Nonuniform Distributions of Patterns of Sequences of Primes in Prime Moduli

David Wu

Abstract

For positive integers q, Dirichlet's theorem states that there are infinitely many primes in each reduced residue class modulo q. Extending a proof of Dirichlet's theorem shows that the primes are equidistributed among the $\varphi(q)$ reduced residue classes modulo q. This project considers patterns of sequences of consecutive primes $(p_n, p_{n+1}, \ldots, p_{n+k})$ modulo q. Numerical evidence suggests a preference for certain prime patterns. For example, computed frequencies of the pattern (a, a) modulo q up to x are much less than the expected frequency $\pi(x)/\varphi(q)^2$. We begin to rigorously connect the Hardy-Littlewood prime k-tuple conjecture to a conjectured asymptotic formula for the frequencies of prime patterns modulo q. We extend a data gathering procedure to estimate prime patterns up to 10^{18}, an improvement of 8 orders of magnitude over previous methods. Using the extended range of data, a possible lower order term in the conjectured formula is identified via curve fitting. We begin to extend a numerical model to reduce the uncertainty in the predictions of these biases in prime patterns. The improved numerical could guide future progress towards understanding implications of the Hardy-Littlewood prime k-tuple conjecture.

1 Introduction

Analytic number theory uses real and complex analysis techniques to prove properties about the integers. The link between analysis and prime numbers is perhaps surprising; after all, analysis deals with continuous variables, while prime numbers are discrete. However, many properties of prime numbers are encoded in the properties of special functions. For example, the behavior of the zeroes of the Riemann zeta function strengthens a famous asymptotic formula known as the Prime Number Theorem [1]. The Riemann Hypothesis, one of the most well-known open problems in number theory, conjectures that all nontrivial zeroes of the Riemann zeta function have real part $\frac{1}{2}$; the Riemann Hypothesis would imply a stronger form of the Prime Number Theorem.

However, the Riemann zeta function is only one of a more general class of functions, the Dirichlet L-functions. Peter Dirichlet [2] used these L-functions to prove that arithmetic progressions with coprime first term and common difference contain an infinite number of primes: this is Dirichlet's theorem. Dirichlet's use of L-functions invoked the realm of analysis to prove statements about integers, thus beginning the study of analytic number theory.

The $\varphi(d)$ classes of residues modulo d coprime to d are referred to as the reduced residue classes, where $\varphi(n)$ is Euler's totient function. For example, the set of residues congruent to 1 modulo 4 is a reduced residue class. Applying Dirichlet's theorem to the arithmetic progression with first term 1 and common difference 4 shows that there are infinitely many primes in the reduced residue class 1 modulo 4. A natural followup question asks how prime sequences are distributed among the reduced residue classes modulo d.

Before we discuss the distribution of primes among reduced residue classes, we introduce a few key concepts and definitions in analytic number theory. Let $\pi(x)$ be the prime counting function, i.e. the number of primes less than or equal to x. Furthermore, let $p(x) \sim q(x)$ denote asymptotic equivalence, i.e. $\lim_{x\to\infty} \frac{p(x)}{q(x)} = 1$. We also make extensive use of big $\mathcal{O}$ notation. We say $f(x) = \mathcal{O}(g(x))$ if there exists some absolute constant C such that $|f(x)| \leq C|g(x)|$ for sufficiently large x. The similar notation $f(x) = \mathcal{O}_n(g(x))$ means the constant C in the definition of $\mathcal{O}(g(x))$ depends on n.

A key concept in analytic number theory is to compare a discrete function such as $\pi(x)$ to a continuous function such as the logarithmic integral $\mathrm{li}(x) = \int_2^x \frac{dt}{\log t}$. The famous Prime Number Theorem (PNT) states that $\pi(x) \sim \mathrm{li}(x)$, and Schoenfeld [3] showed that the Riemann Hypothesis (RH) implies that $|\pi(x) - \mathrm{li}(x)| < \frac{\sqrt{x}\log x}{8\pi}$, for $x \geq 2657$.

We introduce notation analogous to $\pi(x)$ for the purposes of this discussion following Lemke Oliver and Soundararajan's notation [4]. Let p_n refer to the nth prime when the primes are listed in increasing order, the pattern $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ be a vector of length k,

and $q \geq 3$ be a positive integer. Define

$$\pi(x; q, \mathbf{a}) = \#\{p_n \leq x : p_{n+i-1} \equiv a_i \pmod{q} \text{ for } 1 \leq i \leq k\}.$$

This notation counts the number of consecutive prime sequences that follow the pattern $\mathbf{a}$ modulo q. Using this notation, the PNT for arithmetic progressions applied to the simple case where $\mathbf{a} = (a)$ yields

$$\pi(x; q, a) \sim \frac{\mathrm{li}(x)}{\varphi(q)}. \tag{1.1}$$

Although (1.1) shows that primes are roughly equidistributed among the reduced residue classes modulo q, Chebyshev [5] observed that there are almost always more primes of the form $4k+3$ than of the form $4k+1$; this bias was explained by Rubinstein and Sarnak [6] to arise from the error term of $\mathcal{O}(x^{1/2+\epsilon})$ in the PNT when assuming the RH. Chebyshev's bias is one of the first mentions of nonuniform behavior of the primes when reduced modulo q.

Larger biases manifest when the length of $\mathbf{a}$ is greater than or equal to 2 that cannot be solely attributed to error terms of size $\mathcal{O}(x^{1/2+\epsilon})$. In [4] the frequencies of consecutive prime pairs modulo 10 are tabulated, and it was observed that $\pi(10^8; 10, (1,1)) \approx 4.62 \times 10^6$ and $\pi(10^8; 10, (9,1)) \approx 7.99 \times 10^6$, both of which are very different than the expected frequency of $10^8/\varphi(10)^2 = 6.25 \times 10^6$ predicted by naively generalizing (1.1) by replacing $\varphi(q)$ with $\varphi(q)^2$.

While it is known that primes are roughly equidistributed among reduced residue classes according to (1.1), it is not known whether for arbitrary $\mathbf{a}$ with the length of $\mathbf{a}$ at least 2, the modified prime counting function $\pi(x; q, \mathbf{a})$ tends to infinity as x tends to infinity. Shiu [7] proved that $\pi(x; q, (a, a, \ldots, a))$ tends to infinity as x tends to infinity, and Maynard [8] strengthened this to $\pi(x; q, (a, a, \ldots, a)) > C\pi(x)$ for some constant C and sufficiently large x.

We explain the preferences for certain prime patterns by appealing to conjectural statements similar in nature to the PNT. For example, the Hardy-Littlewood prime k-tuple conjecture states the density of specific tuples such as twin primes $(p, p+2)$ and twin sexy primes $(p, p+2, p+6)$ in a form analogous to that of the PNT. By appropriately combining specific cases of the Hardy-Littlewood prime k-tuple conjecture, we obtain conjectures about the density of the patterns modulo q.

As an example of how specific prime tuples relate to prime patterns, consider $q = 3$ and the pattern $\mathbf{a} = (1,1)$. Then we are restricting our consideration to consecutive primes (p_1, p_2) where $p_1 \equiv p_2 \equiv 1 \pmod 3$. For $m \equiv 1 \pmod 3$, these patterns include specific tuples of the form $(m, m+6), (m, m+12), (m, m+18)$, and so on. The densities of these specific tuples can be analyzed with the Hardy-Littlewood prime k-tuple conjectures. In this manner, we obtain conjectures that partially account for the observed preferences for certain

patterns modulo q.

Lemke Oliver and Soundararajan [4] provide a conjectural explanation for the biases for certain prime patterns. However, their heuristic argument omits lower order terms that cause their conjectured form to not be in agreement with the data at smaller values of x. We expand the conjecture to include further terms and begin rigorously connecting the Hardy-Littlewood prime k-tuple conjecture and the main conjecture in [4].

In Section 2, we lay out the definitions and notation important to our discussion. In Section 3, we determine the lower order terms by tightening the asymptotics in the heuristic in [4]. In Section 4, we account for discarded terms in the asymptotic formula for the conjectured behavior to extend the form of an integral to more closely fit the actual behavior of prime patterns and extend our data gathering capabilities by 8 orders of magnitude. We identify a plausible lower order term for the conjectured formula. In Section 5, we conclude and pose questions for future investigation.

2 Preliminaries

We begin with the statement of the Hardy-Littlewood prime k-tuple conjecture. Heuristically, the conjecture generalizes the PNT by assuming the probability of an integer n being prime as roughly $\frac{1}{\log n}$. While the integrand is derived by assuming primality is independent, the constant in front of the integral corrects for this assumption.

Conjecture 2.1 (Hardy-Littlewood prime k-tuple conjecture)**.** *Let $\mathcal{H}$ be a finite set of nonnegative integers and $\pi(x, \mathcal{H})$ denote the number of integers $n \leq x$ such that $n + h$ is a prime for all h in $\mathcal{H}$. Furthermore, let $\nu_p(\mathcal{H})$ denote the number of residue classes occupied by the members of $\mathcal{H}$ modulo p. Then we have that*

$$\pi(x, \mathcal{H}) = \mathfrak{S}(\mathcal{H}) \int_2^x \frac{dt}{(\log t)^{|\mathcal{H}|}} + \mathcal{O}(x^{1/2+\epsilon}),$$

where the singular series is defined as

$$\mathfrak{S}(\mathcal{H}) = \prod_{p \text{ prime}} \frac{1 - \frac{\nu_p(\mathcal{H})}{p}}{(1 - \frac{1}{p})^{|\mathcal{H}|}}.$$

The singular series is modified in [9] to an inclusion-exclusion form

$$\mathfrak{S}_0(\mathcal{H}) = \sum_{\mathcal{T} \subset \mathcal{H}} (-1)^{|\mathcal{H} \setminus \mathcal{T}|} \mathfrak{S}(\mathcal{T}).$$

In [4], Lemke Oliver and Soundararajan modify the singular series to range over primes p not dividing q to account for the prime patterns modulo q as follows.

Definition 2.1. The modified singular series $\mathfrak{S}_q(\mathcal{H})$ is defined to be

$$\mathfrak{S}_q(\mathcal{H}) = \prod_{p \nmid q} \frac{1 - \frac{\nu_p(\mathcal{H})}{p}}{(1 - \frac{1}{p})^{|\mathcal{H}|}}.$$

Lemke Oliver and Soundararajan [4] introduce the same inclusion-exclusion form $\mathfrak{S}_{q,0}$ involving alternating sums of $\mathfrak{S}_q$ is defined to introduce cancellations that lead to Conjecture 2.2.

Let $q \geq 3$ be a positive integer and a and b be reduced residue classes modulo q. Set $h \equiv b - a \pmod{q}$. Also, let p_n be the nth prime. We are specifically interested in the case where $p_n \equiv a \pmod{q}$ and $p_{n+1} = p_n + h$; this guarantees $p_{n+1} \equiv b \pmod{q}$. Let $1_{\mathcal{P}}(x)$ be the prime indicator function, defined to be 1 if x is prime and 0 otherwise; Lemke Oliver and Soundararajan [4] start with the statement that

$$\pi(x; q, a, b) = \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} 1_{\mathcal{P}}(n) 1_{\mathcal{P}}(n+h) \prod_{\substack{0 < t < h \\ (t+a,q)=1}} (1 - 1_{\mathcal{P}}(n+t)). \tag{2.1}$$

Following a series of manipulations and using a conjecture similar to the Hardy-Littlewood prime k-tuple conjecture, they conjecture the following asymptotic for $\pi(x; q, (a, b))$ (see [4, E4449–E4450] for more details).

Conjecture 2.2 (Lemke Oliver & Soundararajan [4])**.** *Let*

$$\alpha(y) = 1 - \frac{q}{\varphi(q) \log y} \quad \text{and} \quad \epsilon_q(a, b) = \#\{0 < t < h : (t + a, q) = 1\} - \frac{\varphi(q)}{q} h.$$

Then

$$\pi(x; q, (a, b)) \sim \frac{1}{q} \int_2^x \alpha(y)^{\epsilon_q(a,b)} \left(\frac{q}{\varphi(q)\alpha(y) \log y} \right)^2 \mathcal{D}(a, b; y) dy,$$

where $\mathcal{D}(a, b; y)$ is defined to be

$$\sum_{\substack{h > 0 \\ h \equiv b-a \pmod{q}}} \sum_{\mathcal{A} \subset \{0,h\}} \sum_{\substack{\mathcal{T} \subset [1,h-1] \\ (t+a,q)=1 \ \forall t \in \mathcal{T}}} (-1)^{|\mathcal{T}|} \mathfrak{S}_{q,0}(\mathcal{A} \cup \mathcal{T}) \left(\frac{q}{\varphi(q)\alpha(y) \log y} \right)^{|\mathcal{T}|} \alpha(y)^{h\varphi(q)/q}.$$

We analyze the growth of $\mathcal{D}(a, b; y)$. For readability purposes, define $\log_k x$ to be $\underbrace{\log \log \ldots \log}_{k \text{ logs}} x$, where $\log x$ is the natural logarithm.

3 A Closer Analysis of the Conjecture

We provide more precise asymptotics for $\pi(x;q,(a,b))$ as in Conjecture 2.2. Because $q=2$ is trivial, we only consider the case where q is an odd prime. However, the results readily generalize to composite q. In particular, we are interested in $\mathcal{D}(a,b;y)$, which is equal to

$$\sum_{\substack{h>0\\ h\equiv b-a \pmod q}} \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\substack{\mathcal{T}\subset[1,h-1]\\ (t+a,q)=1\ \forall t\in\mathcal{T}}} (-1)^{|\mathcal{T}|}\mathfrak{S}_{q,0}(\mathcal{A}\cup\mathcal{T})\left(\frac{q}{\varphi(q)\alpha(y)\log y}\right)^{|\mathcal{T}|}\alpha(y)^{h\varphi(q)/q}, \tag{3.1}$$

in accordance with [4]. Lemke Oliver and Soundararajan heuristically argue that the relevant terms in (3.1) are those where $\mathcal{A}=\mathcal{T}=\varnothing$ and $|\mathcal{A}|+|\mathcal{T}|=2$.

We convert (3.1) into a form more friendly to partitioning by the size of $\mathcal{T}$. Define for convenience

$$z=z(q,y)=\frac{q}{\varphi(q)\alpha(y)\log y}$$

and

$$g=g(q,y)=\alpha(y)^{\varphi(q)/q}.$$

We rewrite the innermost sum of (3.1) as a sum over ℓ element subsets of $[1,h-1]$ where ℓ ranges from 0 to $h-1$ to obtain

$$\mathcal{D}(a,b;y)=\sum_{\substack{h>0\\ h\equiv b-a \pmod q}} g^h \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\ell=0}^{h-1}(-z)^\ell \sum_{\substack{\mathcal{T}\subset[1,h-1]\\ (t+a,q)=1\ \forall t\in\mathcal{T}\\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\mathcal{A}\cup\mathcal{T}). \tag{3.2}$$

Evaluating (3.2) is difficult because the terms are unwieldy when h is large. However, recalling the role of h in (2.1), we see that large h correspond to large prime gaps. Lemma A.3 constrains the behavior of large prime gaps and hence of (3.2) when h is large.

Let c be a sufficiently large positive integer depending on n and define $M=c\log_2 y$. We split the outermost sum over h in (3.2) into two regions: One with $0<h\le M\log y$ and one with $h>M\log y$. The sum where $h>M\log y$ counts contributions where $g_n>M\log y$. However, this portion of the sum can only contribute if its terms exist at all, therefore, the sum where $h>M\log y$ is bounded above by the probability that $g_n>M\log y$. Hence, by Lemma A.3, the sum where $h>M\log y$ is bounded above by $\frac{1}{\log^c y}$. Thus, by controlling c, we can discard the portion of the sum where $h>M\log y$. For the remainder of this paper, we consider $h\le M\log y$.

For $n=0,1,2$, Lemke Oliver and Soundararajan define $\mathcal{D}_n(a,b;y)$ to be the terms obtained from (3.1) where $|\mathcal{T}|=n$ and $\mathcal{A}=\mathcal{T}=\varnothing$ or $|\mathcal{A}|+|\mathcal{T}|=2$. However, note that

$\mathcal{D}_n(a,b;y)$ is precisely the term obtained by isolating the $\ell = n$ term in (3.2). Starting the sum over ℓ in (3.2) at $\ell = n$ rather than $\ell = 0$ is the first step towards investigating $\mathcal{D}_n(a,b;y)$. Define $\mathcal{D}_{\geq n}(a,b;y) = \sum_{i \geq n}^{M \log y} \mathcal{D}_i(a,b;y)$. Written explicitly, the terms of (3.2) we are interested in are

$$\mathcal{D}_{\geq n}(a,b;y) = \sum_{\substack{0<h\leq M\log y \\ h\equiv b-a \pmod q}} g^h \sum_{\mathcal{A}\subset\{0,h\}} \sum_{\ell=n}^{h-1} (-z)^\ell \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1\ \forall t\in\mathcal{T} \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\mathcal{A}\cup\mathcal{T}). \tag{3.3}$$

Furthermore, define

$$A_{h,\ell} = \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1 \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\mathcal{T}), \qquad B_{h,\ell} = \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1 \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\{0\}\cup\mathcal{T}),$$

$$C_{h,\ell} = \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1 \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\{h\}\cup\mathcal{T}), \qquad D_{h,\ell} = \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1 \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\{0,h\}\cup\mathcal{T}).$$

We partition the summation in (3.3) into four terms $S_\varnothing$, $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$, based on $\mathcal{A}$. For example,

$$S_\varnothing = \sum_{\substack{0<h<M\log y \\ h\equiv b-a \pmod q}} g^h \sum_{\ell=n}^{h-1} (-z)^\ell A_{h,\ell}, \tag{3.4}$$

with $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$ defined analogously with sums over $B_{h,\ell}$, $C_{h,\ell}$, and $D_{h,\ell}$, respectively.

In order to handle $A_{h,\ell}$, $B_{h,\ell}$, $C_{h,\ell}$, and $D_{h,\ell}$, we modify the following result of Montgomery and Soundararajan [9], which states the average order of $\mathfrak{S}_0$. They show that

$$\sum_{\substack{\mathcal{T}\subset[1,h] \\ |\mathcal{T}|=\ell}} \mathfrak{S}_0(\mathcal{T}) = \frac{\mu_\ell}{\ell!}(-h\log h + Ah)^{\ell/2} + \mathcal{O}(h^{\ell/2-1/7\ell+\epsilon}), \tag{3.5}$$

where μ_ℓ is the $\ell^{\text{ th}}$ moment of the standard normal distribution and A is an absolute constant between -1 and 0. We expect that $\sum \mathfrak{S}_{q,0}(\mathcal{T})$ has a similar growth rate, up to minor corrections such as the exact value of A and leading factors depending on q. Moreover, these arguments used to justify Theorem 3.1 are expected to be robust against these modifications.

We prove the following theorem concerning the growth rates of $S_\varnothing$, $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$, which proves a weaker version of the claim in [4] that $\mathcal{D}_n(a,b;y)$ is $\mathcal{O}_n\Big(\frac{(\log_2 y)^{n/2}}{(\log y)^{n/2-1}}\Big)$.

Theorem 3.1. *Assuming that (3.5) holds in a similar form for $\mathfrak{S}_{q,0}$, we have that $S_\varnothing$,*

$S_{\{0\}} \log y$, $S_{\{h\}} \log y$, and $S_{\{0,h\}} (\log y)^2$ are all

$$\mathcal{O}_n\left(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\right).$$

In particular, $\mathcal{D}_n(a,b;y)$ and $\mathcal{D}_{\geq n}(a,b;y)$ are both $\mathcal{O}_n\left(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\right)$, allowing us to truncate $\mathcal{D}(a,b;y)$ at specific values of n and control the error terms in Conjecture 2.2.

We defer the proofs of Lemmas A.1-A.4 used in the proof of Theorem 3.1 to Appendix A.

Proof. We begin by evaluating $S_\varnothing$ according to (3.4). We are interested in the case where q is prime, and thus $\varphi(q) = q-1$ and $\alpha(y) = 1 - \frac{q}{(q-1)\log y}$. Because q is an odd prime, $\frac{\varphi(q)}{q} \geq \frac{2}{3}$. Thus,

$$1 - \frac{3}{2\log y} \leq \alpha(y) < 1 - \frac{1}{\log y}.$$

For sufficiently large y, the definition of z gives

$$z = \frac{q}{\varphi(q)\alpha(y)\log y} < \frac{3}{2(1-\frac{3}{2\log y})\log y} < \frac{3}{\frac{3}{2}\log y} = \frac{2}{\log y}.$$

Appealing to our conjectured form for $\sum \mathfrak{S}_{q,0}$ according to (3.5), we replace $A_{h,\ell}$ in (3.4) with $\frac{\mu_\ell}{\ell!}(-h\log h + Ah)^{\ell/2} + \mathcal{O}(h^{\ell/2-1/7\ell+\epsilon})$. Note that $\mu_\ell = 0$ when ℓ is odd, so we analyze the sum based on the parity of ℓ.

Case 1: ℓ is even. For convenience, define $m = \ell/2$. We split the single sum over h into a sum over j, k and h and swap the order of summation so that (3.4) is less than

$$\sum_{m=\frac{n}{2}}^{M\log y - 1} \sum_{j=0}^{\frac{M}{\log_3 y}} \sum_{k=j\log_3 y}^{(j+1)\log_3 y - 1} \sum_{\substack{h=k\log y+1 \\ h\equiv b-a \pmod q}}^{(k+1)\log y} g^h \left(\frac{2}{\log y}\right)^{2m} \left(\frac{\mu_{2m}}{(2m)!}(-h\log h + Ah)^m\right). \tag{3.6}$$

We bound (3.6) above by a series of substitutions. Define $B = -A > 0$ and take the absolute value of the terms of (3.6). Lemma A.1 implies $(h\log h + Bh)^m$ has an upper bound of $2^m[(h\log h)^m + (Bh)^m]$, where we include the extra factor of 2 for convenience. Because $g = \left(1 - \frac{q}{\varphi(q)\log y}\right)^{\varphi(q)/q}$ and $\frac{\varphi(q)}{q} \geq \frac{2}{3}$, We have

$$g < \left(1 - \frac{3}{2\log y}\right)^{2/3} < e^{-2/(3\log y)}.$$

Thus, g^h has an upper bound of $e^{-2j\log_3 y \log y/(3\log y)} = (\log_2 y)^{-2j/3}$. We then maximize all instances of h by replacing h with $h_{\max} = (k+1)\log y$ and remove the sum over h by

multiplying the summand by $\log y$. Finally, note that $\mu_{2m} = (2m-1)!!$, so $\frac{\mu_{2m}}{(2m)!} = \frac{1}{2^m m!}$. These substitutions yield

$$\sum_{m=\frac{n}{2}}^{M\log y-1} \sum_{j=0}^{\frac{M}{\log_3 y}} \sum_{k=j\log_3 y}^{(j+1)\log_3 y-1} \frac{(\log y)^{1-2m}}{(\log_2 y)^{2j/3}} \left(\frac{2^{2m}}{2^m m!} (2^m[(h_{\max}\log h_{\max})^m + (Bh_{\max})^m]) \right). \tag{3.7}$$

Applying Lemma A.1 to $(\log h_{\max})^m = (\log(k+1) + \log_2 y)^m$ implies

$$(h_{\max}\log h_{\max})^m \le (2(k+1)\log y)^m[(\log(k+1))^m + (\log_2 y)^m]. \tag{3.8}$$

Substituting (3.8) into (3.7), distributing the factor of $(2/\log y)^{2m}$, and cancelling the factor of 2^m yields

$$\sum_{m=\frac{n}{2}}^{M\log y-1} \sum_{j=0}^{\frac{M}{\log_3 y}} \sum_{k=j\log_3 y}^{(j+1)\log_3 y-1} \frac{\log y}{(\log_2 y)^{2j/3}} \left(\frac{1}{m!} \left[\left(\frac{8(k+1)\log(k+1)}{\log y} \right)^m + \left(\frac{8(k+1)\log_2 y}{\log y} \right)^m \right] \right).$$

Again, we maximize k and remove the sum over k by multiplying by $\log_3 y$, leaving

$$\sum_{m=\frac{n}{2}}^{M\log y-1} \sum_{j=0}^{\frac{M}{\log_3 y}} \frac{\log y \log_3 y}{(\log_2 y)^{2j/3}} \left(\frac{1}{m!} \left[\left(\frac{8((j+1)\log_3 y)(\log(j+1) + \log_4 y))}{\log y} \right)^m + \left(\frac{8(j+1)\log_2 y \log_3 y}{\log y} \right)^m \right] \right). \tag{3.9}$$

We split the sum in (3.9) up into four cases based on the value of j.

Case 1A: $j = 0$. When $j = 0$, the sum in (3.9) becomes

$$\log y \log_3 y \sum_{m=\frac{n}{2}}^{M\log y-1} \frac{1}{m!} \left[\left(\frac{8\log_3 y \log_4 y}{\log y} \right)^m + \left(\frac{8\log_2 y \log_3 y}{\log y} \right)^m \right]. \tag{3.10}$$

Note that (3.10) is a truncated Taylor polynomial of e^x. We show that (3.10) is $\mathcal{O}(f(n))$, where $f(n)$ is the first term of the truncated Taylor polynomial. With this in mind, because the summation in (3.10) is a truncated series of positive terms, it is less than the value of the complete Taylor series $e^{8\log_3 y \log_4 y/\log y} + e^{8\log_2 y \log_3 y/\log y}$.

Simplifying and noting that $(\log_a y)^b$ is $\mathcal{O}(\log y)$ for any $a \ge 2$ and $b \ge 0$, Lemma A.2, whose statement and proof can be found in Appendix A, implies that the expression is $\mathcal{O}(1)$.

Because the Taylor series is $\mathcal{O}(1)$, the growth rate of (3.10) for varying n is determined by the first term. Hence, (3.10) is

$$\mathcal{O}_n\left(\frac{(\log_2 y)^{n/2}(\log_3 y)^{n/2+1}}{(\log y)^{n/2-1}}\right). \tag{3.11}$$

Case 1B: $j = 1$. Analyzing the $j = 1$ term follows similar logic; the asymptotic we obtain is also

$$\mathcal{O}_n\left(\frac{(\log_2 y)^{n/2}(\log_3 y)^{n/2+1}}{(\log y)^{n/2-1}}\right). \tag{3.12}$$

Case 1C: $2 \le j < \frac{3\log_2 y}{2\log_3 y}$. Because $j \ge 2$, we know

$$(\log_2 y)^{-2j/3} < (\log_2 y)^{-4/3} < \frac{1}{\log_2 y}.$$

We also know that $j+1 \le \frac{3\log_2 y}{2\log_3 y}$. Maximizing $(\log_2 y)^{-2j/3}$ and $j+1$, removing the summation by multiplying by $\frac{3\log_2 y}{2\log_3 y}$, and cancelling $\log\left(\frac{\log_2 y}{\log_3 y}\right)$ with $\log_4 y$ yields

$$\frac{3\log y}{2}\sum_{m=\frac{n}{2}}^{M\log y-1}\frac{1}{m!}\left[\left(\frac{12\log_2 y\log_3 y}{\log y}\right)^m+\left(\frac{12(\log_2 y)^2}{\log y}\right)^m\right]. \tag{3.13}$$

As before, the sum in (3.13) is a truncated Taylor series that is $\mathcal{O}(1)$. Hence, (3.13) is

$$\mathcal{O}_n\left(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\right). \tag{3.14}$$

Case 1D: $\frac{3\log_2 y}{2\log_3 y} \le j \le \frac{M}{\log_3 y}$. When $j > \frac{3\log_2 y}{2\log_3 y}$, the factor $(\log_2 y)^{-2j/3}$ is no greater than $(\log_2 y)^{-\log_2 y/\log_3 y} = \frac{1}{\log y}$. Substituting for $(\log_2 y)^{-j}$ with $\frac{1}{\log y}$ and $j+1$ with $\frac{M}{\log_3 y}$, which is allowed because $\frac{M}{\log_3 y}+1$ is the same size as $\frac{M}{\log_3 y}$, the summation in (3.9) becomes, after simplification,

$$\sum_{m=\frac{n}{2}}^{M\log y-1}\sum_{j=\frac{\log_2 y}{\log_3 y}}^{\frac{M}{\log_3 y}}\log_3 y\left(\frac{1}{m!}\left[\left(\frac{8M\log M}{\log y}\right)^m+\left(\frac{8M\log_2 y}{\log y}\right)^m\right]\right). \tag{3.15}$$

We remove the summation in (3.15) by multiplying the summand by $\frac{M}{\log_3 y}$, truncate the resulting Taylor series, and apply Lemma A.2 to obtain the final contribution from this case as

$$\frac{(8c)^{n/2}}{(n/2)!}\left[\frac{(c\log_2 y)^{n/2+1}(\log_3 y)^{n/2}}{(\log y)^{n/2}}+\frac{(\log_2 y)^{n+1}}{(\log y)^{n/2}}\right]=\mathcal{O}_n\left(\frac{(\log_2 y)^{n+1}}{(\log y)^{n/2}}\right). \tag{3.16}$$

Case 2: ℓ is odd. We proceed analogously to the even ℓ case, noting that if an arbitrary function f is $\mathcal{O}(h^{\ell/2-1/7\ell+\epsilon})$, then f is also $\mathcal{O}(h^{\ell/2})$. Therefore, for odd ℓ, (3.4) is less than

$$\sum_{k=0}^{M-1} \sum_{\substack{h=k\log y+1 \\ h\equiv b-a \pmod q}}^{(k+1)\log y} g^h \sum_{\substack{\ell=n \\ \ell \text{ odd}}}^{h-1} (-z)^\ell \mathcal{O}(h^{\ell/2}). \tag{3.17}$$

For $\ell \in [0, M-1]$, let C_ℓ be the implied constant in the $\mathcal{O}(h^{\ell/2})$ term. Defining $C_{\max} = \max\{C_\ell\}$ allows us to pull $-C_{\max}$ out of the sum and remove the big $\mathcal{O}$ notation. We also switch the order of sums in (3.17) to obtain

$$-C_{\max} \sum_{\substack{\ell=n \\ \ell \text{ odd}}}^{M\log y} \sum_{h>\max\{\ell,\log y\}}^{M\log y} g^h z^\ell h^{\ell/2}. \tag{3.18}$$

Since $h \geq \log y$, we know $g^h \leq e^{-2h/3\log y}$. It thus follows that $z < \frac{2}{\log y}$ and $h \leq M\log y$. Thus, maximizing g^h, z^ℓ, and $h^{\ell/2}$ implies that (3.18) has an upper bound of

$$-C_{\max} \sum_{\substack{\ell=n \\ \ell \text{ odd}}}^{M\log y} \left(\frac{2}{\log y}\right)^\ell (M\log y)^{\ell/2} \sum_{h>\max\{\ell,\log y\}}^{M\log y} e^{-2h/3\log y}.$$

The sum over h is a geometric series that is less than $\frac{e^{-2/3}}{1-e^{-2/3\log y}}$, which is less than $\log y$ for $\log y > 1$. Next, we distribute the $\left(\frac{2}{\log y}\right)^\ell$ into $(M\log y)^{\ell/2}$ and sum the resulting geometric series; this yields

$$-C_{\max}\log y \frac{(4M/\log y)^{n/2}(1-(4M/\log y)^{M\log y+1})}{1-M/\log y}. \tag{3.19}$$

For large y, both $1-(M/\log y)^{M\log y+1}$ and $1-M/\log y$ are $\mathcal{O}(1)$. Thus, (3.19) becomes

$$-C_{\max}\frac{M^{n/2}}{(\log y)^{n/2-1}} = \mathcal{O}_n\left(\frac{(\log_2 y)^{n/2}}{(\log y)^{n/2-1}}\right). \tag{3.20}$$

Note that for sufficiently large y, each of the cases based on j are smaller than (3.14). Thus, the contributions from Cases 1A, 1B, 1D, and 2 as stated in (3.11), (3.12), (3.16), and (3.20), respectively, are all smaller than the contribution from Case 1C as stated in (3.14). Therefore, $S_\varnothing = \mathcal{O}_n\left(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\right)$, as desired.

Lemma A.4 implies that summations of $B_{h,\ell}$, $C_{h,\ell}$, or $D_{h,\ell}$ are closely related to summations of $A_{h,\ell}$. In order to take advantage of the cancellation suggested by the form

$B_{h-1,\ell-1} = A_{h,\ell} - A_{h-1,\ell}$, we consider the sign of $A''_{h,\ell} = \frac{\partial^2}{\partial h^2} A_{h,\ell}$. Namely, if $A''_{h,\ell} > 0$, then

$$A'_{h-1,\ell} < A_{h,\ell} - A_{h-1,\ell} < A'_{h,\ell}.$$

Otherwise, if $A''_{h,\ell} < 0$, then

$$A'_{h-1,\ell} > A_{h,\ell} - A_{h-1,\ell} > A'_{h,\ell}.$$

Regardless of the sign of $A''_{h,\ell}$, we insert the appropriate upper bound given by either $A'_{h,\ell}$ or $A'_{h-1,\ell}$ into $S_{\{0\}}$ and $S_{\{h\}}$. In evaluating $S_{\{0,h\}}$, we take $A'''_{h,\ell}$ and use appropriate bounds for $A_{h,\ell} - 2A_{h-1,\ell} + A_{h-2,\ell}$.

We proceed to evaluate $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$ in an analogous manner to the method of evaluating $S_{\varnothing}$. In loose terms, taking k derivatives of $A_{h,\ell}$ corresponds to adding a factor of $(\log y)^k$ to the denominator of the asymptotic in Theorem 3.1, thus leading to $S_{\{0\}} \log y$, $S_{\{h\}} \log y$, and $S_{\{0,h\}} (\log y)^2$.

Recall that from the definition of $S_{\varnothing}$, $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$, the relevant contribution to $\mathcal{D}_{\geq n}(a, b; y)$, after discarding terms where $h > M \log y$, is $S_{\varnothing} + S_{\{0\}} + S_{\{h\}} + S_{\{0,h\}}$. Therefore, $\mathcal{D}_{\geq n}(a, b; y)$ is $\mathcal{O}_n\Big(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\Big)$ as well. Since

$$\mathcal{D}_n(a, b; y) = \mathcal{D}_{\geq n+1}(a, b; y) - \mathcal{D}_{\geq n}(a, b; y),$$

it is also $\mathcal{O}_n\Big(\frac{(\log_2 y)^n}{(\log y)^{n/2-1}}\Big)$. Thus, the theorem is proved. □

4 Numerical Results

The following simplified asymptotic for the case $\pi(x; 3, (a, b))$ is provided in [4]:

$$\pi(x; 3, (a, b)) = \frac{\mathrm{li}(x)}{4}\left(1 \pm \frac{1}{2 \log x} \log\left(\frac{2\pi \log x}{q}\right)\right) + \mathcal{O}\left(\frac{x}{(\log x)^{11/4}}\right), \tag{4.1}$$

with the plus or minus sign being plus if $a \neq b$ and minus if $a = b$.

We compare (4.1) to the actual behavior of the primes, and find that because the approximations that were necessary to arrive at (4.1), the data deviate from the conjectured form. Using SageMath's `find_fit` function suggested a possible lower order term of size $\mathcal{O}\Big(\frac{(\log_2 x)^2}{(\log x)^2}\Big)$. We modified the data gathering process to approximate values of $\pi(x; q, (a, b))$ for $x \leq 10^{18}$. Finally, we include more terms in the approximation of $\mathcal{D}(a, b; y)$ to improve its accuracy in future work. Graphs may be found in Appendix B.

Lemke Oliver and Soundararajan [4] gathered values of $\pi(x; q, (a, b))$ up to $x = 10^{12}$.

We gathered data up to $x = 10^{18}$. We gathered complete raw data using SageMath for $1 \le x \le 10^{10}$. For $10^{10} < x \le 10^{18}$, a sampling technique was used to approximate the ratio $\pi(x; q, (a, b))/\pi(x)$. Lemke Oliver's `C++` code counts prime patterns in fixed intervals $[X, Y)$; the program was modified to only consider the first 10^8 primes larger than X. We used $X = 10^{b_1}$ and $Y = 10^{b_1+1}$ for $10 \le b_1 \le 18$. The program estimated pattern frequencies at $X + i \cdot 10^{b_1}$ for $1 \le i \le 9$.

Theorem 3.1 shows that the contributions $S_\varnothing$, $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$ to $\mathcal{D}(a, b; y)$ decline quickly with n. After dividing by $\mathrm{li}(x)$, the main terms in the main conjecture in [4] are of size $\mathcal{O}(1)$, $\mathcal{O}(\frac{\log\log x}{\log x})$, and $\mathcal{O}(\frac{1}{\log x})$. When $|\mathcal{T}| = 6$, Theorem 3.1 implies that $S_\varnothing \in \mathcal{O}\Big(\frac{(\log_2 y)^6}{(\log y)^2}\Big)$. Thus, for $n \ge 6$, $S_\varnothing$, $S_{\{0\}}$, $S_{\{h\}}$, and $S_{\{0,h\}}$ make negligible contributions to $\mathcal{D}$. This implies that long range correlations between prime patterns are negligible, which in turn implies that even though we only take the first 10^8 primes after X, the sample can be reasonably assumed to be unbiased.

Following Cramér's model, we model primality as a binomial event with x being prime with probability $\frac{1}{\log x}$ and assume that primality of x and y are independent events. Then the standard deviation of our sampling distribution is proportional to $\frac{1}{\sqrt{C}}$, where C is the number of primes sampled in order to estimate the frequency of $\pi(x; q, (a, b))$ at x.

For each point estimate at $X + i \cdot 10^{b_1}$, we sampled with $C = 10^8$, giving a precision of roughly 10^{-4}. The sample gives a sampling frequency

$$f_{a,b} = \frac{\pi(x + x_0; q, (a, b)) - \pi(x; q, (a, b))}{\pi(x + x_0) - \pi(x)},$$

where $\pi(x + x_0) - \pi(x) = 10^7$. In a crude sense, the sampling frequency $f_{a,b}$ is the derivative of $\pi(x; q, (a, b))$, so we used a Riemann sum with 10 equally spaced subintervals to estimate $\pi(X + i \cdot 10^{b_1}; q, (a, b))$ from $f_{a,b}$. We thus computed

$$\frac{\sum_{\beta=1}^{b_1} \sum_{\alpha=1}^{9} f_{a,b}[\mathrm{li}((\alpha + 1) \cdot 10^\beta) - \mathrm{li}(\alpha \cdot 10^\beta)]}{\mathrm{li}(9 \cdot 10^{b_1})}. \tag{4.2}$$

Note that (4.2) approximates the the ratio $\frac{\pi(10^{b_1+1}; q, (a,b))}{\pi(x)}$ and hence allows us to extend our data to $x = 10^{18}$.

We restate the conjectured form for $\pi(x; q, (a, b))$ as in Conjecture 2.2 for convenience as

$$\pi(x; q, (a, b)) \sim \frac{1}{q} \int_2^x \alpha(y)^{\epsilon_q(a,b)} \left(\frac{q}{\varphi(q)\alpha(y) \log y} \right)^2 \mathcal{D}(a, b; y) dy. \tag{4.3}$$

The numerical model in [4] is evaluated by partitioning $\mathcal{D}(a, b; y)$ into $\sum_n \mathcal{D}_n(a, b; y)$ and discarding $\mathcal{D}_n(a, b; y)$ for $n \ge 3$. Thus $\mathfrak{S}_{q,0}$ is estimated only for zero and two term sets to

approximate $\mathcal{D}(a,b;y)$. For example, in [4], only the zero and two term sets for $\mathcal{D}_1(a,b;y)$ are considered. Lemke Oliver and Soundararajan then write

$$\mathcal{D}_1(a,b;y) \approx -\frac{q}{\varphi(q)\alpha(y)\log y} \sum_{\substack{h>0 \\ h\equiv b-a \pmod q}} \sum_{\substack{t\in[1,h-1] \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{0,t\}) + \mathfrak{S}_{q,0}(\{t,h\}). \tag{4.4}$$

However, Theorem 3.1 suggests that only considering zero and two term sets may not accurate enough. Hence, we add terms to $\mathcal{D}_0$, $\mathcal{D}_1$, and $\mathcal{D}_2$, as well as truncating at $\mathcal{D}_5$ instead of $\mathcal{D}_2$ to approximate $\mathcal{D}$ in (4.3). For example, recalling that $\mathcal{D}_1(a,b;y)$ contains all terms of (3.1) with $|\mathcal{T}|=1$, we write

$$\mathcal{D}_1(a,b;y) = -\frac{q}{\varphi(q)\alpha(y)\log y} \sum_{\substack{h>0 \\ h\equiv b-a \pmod q}} \sum_{\substack{t\in[1,h-1] \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{0,t\}) + \mathfrak{S}_{q,0}(\{t,h\}) + \mathfrak{S}_{q,0}(\{0,t,h\}).$$

This is essentially (4.4) but with three term sets included. The values of singular series $\mathfrak{S}_{q,0}(\mathcal{H})$ were computed up to five term sets with $\max \mathcal{H} \leq 150$ and prepared for future work numerically integrating (4.3) by truncating at $\mathcal{D}_5(a,b;y)$.

5 Conclusion

Assuming an asymptotic formula for $\mathfrak{S}_{q,0}$ similar to (3.5), we proved asymptotic formulas for terms of $\mathcal{D}(a,b;y)$ and justified discarding certain terms to create a numerical model. We plan to complete the numerical model and generate predictions that are more consistent with the actual prime frequencies than Lemke Oliver and Soundararajan's model. We modified the data gathering algorithm to extend the data by eight orders of magnitude and used the increased amount of data to identify further terms in Lemke Oliver and Soundararajan's main conjecture.

There are several possible avenues of exploration. Lemke Oliver and Soundararajan do not directly use the Hardy-Littlewood prime k-tuple conjecture; so a rigorous argument is needed to fully show that the biases can be explained by the Hardy-Littlewood prime k-tuple conjecture. A finer argument would likely show that the growth rate of $\mathcal{D}_n(a,b;y)$ is closer to the growth rate proposed by Lemke Oliver and Soundararajan. The data gathering method can be tweaked to gather more data to greater accuracy by increasing the number of primes sampled, fully randomizing the sampling process, and using arbitrary size integers to bypass the artificial 2^{64} size limit imposed by `C++`.

6 Acknowledgments

I would like to express my deepest appreciation to my mentor Mr. Robert Burklund for his extremely helpful guidance. I am very grateful for the extensive advice and feedback provided by my tutor, Dr. John Rickert, the head mentor, Dr. Tanya Khovanova, and other RSI staff. Also, thanks to Professors David Jerison, Ankur Moitra, and Slava Gerovitch who coordinated the MIT math mentorship program. I would like to thank Professor Robert Lemke Oliver for his helpful comments on my ideas. I would also like to thank Professor Lawrence Washington for sacrificing his own time to explain the background of the project. I would like to thank my fellow RSI students who aided me throughout the project, especially Harshal Sheth and Jordan Lee. I would like to thank my parents for their constant support. Many thanks to my Senior Research Project teacher, Ms. Angelique Bosse, and my fellow classmates who helped me improve my paper.

Finally, I would like to express my gratitude to the generosity to my sponsors who made it possible for me to attend the Research Science Institute, including the Department of Defense, the Center for Excellence in Education, the Massachusetts Institute of Technology, Her Excellency Bahia El Hariri, Mr. John Yochelson, Dr. Noreen Hynes, Mr. Dale P. Bennett, Admiral Michael S. Rogers, Ms. Donna Cooper, Dr. Yan Shi, Dr. Pam Krahl, Mr. and Mrs. Bhanu Durvasula, Mr. and Mrs. Wayne Kamitaki, Mr. Marli Pasternak and Mrs. Art Pasternak, Mr. Jerome H. Powell and Ms. Elissa A. Leonard, Mr. Eamon Walsh, Mrs. Susan S. Lee, Professors Joseph and Nell Sedransk, and Ms. Wendy Kershner.

A Proofs of the Lemmas

We now prove the lemmas that were used to prove the main theorem.

Lemma A.1. *Let a and b be nonnegative real numbers and n be a positive integer. Then*

$$(a+b)^n \leq 2^{n-1}(a^n+b^n).$$

Proof. Since x^n is convex for nonnegative x and positive integers n, Jensen's inequality yields $(\frac{a}{2}+\frac{b}{2})^n \leq \frac{1}{2}a^n + \frac{1}{2}b^n$. Clearing denominators gives $(a+b)^n \leq 2^{n-1}(a^n+b^n)$, as desired. □

Lemma A.2. *For any real constants a and c, we have*

$$\lim_{x\to\infty} (\log x)^{c(\log_2 x)^a/\log x} = 1.$$

Proof. Since the limit

$$L = \lim_{x\to\infty} (\log x)^{(\log_2 x)^a / \log x}$$

does not depend on c, we set $c = 1$; if we prove $L = 1$ then certainly $L^c = 1$ and the lemma follows. Taking logarithms, it suffices to show that

$$\lim_{x\to\infty} \frac{(\log_2 x)^a}{\log x} = 0.$$

However, any power of $\log_2 t$ grows slower than $\log t$ for all sufficiently large t, so the limit indeed equals 0. The lemma is thus proved. □

Lemma A.3. *Let N be a real number and $\mathbb{P}[g_n > x]$ denote the probability that the gap g_n between the nth and $(n+1)$th prime is greater than x for $1 \le n \le N$. Then*

$$\lim_{N\to\infty} \mathbb{P}[g_n > c \log_2 p_N \log p_n] < \frac{1}{(\log N)^c}.$$

We sketch the details of the proof here. Although not fully rigorous, we expect the key ingredients of the proof to be present.

Proof. n the following proof, we omit the limits as N goes to infinity for readability. Gallagher [10] showed that

$$\mathbb{P}[1 \le n \le N \mid g_n > \lambda \log p_n] < e^{-\lambda}.$$

Setting $\lambda = c \log_2 p_N$ implies $\mathbb{P}[g_n > c \log_2 p_N \log p_n] < \frac{1}{(\log p_N)^c}$. By the PNT, $p_N \sim N \log N$, so

$$\frac{1}{(\log p_N)^c} < \frac{1}{(\log(N \log N))^c}.$$

Then, since $\log(N \log N) = \log N + \log_2 N$, we have

$$\frac{1}{(\log(N \log N))^c} < \frac{1}{(\log N)^c}.$$

Thus,

$$\mathbb{P}[g_n > c \log_2 p_N \log p_n] < \frac{1}{(\log N)^c},$$

as desired. □

Lemma A.4. *The sums over subsets of $[1, h-1]$ of size ℓ, given by $A_{h,\ell}$, $B_{h,\ell}$, $C_{h,\ell}$, and*

$D_{h,\ell}$, satisfy the following relations:

$$B_{h-1,\ell-1} = C_{h-1,\ell-1} = A_{h,\ell} - A_{h-1,\ell},$$
$$D_{h-1,\ell-1} = A_{h,\ell} - 2A_{h-1,\ell} + A_{h-2,\ell}.$$

Proof. Note that $A_{h,\ell}$ is a sum that ranges over all subsets $\mathcal{T}$ of $[1, h-1]$. We can partition this sum by $\max\{\mathcal{T}\}$. Setting $m = \max\{\mathcal{T}\}$, we can write

$$A_{h,\ell} = \sum_{m=\ell}^{h-1} \sum_{\substack{\mathcal{T}\in[1,m-1] \\ |T|=\ell-1 \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{m\} \cup \mathcal{T}). \tag{A.1}$$

From Definition 2.1, $\mathfrak{S}_{q,0}(\mathcal{T}) = \mathfrak{S}_{q,0}(s - \mathcal{T})$ for any integer s. Using the translational invariance of $\mathfrak{S}_{q,0}$ and noting that

$$C_{m,\ell-1} = \sum_{\substack{\mathcal{T}\in[1,m-1] \\ |T|=\ell-1 \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{m\} \cup \mathcal{T}),$$

we can rewrite (A.1) as $A_{h,\ell} = \sum_{m=\ell}^{h-1} C_{m,\ell-1}$. Now consider $A_{h,\ell} - A_{h-1,\ell}$; every term in this difference cancels except $C_{h-1,\ell-1}$, so $A_{h,\ell} - A_{h-1,\ell} = C_{h-1,\ell-1}$, as desired.

Similarly, we can partition a sum over subsets $\mathcal{T}$ of $[1, h-1]$ to a sum over sets $\mathcal{T}$ whose minimum value is m. Thus

$$A_{h,\ell} = \sum_{m=\ell}^{h-1} \sum_{\substack{\mathcal{T}\in[m-\ell+1,h-1] \\ |T|=\ell-1 \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{m-\ell\} \cup \mathcal{T}).$$

Translational invariance implies that

$$B_{m,\ell-1} = \sum_{\substack{\mathcal{T}\in[m-\ell+1,h-1] \\ |T|=\ell-1 \\ (t+a,q)=1}} \mathfrak{S}_{q,0}(\{m-\ell\} \cup \mathcal{T}),$$

so $A_{h,\ell} - A_{h-1,\ell}$ telescopes as before and only $B_{h-1,\ell-1}$ remains. Therefore, $A_{h,\ell} - A_{h-1,\ell} = B_{h-1,\ell-1}$ as well.

Finally, in order to relate $A_{h,\ell}$ to $D_{h,\ell}$, we write the sum over subsets of $[1, h-1]$ as a sum over m and over sets $\mathcal{T}$ where $\max\{\mathcal{T}\} - \min\{\mathcal{T}\} = m$. By translational invariance, because there are $h - m + 1$ possibilities for $\min\{\mathcal{T}\}$, there are $h - m + 1$ copies of $\mathfrak{S}_{q,0}(\{1, m\} \cup \mathcal{T})$.

Hence, the definition for $A_{h,\ell}$ can be rewritten as

$$A_{h,\ell} = \sum_{m=\ell}^{h-1} \sum_{\substack{\mathcal{T}\in[2,m-1] \\ |T|=\ell-2 \\ (t+a,q)=1}} (h-m+1)\mathfrak{S}_{q,0}(\{1,m\}\cup\mathcal{T}). \tag{A.2}$$

Recall that $D_{h,\ell} = \sum_{\substack{\mathcal{T}\subset[1,h-1] \\ (t+a,q)=1 \\ |\mathcal{T}|=\ell}} \mathfrak{S}_{q,0}(\{0,h\}\cup\mathcal{T})$, so $A_{h,\ell} - A_{h-1,\ell} = \sum_{m=\ell}^{h-1} D_{m,\ell-1}$. Therefore, we express (A.2) as $A_{h,\ell} = \sum_{m=\ell}^{h-1}(h-m+1)D_{m,\ell-2}$. Note that the sum telescopes when two successive differences are taken. What remains is $D_{h-1,\ell-2} = (A_{h,\ell} - A_{h-1,\ell}) - (A_{h-1,\ell} - A_{h-2,\ell}) = A_{h,\ell} - 2A_{h-1,\ell} + A_{h-2,\ell}$, as desired. □

B Plots of Data and Model

This appendix contains the plots of raw data, extended data, curve fitting.

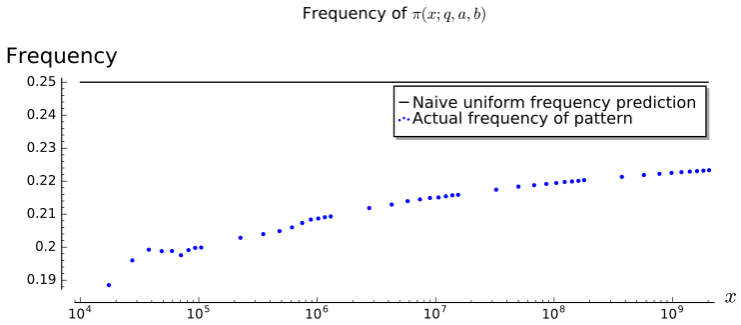

Figure 1: The proportion $\frac{\pi(x;3,(1,1))}{\pi(x)}$ for $x_0 \le x \le x_1$ where $\pi(x_0) = 10^4$ and $\pi(x_1) = 10^9$.

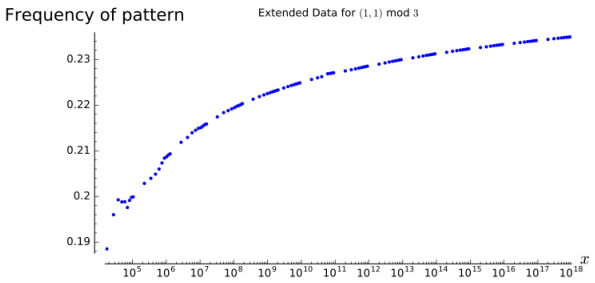

Figure 2: The extended data for $(1, 1)$ modulo 3. The slight bump at $5 \cdot 10^{10}$ is due to combining the raw and sampled data.

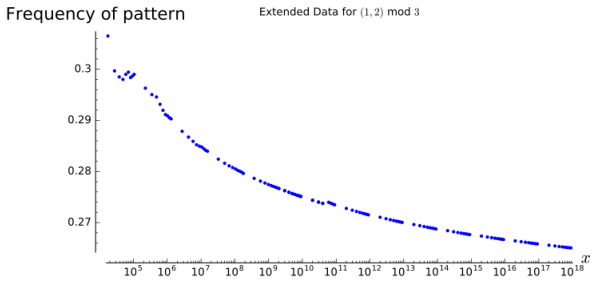

Figure 3: The extended data for $(1, 2)$ modulo 3. The slight bump at $5 \cdot 10^{10}$ is due to stitching the raw data and sampled data together.

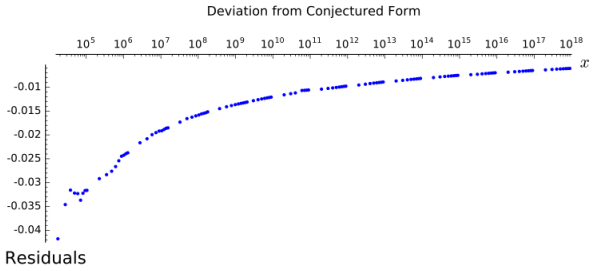

Figure 4: The residuals when (4.1) is subtracted from $\pi(x; 3, (1, 1))$ for $x_0 \leq x \leq x_1$ where $\pi(x_0) = 10^4$ and $\pi(x_1) = 10^{18}$, using the extended data.

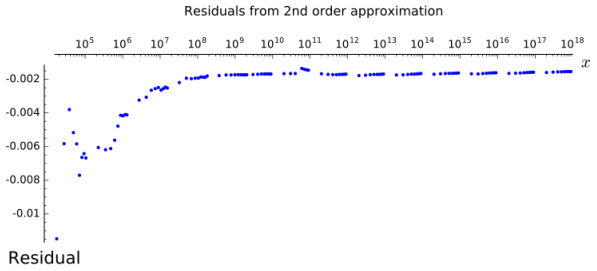

Figure 5: The residuals when curve fitted terms of size $\mathcal{O}((\log_2 y)^2/(\log y)^2)$ and (4.1) are subtracted from $\pi(x; 3, (1, 1))$ for $x_0 \leq x \leq x_1$ where $\pi(x_0) = 10^4$ and $\pi(x_1) = 10^{18}$, using the extended data.

References

[1] T. M. Apostol, *Introduction to Analytic Number Theory*. Springer, 1998.

[2] P. G. L. Dirichlet, *Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält*, p. 342–359. Cambridge Library Collection - Mathematics, Cambridge University Press, 2013.

[3] L. Schoenfeld, "Sharper Bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. ii," *Mathematics of Computation*, vol. 30, no. 134, pp. 337–360, 1976.

[4] R. J. L. Oliver and K. Soundararajan, "Unexpected biases in the distribution of consecutive primes," *Proceedings of the National Academy of Sciences*, 2016.

[5] "Lettre de M. le Professeur Tchébychev à M. Fuss sur un nouveaux théorème relatif aux nombres premiers contenus dans les formes $4n + 1$ et $4n + 3$," 1853.

[6] M. Rubinstein and P. Sarnak, "Chebyshev's bias," *Experiment. Math.*, vol. 3, no. 3, pp. 173–197, 1994.

[7] D. K. Shiu, "Strings of congruent primes," *Journal of the London Mathematical Society*, vol. 61, no. 2, pp. 359–373, 2000.

[8] J. Maynard, "Dense clusters of primes in subsets," *Compositio Mathematica*, vol. 152, no. 7, pp. 1517–1554, 2016.

[9] H. L. Montgomery and K. Soundararajan, "Primes in short intervals," *Communications in Mathematical Physics*, 2004.

[10] P. Gallagher, "On the distribution of primes in short intervals," *Mathematika*, vol. 23, no. 1, pp. 4–9, 1976.