Is There a Limiting Distribution of Maximal Gaps Between Primes?

Alexei Kourbatov   www.JavaScripter.net/math

Dedicated to the 50th Anniversary of the Online Encyclopedia of Integer Sequences and the 75th Birthday of N. J. A. Sloane

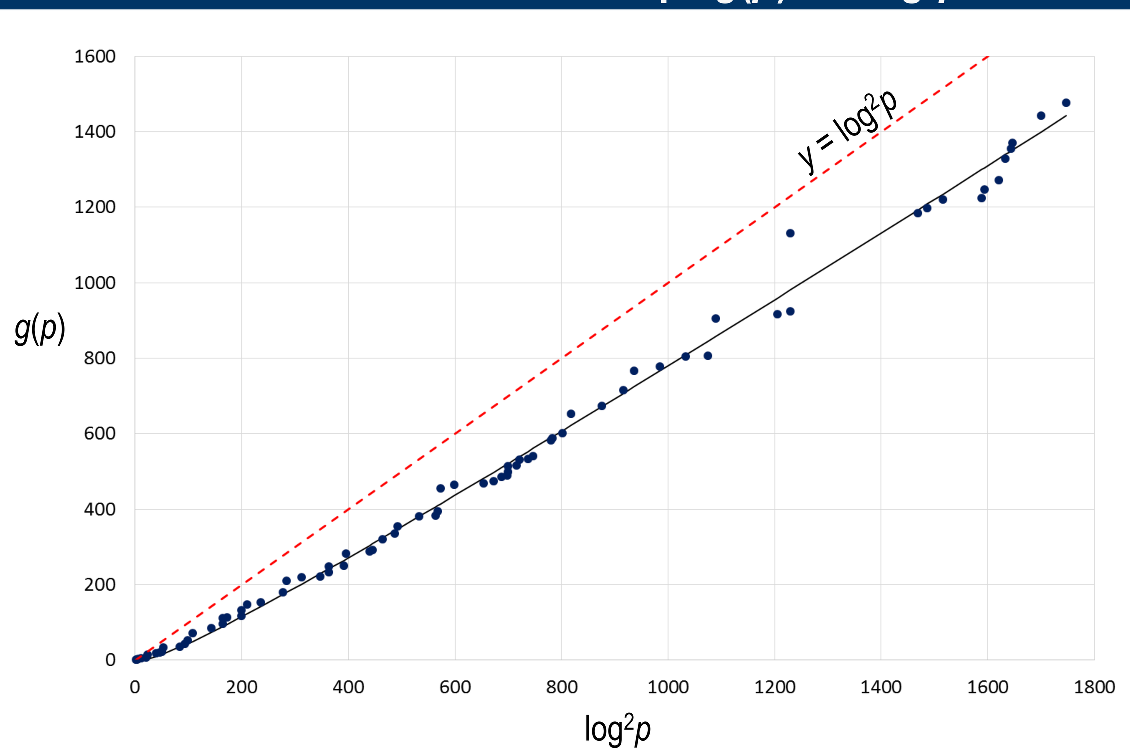
Maximal Gaps Between Primes

Prime gaps are distances between consecutive prime numbers. **Maximal prime gaps** are those greater than all preceding gaps:

Consecutive Primes		Gap	Consecutive Primes		Gap
2	3	1	25056082087	25056082543	456
3	5	2	42652618343	42652618807	464
7	11	4	127976334671	127976335139	468
23	29	6	182226896239	182226896713	474
89	97	8	241160624143	241160624629	486
113	127	14	297501075799	297501076289	490
523	541	18	303371455241	303371455741	500
887	907	20	304599508537	304599509051	514
1129	1151	22	416608695821	416608696337	516
1327	1361	34	461690510011	461690510543	532
9551	9587	36	614487453523	614487454057	534
15683	15727	44	738832927927	738832928467	540
19609	19661	52	1346294310749	1346294311331	582
31397	31469	72	1408695493609	1408695494197	588
155921	156007	86	1968188556461	1968188557063	602
360653	360749	96	2614941710599	2614941711251	652
370261	370373	112	7177162611713	7177162612387	674
492113	492227	114	13829048559701	13829048560417	716
1349533	1349651	118	19581334192423	19581334193189	766
1357201	1357333	132	42842283925351	42842283926129	778
2010733	2010881	148	90874329411493	90874329412297	804
4652353	4652507	154	171231342420521	171231342421327	806
17051707	17051887	180	218209405436543	218209405437449	906
20831323	20831533	210	1189459969825483	1189459969826399	916
47326693	47326913	220	1686994940955803	1686994940956727	924
122164747	122164969	222	1693182318746371	1693182318747503	1132
189695659	189695893	234	43841547845541059	43841547845542243	1184
191912783	191913031	248	55350776431903243	55350776431904441	1198
387096133	387096383	250	80873624627234849	80873624627236069	1220
436273009	436273291	282	203986478517455989	203986478517457213	1224
1294268491	1294268779	288	218034721194214273	218034721194215521	1248
1453168141	1453168433	292	305405826521087869	305405826521089141	1272
2300942549	2300942869	320	352521223451364323	352521223451365651	1328
3842610773	3842611109	336	401429925999153707	401429925999155063	1356
4302407359	4302407713	354	418032645936712127	418032645936713497	1370
10726904659	10726905041	382	804212830686677669	804212830686679111	1442
20678048297	20678048681	384	1425172824437699411	1425172824437700887	1476
22367084959	22367085353	394	(OEIS A000101, A002386, A005250)		

Computations of Oliveira e Silva et al. (2014) show that there are 75 maximal prime gaps below $4\cdot 10^{18}$ (given in the above table).

Plot of Maximal Prime Gaps $g(p)$ vs. $\log^2 p$



The trend (black curve) is $a(\log(\mathrm{li}\,p) - b)$, where $a = p/(\mathrm{li}\,p)$ is the estimated average gap between primes below p, and $b \approx 2.7$.

Cramér's Probabilistic Model of Primes

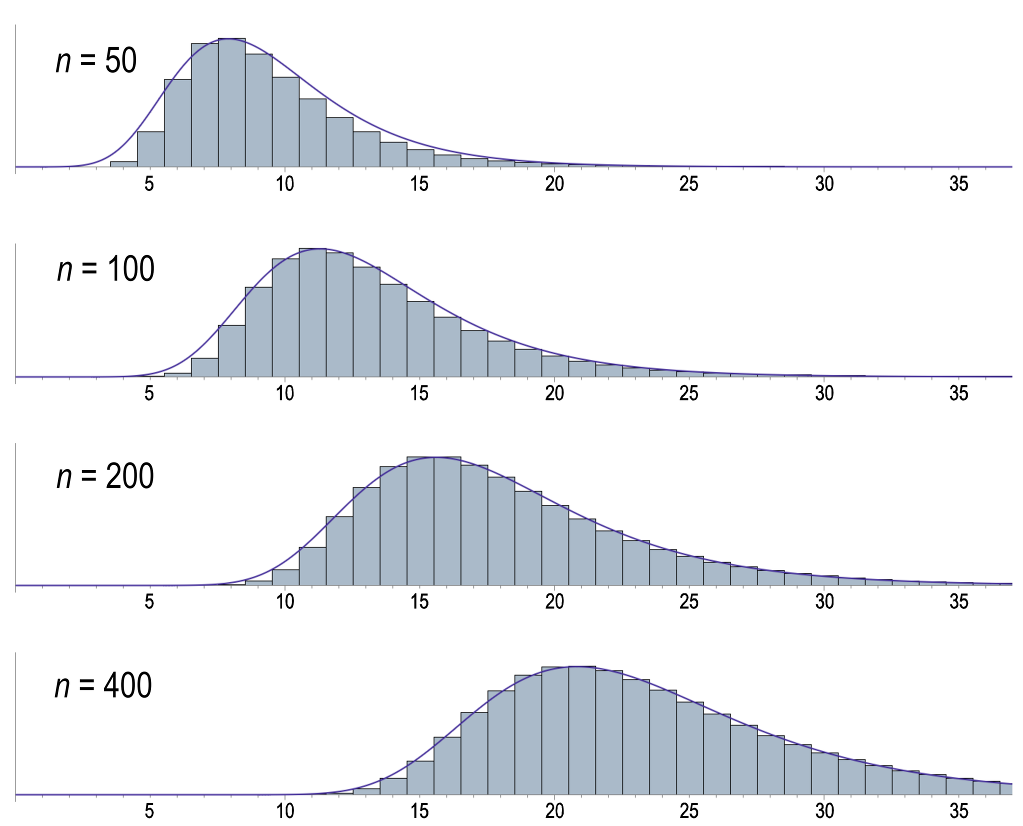
Cramér (1936) used this observation (well known since Gauss): the chance for n to be a prime is about $1/(\log n)$.

"Let $U_1, U_2, U_3, \ldots$ be an infinite series of urns containing black and white balls, the chance of drawing a white ball from U_n being $1/(\log n)$ for $n > 2$, while the composition of U_1 and U_2 may be arbitrarily chosen. We now assume that one ball is drawn from each urn, so that an infinite series of alternately black and white balls is obtained. If P_n denotes the number of the urn from which the n-th white ball in the series was drawn, the numbers $P_1, P_2, \ldots$ will form an increasing sequence of integers, and we shall consider the class C of all possible sequences $\{P_n\}$. Obviously the sequence S of ordinary prime numbers $\{p_n\}$ belongs to this class."

Thus, in Cramér's model, white balls represent "primes" and black balls represent "composites." Cramér's model is underdetermined: the content of urns U_1 and U_2 is arbitrary. We will assume that urns U_1 and U_2 never produce black balls. Still, it is not guaranteed that there are any "primes" greater than 2 at all. To circumvent this, for Cramér's model we define maximal prime gap as 1 plus the longest run of black balls (allowing runs of length zero).

Exact Distribution of Maximal Gaps in Cramér's Model

Here are exact distributions of maximal prime gaps in Cramér's model with n = 50, 100, 200, 400 urns. Smooth curves show the Gumbel distributions with scale $a = n/(\mathrm{li}\,n)$ and mode $\mu = a\log(\mathrm{li}\,n)$.



Side Note: Distribution of Maxima – the Gumbel Distribution

Suppose that a Geiger counter clicks, on average, once per second. If our observations continue for T = 1 hour, what is the longest interval of silence, without the Geiger counter clicking?

If the average time between clicks is a, then the expected total number of clicks will be $N \approx T/a$. A good way to describe the wait times t between clicks is the ***exponential distribution***, with the cumulative distribution function (cdf) $1 - e^{-t/a}$. Therefore, the distribution of the maxima of N wait times has this cdf: $(1 - e^{-t/a})^N = (1 - (1/N)e^{-(t-\mu)/a})^N$, with $\mu = a \log N \approx a \log(T/a)$. For large N, we have a limiting distribution: cdf $\exp(-e^{-(t-\mu)/a})$. This distribution of maxima is called the ***Gumbel distribution***. The distribution has two parameters: ***mode*** μ and ***scale*** a.

Maximal Gaps Between Cramér's Random "Primes" Have a Limiting Distribution – the Gumbel Distribution

A function $\ell(t) > 0$ is ***slowly varying*** if it is defined for positive t and $\ell(\lambda t)/\ell(t) \to 1$ as $t \to \infty$ for any fixed $\lambda > 0$.

Theorem. Consider a sequence of tosses of biased coins with tails probability $\ell(k)$ at the k-th toss, where $\ell(t)$ is a smooth, slowly varying, monotonically decreasing function, $0 < \ell(k) \le 1$, and $\ell(t) \to 0$ as $t \to \infty$. Then, after a large number n of tosses, the limiting distribution of the longest runs of heads R_n is the Gumbel distribution: there exist positive a_n and b_n such that

$$\lim_{n\to\infty} \mathrm{P}(R_n \le x \equiv a_n z + b_n) = \exp(-e^{-z}),$$

where

$$\text{scale } a_n \sim n / \mathrm{E}\Pi(n), \quad \text{mode } b_n \sim n \log(\mathrm{E}\Pi(n)) / \mathrm{E}\Pi(n),$$

and $\mathrm{E}\Pi(n)$ is the mathematical expectation of the total number of tails $\Pi(n)$ observed during the first n tosses:

$$\mathrm{E}\Pi(n) = \ell(1) + \ell(2) + \ldots + \ell(n).$$

Corollary. In Cramér's probabilistic model with n urns, the limiting distribution of maximal prime gaps G_n is the Gumbel distribution: there exist positive a_n and b_n such that

$$\lim_{n\to\infty} \mathrm{P}(G_n \le x \equiv a_n z + b_n) = \exp(-e^{-z}),$$

where

$$\text{scale } a_n \sim n/(\mathrm{li}\,n), \quad \text{mode } b_n \sim n \log(\mathrm{li}\,n)/(\mathrm{li}\,n),$$

$\mathrm{li}\,n$ is the logarithmic integral of n.

Mode of Maximal Prime Gaps – Probabilistic Model

For Cramér's probabilistic model with n urns, computations up to $n = 10^7$ show that the **most probable value** (**mode**) μ_n of maximal prime gaps (OEIS A235402) satisfies the inequalities

$$n \log(\mathrm{li}\,n)/(\mathrm{li}\,n) - \log n < \mu_n \le \lceil n \log(\mathrm{li}\,n)/(\mathrm{li}\,n) \rceil.$$

An easy-to-prove upper bound is $\mu_n < n\log(\mathrm{li}\,n)/(\mathrm{li}\,n) + O(\log n)$; moreover, computations even suggest $\mu_n < n\log(\mathrm{li}\,n)/(\mathrm{li}\,n) + 1$. Note that, in Cramér's model with n urns, the width of the distribution of maximal gaps is $O(\log n)$, while the above upper bound for the mode μ_n of maximal gaps is

$$\mu_n < n \log(\mathrm{li}\,n)/(\mathrm{li}\,n) + O(\log n) = \log^2 n - O(\log n \log\log n).$$

As n grows, so does the distance from μ_n to $\log^2 n$, and an ever larger percentage of maximal prime gaps end up below $\log^2 n$. When $n \to \infty$, not only do we see that Cramér's conjecture is true with probability 1, but also that 99.99+ percent of maximal prime gaps are below $\log^2 n$.

Notes: Wolf (2011) develops a more sophisticated model for prime gaps, expressing the gap sizes in terms of the prime counting function $\pi(n)$ rather than $\mathrm{li}\,n$. Interestingly, in Wolf's model we also have

$$\text{maximal prime gaps} = \log^2 n - O(\log n \log\log n)$$

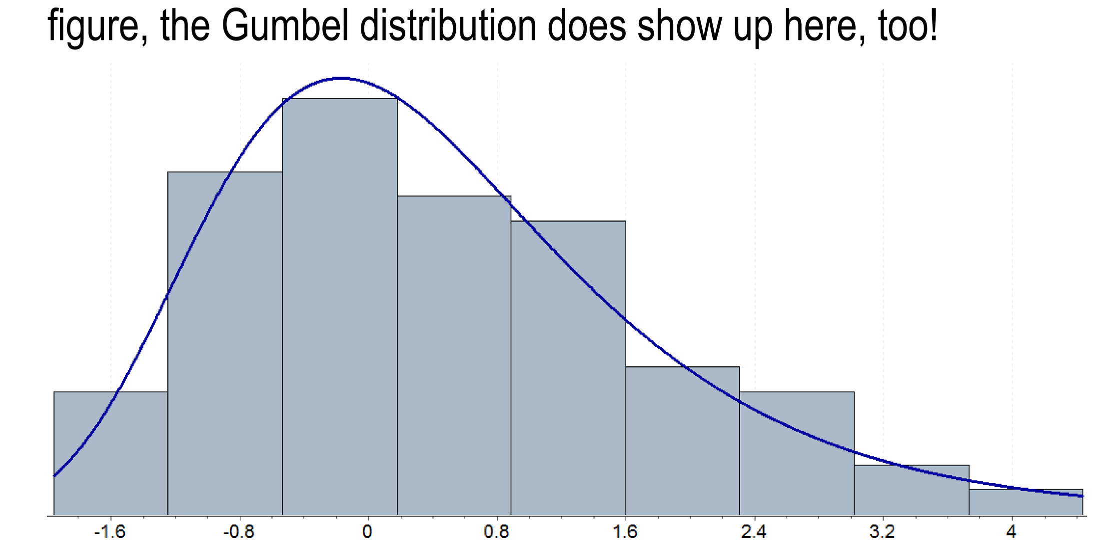
(hypothetically, save for very rare exceptional cases).

On the other hand, Granville (1995) heuristically argues that there exist unusually large gaps $g(p)$ such that

$$\limsup_{p\to\infty} g(p)/\log^2 p \ge 2e^{-\gamma} \approx 1.1229\ldots$$

Does This Work for True Primes Too?

Maximal gaps between *true primes*, after proper rescaling, are also well approximated by the Gumbel distribution. (Alas, we cannot prove much for true primes. But see the next section for additional related conjectures.) Let $a = p/(\mathrm{li}\,p) \approx \log p - 1$ denote the estimated average gap between primes up to p. For each maximal prime gap, consider the end-of-gap prime p and the gap size $g(p)$. Let us make a histogram of rescaled (*standardized*) maximal gaps: subtract the trend $a(\log(\mathrm{li}\,p) - b)$ from the actual gap sizes $g(p)$, and then divide the result by the "natural scale" a. The histogram thus illustrates how far above or below the trend curve we can find the maximal gap sizes; and most of them turn out to be within $2a$ from the trend. As we see in the following figure, the Gumbel distribution does show up here, too!



Related Conjectures for Prime Gaps

Cramér's Conjecture: Prime gaps $g(p)$ grow no faster than $\log^2 p$: we have $g(p) = O(\log^2 p)$.

Strengthened Cramér's Conjecture: If the prime gap $g(p)$ ends at the prime p, then $g(p) < \log^2 p$.

Shanks' Conjecture: If $g(p)$ is a ***maximal*** prime gap that ends at the prime p, then $g(p) \sim \log^2 p$ as $p \to \infty$.

Firoozbakht's Conjecture: If p_k is the k-th prime, then the sequence $(p_k)^{1/k}$ is decreasing. Equivalently, $(p_{k+1})^k < (p_k)^{k+1}$.

Note: We do not know if these statements are true or false! Granville (1995) casts doubt on the last three conjectures.

References

Cramér, H. (1936). On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.* 2, 23–46.

Granville, A. (1995). Harald Cramér and the distribution of prime numbers. *Scandinavian Actuarial Journal*, 1, 12–28.

Kourbatov, A. (2013). Maximal gaps between prime k-tuples: a statistical approach. *Journal of Integer Sequences*, 16, Article 13.5.2. http://arxiv.org/abs/1301.2242

Kourbatov, A. (2014). The distribution of maximal prime gaps in Cramér's probabilistic model of primes. *Int. Journal of Statistics and Probability*, 3, No. 2, 18–29. http://arxiv.org/abs/1401.6959

Oliveira e Silva, T., Herzog, S., & Pardi, S. (2014). Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4\cdot 10^{18}$, *Math. Comp.* 83, 2033–2060.

Wolf, M. (2011). Some heuristics on the gaps between consecutive primes, preprint. http://arxiv.org/abs/1102.0481