On the Distribution of Maximal Gaps Between Primes in Residue Classes

Alexei Kourbatov
www.JavaScripter.net/math
akourbatov@gmail.com

Abstract

Let $q > r \geq 1$ be coprime positive integers. We empirically study the record gaps $G_{q,r}(x)$ between primes $p \leq x$ of the form $p = qn + r$. Extensive computations suggest that $G_{q,r}(x) < \varphi(q)\log^2 x$ almost always; more precisely, $G_{q,r}(x) \sim a(\log(\operatorname{li} x/\varphi(q))+b)$, where $a = \varphi(q)x/\operatorname{li} x$ is the expected average gap between primes $p = qn + r \leq x$, and $b = O((\log q)^{1/2} \log\log x)$ is a correction term. The distribution of $G_{q,r}(x)$ near its trend is close to the Gumbel extreme value distribution. However, the question whether there exists a *limiting distribution* of $G_{q,r}(x)$ is open.

1 Introduction

Let q and r be fixed positive integers such that $1 \leq r < q$ and $\gcd(q, r) = 1$. Dirichlet proved in 1837 that the integer $qn + r$ is prime infinitely often; this is Dirichlet's theorem on arithmetic progressions. The prime number theorem tells us that the total number of primes $\leq x$ is asymptotic to $\operatorname{li} x$, i.e., $\pi(x) \sim \operatorname{li} x$. Moreover, the generalized Riemann hypothesis implies that primes are distributed approximately equally among the $\varphi(q)$ residue classes modulo q corresponding to specific values of r. Thus each residue class contains a positive proportion, about $\frac{1}{\varphi(q)}$, of all primes below x. Accordingly, the average prime gap below x is about $x/\operatorname{li} x \sim \log x$, while the average gap between primes $p = qn + r \leq x$ is about $\varphi(q)x/\operatorname{li} x \sim \varphi(q)\log x$.

In this paper we empirically study the growth and distribution of the values of the function $G_{q,r}(x)$, the record (maximal) gap between primes of the form $qn + r$ below x, for $x < 10^{12}$. The special case $G_{2,1}(x)$, i.e. the maximal gaps between primes below x, has been studied by many authors [1, 3, 4, 7, 8, 13, 14, 18, 21, 22]. Computational experiments investigating the actual distribution of (properly rescaled) values $G_{q,r}(x)$ are of interest, in part, because in Cramér's *probabilistic model of primes* [3] there exists a limiting distribution of maximal prime gaps [9], namely, the Gumbel extreme value distribution.

The fact that values of certain arithmetic functions have limiting distributions is among the most beautiful results in number theory. A well-known example is the limiting distribution of $\omega(n)$, the number of distinct prime factors of n. The *Erdős-Kac theorem* states that, roughly speaking, the values of $\omega(n)$ for $n \le x$ follow the *normal distribution* with the mean $\log\log x$ and standard deviation $\sqrt{\log\log x}$, as $x \to \infty$. Note that it is virtually impossible to observe the Erdős-Kac normal curve for $\omega(n)$ in a computational experiment because such an experiment would involve factoring a lot of gigantic values of n. By contrast, computations described in Section 3 and *Appendix* are quite manageable and do allow one to draw maximal gap histograms, which turn out to closely fit the Gumbel distribution. These results, together with [9], support the hypothesis that there is a limiting distribution of (properly rescaled) values of $G_{q,r}(x)$. However, a formal proof or disproof of existence of a Gumbel limit law for maximal gaps between primes in residue classes seems beyond reach.

2 Notation and abbreviations

p_k	the k-th prime; $\{p_k\} = \{2, 3, 5, 7, 11, \ldots\}$
$\pi(x)$	the prime counting function: the total number of primes $p_k \le x$
$\pi_{q,r}(x)$	the prime counting function in residue class r modulo q: the total number of primes $p = qn + r \le x$, $n \in \mathbb{N}^0$
$\gcd(q, r)$	the greatest common divisor of q and r
$\varphi(n)$	Euler's φ function: the number of positive $m \le n$ with $\gcd(m, n) = 1$
$G(x)$	the record (maximal) gap between primes $\le x$
$G_{q,r}(x)$	the record (maximal) gap between primes $p = qn + r \le x$
$a = a(q, x)$	the expected average gap between primes $p = qn + r \le x$; defined as $a = \varphi(q)x/\operatorname{li} x$; used in the $G_{q,r}(x)$ trend equation (1)
$b = b(q, x)$	the correction term in equation (1)
i.i.d.	independent and identically distributed
cdf	cumulative distribution function
pdf	probability density function
$\mathrm{Exp}(x; \alpha)$	the exponential distribution cdf: $\mathrm{Exp}(x; \alpha) = 1 - e^{-x/\alpha}$
$\mathrm{Gumbel}(x; \alpha, \mu)$	the Gumbel distribution cdf: $\mathrm{Gumbel}(x; \alpha, \mu) = e^{-e^{-\frac{x-\mu}{\alpha}}}$
α	the *scale parameter* of exponential/Gumbel distributions, as applicable
μ	the *location parameter* (*mode*) of the Gumbel distribution
γ	the Euler-Mascheroni constant: $\gamma = 0.57721\ldots$
$\log x$	the natural logarithm of x
$\operatorname{li} x$	the logarithmic integral of x: $\operatorname{li} x = \int_0^x \frac{dt}{\log t} = \int_2^x \frac{dt}{\log t} + 1.04516\ldots$

3 Numerical results

Using the PARI/GP program `maxgap.gp` (see *Appendix*) we have computed the values of $G_{q,r}(x)$ for many different values of q in the range from 4 to 15000 for $x < 10^{12}$. We used all admissible values of $r \in [1, q-1]$, $\gcd(q, r) = 1$, to assemble a complete data set of maximal gaps for a given q. Below we summarize these numerical results.

Table 1. Example: record gaps between primes $p = 1000n + 1$

Start of gap	End of gap (p)	Gap $G_{q,r}(p)$	Rescaled gap w, eq. (7)
3001	4001	1000	−0.7152793957
4001	7001	3000	−0.3807033581
9001	13001	4000	−0.5014278923
28001	51001	23000	3.3856905292
294001	318001	24000	1.4131420099
607001	633001	26000	1.0363075229
4105001	4132001	27000	−0.8849971087
5316001	5352001	36000	0.3933925543
14383001	14424001	41000	0.0360536511
26119001	26163001	44000	−0.2102813328
46291001	46336001	45000	−0.7494765793
70963001	71011001	48000	−0.8198719265
95466001	95515001	49000	−1.0343372505
114949001	115003001	54000	−0.5445782196
229690001	229752001	62000	−0.2996631434
242577001	242655001	78000	1.8227176550
821872001	821958001	86000	1.1782430130
3242455001	3242545001	90000	−0.2118232294
7270461001	7270567001	106000	0.5535925382
11281191001	11281302001	111000	0.5044228464
32970586001	32970700001	114000	−0.6657570997
50299917001	50300032001	115000	−1.1413988974
63937221001	63937353001	132000	0.3133056088
92751774001	92751909001	135000	0.0905709933
286086588001	286086729001	141000	−0.9100787710
334219620001	334219767001	147000	−0.5413814289
554981875001	554982043001	168000	0.7559934308
1322542861001	1322543032001	171000	−0.2444699459
2599523890001	2599524073001	183000	−0.1376484006
4651789531001	4651789729001	198000	0.3405219324
7874438322001	7874438536001	214000	0.9493170656
8761032430001	8761032657001	227000	1.9158059981

3.1 The growth trend of maximal gaps $G_{q,r}(x)$

Let us begin with a simple example. For $q = 1000$, $r = 1$, running the program `maxgap.gp` produces the results shown in Table 1. It is easy to check that all gaps $G_{q,r}(p)$ in the table satisfy the inequality $G_{q,r}(p) < \varphi(q)\log^2 p$, suggesting several possible generalizations of *Cramér's conjecture* (see Sect. 5.2).

Figure 1 shows all values of $G_{q,r}(p)$ for $q = 1000$, $\forall r \in [1, q-1]$, $\gcd(q, r) = 1$, $p < 10^{12}$. The horizontal axis $\log^2 p$ reflects the actual end-of-gap prime $p = qn + r$ of each maximal gap. The results for other values of q closely resemble Fig. 1.

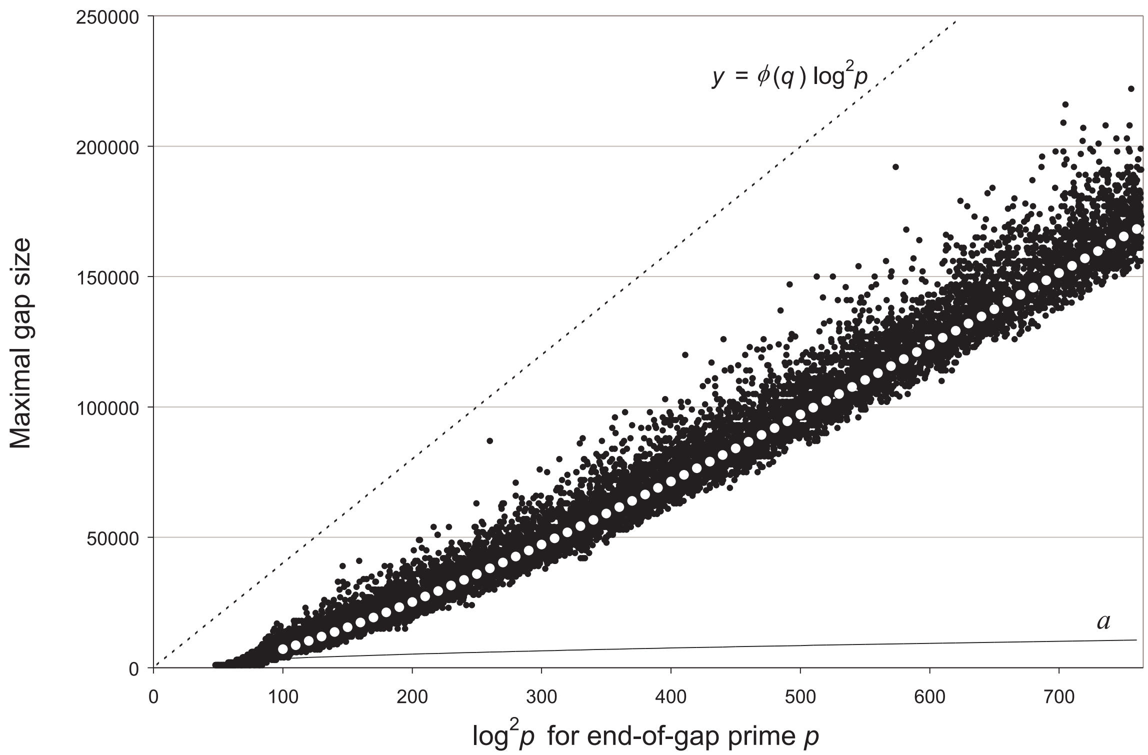

Figure 1: Record gaps $G_{q,r}(p)$ between primes $p = qn + r < 10^{12}$ in residue classes mod q, with $q = 1000$, $\varphi(q) = 400$, $\gcd(q, r) = 1$. Plotted (bottom to top): *average gaps* $a = \frac{\varphi(q)\cdot p}{\operatorname{li} p}$ between primes $\le p$ in residue classes mod q (solid curve); *trend curve* T of eq. (1) (white dotted curve); the conjectural (a.s.) upper bound for $G_{q,r}(p)$: $y = \varphi(q)\log^2 p$ (dashed line).

The vast majority of record gaps $G_{q,r}(x)$ are near a smooth trend curve T:

$$G_{q,r}(x) \;\sim\; T(q, x) = a \cdot (\log \frac{\operatorname{li} x}{\varphi(q)} + b), \tag{1}$$

where a is the *expected average gap* between primes in a residue class mod q, defined as

$$a \;=\; a(q, x) \;=\; \varphi(q)\frac{x}{\operatorname{li} x}, \tag{2}$$

and b is a correction term. Clearly, for $b = 0$ or for any negative $b = o(\log x)$, we have

$$T(q, x) \lesssim \varphi(q) \log^2 x \qquad \text{and} \qquad a(q, x) \lesssim \varphi(q) \log x \qquad \text{as } x \to \infty.$$

We can heuristically derive equations (1) and (2) as follows. *Extreme value theory* predicts that, for N consecutive events occurring at i.i.d. random intervals with cdf $\mathrm{Exp}(\xi; \alpha) = 1 - e^{-\xi/\alpha}$ (i.e. at average intervals α), the most probable maximal interval between events is about $\alpha \log N$, while the width of extreme value distribution is $O(\alpha)$; see [6]. As $N \to \infty$, the distribution of maximal intervals approaches the Gumbel distribution $\mathrm{Gumbel}(x; \alpha, \mu)$ with scale α and mode $\mu = \alpha \log N$. In eqs. (1), (2) we simply take $N = \pi_{q,r}(x) \approx \mathrm{li}\, x/\varphi(q)$ and $\alpha \approx a(q, x) \approx x/N \approx \varphi(q)x/\mathrm{li}\, x$. However, extreme value theory alone cannot accurately predict the actual behavior of the correction term b in (1). It is reasonable to expect that, similar to a, the term b is a function of both q and $\frac{x}{\mathrm{li}\, x}$ (the average prime gap below x). Indeed, empirically[1] we have

$$b = b(q, x) = c_0 - c_1 \log \frac{x}{\mathrm{li}\, x}, \tag{3}$$

where the approximate values[2] of c_0 and c_1 are

$$c_0 \approx 2.7\sqrt{\log \varphi(q)} - 1.2, \tag{4}$$

$$c_1 \approx 0.57\sqrt{\log \varphi(q)} + 1. \tag{5}$$

Computations show that formulas (1)–(5) are applicable at least for $20 \le \varphi(q) \le 15000$, $10^6 < x < 10^{12}$. (For smaller q the data become scarce.) As we can see in Figure 1, these formulas reflect the actual trend of $G_{q,r}(x)$ quite well.

Equations (3)–(5) mean that the correction term $b = b(q, x)$ is unbounded and (eventually) negative: $b(q, x) \to -\infty$ as $x \to \infty$, $q = \text{const}$. Combining (3)–(5) with the inequality $\varphi(q) < q$ for $q > 1$, we can also conclude that

$$b = b(q, x) = O(\sqrt{\log q} \log \log x). \tag{6}$$

[1] It quickly became apparent that a bounded correction term $|b| \le 1$ did ***not*** work in $G_{q,r}(x)$ trend (1). But if instead of $G_{q,r}(x)$ we look at the trend of record gaps in an increasing sequence of numbers n at *random intervals* $g = \lceil \xi \rceil$, where ξ is a random variable with cdf $\mathrm{Exp}(x; \varphi(q) \log n)$, then $|b| < 1$ in the trend equation does work. (We have verified this in computational experiments using the program `randomgap.gp`; see *Appendix*.)

[2] We leave it up to interested readers to ponder the possible expressions of c_0 and c_1 in terms of e, γ or combinations thereof. (For instance, could eq. (5) take a more precise form $c_1 = \gamma\sqrt{\log \varphi(q)} + 1$ — or possibly $c_1 = e^{-\gamma}\sqrt{\log \varphi(q)} + 1$?) However, the numerical values of c_0 and c_1 are perhaps not as interesting as the *unbounded growth* of the correction term $b = O(\sqrt{\log q} \log \log x)$ in (1). The situation here resembles Legendre's approximation to the prime counting function, $\pi(x) \sim \frac{x}{\log x - 1.08366}$ [19, A228211], where the numerical value of the constant 1.08366 is of historical interest only. Nevertheless, for the purpose of distribution fitting in Section 3.2 we do need *some* realistic formula for the $G_{q,r}(x)$ trend, and eqs. (1)–(5) do this job well, at least for $20 \le \varphi(q) \le 15000$, $10^6 < x < 10^{12}$.

3.2 The distribution of maximal gaps

We have thus found that the maximal gaps between primes in each residue class are mainly observed within a strip of increasing width $O(a)$ around the trend curve $T(q,x)$ of eq. (1), where $a = a(q,x)$ is the expected average gap between primes in the respective residue class. Now let us take a closer look at the distribution of maximal gaps in the neighborhood of this trend curve. We perform a rescaling transformation similar to [7, Sect. 5.2]: subtract the trend $T(q,x)$ from actual gap sizes, and then divide the result by the "natural unit" a. All record gap values $G_{q,r}(p)$ are mapped to *standardized values* w:

$$G_{q,r}(x) \quad \to \quad w = \frac{G_{q,r}(x) - T(q,x)}{a}, \qquad \text{where} \;\; a = a(q,x) = \frac{\varphi(q)x}{\operatorname{li} x}. \tag{7}$$

Figure 2 shows the histograms of standardized maximal gaps for $q = 10007$. (Histograms for other q look similar to Fig. 2.) We can see at once that the histograms and fitting distributions are skewed: the right tail is longer and heavier. This skewness is a well-known feature of extreme value distributions. Among all two-parameter distributions supplied by the distribution fitting software [12], the best fit is the *Gumbel distribution.* This opens up the question whether the Gumbel distribution is the limit law for properly rescaled sequences of the $G_{q,r}(x)$ values as $x \to \infty$; cf. [8, 9]. Does such a limiting distribution exist at all?

If we look at three-parameter distributions, then one of the best fits is the *Generalized Extreme Value* (GEV) distribution, which includes the Gumbel distribution as a special case. The *shape parameter* in the best-fit GEV distributions is very close to zero; note that the Gumbel distribution is a GEV distribution whose shape parameter is exactly zero.

There are notable differences between our situation shown in Fig. 2 and the distributions of maximal gaps between *prime k-tuples* shown in [7, Fig. 4] — even though in either case the Gumbel distribution is a good fit, with the majority of record gaps occurring within $\pm 2a$ of the respective trend curve.

- In our case (Figs. 1 and 2), the available data is not scarce — quite the opposite: there are thousands of data points available. On the other hand, data on maximal gaps between prime k-tuples is scarce; at present we have fewer than 100 data points for any given type of prime k-tuple [7, Tables 2–4].

- In case of maximal gaps between prime k-tuples, $k \geq 2$, a smaller correction term $|b| \leq 1$ usually works well; see [7, Sect. 4-5]. (In our `randomgap.gp` model runs, even $b \approx 0$ did work; see *Appendix.*) By contrast, for maximal gaps between primes in a residue class, the trend equations (1)–(5) involve an unbounded correction term $b = O(\sqrt{\log q} \log\log x)$.

As noted by Brent [2], primes seem to be less random than twin primes. We can add that, likewise, record gaps between primes in a residue class seem to be somewhat less random than those for prime k-tuples. Our record gaps analysis also shows that primes $p = qn + r$ do not go quite as far from each other as in the `randomgap.gp` model. Pintz [15] discusses various other aspects of the "random" and not-so-random behavior of primes.

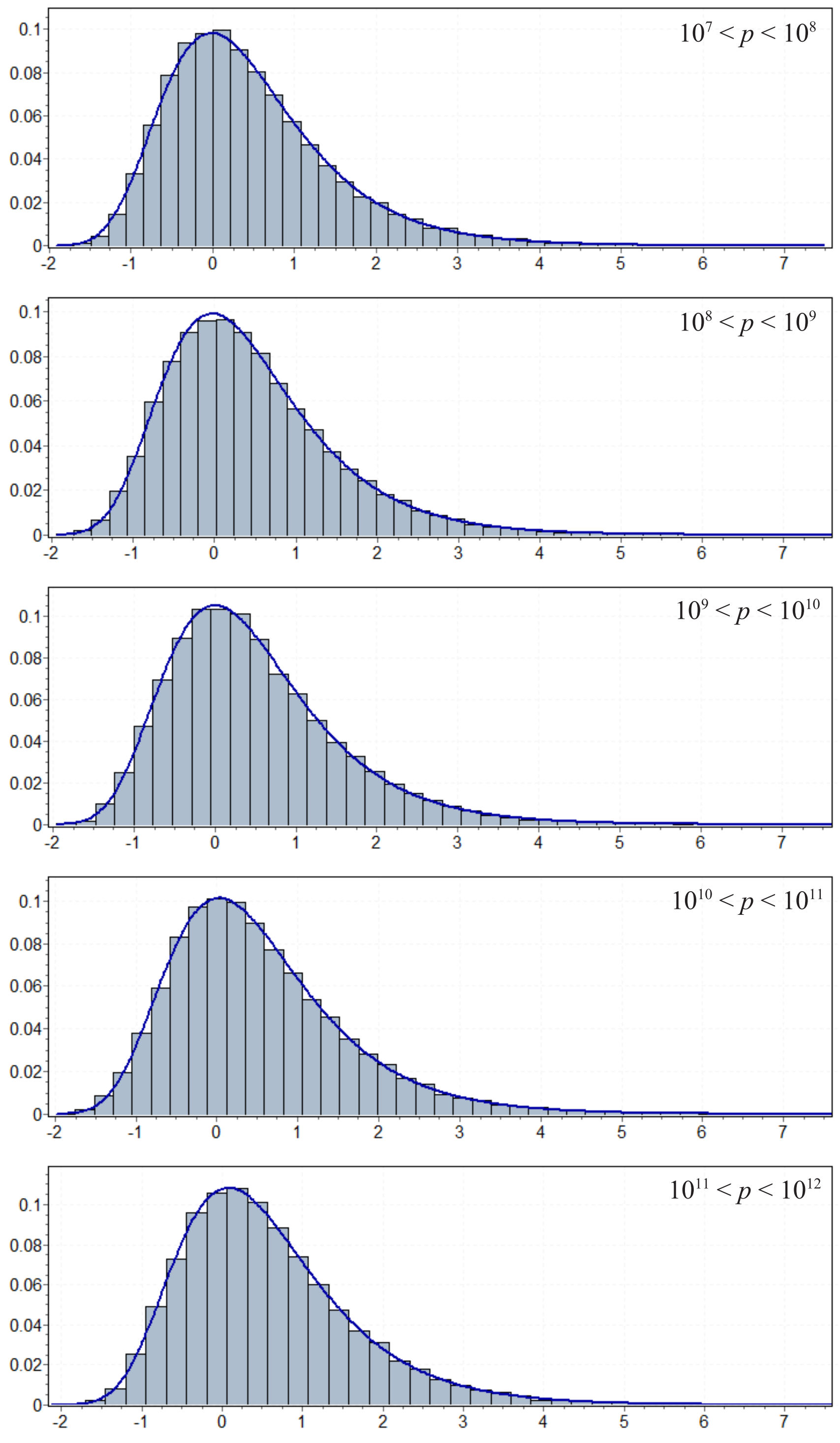

Figure 2: Histograms of rescaled maximal gaps w and fitted Gumbel distributions (pdf) for 8-, 9-, 10-, 11-, and 12-digit primes $p = qn + r$, with $q = 10007$, $1 \leq r < q$.

4 A new conjecture on $G_{q,r}(x)$ trend. Wolf's conjecture

Our extensive computations and eqs. (1)–(5) suggest the following interesting conjecture describing the growth of maximal gaps between primes in residue classes.

Conjecture on the trend of $G_{q,r}(x)$. For any $q > r \geq 1$ with $\gcd(q,r) = 1$, there exist real numbers $c_0 = c_0(q)$ and $c_1 = c_1(q)$ such that

$$G_{q,r}(x) \;\sim\; T(q,x) = \frac{\varphi(q)x}{\operatorname{li} x}\left(\log \frac{\operatorname{li} x}{\varphi(q)} - c_1 \log \frac{x}{\operatorname{li} x} + c_0\right) \qquad \text{as } x \to \infty, \tag{8}$$

and the difference $G_{q,r}(x) - T(q,x)$ changes its sign infinitely often.

On the other hand, Marek Wolf [21, 22] gives the following approximation for $G(x)$, the *maximal prime gap below* x, in terms of the prime counting function $\pi(x)$.

Wolf's conjecture:

$$G(x) \;\sim\; g(x) = \frac{x}{\pi(x)}(2\log \pi(x) - \log x + c) \qquad [21, \text{ p. } 11]. \tag{9}$$

Wolf's reasoning also implies that the difference $G(x) - g(x)$ changes its sign infinitely often.

Conjectures (8) and (9) are related. Indeed, for $q = 2$ equation (5) gives $c_1 = 1$; therefore, using the fact that $\varphi(2) = 1$ and applying the approximations

$$\frac{x}{\pi(x)} \approx \frac{x}{\operatorname{li} x} \approx \log x - 1, \qquad \log \pi(x) \approx \log \operatorname{li} x \approx \log x - \log\log x,$$

we can rewrite both (8) and (9) in this common form:

$$G_{2,1}(x) \;\sim\; T(2,x) = \log^2 x - 2\log x \log\log x + O(\log x) \qquad \text{as } x \to \infty. \tag{10}$$

So the maximal prime gap predictions (8) and (9) agree up to $O(\log x)$. Moreover, one can ensure that (8) and (9) agree up to $o(\log x)$ using an appropriate choice of c_0 for $q = 2$.

5 Generalizations of some familiar conjectures corroborated by experimental data

In this section we discuss several generalizations of familiar conjectures that are related to (and in some cases suggested by) the numerical results of Section 3. This discussion places our computational experiments in a broader context. Some of the conjectures (Sect. 5.2) have been proposed by the author earlier on the *PrimePuzzles.net* website [17]. We always assume that $1 \leq r < q$ and $\gcd(q,r) = 1$.

5.1 Generalized Riemann hypothesis

The Riemann hypothesis is a statement about non-trivial zeros of the Riemann ζ-function. An "elementary" reformulation of the RH is the *prime number theorem* with an $O(x^{1/2+\varepsilon})$ error term. It states that $\pi(x)$, the total number of primes $\le x$, satisfies

$$\pi(x) \;=\; \operatorname{li} x + O(x^{1/2+\varepsilon}) \qquad \text{for any } \varepsilon > 0, \qquad \text{as } x \to \infty.$$

The generalized Riemann hypothesis is a similar statement concerning non-trivial zeros of Dirichlet's L-functions. An "elementary" reformulation of the GRH says that $\pi_{q,r}(x)$, the total number of primes $p = qn + r \le x$, $n \in \mathbb{N}^0$, satisfies

$$\pi_{q,r}(x) \;=\; \frac{\operatorname{li} x}{\varphi(q)} + O_q(x^{1/2+\varepsilon}) \qquad \text{for any } \varepsilon > 0, \qquad \text{as } x \to \infty.$$

Roughly speaking, the GRH means that for large x the numbers $\pi_{q,r}(x)$ and $\lfloor \operatorname{li} x/\varphi(q) \rfloor$ almost agree in the left half of their digits. Thus the GRH justifies our assumption that average gaps between primes $\le x$ in each residue class mod q are nearly the same size, $\varphi(q)x/\operatorname{li} x$. Consequently, it is reasonable to expect that maximal gaps in each residue class have about the same growth trend and obey approximately the same distribution around their trend. This is indeed observed in our computational experiments.

(A weaker unconditional result similar to the GRH is the Siegel–Walfisz theorem.)

5.2 Generalizations of Cramér's conjecture

If $G(x)$ is the maximal gap between *primes below* x, Cramér [3] conjectured in the 1930s that $G(x) = O(\log^2 x)$. Clearly $G(x) = G_{2,1}(x)$ for all $x \ge 5$. We give four conjectures (from most to least plausible), each of which can be considered a generalization of Cramér's conjecture to maximal gaps $G_{q,r}(p)$ between primes in a residue class modulo q. Everywhere we assume that the maximal gap $G_{q,r}(p)$ ends at the prime p.

5.2.1 "Big-O" formulation

Generalized Cramér conjecture A. For any $q > r \ge 1$ with $\gcd(q, r) = 1$, we have

$$G_{q,r}(p) \;=\; O(\varphi(q) \log^2 p).$$

This weakest generalization of Cramér's conjecture seems quite likely to be true.

5.2.2 "Almost-all" formulation

Generalized Cramér conjecture B. *Almost all* maximal gaps $G_{q,r}(p)$ satisfy

$$G_{q,r}(p) \;<\; \varphi(q) \log^2 p \qquad \text{for any } q > r \ge 1 \text{ with } \gcd(q, r) = 1.$$

For $q = 2$ and $r = 1$, the above two generalizations of Cramér's conjecture are compatible with the heuristics of Granville [5]. On the other hand, the stronger formulations in subsections 5.2.3 and 5.2.4 contradict Granville's heuristic reasoning which suggests

$$\limsup_{p\to\infty} \frac{G_{2,1}(p)}{\log^2 p} \ \geq\ 2e^{-\gamma} = 1.12291\ldots \qquad [5,\text{ p. }24].$$

5.2.3 "Limit superior" formulation

Generalized Cramér conjecture C. For any $q > r \geq 1$ with $\gcd(q, r) = 1$, we have

$$\limsup_{p\to\infty} \frac{G_{q,r}(p)}{\varphi(q)\log^2 p} = 1.$$

5.2.4 "Naive" formulation

Generalized Cramér conjecture D. For any $q > r \geq 1$ with $\gcd(q, r) = 1$, we have

$$G_{q,r}(p) \ <\ \varphi(q)\log^2 p \qquad \text{if } p \text{ is large enough.}$$

Here p is the prime at the end of the maximal gap $G_{q,r}(p)$.

The latter generalization of Cramér's conjecture seems a little far-fetched. Nevertheless, we are yet to see even a single counterexample with $G_{q,r}(p) > \varphi(q)\log^2 p$. In February 2016, conjecture D has been posted at the website *PrimePuzzles.net* [17]. This website is frequented by many computation-savvy people; yet no one came up with a counterexample.

5.3 Generalizations of Shanks conjecture

As above, let $G(p)$ be the maximal prime gap ending at the prime p. Shanks [18] conjectured that the infinite sequence of maximal prime gaps satisfies the asymptotic equivalence

$$G(p) \sim \log^2 p \qquad \text{as } p \to \infty \qquad [18,\text{ p. }648].$$

Our numerical experiments suggest the following natural generalizations of the Shanks conjecture to describe the behavior of $G_{q,r}(p)$.

5.3.1 "Almost-all" formulation

Generalized Shanks conjecture I. For any $q > r \geq 1$ with $\gcd(q, r) = 1$, there exists an infinite sequence S that comprises *almost all* maximal gaps $G_{q,r}(p)$ such that every gap in S satisfies the asymptotic equality

$$G_{q,r}(p) \ \sim\ \varphi(q)\log^2 p \qquad \text{as } p \to \infty.$$

For $q = 2$ and $r = 1$, this formulation is compatible with heuristics of Granville [5] implying that there should exist an *exceptional infinite subsequence* of $G_{2,1}(p)$ satisfying

$$G_{2,1}(p) \;\sim\; M\log^2 p \qquad \text{for some } M \ge 2e^{-\gamma} > 1.$$

Indeed, our "almost-all" formulation simply means that any exceptional subsequence is very thin; that is, a zero proportion of maximal gaps $G_{2,1}(p)$ have exceptionally large sizes predicted by Granville.

5.3.2 Strong formulation

Generalized Shanks conjecture II. *All* maximal gaps $G_{q,r}(p)$ satisfy

$$G_{q,r}(p) \;\sim\; \varphi(q)\log^2 p \qquad \text{as } p \to \infty.$$

This strong formulation contradicts Granville's heuristics cited above.

5.4 Generalizations of Firoozbakht's conjecture

Firoozbakht [16] conjectured that $(p_k^{1/k})_{k\in\mathbb{N}}$ is a decreasing sequence. Equivalently,

$$p_{k+1} \;<\; p_k^{1+1/k} \qquad \text{for all } k \ge 1. \tag{11}$$

The conjecture has been verified for all $p_k < 4\cdot 10^{18}$ [10]. An independent verification was also performed by Wolf (unpublished); see [20]. Firoozbakht's conjecture implies

$$p_{k+1} - p_k \;<\; \log^2 p_k - \log p_k - 1 \qquad \text{for all } k > 9 \qquad [11, \text{Th. } 1]. \tag{12}$$

5.4.1 The Sun–Firoozbakht conjecture for primes in residue classes

Z.-W. Sun [20, Conjecture 2.3] generalized Firoozbakht's conjecture (11) as follows:

Let $q > r \ge 1$ be positive integers with r odd, q even and $\gcd(r, q) = 1$. Denote by $p_n(r, q)$ the n-th prime in the progression r, $r + q$, $r + 2q \dots$ Then there exists $n_0(r, q)$ such that the sequence $(p_n(r, q)^{1/n})_{n\ge n_0(r,q)}$ is strictly decreasing. In particular, one can take $n_0(r, q) = 2$ for $q \le 45$.

Remark. If the latter conjecture is true, $n_0(r, q)$ may be quite large. For example, with $q = 1168$ and $r = 141$ we must have $n_0(r, q) > 1893$. Indeed,

$$17010893^{1/1893} < 17163901^{1/1894}$$

(17010893 and 17163901 are the 1893rd and 1894th primes $p \equiv 141 \bmod 1168$).

Assuming the GRH and reasoning along the lines of [11], we can prove that the Sun-Firoozbakht conjecture implies

$$\frac{p_{n+1}(r, q) - p_n(r, q)}{\varphi(q)} \;<\; \log^2 p_n(r, q) - \log p_n(r, q) - 1 \qquad \text{for all } n \text{ large enough.} \tag{13}$$

This in turn implies the generalized Cramér conjectures 5.2.1, 5.2.2, and 5.2.4, as well as a modified form of 5.2.3 with $\limsup\limits_{p\to\infty} G_{q,r}(p)/(\varphi(q)\log^2 p) \le 1$.

5.4.2 "Almost all" form of the Sun–Firoozbakht conjecture

The Sun–Firoozbakht conjecture of sect. 5.4.1 might be excessively strong. A weaker, more plausible statement is the "almost all" formulation below.

For any coprime $q > r \ge 1$, the endpoints $p_n(r,q)$, $p_{n+1}(r,q)$ of ***almost all*** record gaps $G_{q,r}(x)$ satisfy inequality (13).

Here, again, $p_n(r,q)$ denotes the n-th prime in the arithmetic progression r, $r+q$, $r+2q\ldots$

6 Appendix: Details of computational experiments

Interested readers can reproduce and extend our results using the programs below.

6.1 PARI/GP program maxgap.gp

```
\\ Usage example: q=1000;for(r=1,q-1,if(gcd(q,r)==1,maxgap(q,r,1e12)))

default(realprecision,11)

\\ li(x) computes the logarithmic integral of x
li(x) = return(real(-eint1(-log(x))))

\\ pmin(q,r) computes the least prime p = qn + r, for n=0,1,2,3,...
pmin(q,r) = forstep(p=r,1e11,q, if(isprime(p), return(p)))

\\ maxgap(q,r,end) computes maximal gaps g between primes p = qn + r
\\ as well as rescaled gap values (w and h).
\\ Results are written on screen and in the c:\xgap folder.
\\ Computation ends when primes exceed the end parameter.
maxgap(q,r,end) = {
  re=0;
  s=pmin(q,r);
  t=eulerphi(q);
  SqrtLogPhi=sqrt(log(t));
  while(s<end,
    m=s+re; p=m+q;
    while(!isprime(p), p+=q);
    while(!isprime(m), m-=q);
    g=p-m;
    if(g>re,
      re=g; Lip=li(p); a=t*p/Lip;
      h=g/a-log(Lip/t);
      w=g/a-log(Lip/t)+(0.57*SqrtLogPhi+1)*log(p/Lip)-2.7*SqrtLogPhi+1.2;
      f=ceil(log(p)/log(10));
      write("c:\\xgap\\1e"f"_"q".txt",
            w"  "h"  "g" "m" "p" q="q" r="r);
      print(w"  "h"  "g" "m" "p" q="q" r="r);
      if(g/t>log(p)^2, write("c:\\xgap\\1e"f"_"q".txt","extra large"));
    );
    s=p;
  )
}
```

6.2 PARI/GP program randomgap.gp

```
\\ Usage example: q=1000;for(r=1,q-1,if(gcd(q,r)==1,randomgap(q,r,1e10)))

default(realprecision,11)

\\ exprv(m) returns an exponential random variable with mean m
exprv(m) = return(-m*log(random(1.0)))

\\ randomgap(q,r,end) writes to c:\ygap\ a set of files with record gaps
\\ in a growing sequence of integers p separated by "random" gaps which
\\ are exponentially distributed with mean m = phi(q)*log(p).
\\ The parameter r is included to mimic maxgap(q,r,end).
randomgap(q,r,end) = {
  re=0; p=max(2,r); t=eulerphi(q);
  while(p<end,
    g=ceil(exprv(t*log(p)));
    s=p;
    p+=g;
    if(g>re,
      re=g;
      Lip=real(-eint1(-log(p)));  \\li(p)
      a=t*p/Lip;
      h=g/a-log(Lip/t);
      f=ceil(log(p)/log(10));
      write("c:\\ygap\\rand"f"_"q".txt",
            h"  "g" "s" "p" q="q" r="r);
      print(h"  "g" "s" "p" q="q" r="r);
    )
  )
}
```

6.3 Notes on distribution fitting

To study the distributions of standardized maximal gaps $G_{q,r}(p)$ we used the distribution fitting software `EasyFit` [12]. Data files created with `maxgap.gp` or `randomgap.gp` are easily imported into `EasyFit`: from the *File* menu, choose *Open*, select the data file, then specify *Field Delimiter = space*, click *Update*, then *OK*.

Caution: PARI/GP writes real numbers near zero in a mantissa-exponent format *with a space* preceding the exponent (e.g. `1.7874829515 E-5`), whereas `EasyFit` expects such numbers *without a space* (e.g. `1.7874829515E-5`). Therefore, before importing into `EasyFit`, search the data files for `" E-"` and replace all occurrences with `"E-"`.

7 Acknowledgments

I am grateful to all contributors and editors of the websites *OEIS.org* and *PrimePuzzles.net*. Thanks also to Marek Wolf and Zhi-Wei Sun for putting forward many interesting conjectures [20, 21, 22]. I am especially grateful to Prof. Wolf for his very useful comments via email.

References

[1] R. C. Baker, G. Harman, J. Pintz, The difference between consecutive primes, II. *Proceedings of the London Mathematical Society.* **83** (3) (2001), 532–562.

[2] R. P. Brent, Twin primes (seem to be) more random than primes, a talk at *The Second Number Theory Down Under Conference*, Newcastle, NSW, Australia, Oct. 24-25, 2014. `http://maths-people.anu.edu.au/~brent/pd/twin_primes_and_primes.pdf`

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

[4] K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, Long gaps between primes, arXiv preprint arXiv:1412.5029 (2014). `http://arxiv.org/abs/1412.5029`

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

[6] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, 1958. Dover, 2004.

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

[8] A. Kourbatov, Is there a limiting distribution of maximal gaps between primes? Poster presentation at *DIMACS Conference on Challenges of Identifying Integer Sequences*, Oct. 9-10, 2014. `http://www.javascripter.net/math/publications/AKourbatovOEIS50poster.pdf`

[9] A. Kourbatov, The distribution of maximal prime gaps in Cramér's probabilistic model of primes, *Int. Journal of Statistics and Probability* **3** (2) (2014), 18–29. arXiv:1401.6959

[10] A. Kourbatov, Verification of the Firoozbakht conjecture for primes up to four quintillion, *Int. Math. Forum* **10** (2015), 283–288. `http://arxiv.org/abs/1503.01744`

[11] A. Kourbatov, Upper bounds for prime gaps related to Firoozbakht's conjecture, *Journal of Integer Sequences* **18** (2015), Article 15.11.2. `http://arxiv.org/abs/1506.03042`

[12] MathWave Technologies, *EasyFit — Distribution Fitting Software*, the company web site at `http://www.mathwave.com/easyfit-distribution-fitting.html` (2016).

[13] T. R. Nicely, First occurrence prime gaps, preprint, 2014. Available at `http://www.trnicely.net/gaps/gaplist.html`.

[14] T. Oliveira e Silva, S. Herzog, and S. Pardi, Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$, *Math. Comp.* **83** (2014), 2033-2060. http://www.ams.org/journals/mcom/2014-83-288/S0025-5718-2013-02787-1/

[15] J. Pintz, Cramér vs Cramér: On Cramér's probabilistic model of primes. *Functiones et Approximatio*, **37** (2) (2007), 361–376.

[16] C. Rivera (ed.), Conjecture 30. The Firoozbakht Conjecture, 2002.
Available at http://www.primepuzzles.net/conjectures/conj_030.htm.

[17] C. Rivera (ed.), Conjecture 77. Gaps between primes of the form $p = qn + r$, 2016.
Available at http://www.primepuzzles.net/conjectures/conj_077.htm.

[18] D. Shanks, On maximal gaps between successive primes. *Math. Comput.* 18 (1964), 646–651.

[19] N. J. A. Sloane (ed.), *The On-Line Encyclopedia of Integer Sequences*, 2015. Published electronically at http://oeis.org/.

[20] Z.-W. Sun, Conjectures involving arithmetical sequences. In: S. Kanemitsu, H. Li, J. Liu (eds.) *Number Theory: Arithmetic in Shangri-La. Proc. 6th China-Japan Seminar* (Shanghai, August 15-17, 2011), World Sci., Singapore, 2013, pp. 244–258. http://arxiv.org/abs/1208.2683

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

[22] M. Wolf, Nearest neighbor spacing distribution of prime numbers and quantum chaos, *Phys. Rev. E* **89**, 022922 (2014). http://arxiv.org/abs/1212.3841

2010 *Mathematics Subject Classification*: 11N05.
Keywords: Cramér conjecture, prime gap, residue class, Shanks conjecture.

(Concerned with sequences A005250, A084162, A235402, A268799, A268925, A268928, A268984, A269234, A269238, A269261, A269420, A269424, A269513, A269519.)
