaps of length} \geqslant \lambda \log N\} \approx e^{-\lambda}.$$



Actual prime gap statistics, $p_n < 4 \cdot 10^{18}$

Gallagher, 1976. Prime $k-$tuples conjecture $\Rightarrow$ exponential prime gap distribution

General Cramér's-type model: random darts

Choose N random points in $[0,1]$ (random darts)

Theorem (classical?)

W.h.p., the max. gap is $\sim \frac{\log N}{N}$.

Proof idea (Rényi). The $N+1$ gaps have distribution

$$\overset{d}{=} \left(\frac{E_1}{S}, \dots, \frac{E_{N+1}}{S} \right), \quad S = E_1 + \cdots + E_{N+1},$$

where each E_i has exponential distribution, $\mathbb{P}(E_i \leqslant x) = 1 - e^{-x}$.
W.h.p., $S \sim N$. Also,

$$\mathbb{P}\left(\max E_i \leqslant \log N + u\right) = \left(1 - \frac{e^{-u}}{N}\right)^{N+1} \sim \exp\{-e^{-u}\},$$

the Gumbel extreme value distribution.

Random darts and Cramér's model

Choose N random points in $[0, x]$ (random darts)

$$\mathbb{P}\left(\text{max gap} \leqslant \frac{x(\log N + u)}{N}\right) \approx \exp\{-e^{-u}\}.$$

Probabilistic primes; $N = \text{li}(x) + O(x^{1/2+\varepsilon})$

$$\mathbb{P}\left(\max_{P_n \leqslant x} P_{n+1} - P_n \leqslant \frac{x(\log \text{li}(x) + u)}{\text{li}(x)}\right) \approx \exp\{-e^{-u}\}.$$

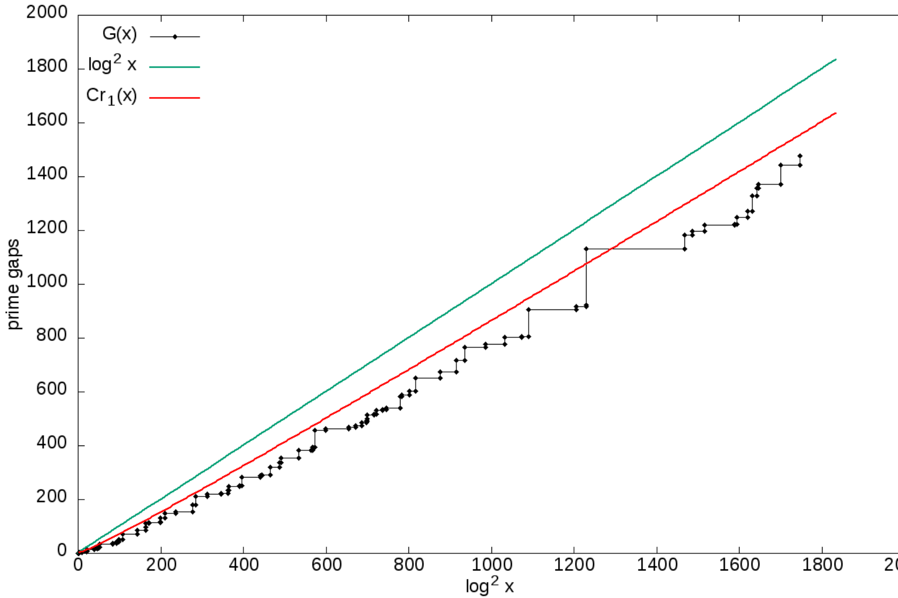
Theorem

For Cramér's probabilistic primes,

$$\max_{P_n \leqslant x} P_n - P_{n-1} \lesssim \text{Cr}_1(x) := \frac{x \log \text{li}(x)}{\text{li}(x)} \approx (\log x)(\log x - \log_2 x).$$

Q1: Does this explain the data for actual primes?

Data vs. refined Cramér conjecture



General Cramér's-type model: random darts

Choose N random points in $[0, x]$ (random darts)
Expected maximal gap is of size $\approx \frac{x \log N}{N}$

Prime k-tuples. Let $f_1, \dots, f_k$ be distinct, irreducible polynomials $f_i : \mathbb{Z} \to \mathbb{Z}$ with pos. leading coeff., degrees d_i, and $f_1 \cdots f_k$ has no fixed prime factor.

Conjecture (Bateman-Horn)

$$\#\{n \leqslant x : f_1(n), \dots, f_k(n) \text{ all prime}\} \sim C \ \mathrm{li}_k(x),$$

where $C = C(f_1, \dots, f_k) > 0$ is constant and $\mathrm{li}_k(x) = \int_2^x \frac{dt}{(\log t)^k}$

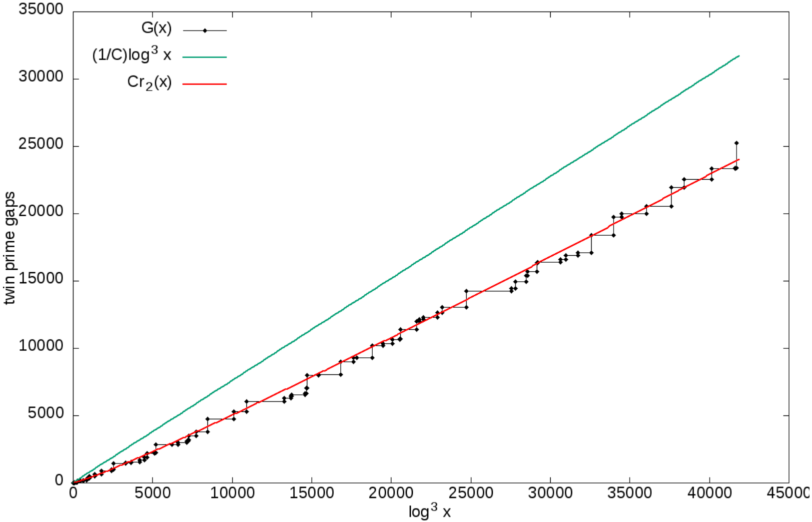
Refined Cramér conjecture

The largest gap in $\{n \leqslant x : f_1(n), \dots, f_k(n) \text{ all prime}\}$ is

$$\lesssim \mathrm{Cr}_k(x) := \frac{x \log(C \ \mathrm{li}_k(x))}{C \ \mathrm{li}_k(x)} \approx \frac{(\log x)^k}{C}(\log x - k \log_2 x)$$

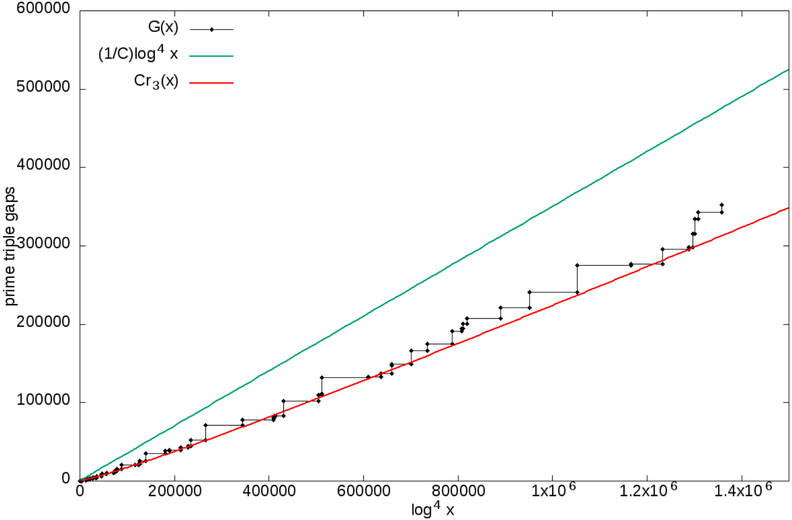
Twin prime gaps

$$(f_1, f_2) = (n, n+2), \qquad C \approx 1.32032$$



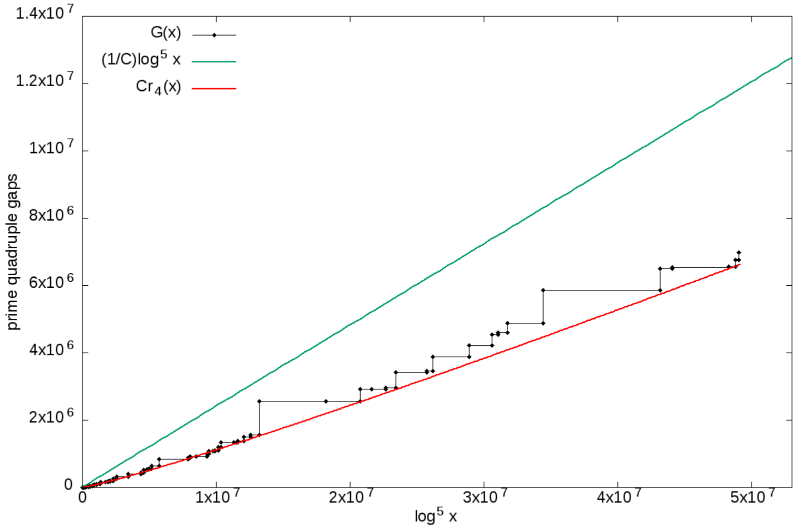
Prime triplet gaps

$$(f_1, f_2, f_3) = (n, n+2, n+6) \qquad C \approx 2.85825$$



Prime quadruplet gaps

$$(f_1, \ldots, f_4) = (n, n+2, n+6, n+8) \qquad C \approx 4.15118$$



Cramér's model defect: global distribution of primes

Theorem (Cramér, 1936 ("Probabilistic RH"))

With probability 1, $\Pi(x) := \#\{P_n \leqslant x\} = li(x) + O(x^{1/2+\varepsilon})$.

J. Pintz observed the following:

Theorem

$$\mathbb{E}(\Pi(x) - \mathrm{li}(x))^2 \sim \frac{x}{\log x},$$

contrast with

Theorem (Cramér,1920)

On R.H.,

$$\frac{1}{x}\int_x^{2x} |\pi(t) - \mathrm{li}(t)|^2 \, dt \ll \frac{x}{\log^2 x}$$

Cramér's model defect: small gaps

Theorem. With probability 1,

$$\#\{n : n, n+1 \in \mathscr{P}\} = \infty$$

This does not hold for real primes!

Theorem. With probability 1,

$$\#\{n \leqslant x : n, n+2 \in \mathscr{P}\} \sim \frac{x}{\log^2 x}.$$

Conjecture (Hardy-Littlewood, 1923).

$$\#\{n \leqslant x : n, n+2 \text{ prime}\} \sim C\frac{x}{\log^2 x},$$

where $C = 2\prod_{p>2}(1 - 1/(p-1)^2) \approx 1.3203$

Granville's refinement of Cramér's model

Cramér's model major defect: Cramér primes are equidistributed modulo small primes like 2,3,..., whereas real primes are not.

This shows up in asymptotics for prime k-tuples, and for counts of primes in very short intervals: w.h.p.,

$$\Pi(x+y) - \Pi(x) \sim \frac{y}{\log x} \quad (y/\log^2 x \to \infty)$$

By contrast,

Theorem (H. Maier, 1985)

$\forall M > 1$,

$$\limsup_{x\to\infty} \frac{\pi(x+\log^M x) - \pi(x)}{\log^{M-1} x} > 1.$$

Granville's refinement of Cramér's model, II

Let $T = \varepsilon \log x, \quad Q_T = \prod_{p \leqslant T} p = x^{o(1)}.$

Real primes live in

$$\mathcal{S}_T := \{n \in \mathbb{Z} : (n, Q_T) = 1\}$$

For $n \in \mathcal{S}_T \cap (x, 2x]$, define the random variables

$$Z_n : \quad \mathbb{P}(Z_n = 1) = \theta / \log n, \quad \mathbb{P}(Z_n = 0) = 1 - \theta / \log n,$$

where $1/\theta = \phi(Q_T)/Q_T \sim e^{-\gamma}/\log T$ is the density of $\mathcal{S}_T$. That is,

$$\frac{\theta}{\log n} = \text{ conditional prob. that } n \text{ is prime given that } n \in \mathcal{S}_T.$$

Granville's refinement of Cramér's model, III

$$\mathcal{S}_T := \{n \in \mathbb{Z} : (n, Q_T) = 1\}$$

Let $y = c \log^2 x$, take special values of $m \in (x, 2x]$, namely those with with $Q_T | m$. Since $y = T^{2+o(1)}$,

$$\#\left([m, m+y] \cap \mathcal{S}_T\right) = \#\left([0, y] \cap \mathcal{S}_T\right) \sim \frac{y}{\log y}. \qquad \text{(sp)}$$

By contrast, for a *typical* $m \in \mathbb{Z}$,

$$\#\left([m, m+y] \cap \mathcal{S}_T\right) \sim \theta^{-1} y \sim 2e^{-\gamma} \frac{y}{\log y}. \qquad \text{(ty)}$$

Note $2e^{-\gamma} = 1.1229\ldots > 1$, so the intervals in (sp) are deficient in sifted numbers.

Granville's refinement of Cramér's model, IV

$T = \varepsilon \log x$, $y = c \log^2 x$,

$$\#\left([m, m+y] \cap \mathcal{S}_T\right) = \#\left([0, y] \cap \mathcal{S}_T\right) \sim \frac{y}{\log y}. \qquad \text{(sp)}$$

Get

$$\mathbb{P}\big(Z_n = 0 : n \in [m, m+y] \cap \mathcal{S}_T\big) \approx (1 - \theta/\log x)^{y/\log y}$$
$$\approx e^{-c(e^{\gamma}/2)\log x}.$$

Therefore, gaps of size $\geqslant (2e^{-\gamma} + o(1))(\log x)^2$ exist w.h.p.

Computing secondary terms; get gaps of size

$$2e^{-\gamma}(\log x)^2 + A(\varepsilon)\frac{(\log x)^2}{\log_2 x} + \cdots, \qquad A(\varepsilon) \to \infty (\varepsilon \to 0).$$

Project: work out the secondary term; compare with data.

Proving large gaps: Jacobsthal's function

$$\mathcal{S}_T = \{n \in \mathbb{Z} : (n, Q_T) = 1\}, \quad Q_T = \prod_{p \leqslant T} p.$$

Main goal: Find $J(T)$, the largest gap in $\mathcal{S}_T$.

$$G(2Q_T) \geqslant J(T), \qquad G(x) := \max_{p_n \leqslant x} p_{n+1} - p_n.$$

Since $Q_T \approx e^T$, get $G(x) \gtrsim J(\log x)$.

Trivial: Avg. gap is $\sim e^\gamma \log T$;$J(T) \geqslant T - 2$ ($[2, T] \cap \mathcal{S}_T = \emptyset$)

Lower bound (FGKMT, 2018). $J(T) \gg T \frac{\log T \log_3 T}{\log_2 T}$.

Upper bound (Iwaniec, 1978). $J(T) \ll T^2 (\log T)^2$.

Conjecture (Maier-Pomerance, 1990). $J(T) = T(\log T)^{2+o(1)}$.

Random dart model prediction: $J(T) \sim T \frac{Q_T}{\phi(Q_T)} \sim e^\gamma T \log T$.

Finding large gaps in $\mathcal{S}_T$

Covering: $J(T)$ is the largest y so that there are $a_2, a_3, a_5, \ldots$ with

$$\{a_p \mod p : p \leqslant T\} \supseteq [0, y]$$

Classical 3-stage-process (Westzynthius's-Erdős-Rankin)

1. Take $a_p = 0$ for $p \in (z, x/2] \cap [2, 2y/x]$. Uncovered: z-smooth numbers (few for appropriate z) and primes; Total $\sim y/\log y$ numbers uncovered.**Far better that typical choice, which leaves about $y \prod_{z<p\leqslant x/2}(1-1/p) \sim y\frac{\log x}{\log z}$ uncovered numbers**
2. Greedy choice for a_p, $p \in (2y/x, z]$; Unconvered: $\lesssim (y/\log y)\frac{\log z}{\log(2y/x)}$ numbers. Want this to be $\leqslant \frac{x}{4\log x}$.
3. use each a_p for $p \in (x/2, x]$ to cover the remaining uncovered elements of $[1, y]$, one element for each p.

If fewer than $\pi(x) - \pi(x/2) \sim \frac{x}{2\log x}$ elements left after stages 1-2, then succeed!

Lower bounds for $J(T)$: prime k-tuples

Lower bound (FGKMT, 2018). $J(T) \gg T\frac{\log T \log_3 T}{\log_2 T}$.

Conjecture (Maier-Pomerance, 1990). $J(T) = T(\log T)^{2+o(1)}$.

Random dart model prediction: $J(T) \sim T\frac{Q_T}{\phi(Q_T)} \sim e^{\gamma} T \log T$.

Maier-Pomerance: In Step 3, show that many $a_p \mod p$ can cover two remining elements; uses "twin-prime on average" results.

FGKMT: Use new prime detecting sieve (GPY-Maynard-Tao) to find $a_p \mod p$ which cover many remaining elements.

Assuming a uniform H-L prime k-tuples conjecture: Cover even more remaining elements with a_p's. Improve lower bound to $J(T) \gg T(\log T)^{1+c}$.

Open Problems

1. Select a residue $a_p \in \mathbb{Z}/p\mathbb{Z}$ for each $p \leqslant x$, let
$$S = [0, x] \setminus \bigcup_{p \leqslant x} (a_p \mod p).$$
I. When $a_p = 0$ for all p, $|S| = 1$ (extremal case).
II. A random choice yields $|S| \sim x(e^{-\gamma}/\log x)$.
III. Another construction (??) gives $|S| \sim x/\log x$.
IV. (sieve) Any choice leaves $|S| \ll x/\log x$.
Q. Can one do better than III? $|S| \geqslant (1+\delta)x/\log x$?

2. For each prime $p \leqslant \sqrt{x}$, choose a residue $a_p \mod p$, and let
$$\mathcal{S} = [0, x] \setminus \bigcup_{p \leqslant \sqrt{x}} (a_p \mod p).$$
I. When $a_p = 0$ for all p, $|S| \sim x/\log x$.
II. A random choice yields $|S| \sim x(2e^{-\gamma}/\log x)$.
Q. Are these the extreme cases?

References

- H. Cramér, *Some theorems concerning prime numbers*, Ark. Mat. Astr. Fys. **15** (1920), 1–33.
- H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 396–403.
- Andrew Granville, *Harald Cramér and the distribution of prime numbers*, Scandanavian Actuarial J. **1** (1995), 12–28.
- Alexei Kourbatov, The distribution of maximal prime gaps in Cramér's probabilistic model of primes, Inter. J. Stat. Prob. **3**, No. 2 (2014), 18–29.
- J. Pintz, *Cramér vs. Cramér. On Cramér's Probabilistic Model for primes*, Func. Approx. Comment. Math. **37** (2007), 361–376.
- T. Oliveira e Silva, *Gaps between consecutive primes* (web page), `http://sweet.ua.pt/tos/gaps.html`