Prime Number Races

Daniel Pareja
dpareja@math.ubc.ca

Abstract.

In this paper, we discuss the idea of a prime number race: what is it, what do we need to study it, and, most importantly, who wins? To answer these questions, we consider various notions of the density of a set of integers, and use these to say how often one "contestant" is winning.

I. Introduction

Pick a positive integer, at least 3; we'll call it q. Now pick two other positive integers, less than q, and not equal; we'll call them a_1 and a_2. (This is why we want $q \geq 3$: if not, we don't have a race!) Make sure that both a_1 and a_2 are coprime to q, and we're ready for a prime number race.

We define the function $\pi(x; q, a) = \#\{p \leq x : p \,\text{prime}, p \equiv a \,(\text{mod}\, q)\}$. A prime number race compares the functions $\pi(x; q, a_1)$ and $\pi(x; q, a_2)$ at each natural number n. We want to know "how often" $\pi(x; q, a_1) > \pi(x; q, a_2)$.

It will also be of interest to us to examine races with more "contestants". More specifically, instead of choosing only a_1 and a_2, we choose i integers in $[1, q-1]$, where $a_j \neq a_k$ for $j \neq k$, and each a_j is coprime to q. We then ask "how often" $\pi(x; q, a_1) > \pi(x; q, a_2) > \ldots > \pi(x; q, a_i)$.

To prove necessary results, it will be necessary at times to assume two strong hypotheses: the Generalized Riemann Hypothesis (GRH) and the Grand Simplicity Hypothesis (GSH). We will state these presently.

II. Preliminaries

a. Primes in Arithmetic Progressions

One of our first assumptions was that a_1 was coprime to q. We can see that if this is not the case, the function $\pi(x; q, a_1)$ is not very interesting: if a_1 is prime, $\pi(x; q, a_1) = 1$ for all $x \geq a_1$, and if a_1 is not prime, the function is identically zero. However, even in the coprime case, we have not ruled out that $\pi(x; q, a_1)$ is not constant for all x sufficiently large. That this is not the case is given by Dirichlet's Theorem on Primes in Arithmetic Progressions (Corollary 4.10, [7]):

Theorem (Dirichlet). If $(a, q) = 1$, then there are infinitely many primes $p \equiv a \,(\text{mod}\, q)$.

From this we see that $\lim_{x \to \infty} \pi(x; q, a_1) = \infty$.

It is because of Dirichlet's theorem that we concern ourselves only with the case where the modulus and remainder are coprime.

b. Dirichlet Characters and Dirichlet L-functions

The proof of Dirichlet's theorem usually presented in courses on analytic number theory involves Dirichlet characters and their associated Dirichlet L-functions. (Selberg [11] gave a proof of this theorem that does not use Dirichlet L-functions.) As we will use these objects many times in this paper, we define them now.

Consider the ring $\mathbb{Z}/q\mathbb{Z}$; we call its elements the *residue classes* modulo q. An element $a \in \mathbb{Z}/q\mathbb{Z}$ is a unit if there exists $b \in \mathbb{Z}/q\mathbb{Z}$ with $ab = 1$. (Since $\mathbb{Z}/q\mathbb{Z}$ is commutative, it follows that b is also a unit.) These units form an abelian group under the multiplication in $\mathbb{Z}/q\mathbb{Z}$; we denote this group $(\mathbb{Z}/q\mathbb{Z})^\times$. We call the elements of $(\mathbb{Z}/q\mathbb{Z})^\times$ the *reduced residue classes* modulo q. We denote the residue class of $a \in \mathbb{Z}$ modulo q by $[a]_q$, and we know that $[a]_q = \{qn + a : n \in \mathbb{Z}\}$.

Definition. A *Dirichlet character modulo* q is a function $\chi : \mathbb{Z} \to \mathbb{C}$ that is totally multiplicative, supported on the reduced residue classes modulo q, and has period q.

We can also think of a Dirichlet character as an irreducible representation of the group $(\mathbb{Z}/q\mathbb{Z})^\times$ extended to $\mathbb{Z}$. (For more on this approach, see [12] and especially Section 4.2 in [7].) We can weaken the hypothesis of total multiplicativity to multiplicativity; see Theorem 4.7 in [7].

All nonzero values of Dirichlet characters are roots of unity. We call a Dirichlet character complex if at least one of its values has nonzero imaginary part; otherwise it is real. The real character that is identically 1 (except where the definition requires that it be zero) is the principal character, and we denote it χ_0; other real characters are quadratic. Note that if χ is quadratic, $\chi^2 = \chi_0$, and that this is not the case for complex characters.

Definition. Let χ be a Dirichlet character modulo q. The *Dirichlet L-function associated to* χ is $L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$ wherever this series converges.

We can show that $L(s, \chi)$ converges absolutely for $\mathrm{Re}(s) > 1$, and if $\chi \neq \chi_0$, $L(s, \chi)$ converges for $\mathrm{Re}(s) > 0$ (but does not converge absolutely for $\mathrm{Re}(s) \in (0, 1]$). $L(s, \chi_0)$ has a simple pole of residue $1/\phi(q)$ at $s = 1$. Dirichlet's theorem can be shown by proving that $L(1, \chi) \neq 0$ for all Dirichlet characters χ. Additionally, in the region of absolute convergence, $L(s, \chi)$ has an Euler product: $L(s, \chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}$, where the product runs over all primes.

c. The Generalized Riemann Hypothesis and the Grand Simplicity Hypothesis

From the fact that $L(s, \chi)$ is equal to a convergent product for $\mathrm{Re}(s) > 1$, it is nonzero in that region. From the functional equation for $L(s, \chi)$ (see Section 10.1 in [7]), $L(s, \chi)$ has "trivial" zeros on the nonpositive real axis. (It may also be the case that $L(s, \chi)$ has zeros on the imaginary axis. These only arise when χ is an imprimitive character–see Section 9.1 in [7]–and result from factors of the form $(1 - \chi(p)p^{-s})$ being the only difference between the L-function associated to χ and that associated to the primitive character inducing χ.) All other zeros of $L(s, \chi)$ are nontrivial zeros, with real part in $(0, 1)$, the critical strip. We use the notation $\rho = \beta + i\gamma$ to denote a nontrivial zero of $L(s, \chi)$.

Conjecture (Generalized Riemann Hypothesis). If $\rho = \beta + i\gamma$ is a nontrivial zero of $L(s, \chi)$, then $\beta = 1/2$.

Conjecture (Grand Simplicity Hypothesis, as in [10]). The set $\{\gamma : L(\frac{1}{2} + i\gamma) = 0,\ \chi \text{ primitive},\ \gamma \geq 0\}$ is linearly independent over $\mathbb{Q}$.

Making these two assumptions will facilitate later computations.

d. The functions $\psi(x; q, a)$ and $\theta(x; q, a)$, and the Prime Number Theorem for Arithmetic Progressions

The Prime Number Theorem for Arithmetic Progressions can be stated as an asymptotic: If $(a,\ q) = 1$, then $\pi(x; q, a) \sim \frac{\mathrm{li}(x)}{\phi(q)}$, where $\mathrm{li}(x) = \int_2^x \frac{dt}{\log(t)}$. In essence, each residue class modulo q containing infinitely many primes gets its "fair share", as there are $\phi(q)$ such residue classes. However, the error term hidden by this asymptotic is of great interest here. It turns out to be more natural to study two functions related to $\pi(x; q, a)$.

Definition. $\theta(x; q, a) = \sum_{\substack{p \leq x \\ p \equiv a\ (q)}} \log p.$

The Prime Number Theorem for Arithmetic Progressions can be stated in the following form: If $(a,\ q) = 1$, then $\theta(x; q, a) \sim \frac{x}{\phi(q)}$.

However, θ is not the function from which we will be able to derive the most useful information. Rather, it is $\psi(x; q, a)$ that will do this.

Definition. The *von Mangoldt lambda function* is defined by $\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^r \\ 0, & \text{else.} \end{cases}$

$\Lambda(n)$ detects when a number n is a power of a prime, p^r, and its value can be thought of as $\frac{1}{r} \log p^r$.

Definition. $\psi(x; q, a) = \sum_{\substack{n \leq x \\ n \equiv a\ (q)}} \Lambda(n).$

The Prime Number Theorem for Arithmetic Progressions can also be stated as follows: If $(a,\ q) = 1$, then $\psi(x; q, a) \sim \frac{x}{\phi(q)}$. We do have an explicit formula for this function, which we will derive later, and which, when combined with assumptions regarding the zeros of $L(s,\ \chi)$ for characters $\chi \pmod q$, will facilitate our computations.

III. A Prime Number Race

Let's examine the functions $\pi(x; 4, 1)$ and $\pi(x; 4, 3)$. Each of these is asymptotic to $\frac{1}{2}\mathrm{li}(x)$, but is it the case that the "size" of the set where $\pi(x; 4, 1) > \pi(x; 4, 3)$ (and vice versa) is asymptotically $\frac{1}{2}$? As we will see later, the answer to this question is no. To provide some evidence for why this might be so, let us examine the values of these functions at various points. (The values in the following table are from [8] and [9].)

x	$\pi(x;4,1)$	$\pi(x;4,3)$
10	1	2
100	11	13
1000	80	87
10000	609	619
100000	4783	4808
1000000	39175	39322
10000000	332180	332398
100000000	2880504	2880950
1000000000	25423491	25424042
10000000000	227523275	227529235
100000000000	2059020280	2059034532
1000000000000	18803924340	18803987677

While we have only examined these functions at very specific points (10^n for $1 \leq n \leq 12$), we see that $\pi(x;4,3)$ seems to be maintaining a lead over $\pi(x;4,1)$, small compared to x, but nonetheless a lead. Leech [4] found that $\pi(26861;4,1) = 1473$, while $\pi(26861;4,3) = 1472$. However, 26863 is also prime, and congruent to 3 modulo 4, and the next prime after that one is 26879, also congruent to 3 modulo 4. (Exploring a bit further, we find that the next few primes are 26881, 26891, 26893, 26903, and 26921, so that while the race is tied at some of these values, $\pi(x;4,1)$ does not retake the lead.) He also found a more extreme difference at 623681: $\pi(623681;4,1) = 25444$, while $\pi(623681;4,3) = 25436$, and notes that except for 26861 and various values between 616000 and 634000, there are no other x where $\pi(x;4,1) > \pi(x;4,3)$ and x is less than or equal to 3000000.

We will see later that in a sense we will soon define, $\pi(n;4,3) > \pi(n;4,1)$ is true for over 99% of integers n.

IV. Probability and Density

a. Probability

Ideally, we would like to define a "nice" probability measure on $\mathbb{N}$. We first recall the definition of a probability measure (from [1]).

Definition. Let Ω be a set. We say that a family $\mathcal{F}$ of subsets of Ω is a *σ-algebra* on Ω if:
i. $\mathcal{F}$ is not empty.
ii. $\mathcal{F}$ is closed under complements, that is, $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$, and
iii. $\mathcal{F}$ is closed under countable union, that is, if $\{A_i\}_{i=1}^{\infty}$ is a sequence of sets in $\mathcal{F}$, then $\bigcup_{i \geq 1} A_i \in \mathcal{F}$.

The ordered pair $(\Omega, \mathcal{F})$ is called a measurable space. When Ω is countable, we usually take $\mathcal{F}$ to be its power set. Another important σ-algebra is the *Borel σ-algebra*, defined as follows (see, for instance, p. 22 in [3]):

Definition. Let X be a topological space. The Borel σ-algebra on X, denoted $\mathcal{B}_X$, is the smallest σ-algebra containing all open sets in X, that is, it is the intersection of all σ-algebras containing all open sets in X.

We now define the notion of a *probability measure* on a measurable space.

Definition. A probability measure P on the measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to \mathbb{R}$ such that:
i. $P(A) \geq P(\emptyset) = 0$ for all sets $A \in \mathcal{F}$,
ii. if $\{A_i\}_{i=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{F}$, that is, with $A_i \cap A_j = \emptyset$ whenever $i \neq j$, $P(\bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} P(A_i)$, and
iii. $P(\Omega) = 1$.

One immediate consequence of this definition is that if A, $B \in \mathcal{F}$ and $A \subset B$, then $P(A) \leq P(B)$, and another is that $P(A^c) = 1 - P(A)$.

We also define what it means for two elements of $\mathcal{F}$, which we call events, to be independent.

Definition. Two events A, $B \in \mathcal{F}$ are independent if $P(A \cap B) = P(A)P(B)$.

We see that if A and B are independent, so are A^c and B. We see this by using the fact that $B = (A \cap B) \cup (A^c \cap B)$, and the union is one of disjoint sets. Thus, taking probabilities, using independence, and rearranging we find that $P(A^c \cap B) = P(B)(1 - P(A)) = P(A^c)P(B)$. It follows that if A and B are independent, so are A^c and B^c.

The "nicest" probability measure we could place on $\mathbb{N}$ would be a uniform measure: one where $P(\{m\}) = P(\{n\})$ for all m, $n \in \mathbb{N}$. But this is clearly not possible, since $\mathbb{N}$ can be written as the union of the sets $\{n\}$ for each $n \in \mathbb{N}$, and these sets are pairwise disjoint. If each of these sets has measure 0, then by the second axiom $P(\mathbb{N}) = 0$, and if each has measure $\alpha > 0$, then again by the second axiom $P(\mathbb{N}) = \infty$, and in either case this contradicts the third axiom.

But while we cannot define a uniform probability measure on $\mathbb{N}$, we may still be able to define a probability measure that preserves some of our intuitions about the integers: for instance, that the probability that a number is divisible by an integer a is $1/a$. It turns out that we cannot do this. We follow the exposition in Chapter III.1 in [13].

Theorem (Theorem 1 from Chapter III.1 in [13]). Let $a\mathbb{N} = \{k \in \mathbb{N} : a \mid k\}$. There does not exist a probability measure P on $\mathbb{N}$ such that for all $a \in \mathbb{N}$, $P(a\mathbb{N}) = 1/a$.

To prove this theorem, we will need one of Mertens' formulas. We omit the proof.

Lemma (Theorem 2.7e in [7]). For $x \geq 2$, $\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} = e^{\gamma} \log x + O(1)$, where γ is Euler's constant: $\gamma = \lim_{x \to \infty} (\sum_{n \leq x} \frac{1}{n} - \log x)$.

Proof of the Theorem. We employ contradiction. Assume such a measure exists.

We know that if $(a, b) = 1$, $a\mathbb{N} \cap b\mathbb{N} = ab\mathbb{N}$, and so the events $a\mathbb{N}$ and $b\mathbb{N}$ are in this case independent. Thus so are $a\mathbb{N}^c$ and $b\mathbb{N}^c$, and hence

$$P(a\mathbb{N}^c \cap b\mathbb{N}^c) = \left(1 - \frac{1}{a}\right)\left(1 - \frac{1}{b}\right)$$

when $(a, b) = 1$. By induction we obtain that for any integers m, n with $m < n$, that, with p

denoting a prime,

$$P(\{m\}) \leq P\left(\bigcap_{m<p\leq n} p\mathbb{N}^c\right) = \prod_{m<p\leq n}\left(1-\frac{1}{p}\right).$$

But by the Lemma, since we may take n to be as large as desired, the product can be made arbitrarily close to 0. Hence for any $m \in \mathbb{N}$, $P(\{m\}) = 0$, but, as above, this contradicts the definition of a probability measure. Q.E.D.

We also note that while intuitively we would like the probability that an integer n is greater than a given integer N to be 1, that is, that $P(\{n \in \mathbb{N} : n > N\}) = 1$, since this set contains all the integers save only finitely many, this would require that for all $n \leq N$, $P(\{n\}) = 0$. But again, since we would want this to be true for all N, we have the same contradiction. (Alternatively, we would like the probability that an integer n lies in some given finite subset of $\mathbb{N}$ to be zero, but by the same argument this cannot happen.) It is for this reason that we have the notion of the density of a set of integers.

b. Density

By using density, we preserve our intuitions that the "probability" of $a\mathbb{N}$ is $1/a$ and that a finite set has "probability" zero, at the cost of various properties of probability measures. We may define the density of a set of integers generally as follows:

Definition. Let $\{\lambda_n\}_{n=1}^{\infty}$ be a sequence of nonnegative real numbers with $\sum_{n=1}^{\infty}\lambda_n = \infty$. Let $A \subseteq \mathbb{N}$ and define the *density* $d(A)$ of A as the limit (if it exists), of

$$d(A;x) := \left(\sum_{a\leq x,\, a\in A}\lambda_a\right)\left(\sum_{n\leq x}\lambda_n\right)^{-1},$$

as $x \to \infty$.

Certainly any finite set has $d(A) = 0$, and so this is not a probability measure on $\mathbb{N}$.

Since we define density as a limit, we can replace the sum $\sum_{n\leq x}\lambda_n$ by any function $f(x)$ asymptotic to it. For instance, since $\sum_{n\leq x} 1 = x + O(1)$, we can define the *natural density* of a set $A \subseteq \mathbb{N}$ by

$$\mathbf{d}(A) = \lim_{x\to\infty}\frac{1}{x}\left(\#\{a \leq x : a \in A\}\right).$$

We also define the upper and lower asymptotic densities of A, denoted $\overline{\mathbf{d}}(A)$ and $\underline{\mathbf{d}}(A)$ respectively, by replacing the limit in the above with lim sup or lim inf respectively.

Let us note the following:

i. If $A = \{n \in \mathbb{N} : n \equiv a \,(mod\, q)\}$, then since $\#\{a \leq x : a \in A\} = [x/q] + O(1)$, any arithmetic progression modulo q has a natural density, and it is $1/q$, as our intuition says it is.

ii. The sequence $a_1 < a_2 < \ldots$ has natural density $\alpha \in [0,1]$ if and only if $\lim_{n\to\infty} n/a_n = \alpha$.

iii. Not all sequences have a natural density. Consider the sequence A of integers n with first digit 1 in their base 10 expansion. Then $A = \bigcup_{k=0}^{\infty} \left\{n : 10^k \leq n < 2 \cdot 10^k\right\}$. Define $A(x) := \#\{a \leq x : a \in A\}$. Then for $m \geq 1$, we have that $A(10^m - 1) = \frac{1}{9}(10^m - 1)$ while $A(2 \cdot 10^m - 1) = \frac{5}{9}(2 \cdot 10^m - 1) + \frac{4}{9}$. Hence $\underline{\mathbf{d}}(A) \leq 1/9$, $\overline{\mathbf{d}}(A) \geq 5/9$, and so $d(A)$ does not exist. We will return to this example later.

iv. If we denote by ν_N the probability measure on $\mathbb{N}$ assigning probability $1/N$ to each of the first N positive integers, we have that for any sequence A, $\mathbf{d}(A) = \lim_{N\to\infty} \nu_N(A)$, assuming both exist, so that while natural density itself is not a probability measure, it is the limit of probability measures, and it is these measures that have properties close to those we expect intuitively.

As we noted above, the sum $\sum_{n\leq x} \lambda_n$ can be replaced by any function $f(x)$ asymptotic to it. We have examined the case $\lambda_n = 1$. We can also consider other values for λ_n. We list some possible choices, with smooth asymptotic $f(x)$, below.

#	λ_n	$f(x)$
1	1	x
2	$1/n$	$\log x$
3	$\begin{cases} 1, & \text{if } n \text{ is prime} \\ 0, & \text{else} \end{cases}$	$\frac{x}{\log x}$, $\mathrm{li}(x)$
4	$\begin{cases} \frac{1}{p}, & \text{if } n \text{ is a prime } p \\ 0, & \text{else} \end{cases}$	$\log\log x$
5	$\begin{cases} \log p, & \text{if } n \text{ is a prime } p \\ 0, & \text{else} \end{cases}$	x
6	$\begin{cases} \frac{\log p}{p}, & \text{if } n \text{ is a prime } p \\ 0, & \text{else} \end{cases}$	$\log x$
7	$\Lambda(n)$	x
8	$\frac{\Lambda(n)}{n}$	$\log x$
9	$\begin{cases} 1, & \text{if } n \equiv a \pmod q \\ 0, & \text{else} \end{cases}$	$\frac{x}{q}$

(2 follows from Corollary 1.15 in [7]; 3, 5, and 7 from Theorem 6.9 in the same; 4, 6, and 8 follow from Theorem 2.7 there.)

We will use choice 2 in our discussion of how often a "contestant" in a prime number race is winning. It is with this notion of density that we make statements such as "$\pi(n; 4, 3) > \pi(n; 4, 1)$ is true for over 99% of integers n". We call this density *logarithmic density*.

We denote the logarithmic density of a sequence A by $\delta(A)$, and the upper and lower logarithmic densities (again defined by replacing the limit in the definition by lim sup or lim inf respectively) by $\overline{\delta}(A)$ and $\underline{\delta}(A)$ respectively.

It turns out that the existence of logarithmic density is a weaker statement than the existence of natural density. We make this precise as follows.

Theorem (Theorem 2 from Chapter III.1 in [13]). For any sequence $A \in \mathbb{N}$, $\underline{\mathbf{d}}(A) \leq \underline{\delta}(A) \leq \overline{\delta}(A) \leq \overline{\mathbf{d}}(A)$, and so if A has a natural density, it has a logarithmic density and the two are equal.

Proof. Define $A(x) := \sum_{a \leq x,\, a \in A} 1$ and $L(x) := \sum_{a \leq x,\, a \in A} 1/a$. Using integration by parts on $L(x)$, we obtain $L(x) = \frac{A(x)}{x} + \int_1^x \frac{A(t)}{t^2} dt$, for $x \geq 1$.

Now let $\epsilon > 0$. Then we have some $t_0(\epsilon)$ such that for all $t > t_0(\epsilon)$, $\underline{\mathbf{d}}(A) - \epsilon \leq \frac{A(t)}{t} \leq \overline{\mathbf{d}}(A) + \epsilon$. Substituting this into our formula for $L(x)$, with $x > t_0$, we obtain

$$(\underline{\mathbf{d}}(A) - \epsilon) \log(x/t_0) \leq L(x) \leq 1 + \log t_0 + \left(\overline{\mathbf{d}}(A) + \epsilon\right) \log(x/t_0).$$

(Note that $\frac{A(t)}{t}$ is bounded below trivially by 0 and above trivially by 1.) Taking x to ∞ (using $\liminf$ and $\limsup$ if necessary) and then ϵ to 0 gives the stated inequalities. Q.E.D.

To show that a set can have a logarithmic density without having a natural density, we consider the set $A = \bigcup_{k=0}^{\infty} \left\{n : 10^k \leq n < 2 \cdot 10^k\right\}$ again. We have seen that this set does not have a natural density. However, it does have a logarithmic density. To see this, consider $L(x)$ as defined in the proof of the above theorem. Then

$$L(x) = \sum_{a \leq x} \frac{1}{a} = \sum_{0 \leq k \leq \log x/\log 10} \sum_{\substack{10^k \leq n < 2 \cdot 10^k \\ n \leq x}} \frac{1}{n} = \sum_{0 \leq k \leq \log x/\log 10} \left(\log 2 + O\left(\frac{1}{k+1}\right)\right)$$
$$+ O(1) = \frac{\log 2}{\log 10} \log x + O(\log \log x),$$

from which it follows that $\delta(A) = \frac{\log 2}{\log 10}$.

From this we see that while "proper" probability theory will not yield the results we expect in the case we want, we have a way of recovering our intuitive results, through natural density. Even when that fails to exist, we can look to see if the logarithmic density exists, as it coincides with natural density when the latter exists.

V. The Explicit Formula

It turns out that there is a formula for $\psi(x; q, a)$ that depends upon the zeros of $L(s, \chi)$ over all characters $\chi \bmod q$. To derive this formula, we examine a function more closely related to $L(s, \chi)$, $\psi(x, \chi)$.

Definition. $\psi(x, \chi) = \sum_{n \leq x} \chi(n)\Lambda(n)$.

We have an "orthogonality" relationship between characters that allows us to detect a given reduced residue class:

$$\frac{1}{\phi(q)} \sum_{\chi} \overline{\chi}(a)\chi(n) = \begin{cases} 1, & \text{if } n \equiv a \pmod q \\ 0, & \text{else,} \end{cases}$$

and from this we see that $\psi(x;q,a) = \frac{1}{\phi(q)}\sum_{\chi}\overline{\chi}(a)\psi(x,\chi)$.

We define the related function $\psi_0(x,\chi)$ by modifying the sum defining $\psi(x,\chi)$ slightly: if x is an integer such that $\Lambda(x) \neq 0$, count the last summand with only half the weight, that is, make it $(\chi(x)\Lambda(x))/2$ rather than $\chi(x)\Lambda(x)$. Then as long as $\chi \neq \chi_0$, and q is fixed, we have

$$\psi_0(x,\chi) = -\sum_{\rho}\frac{x^\rho}{\rho} + O\left(\log x\right), \tag{1}$$

where the sum runs over all nontrivial zeros ρ of $L(s,\,\chi)$, from Corollary 12.11 and (12.13) in [7]. We may replace $\psi_0(x,\chi)$ with $\psi(x,\chi)$, as the difference is $O\left(\log x\right)$.

For the principal character χ_0, we have the same formula except for an additional term x. These terms arise from Perron's formula (Theorem 5.1 in [7]), taking $\alpha(s)$ to be $-\frac{L'}{L}(s,\,\chi)$. Plugging these into our formula for $\psi(x;q,a)$ in terms of $\psi(x,\chi)$, we obtain an explicit formula for $\psi(x;q,a)$:

$$\psi(x;q,a) = \frac{x}{\phi(q)} - \frac{1}{\phi(q)}\sum_{\chi}\left(\overline{\chi}(a)\sum_{\rho}\frac{x^\rho}{\rho}\right) + O(\log x), \tag{2}$$

where the outer sum runs over the characters χ modulo q, and the inner sum runs over the nontrivial zeros of $L(s,\,\chi)$, that is, those with $0 < \mathrm{Re}(\rho) < 1$.

From this, one can derive that under the Generalized Riemann Hypothesis, we have a formula for the difference between $\pi(x;q,a_1)$ and $\pi(x;q,a_2)$:

$$\frac{\pi(x;q,a_1) - \pi(x;q,a_2)}{\sqrt{x}/\log x} = \frac{1}{\phi(q)}\left(\left(\sum_{\chi}(\overline{\chi}(a_1) - \overline{\chi}(a_2))\sum_{\rho=\frac{1}{2}+i\gamma}\frac{x^{i\gamma}}{\rho}\right) + N(a_2,q) - N(a_1,q)\right) + O(1), \tag{3}$$

where the outer sum runs over the characters χ modulo q, the inner sum runs over the nontrivial zeros of $L(s,\,\chi)$ (all of which have real part $1/2$ under GRH), and

$$N(a,q) = \#\left\{m \in \mathbb{Z}/q\mathbb{Z} :\, m^2 \equiv a \,(\mathrm{mod}\, q)\right\}.$$

The presence of the term $N(a_2,q) - N(a_1,q)$ in (3) already indicates some sort of bias in favour of nonsquares modulo q, as it is positive when a_2 is a square and a_1 is not, and negative when a_1 is a square and a_2 is not. It would seem that this bias disappears when a_1 and a_2 are both squares or both nonsquares; we shall discuss this further shortly.

We also have a truncated form of (1) ((2.1) in [10]), with $\gamma = \mathrm{Im}(\rho)$ and where the sum runs over nontrivial zeros ρ of $L(s,\,\chi)$:

$$\psi(x,\chi) = -\sum_{|\gamma|\leq T}\frac{x^\rho}{\rho} + O\left(\frac{x\log^2(xT)}{T} + \log x\right). \tag{4}$$

One natural question is to ask whether the double sum in (3) ever "overcomes" the negative number from $N(a_2,q) - N(a_1,q)$ (when a_1 is a square and a_2 is not). In 1914, Littlewood showed in [5] the following:

Theorem. There are arbitrarily large values of x for which $\pi(x;4,1) > \pi(x;4,3)$; in fact, there are arbitrarily large x for which

$$\frac{\pi(x;4,1) - \pi(x;4,3)}{\sqrt{x}/\log x} \geq \frac{1}{2} \log\log\log x.$$

So $\pi(x;4,3)$ never quite stays in the lead, and in fact $\pi(x;4,1)$ sometimes races out far ahead of $\pi(x;4,3)$. But the values of x for Littlewood's result guarantees this are extremely large. For instance, this lower bound is not attained the first time $\pi(x;4,1)$ is ahead of $\pi(x;4,3)$, at $x = 26861$, for there the left-hand side is less than 0.07 while the right-hand side is greater than 0.4. At the larger gap Leech [4] found, at $x = 623681$, where $\pi(x;4,1) - \pi(x;4,3) = 8$, the left-hand side is less than 0.15, while the right-hand side is greater than 0.45. It is worth noting, however, that $\log\log\log x$ grows extremely slowly; the least integer x for which $\log\log\log x \geq 1$ is $x = 3814280$, and the least integer x for which $\frac{1}{2}\log\log\log x \geq 1$ exceeds 10^{702}, and so it is not implausible that this bound is attained infinitely often.

However we are interested in how often $\pi(x;4,1) > \pi(x;4,3)$. Our discussion of notions of density above will aid us in answering this question.

VI. More About Prime Number Races

With our notions of density and the explicit formula in hand, we are better equipped to handle prime number races. We follow part of the exposition in [10] for now.

The sets whose densities we wish to know are those corresponding to a specific order of the race. That is, if $a_1, a_2, \ldots, a_r \in (\mathbb{Z}/q\mathbb{Z})^\times$, we want to know the density of the set of natural numbers n for which $\pi(n;q,a_1) > \pi(n;q,a_2) > \cdots > \pi(n;q,a_r)$. We do not quite look at this set, but rather at the set of real numbers x for which $\pi(x;q,a_1) > \pi(x;q,a_2) > \cdots > \pi(x;q,a_r)$, and we call this set $P_{q;a_1,a_2,\ldots,a_r}$. We can still define logarithmic density here: for a set P of real numbers, define its logarithmic density $\delta(P)$ by

$$\delta(P) = \lim_{X\to\infty} \frac{1}{\log X} \int_{P\cap[2,X]} \frac{dt}{t},$$

and define the upper and lower densities $\overline{\delta}(P)$ and $\underline{\delta}(P)$ by taking lim sup and lim inf respectively instead of the limit.

To help us compute the logarithmic density of these $P_{q;a_1,a_2,\ldots,a_r}$, we will define a few other functions. First, we define $E_{q;a_1,a_2,\ldots,a_r}(x)$ by

$$E_{q;a_1,a_2,\ldots,a_r}(x) = \frac{\log x}{\sqrt{x}} \left(\phi(q)\pi(x;q,a_1) - \pi(x), \ldots, \phi(q)\pi(x;q,a_r) - \pi(x)\right),$$

for $x \geq 2$. Each coordinate measures how far away $\pi(x;q,a_i)$ is from its "fair share". The normalization $\frac{\log x}{\sqrt{x}}$ is there so that, under the Generalized Riemann Hypothesis, E varies in a controllable way. In fact, it has a limiting distribution in the following sense:

Theorem 1. Assume GRH. Then $E_{q;a_1,a_2,\ldots,a_r}$ has a limiting distribution $\mu_{q;a_1,a_2,\ldots,a_r}$ on $\mathbb{R}^r$, that is,

$$\lim_{X\to\infty} \frac{1}{\log X} \int_2^X f(E_{q;a_1,a_2,\ldots,a_r}(x)) \frac{dx}{x} = \int_{\mathbb{R}^r} f(x) d\mu_{q;a_1,a_2,\ldots,a_r}(x),$$

for all bounded, continuous functions f on $\mathbb{R}^r$.

We note that if $\mu_{q;a_1,a_2,\dots,a_r}$ is absolutely continuous (with respect to Lebesgue measure), then

$$\delta(P_{q;a_1,a_2,\dots,a_r}) = \mu_{q;a_1,a_2,\dots,a_r}\left(\{x \in \mathbb{R}^r : x_1 > x_2 > \cdots > x_r\}\right).$$

However, this would require a version of Theorem 1 that allows for discontinuous integrands, or at least characteristic functions of sufficiently nice sets.

A related problem is that of the number of primes that are quadratic residues as compared to those that are not. Here we must take $q = 4$, $q = p^\alpha$, or $q = 2p^\alpha$, where $\alpha \geq 1$ and p is an odd prime, as for these moduli exactly half of the reduced residues are squares. Then define $\pi_R(x, q)$ to be the number of primes that are squares modulo q and do not exceed x, and define $\pi_N(x, q)$ to be the number of primes that are not squares modulo q and do not exceed x. We define the sets $P_{q;N,R}$ and $P_{q;R,N}$ in a fashion similar to how we defined $P_{q;a_1,a_2,\dots,a_r}$, that is,

$$P_{q;N,R} = \{x \geq 2 : \pi_N(x, q) > \pi_R(x, q)\},$$

$$P_{q;R,N} = \{x \geq 2 : \pi_R(x, q) > \pi_N(x, q)\}.$$

Now we define $E_{q;N,R}(x) := \frac{\log x}{\sqrt{x}}(\pi_N(x, q) - \pi_R(x, q))$, and we can show that $\underline{\delta}(P_{q;N,R})\underline{\delta}(P_{q;R,N}) > 0$, so that the residues win at times. We construct measures $\mu_{q;N,R}$ and $\mu_{q;R,N}$ similarly.

We can also construct measures $\mu^T_{q;a_1,a_2,\dots,a_r}$ that are defined in terms of nontrivial zeros of $L(s, \chi)$, where χ runs over the Dirichlet characters modulo q and the imaginary parts γ_χ of the zeros are at most T in absolute value. It is to facilitate working with these measures that we make the Grand Simplicity Hypothesis regarding linear independence of imaginary parts of zeros of Dirichlet L-functions.

Under GRH and GSH we can find an explicit formula for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}$ of $\mu_{q;a_1,a_2,\dots,a_r}$:

$$\hat{\mu}_{q;a_1,\dots,a_r}(\xi_1, \dots, \xi_r) = \left(\exp\left(i\sum_{j=1}^{r} c(q, a_j)\xi_j\right)\right)\left(\prod_{\substack{\chi \neq \chi_0 \\ \chi \bmod q}} \prod_{\gamma_\chi > 0} J_0\left(\frac{2\left|\sum_{j=1}^{r} \chi(a_j)\xi_j\right|}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}\right)\right), \tag{5}$$

where χ_0 is, as above, the principal character, $c(q, a) = N(a, q) - 1$ (where $N(a, q)$ is the number of square roots of a modulo q), and

$$J_0(z) = \sum_{m=0}^{\infty} \frac{(-1)^m (\frac{z}{2})^{2m}}{(m!)^2},$$

the zeroth Bessel function. If one can compute many zeros of $L(s, \chi)$, one can use this formula to approximate $\mu_{q;a_1,a_2,\dots,a_r}$.

We are also interested in knowing when we may permute the r variables $(x_1, x_2, \dots, x_r)$ in $\mathbb{R}^r$ without changing $\mu_{q;a_1,a_2,\dots,a_r}$, that is to say, whether $\mu_{q;a_1,a_2,\dots,a_r}$ is symmetric. If it is, we say that $(q; a_1, a_2, \dots, a_r)$ is *unbiased*. In this case we have $\delta(P_{q;a_1,a_2,\dots,a_r}) = \frac{1}{r!}$, and in fact we have this equality for any permutation of the parameters a_i. (It is not known whether the converse is true.) Unfortunately, it is rarely the case that our choice of a_i is unbiased:

Theorem 2. Assume GRH and GSH. Then $(q;\, a_1, a_2, \ldots, a_r)$ is unbiased if and only if $r = 2$ and $c(q, a_1) = c(q.a_2)$, or $r = 3$ and there is some $\rho \in (\mathbb{Z}/q\mathbb{Z})^\times$, $\rho \neq 1$, with $\rho^3 = 1$, $a_2 = a_1\rho$ and $a_3 = a_1\rho^2$.

Analysing the symmetry of $\mu_{q;N,R}$ displays a bias toward nonresidues, so that $\delta(P_{q;N,R}) \in (\frac{1}{2}, 1)$. More generally, when we race a square against a nonsquare, there is a bias toward the nonsquare, and this is as we expect. This follows from an explicit formula for the Fourier transform $\hat{\mu}_{q;R,N}$ of $\mu_{q;R,N}$ that is similar to the one given above for $\hat{\mu}_{q;a_1,\ldots,a_r}$; we will give some details later.

However, if we let the modulus q grow, we find that the bias dissipates. In fact, we find the following theorem:

Theorem 3. Assume GRH and GSH. Fix r. Then:

$$\lim_{q\to\infty} \max_{a_1,\ldots,a_r \in (\mathbb{Z}/q\mathbb{Z})^\times} \left| \delta(P_{q;a_1,\ldots,a_r}) - \frac{1}{r!} \right| = 0.$$

This occurs even for the extreme case $\mathrm{P}_{q;N,R}$: as $q \to \infty$, $\delta(P_{q;N,R}) \to \frac{1}{2}$. Indeed we have a "central limit theorem":

Theorem 4. Assume GRH and GSH. Let $\overline{\mu}_{q;N,R}$ be the limiting distribution of $\frac{E_{q;N,R}(x)}{\sqrt{\log q}}$. Then $\overline{\mu}_{q;N,R}$ converges in measure to the Gaussian distribution $\frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ as $q \to \infty$.

We can think of this as meaning that the "variance", $\log q$, grows as q does to swamp the bias.

The proofs of these results are highly technical, and we do not present them in full.

Theorem 1 is shown by way of three lemmas. We assume the Generalized Riemann Hypothesis in the following.

Lemma 1. Let $E(x, q, a) = \frac{\log x}{\sqrt{x}}(\phi(q)\pi(x; q, a) - \pi(x))$. As $x \to \infty$,

$$E(x, q, a) = -c(q, a) + \sum_{\chi \neq \chi_0} \overline{\chi}(a) \frac{\psi(x, \chi)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right).$$

The constant term both here and in (3) seem to be skewing these functions in favour of nonsquares. This result is shown by considering the explicit formulas in V.

Consider (4). If we assume GRH, it can be rewritten as follows:

$$\psi(x, \chi) = -\sqrt{x} \sum_{|\gamma| \leq T} \frac{x^{i\gamma}}{\frac{1}{2} + i\gamma} + O\left(\frac{x \log^2(xT)}{T} + \log x\right). \tag{6}$$

Combining this with Lemma 1, we find that for $T \geq 1$ and $x \in [2, X]$

$$E(x, q, a) = -c(q, a) - \sum_{\chi \neq \chi_0} \overline{\chi}(a) \sum_{|\gamma| \leq T} \frac{x^{i\gamma}}{\frac{1}{2} + i\gamma} + \epsilon_a(x, T, X),$$

where

$$\epsilon_a(x,T,X) = -\sum_{\chi\neq\chi_0}\overline{\chi}(a)\sum_{T\leq|\gamma|\leq X}\frac{x^{i\gamma}}{\frac{1}{2}+i\gamma} + O_q\left(\frac{\sqrt{x}\log^2 X}{X}+\frac{1}{\log x}\right).$$

Now define $y = \log x$. Then $dy = {}^{dx}/_{x}$, and we can state the second lemma.

Lemma 2. Let $T \geq 1$ and $Y \geq \log 2$. Then

$$\int_{\log 2}^{Y}\left|\epsilon_a\left(e^y,T,e^Y\right)\right|^2 dy \ll \frac{Y\log^2 T}{T}+\frac{\log^3 T}{T}.$$

This lemma follows by expanding the integrand as a finite sum, integrating term by term, and using Corollary 14.7 in [7], which states that $N(T,\chi)$, the number of nontrivial zeros of $L(s,\,\chi)$ with imaginary part in $[0,T]$, is given by

$$N(T,\chi) = \frac{T}{2\pi}\log\left(\frac{qT}{2\pi}\right)-\frac{T}{2\pi}+O\left(\log qT\right).$$

We note that the number of nontrivial zeros of $L(s,\,\chi)$ with imaginary part in $[-T,0]$ is $N(T,\overline{\chi})$ by the mirroring of zeros of $L(s,\,\chi)$ and $L(s,\,\overline{\chi})$ in the critical strip (see p. 333 in [7]).

Lastly, we need a result on the limiting distribution of $E(x,q,a)-\epsilon_a(x,T,X)$. We consider a Lipschitz continuous function $f:\mathbb{R}^r\to\mathbb{R}$ with Lipschitz constant c_f, that is, for all $x,\,y\in\mathbb{R}^r$,

$$|f(x)-f(y)| \leq c_f\,|x-y|.$$

Now let $E(y) = (E(e^y,q,a_1),\ldots,E(e^y,q,a_r))$, and consider

$$\frac{1}{Y}\int_{\log 2}^{Y} f(E(y))dy.$$

To facilitate studying this, we define

$$E_j^{(T)}(y) = -c(q,a_j)-\sum_{\chi\neq\chi_0}\overline{\chi}(a_j)\sum_{|\gamma|\leq T}\frac{e^{iy\gamma}}{\frac{1}{2}+i\gamma}$$

and let $E^{(T)}(y) = \left(E_1^{(T)}(y),\ldots,E_r^{(T)}(y)\right)$; we then look at

$$\frac{1}{Y}\int_{\log 2}^{Y} f(E^{(T)}(y))dy.$$

Lemma 3. For each T we have a probability measure ν_T on $\mathbb{R}^r$ with

$$\nu_T(f) := \int_{\mathbb{R}^r} f(x)d\nu_T(x) = \lim_{Y\to\infty}\frac{1}{Y}\int_{\log 2}^{Y} f(E^{(T)}(y))dy$$

for all bounded, continuous functions f on $\mathbb{R}^r$. Furthermore, we have a constant c, depending only on the modulus q, such that the support of ν_T is contained in the ball $B(0, c\log^2 T)$.

From this, letting $\epsilon^{(T)}(y) := E(y) - E^{(T)}(y)$, we have, when f is Lipschitz continuous with Lipschitz constant c_f, that

$$\frac{1}{Y}\int_{\log 2}^{Y} f(E(y))dy = \frac{1}{Y}\int_{\log 2}^{Y} f(E^{(T)}(y))dy + O\left(\frac{c_f}{Y}\int_{\log 2}^{Y}\left|\epsilon^{(T)}(y)\right|dy\right),$$

with the implicit constant depending only on q. By Jensen's inequality (see p. 461 in [3]) this is

$$\frac{1}{Y}\int_{\log 2}^{Y} f(E^{(T)}(y))dy + O\left(\frac{c_f}{\sqrt{Y}}\left(\int_{\log 2}^{Y}\left|\epsilon^{(T)}(y)\right|^2 dy\right)^{\frac{1}{2}}\right),$$

and by Lemma 2, this becomes

$$\frac{1}{Y}\int_{\log 2}^{Y} f(E^{(T)}(y))dy + O\left(c_f\left(\frac{\log T}{\sqrt{T}} + \frac{\log^2 T}{Y\sqrt{T}}\right)\right).$$

Taking lim inf and lim sup as $Y \to \infty$ here gives us that

$$\begin{aligned}\nu_T(f) - O\left(\frac{c_f \log T}{\sqrt{T}}\right) \leq \liminf \frac{1}{Y}\int_{\log 2}^{Y} f(E(y))dy \\ \leq \limsup \frac{1}{Y}\int_{\log 2}^{Y} f(E(y))dy \leq \nu_T(f) + O\left(\frac{c_f \log T}{\sqrt{T}}\right). \quad (7)\end{aligned}$$

Since T can be arbitrarily large, we have that the desired limit exists, so that for Lipschitz continuous functions f,

$$\mu(f) := \lim_{Y\to\infty}\frac{1}{Y}\int_{\log 2}^{Y} f(E(y))dy.$$

This provides the limiting distribution, since the fact that μ is a probability measure on $\mathbb{R}^r$ follows from the fact that ν_T is for all T.

The result about the support of ν_T from Lemma 3 allows us to say something about the behaviour of μ away from the origin. Let $B(0,\lambda)$ be the open ball of radius λ centred at the origin; then for $\lambda = c\log^2 T$ we have, from (7), that

$$\mu\left(B(0,\lambda)^c\right) = \nu_T\left(B(0,\lambda)^c\right) + O\left(\frac{\log T}{\sqrt{T}}\right) = O\left(\frac{\log T}{\sqrt{T}}\right) = O\left(\sqrt{\lambda}e^{-c\sqrt{\lambda}}\right) = O\left(e^{-c_2\sqrt{\lambda}}\right).$$

Here c_2 depends only upon q.

Just as we can, under GRH and the Grand Stability Hypothesis (GSH), find a formula for the Fourier transform of $\mu_{q;a_1,a_2,\ldots,a_r}$, we can find one for the Fourier transform $\hat{\mu}_{q;R,N}$ of $\mu_{q;R,N}$; recall that we require here that $q = 4$, or that there is some prime p and some $r \in \mathbb{N}$ with $q = p^r$ or $q = 2p^r$. For such q there is one real, non-principal Dirichlet character. We denote this character

χ_1, and we denote a nontrivial zero of $L(s, \chi_1)$ by $\frac{1}{2} + i\gamma_{\chi_1}$. Then we may state the formula, which is (3.4) in [10]:

$$\hat{\mu}_{q;R,N}(\xi) = e^{i\xi} \prod_{\gamma_{\chi_1}>0} J_0\left(\frac{2\xi}{\sqrt{\frac{1}{4}+\gamma_{\chi_1}^2}}\right).$$

But J_0 is even, and so this requires that the density function of $\mu_{q;R,N}$ be symmetric about $t = -1$. But this function is entire, so it is not identically zero on $(-1, 0)$, and thus

$$\delta(P_{q;R,N}) = \int_0^\infty d\mu_{q;R,N}(t) < \frac{1}{2}.$$

In other words, the nonsquares "win".

Theorem 2, giving a sufficient condition on $\delta(P_{q;a_1,a_2,\ldots,a_r}) = \frac{1}{r!}$ for any permutation of the a_i, is proved by considering the explicit formula (5) after first proving a symmetry in the argument of the Bessel function. Define

$$B_\chi(\xi_1, \ldots, \xi_r) := \left|\sum_{j=1}^{r} \chi(a_j)\xi_j\right|.$$

Lemma 4. $B_\chi(\xi_1, \ldots, \xi_r)$ is symmetric in its arguments for all χ if and only if $r = 2$ and $c(q, a_1) = c(q.a_2)$, or $r = 3$ and there is some $\rho \in (\mathbb{Z}/q\mathbb{Z})^\times$, $\rho \neq 1$, with $\rho^3 = 1$, $a_2 = a_1\rho$ and $a_3 = a_1\rho^2$.

The proof is a case-by-case calculation; we do note that it shows that if $r \geq 4$ and B_χ is assumed to be symmetric, any three of the a_i would be related as in the $r = 3$ condition, which is a contradiction to their being distinct. (Note that Theorem 2 gives a sufficient condition. It is possible to "get lucky" and have $\delta(P_{q;a_1,a_2,\ldots,a_r}) = \frac{1}{r!}$ for all permutations of the a_j even if the conditions in Theorem 2 fail. We will give some cases in which this does not happen later.)

One interesting consequence of Theorem 2 is that $(q; a_1, a_2, \ldots, a_r)$ is never unbiased in the case where q is a prime congruent to 2 modulo 3, as for such primes, each reduced residue class has a unique cube root.

To prove Theorems 3 and 4, one shows that the Fourier transforms of the measures under consideration converge (as $q \to \infty$) to the Fourier transform of a Gaussian measure (of appropriate dimension); Lévy's theorem ([6]) then implies that the measures themselves converge to a Gaussian measure.

Another result on symmetries was proved (under GRH and GSH) by Feuerverger and Martin in [2]:

Theorem 5. Let q, $r \geq 2$ be integers, and let $a_1, \ldots, a_r$ be distinct reduced residue classes modulo q. Then all of the following hold:
a. Let a_j^{-1} be the multiplicative inverse of a_j. Then $\delta(P_{q;a_1,\ldots,a_r}) = \delta(P_{q;a_1^{-1},\ldots,a_r^{-1}})$.
b. Let b be a reduced residue class modulo q such that for each $1 \leq j \leq r$, we have $c(q, a_j) = c(q, ba_j)$. (For example, this holds if b is a square modulo q.) Then $\delta(P_{q;a_1,\ldots,a_r}) = \delta(P_{q;ba_1,\ldots,ba_r})$.

c. If each a_j is a square modulo q and b is a reduced residue class modulo q, then $\delta(P_{q;a_1,\dots,a_r}) = \delta(P_{q;ba_1,\dots,ba_r})$.
d. If all the a_j are squares modulo q, or they are all nonsquares modulo q, we have $\delta(P_{q;a_1,\dots,a_r}) = \delta(P_{q;a_r,\dots,a_1})$.
e. Let b be a reduced residue class modulo q such that for each $1 \le j \le r$, we have $c(q,a_j) \ne c(q,ba_j)$. (For example, this holds if q is an odd prime power, or twice an odd prime power, and b is a nonsquare modulo q.) Then $\delta(P_{q;a_1,\dots,a_r}) = \delta(P_{q;ba_r,\dots,ba_1})$.

For instance, modulo 8 or 12, the multiplicative group of the reduced residue classes is isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. Thus each reduced residue class is its own multiplicative inverse, and the only square is 1. Hence in determining the symmetries in the prime number race involving the three nonsquares, criteria a, b, c and e cannot apply, but criterion d implies all of the following:

$$\begin{aligned}
\delta(P_{8;3,5,7}) &= \delta(P_{8;7,5,3}),\\
\delta(P_{8;5,3,7}) &= \delta(P_{8;7,3,5}),\\
\delta(P_{8;3,7,5}) &= \delta(P_{8;5,7,3}),\\
\delta(P_{12;5,7,11}) &= \delta(P_{12;11,7,5}),\\
\delta(P_{12;7,5,11}) &= \delta(P_{12;11,5,7}),\\
\delta(P_{12;5,11,7}) &= \delta(P_{12;7,11,5}).
\end{aligned}$$

Feuerverger and Martin also compute these densities explicitly:

Theorem 6. Assume GRH and GSH. With all densities given within 0.000001 of the true value, we have

$$\begin{aligned}
\delta(P_{8;3,5,7}) &= \delta(P_{8;7,5,3}) = 0.1928013,\\
\delta(P_{8;5,3,7}) &= \delta(P_{8;7,3,5}) = 0.1407724,\\
\delta(P_{8;3,7,5}) &= \delta(P_{8;5,7,3}) = 0.1664263,\\
\delta(P_{12;5,7,11}) &= \delta(P_{12;11,7,5}) = 0.1984521,\\
\delta(P_{12;7,5,11}) &= \delta(P_{12;11,5,7}) = 0.1799849,\\
\delta(P_{12;5,11,7}) &= \delta(P_{12;7,11,5}) = 0.1215630.
\end{aligned}$$

Hence we have cases in which the densities are not invariant under all permutations of the a_j, that is, in light of Theorem 2, we do not always "get lucky" and have $\delta(P_{q;a_1,a_2,\dots,a_r}) = \frac{1}{r!}$ for all permutations of the a_j.

They also establish inequalities between various permutations of the a_j provided those residue classes satisfy certain conditions.

Theorem 7. Assume GRH and GSH. Let $q \ge 2$ be an integer, and let N, N', S, S' be distinct reduced residue classes modulo q where N and $N'$are nonsquares, and S and S' are squares. Then:
a. $\delta(P_{q;N,N',S}) > \delta(P_{q;S,N',N})$;
b. $\delta(P_{q;N,S,S'}) > \delta(P_{q;S',S,N})$;
c. $\delta(P_{q;N,S,N'}) > \delta(P_{q;N',S,N})$ if and only if $\delta(P_{q;N,S}) > \delta(P_{q;N',S})$;

d. $\delta(P_{q;S,N,S'}) > \delta(P_{q;S',N,S})$ if and only if $\delta(P_{q;S,N}) > \delta(P_{q;S',N})$.

In other words, nonsquares tend to dominate three-way races, and the behaviour of two-way races between squares and nonsquares dictates the behaviour of the three-way race involving those squares and nonsquares, and vice versa.

Going back to the modulo 4 race, the computation of $\delta(P_{4;N,R})$, among other such densities, is a final result in Rubenstein and Sarnak's paper. The derivation of their formula is most of Section 4 of that paper, and we do not state it here. We state only their final calculations:

$$\begin{aligned}
\delta(P_{3;N,R}) &= 0.9990\ldots, \\
\delta(P_{4;N,R}) &= 0.9959\ldots, \\
\delta(P_{5;N,R}) &= 0.9954\ldots, \\
\delta(P_{7;N,R1}) &= 0.9782\ldots, \\
\delta(P_{11;N,R}) &= 0.9167\ldots, \\
\delta(P_{13;N,R}) &= 0.9443\ldots.
\end{aligned}$$

Already we begin to see the phenomenon Theorems 3 and 4 assert exists in the limit, that as q grows large, the race between the nonsquares and the squares becomes more and more even.

VII. Conclusion

Bet on nonsquares. Just make sure there aren't too many of them.

References

[1] R. Durrett. *Probability: Theory and Examples.* Duxbury Press, Third edition, 2005.

[2] A. Feuerverger and G. Martin. Biases in the Shanks-Rényi Prime Number Race. *Experimental Mathematics*, 9 (4):535–570, 2000.

[3] G. Folland. *Real Analysis: Modern Techniques and Their Applications.* John Wiley & Sons, Second edition, 1999.

[4] J. Leech. Note on the distribution of prime numbers. *J. London Math. Soc.*, 32:56–58, 1957.

[5] J.E. Littlewood. Distribution des nombres premiers. *C. R. Acad. Sci. Paris*, 158:1869–1872, 1914.

[6] P. Lévy. Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *C. R. Acad. Sci. Paris*, 175:854–856, 1922.

[7] H. L. Montgomery and R. C. Vaughan. *Multiplicative Number Theory I. Classical Theory.* Cambridge University Press, 2007.

[8] The Online Encyclopedia of Integer Sequences. Sequence A091098. `http://oeis.org/A091098`, June 2008.

[9] The Online Encyclopedia of Integer Sequences. Sequence A091099. `http://oeis.org/A091099`, June 2008.

[10] M. Rubinstein and P. Sarnak. Chebyshev's bias. *Experimental Mathematics*, 3 (3):173–197, 1994.

[11] A. Selberg. An elementary proof of Dirichlet's theorem about primes in an arithmetic progression. *Annals of Mathematics*, 50 (2):297–304, 1949.

[12] J.-P. Serre. *Linear Representations of Finite Groups*. Springer, 1977.

[13] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*. Cambridge University Press, 1995.