WXML Final Report: Prime Spacing

Jayadev Athreya, Barry Forman, Kristin DeVleming, Sara Ahmed, Astrid Berge

Spring 2017

1 Introduction

Throughout the past academic year, our research group has been interested in the distribution of primes as well as their density. Specifically, we have been investigating **prime intervals**. We define a prime interval to be the number I

$$I = p_{n+1} - p_n - 1$$

where p_n is the nth prime. We are especially interested in the case when I is prime, which we denote as a prime-prime interval.

1.1 The initial problem

We wish to approximate the density of prime-prime intervals relative to all other prime intervals. In doing so, we hope to illuminate patterns in the distribution of prime numbers.

1.2 New directions

In describing the density of prime-prime intervals, we have encountered roadblocks in the form of the bias in distribution of consecutive primes. As a result, we would like to move away from the consecutive case and begin to look at how often $r = p - q - 1$ is prime, where p and q are prime.

$$P(x) \sim \frac{C}{D} \int_2^x \frac{dt}{(\log t)^m}, \qquad C = \prod_p \frac{1 - N(p)/p}{(1 - 1/p)^m}$$

Figure 1: Left: Bateman-Horn conjecture, which provides an asymptotic estimate for how often a given set of polynomials $f_i(x)$ is prime. Right: Formula for C. $N(p)$ denotes the number of solutions to $f_i(x)$ mod p where p is prime.

$$\pi_{m_1, m_2, \ldots, m_k}(x) \sim C(m_1, m_2, \ldots, m_k) \int_2^x \frac{dt}{\ln^{k+1} t},$$

$$C(m_1, m_2, \ldots, m_k) = 2^k \prod_q \frac{1 - \frac{w(q; m_1, m_2, \ldots, m_k)}{q}}{\left(1 - \frac{1}{q}\right)^{k+1}},$$

Figure 2: The Hardy-Littlewood prime k-tuple conjecture, which provides an analogous estimate to the Bateman-Horn conjecture for the probability that a prime p and $p + m_1$, $p + m_2$, $\ldots$ are also prime. We focused primarily on the case p and $p + m_1$, where $m_1 = q + 1$ for some prime q.

2 Progress

2.1 Computational

In order to describe the approximate density of prime-prime intervals, we have utilized the Bateman-Horn conjecture (Figure 1), which reduces to the famous Hardy-Littlewood prime k-tuple conjecture (Figure 2) for the linear polynomial case. To compute the density constant C outputted by the Bateman-Horn conjecture, the *prime gap* rather than prime interval must be considered. We define the prime gap G to be the difference between two consecutive primes, that is:

$$G = p_{n+1} - p_n$$

We are most interested in the case where $G = q + 1$ where q is prime, since that corresponds to the associated prime-prime interval. (That is, for a prime-prime interval $R = p_{n+1} - p_n - 1$, if we add 1 to I we will have the

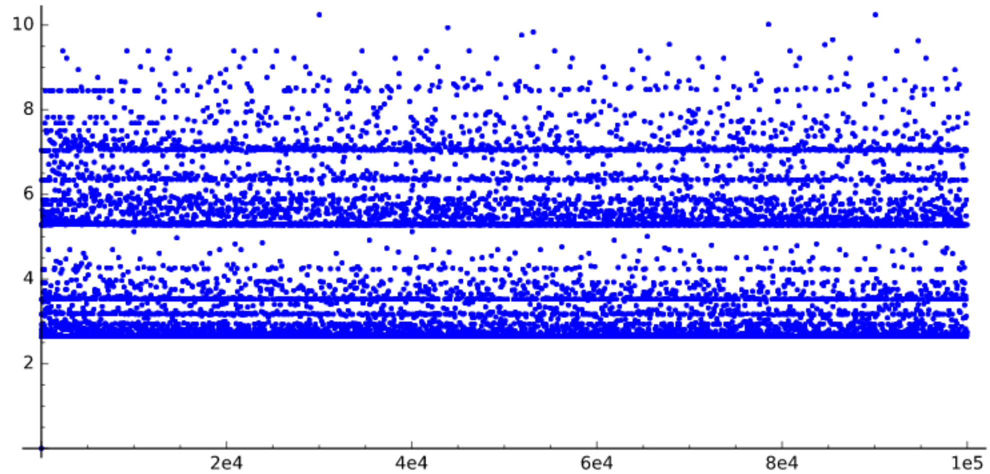

Figure 3: Primes R and their corresponding Hardy-Littlewood constants, with the primes on the horizontal axis. Note clustering at approximately the horizontal line $y \approx 2.6$.

corresponding prime gap $G = R + 1$.)

In further researching prime gaps, we came across previous research on the subject of the frequency of prime gaps. In the paper, entitled Jumping Champions, Odlyzko, et al. present a heuristic argument in support of the idea that primorials, or prime gaps whose prime factorizations consist of distinct consecutive primes (e.g. $30 = 2 * 3 * 5$), appear the most often. In the paper, Odlyzko, et al. argue that 6 is the most common prime gap until approximately $1.7427 * 10^{35}$.

Correspondingly, in order to increase the density of primes produced by a polynomial $f(x) = (x)(x + k)$, the number of solutions to $f(x) \mod p$ for all primes p must be minimized. One way to do this is for k to have a prime factorization consisting of distinct consecutive primes, particularly smaller primes, since those influence the calculation of the conjectured density the most. At each prime p that is an element of the factorization of k, $(x)(x+k)$ will have one solution mod p and two solutions at every other prime. Hence the more distinct consecutive primes in the factorization of k, the smaller $N(p)$ will be. Noting this idea, for primes R up to 10^5, we computed the probability that x and $x + R$ was also prime.

In the set of data that we generated, we found that the highest Hardy-Littlewood constant was ≈ 10.242, associated with the prime-prime interval 30029. We note that $30030 = 2*3*5*7*11*13$. In particular, we note that there is a clustering of Hardy-Littlewood constants around the horizontal line $y \approx 2.6$: that is, a lot of prime-prime intervals have density constants around 2.6. In finding a list of the most common density constants, the number 2.64067318602201 appears 706 times in the list, while the next most common density constant (5.28134637204403) appears just 179 times. This distribution could be due to the fact that we are looking at a small sample up to 10^5 so there may be a bias towards smaller prime factors.

2.2 Theoretical

The Twin Prime Constant

Let P be the set of all primes. The twin prime constant is defined as the following infinite product:

$$\prod_{p\in P}(1-\frac{1}{p})^{-1}(1-\frac{1}{p-1})$$

.

To obtain the $1-\frac{1}{p-1}$ factor, modular arithmetic is used. That is, we use modular arithmetic to determine the probability $p+2$ is divisible by a certain prime given that p is prime.

Consider the following example: Let p be a prime. Then p is 1 or 2 mod 3, so $p+2$ is 3 or 4 mod 3, i.e, $p+2$ is 0 or 1 mod 3. So, the probability that $p+2$ is divisible by 3 is $\frac{1}{2}$. So, the probability that $p+2$ is not divisible by 3 is $1-\frac{1}{2}=\frac{1}{3-1}$
Similarly, p can be 1, 2, 3, or 4 mod 5, so $p+2$ can be 3, 4, 5, 6 mod 5, i.e., $p+2$ can be 0, 1, 3, 4 mod 5. Then, the probability that $p+2$ is divisible by 5 is $\frac{1}{4}$, which means the probability that $p+2$ is not divisible by 5 is $1-\frac{1}{4}=1-\frac{1}{p-1}$

In general, if p and q are prime, and $p \neq q$ then p is 1, 2, 3, ..., $q-p$, ... $q-2$, or $q-1$ mod q. So, $p+2$ is 3, 4, 5, ..., q, ..., $(q-2+p)$, or $(q-1-p)$ mod q. That is, the probability that $p+2$ is divisible by q is $\frac{1}{1-q}$, i.e., the probability that p is not divisible by q is $\frac{1}{q-1}$.

Generalizing to Other Prime Gaps

We can use the same logic for p and $p + (r + 1)$ where p and r are prime.

First consider the following interesting example: Let p be a prime. Then p is 1 or 2 mod 3, so $p + 6$ is either 7 or 8 mod 3, i.e., $p + 6$ is either 1 or 2 mod 3. That is, $p + 6$ has a 0% chance of being divisible by 6, so the probability that $p + 6$ is not divisible by 6 is 1 - 0 = 1.

In general, if $r+1$ has q as a prime factor, then the probability that $p+(r+1)$ is not divisible by q is 1: Let p, q, and r be prime and suppose $p \neq q$. Suppose also that $r + 1$ is of the form $q \cdot a$ for some positive integer a. Then we have that p is 1, 2, 3, ..., $q - p$, ... $q - 2$, or $q - 1$ mod q, so $p + (r + 1) = p + (q \cdot a)$ is $1 + (q \cdot a)$, $2 + (q \cdot a)$, $3 + (q \cdot a)$, ..., $q - p + (q \cdot a)$, ... $q - 2 + (q \cdot a)$, or $q - 1 + (q \cdot a)$ mod q. None of which are multiples of q.

So, the probability that p and $p + (r + 1)$ are both prime is dependent on the prime factorization of $r + 1$

Estimating the Probability that a Prime Interval is Prime

We know that $\pi(x) \approx \frac{x}{log(x)}$, so a simple approximation for the probability that $p + (r + 1)$ is prime if p is prime could be $\frac{x}{log^2(x)}$. Then since p is prime, p is odd unless it is 2, so $p + 2$ is also odd, so we can double this probability. For the specific case of $r + 1 = 2$, we use the twin prime constant to obtain a better approximation of the probability that p and $p + 2$ are prime. So, if r is a prime, then we can approximate the number of prime prime intervals with the following equation:

$$\frac{x}{log^2(x)} \cdot \prod_{p \in P} (1 - \frac{1}{p})^{-1} \cdot \sum_{p \in P} C_{p+1}$$

where C_{p+1} is the product over all primes of the probability that if q is prime $q + p + 1$ is not divisible by that prime.

3 Future directions

We will compare the theoretical conjectured numbers with some of the data we have gotten to see if we are on the correct track with our ideas.
In short, if we think of the twin prime constant as the base (because every prime gap will have 2 as a factor because they are all even), we can obtain the analogous constant for any gap by simply looking at its prime factorization. That is, if a is a prime gap where a has non-even prime factors p, q, and r and we let C_2 , $C_a = C_2 * \frac{1}{1-\frac{1}{p-1}} * \frac{1}{1-\frac{1}{q-1}} * \frac{1}{1-\frac{1}{r-1}}$. So, by having a high number of consecutive prime factors in the prime gap being consider, that maximizes the number of times that gap will show up.