The Local Behavior of Germain Primes

Brad Weir

February 4, 2003

Abstract

A Germain prime is a prime p such that $2p+1$ is also a prime, in which case we call $2p+1$ a co-Germain prime. Germain primes are the subject of modern day interest, because given a co-Germain prime q, $q-1$ has a very large prime factor, namely the Germain prime corresponding to q. The primality testing algorithm of Agrawal, Kayal, and Saxena uses such numbers in a deterministic primality test. If the number of Germain primes up to a number N follows an asymptotic formula conjectured by Hardy and Littlewood, then the primality testing algorithm of Agrawal, Kayal, and Saxena runs in $O(\log^6 n)$ time, where n is the number being tested. We ask two questions about the behavior of Germain primes. First, do Germain primes follow the conjecture of Hardy-Littlewood in local intervals? Second, how are Germain primes distributed in these local intervals?

Contents

List of Figures

1 The local Hardy-Littlewood conjecture

The number of primes up to N is commonly denoted $\pi(N)$, similarly for Germain primes, $\pi_G(N)$ is the number of Germain primes up to N. Hardy and Littlewood [1] conjectured that $\pi_G(N)$ is asymptotic to $N/\log^2 N$. Since the number of primes up to N is asymptotic to $N/\log N$ by the prime number theorem, it is reasonable to suppose the Hardy-Littlewood conjecture.

In this paper we will consider the Hardy-Littlewood conjecture on local intervals. By $\pi_G(N, N+l)$ we denote the number of Germain primes in the interval $[N, N+l]$, and we conjecture that this is asymptotic to the difference of the asymptotic formula at $N+l$ with the asymptotic formula at N.

Conjecture 1 (Local Hardy-Littlewood).

$$\pi_G(N, N+l) \sim 2T_2 \left(\frac{N+l}{\log^2(N+l)} - \frac{N}{\log^2 N} \right),$$

where $T_2 = 0.6601618158 \cdots$ *is the twin primes constant.*

We will derive the Hardy-Littlewood conjecture using a variant of the method used by Hardy-Littlewood that is due primarily to Vinogradov. Our derivation will follow the course of Miller [3] and a number of time refer the reader their for complete proofs. In order to develop an asymptotic formula for π_G we construct a generating function for π_G. We will evaluate its integral on an interval that we decompose into two parts, called the major arcs and the minor arcs. We will not be able to evaluate the integral on the minor arcs, but we will conjecture that the integral over the minor arcs contributes only lower order terms. Finally, we will show that this conjecture implies the Hardy-Littlewood conjecture.

1.1 The generating function

We denote the generating function $f_N(x)$ for π_G to be

$$f_N(x) = \sum_{p_1 \leq N} \sum_{p_2 \leq N} e((-2p_1 + p_2)x) \log p_1 \cdot \log p_2, \tag{1}$$

where $e(x) = e^{2\pi i x}$.

We will find that the value of the integral of this function over the interval $[-1/2, 1/2]$ is equal to $\pi_G(N)$ up to lower order terms. We will evaluate this integral to get an asymptotic formula for π_G.

Let $\mathfrak{U} = [-1/2, 1/2]$,

$$\begin{aligned} \int_{\mathfrak{U}} f_N(x) e(-x) dx &= \sum_{p_1 \leq N} \sum_{p_2 \leq N} \log p_1 \cdot \log p_2 \int_{\mathfrak{U}} e((-2p_1 + p_2 - 1)x) dx, \\ &= \sum_{p_1 \leq N} \sum_{p_2 = 2p_1 + 1} \log p_1 \cdot \log p_2, \\ &= \sum_{\substack{p_1 \leq N \\ 2p_1 + 1 \text{ prime}}} \log p_1 \cdot \log(2p_1 + 1), \end{aligned}$$

where we've used the fact that the integral of $e(ax)$ is 0 when $a \neq 0$ and 1 when $a = 0$.

Now let

$$G(N) = \sum_{\substack{p \leq N \\ 2p+1 \text{ prime}}} \log p \cdot \log(2p + 1).$$

Obviously,

$$G(N) \leq \log^2 N \cdot \pi_G(N)$$

For the bound in the other direction, we can use partial summation to get the formula

$$G(N) \geq \log^2 N \cdot \pi_G(N) - O\left(N \frac{\log \log N}{\log N}\right)$$

Thus, up to lower order terms,

$$\log^2 N \cdot \pi_G(N) \approx \int_{\mathfrak{U}} f_N(x)e(-x)\,dx.$$

1.2 The major and minor arcs

We can find an asymptotic formula for this integral by breaking up $\mathfrak{U}$ into numbers that are well approximated by rationals, and those numbers that are not. These two sections are called, respectively, the major and minor arcs.

Let $Q = \log^B N$, and $1 \leq a \leq q \leq Q$ where $(a, q) = 1$, then the major arc of a/q is,

$$\mathfrak{M}(q, a) = \{\alpha \in \mathfrak{U} : |\alpha - a/q| < Q/N\},$$

where B can be chosen large enough to make certain terms lower order.

So the elements of the major arcs are the real numbers that are within Q/N of a rational that has a denominator no greater than Q. The major arcs are pairwise disjoint, and we define the major arcs $\mathfrak{M}$ as the union of every major arc.

The minor arcs are the intervals in $\mathfrak{U}$, not covered by the major arcs, so

$$\mathfrak{m} = \mathfrak{U} \backslash \mathfrak{M}.$$

From the definition of the major and minor arcs it follows that

$$\int_{\mathfrak{U}} f_N(x)e(-x)\,dx = \int_{\mathfrak{M}} f_N(x)e(-x)\,dx \; + \int_{\mathfrak{m}} f_N(x)e(-x)\,dx \qquad (2)$$

We will show that the major arcs contribute a constant times N^2 plus lower order terms. We will also see by evaluation that $f_N(x)$ is large on the minor arcs. Although the major arcs take up most of $\mathfrak{U}$ it is our hope that there is some type of cancellation on the minor arcs so that $f_N(x)$ is small on these portions of $\mathfrak{U}$. Thus, all we would need to show is that the integral over the minor arcs $f_N(x)$ is at most $o(N^2)$, but this is in fact the most difficult part of the circle method. The following conjecture is equivalent to the local Hardy-Littlewood conjecture

Conjecture 2. *The integral over the major arcs contributes lower order terms, so*

$$\int_{\mathfrak{U}} f_N(x)e(-x)\,dx \sim \int_{\mathfrak{M}} f_N(x)e(-x)\,dx$$

1.3 The generating function at a rational

We now consider the value of f_N at a rational a/q. Splitting the sums up from above equation we get

$$f_N(a/q) = \sum_{p_1 \leq N} e(-2p_1 a/q) \log p_1 \sum_{p_2 \leq N} e(p_2 a/q) \log p_2$$

Dividing the indices of the sum into the congruence classes of q, we get

$$f_N(a/q) = \sum_{r_1=1}^{q} \sum_{\substack{p_1 \leq N \\ p_1 \equiv r_1(q)}} e(-2p_1 a/q) \log p_1 \sum_{r_2=1}^{q} \sum_{\substack{p_2 \leq N \\ p_2 \equiv r_2(q)}} e(p_2 a/q) \log p_2$$

Since $e(k) = 1$ for any integer k, we can replace p_1 with r_1 and p_2 with r_2 and remove them from the inner sums, so

$$f_N(a/q) = \sum_{r_1=1}^{q} e(-2r_1 a/q) \sum_{\substack{p_1 \leq N \\ p_1 \equiv r_1(q)}} \log p_1 \sum_{r_2=1}^{q} e(r_2 a/q) \sum_{\substack{p_2 \leq N \\ p_2 \equiv r_2(q)}} \log p_2 \tag{3}$$

Now $p_1 \equiv r_1 \pmod{q}$ so $p_1 = kq + r_1$, for $0 \leq k < r_1$. If $(r_1, q) = d > 1$, then $p_1 = d(kq' + r_1')$ where $q = dq'$ and $r_1 = dr_1'$. So there is at most one prime p_1 such that $(r_1, q) > 1$, and this prime occurs only when r_1 itself is prime. Thus, the number of values for r_1 such that $(r_1, q) > 1$ is very small. The same is true for r_2.

$$f_N(a/q) = \sum_{\substack{r_1=1 \\ (r_1,q)=1}}^{q} e(r_1(-2a)/q) \sum_{\substack{p_1 \leq N \\ p_1 \equiv r_1(q)}} \log p_1 \sum_{\substack{r_2=1 \\ (r_2,q)=1}}^{q} e(r_2 a/q) \sum_{\substack{p_2 \leq N \\ p_2 \equiv r_2(q)}} \log p_2 \tag{4}$$

Theorem 3 (Siegel-Walfisz). *For $B, C > 0$, $(a, q) = 1$, and $q \leq \log^B N$,*

$$\sum_{\substack{p \leq N \\ p \equiv a(q)}} \log p = \frac{N}{\phi(q)} + O\left(\frac{N}{\log^C N}\right),$$

where the constant depends only on B and C.

Notice that Siegel-Walfisz allows us to make sure that the error is very small, because given any $B > 0$ we can find a $C > 0$ such that the theorem holds. Notice also that although we may make the denominator of the error term very large, since the constant depends on B and C, the constant may become large also, but as N approaches infinity this won't matter. We use this to keep this correction term as a lower order contribution.

A Ramanujan Sum, denoted $c_q(a)$ is

$$c_q(a) = \sum_{\substack{r=1 \\ (r,q)=1}}^{q} e(ra/q)$$

Using the Siegel-Walfisz theorem and the definition of a Ramanujan sum, Eq. (4) becomes

$$f_N(a/q) \approx \frac{N^2}{\phi^2(q)} c_q(-2a) c_q(a),$$

where the approximation holds up to lower order terms.

1.4 The generating function on the major arcs

Now that we have a formula with very good control on the error for the generating function at a rational with a small denominator, we want to know what the value of the generating function is over that rational's major arc, again with a good control on the error.

We now rewrite Eq. (4) as

$$f_N(a/q) = C_q(a)u(0),$$

where

$$\begin{aligned} u(x) &= \sum_{p_1 \le N} \sum_{p_2 \le N} e((-2p_1 + p_2)x), \qquad (5) \\ C_q(a) &= \frac{c_q(-2a)c_q(a)}{\phi^2(q)} \end{aligned}$$

So $C_q(a)u(0)$ agrees perfectly with $f_N(a/q)$ and it is our hope that if α is near a/q then $C_q(a)u(\alpha - a/q)$ is near $f_N(\alpha)$.

Let

$$S_{a,q} = f_N(\alpha) - C_q(a)u(\alpha - a/q).$$

Since both $f_N(\alpha)$ and $u(\alpha - a/q)$ are sums over p_1, p_2 by Eqs. (1) and (5) respectively, we can incorporate their difference into one single sum over

p_1, p_2. Using partial summation, we can separate this sum into an integral and another sum. Evaluating the contribution over this integral and sum we get the next lemma.

Lemma 4. *If* $\alpha \in \mathfrak{M}(q,a)$ *then*

$$f_N(\alpha) = C_q(a)u(\alpha - a/q) + O\left(\frac{N^2}{\log^{C-2B} N}\right).$$

Now by Siegel-Walfisz we can chose C such that $C > 2B$ so that this error term is lower order. A complete proof of this lemma is given by Miller [3].

1.5 The integral over the major arcs

We must now evaluate the integral of $C_q(a)u(\alpha - a/q)$ over the major arcs. Since $C_q(a)$ depends only on a and q we can remove it from the integral and worry about it later. We may break up the integral over the major arcs into several parts and evaluate each part.

Lemma 5.

$$\int_{\mathfrak{M}(q,a)} u(\alpha - a/q)e(-\alpha)\,d\alpha \approx \frac{N}{2}e(-a/q)$$

Again a complete proof of this lemma is given by Miller [3].

Now we compute the value of the integral of the generating function over the major arcs up to lower order terms.

$$\begin{aligned}
\int_{\mathfrak{M}} f_N(x)e(-x)\,dx &\approx 2\sum_{q\le Q}\sum_{\substack{a=1\\(a,q)=1}}^{q}\int_{a/q-Q/N}^{a/q+Q/N} f_N(\alpha)e(-\alpha)\,d\alpha \\
&= 2\sum_{q\le Q}\sum_{\substack{a=1\\(a,q)=1}}^{q}\int_{a/q-Q/N}^{a/q+Q/N} C_q(a)u(\alpha - a/q)e(-\alpha)\,d\alpha \\
&\approx N\sum_{q\le Q}\sum_{\substack{a=1\\(a,q)=1}}^{q} C_q(a)e(-a/q)
\end{aligned}$$

Let

$$\rho_q = \sum_{\substack{a=1\\(a,q)=1}}^{q} C_q(a)e(-a/q).$$

We call the sum of the ρ_q the singular series of π_G,

$$\mathfrak{S}_N = \sum_{q=1}^{Q} \rho_q.$$

Substituting in the above definitions it follows immediately that

$$\int_{\mathfrak{M}} f_N(x)e(-x)\,dx \approx N\mathfrak{S}_N$$

1.6 The singular series

We will now show that ρ_q is a multiplicative sequence, and for any multiplicative sequence

$$\sum_q \rho_q = \prod_p \left(1 + \sum_{k=1}^{\infty} \rho_{p^k}\right) \tag{6}$$

because each q in the left hand sum can be decomposed into its prime factors. We can then eliminate certain elements from the sum over the powers of p on the basis of the properties of ρ_{p^k}. What we are left with is a product over primes that converges to the twin prime constant T_2. Thus, we will come to the identity

$$\mathfrak{S}_N = 2T_2.$$

In order to prove that ρ_q is multiplicative we first need two corollaries.

Corollary 6. *If $r \leq q$ and $(r, q) = 1$, then*

$$r \equiv r_1 q_2 + r_2 q_1 \pmod{q_1 q_2},$$

where $1 \leq r_1 \leq q_1, (r_1, q_1) = 1$, and $1 \leq r_2 \leq q_2, (r_2, q_2) = 1$.

Corollary 7. *If $(q_1, q_2) = 1$, then*

$$C_{q_1}(a) = C_{q_1}(a_1 q_2 + a_2 q_1) = C_{q_1}(a_1 q_2),$$

where $1 \leq a_1 \leq q_1, (a_1, q_1) = 1$, and $1 \leq a_2 \leq q_2, (a_2, q_2) = 1$.

Lemma 8. *ρ_q is multiplicative*

Proof. In order to show that ρ_q is multiplicative, we first show that a Ramanujan sum is multiplicative with respect to q. Throughout this proof, q_1 and q_2 are assumed to be relatively prime.

Using Cor. 6 we can separate $c_{q_1q_2}(a)$ into two different sums, which are equal to $c_{q_1}(a)$ and $c_{q_2}(a)$.

$$\begin{aligned}
c_{q_1q_2}(a) &= \sum_{\substack{r=1 \\ (r,q_1q_2)=1}}^{q_1q_2} e(ra/q_1q_2), \\
&= \sum_{\substack{r_1=1 \\ (r_1,q_1)=1}}^{q_1} \sum_{\substack{r_2=1 \\ (r_2,q_2)=1}}^{q_2} e((r_1q_2 + r_2q_1)a/q_1q_2), \\
&= \sum_{\substack{r_1=1 \\ (r_1,q)=1}}^{q} e(r_1a/q) \sum_{\substack{r_2=1 \\ (r_2,q)=1}}^{q} e(r_2a/q), \\
&= c_{q_1}(a)c_{q_2}(a).
\end{aligned}$$

Since

$$C_{q_1q_2}(a) = \frac{c_{q_1q_2}(a)c_{q_1q_2}(-2a)}{\phi^2(q_1q_2)},$$

and ϕ is obviously multiplicative, $C_q(a)$ is also multiplicative.

The proof follows easily using the fact that $C_q(a)$ is multiplicative, and Cor. 7 repeated times.

$$\begin{aligned}
\rho_{q_1q_2} &= \sum_{\substack{a=1 \\ (a,q_1q_2)=1}}^{q_1q_2} C_{q_1q_2}(a)e(-a/q_1q_2), \\
&= \sum_{\substack{a=1 \\ (a,q_1q_2)=1}}^{q_1q_2} C_{q_1}(a)C_{q_2}(a)e(-a/q_1q_2), \\
&= \sum_{\substack{a=1 \\ (a,q_1)=1}}^{q_1} \sum_{\substack{a=1 \\ (a,q_2)=1}}^{q_2} C_{q_1}(a_1q_2 + a_2q_1)C_{q_2}(a_1q_2 + a_2q_1)e(-(a_1q_2 + a_2q_1)/q_1q_2), \\
&= \sum_{\substack{a=1 \\ (a,q_1)=1}}^{q_1} C_{q_1}(a_1q_2)e(-a_1/q_1) \sum_{\substack{a=1 \\ (a,q_2)=1}}^{q_2} C_{q_2}(a_2q_1)e(-a_2/q_2), \\
&= \rho_{q_1}\rho_{q_2}.
\end{aligned}$$

□

Since ρ_q is multiplicative, Eq. 6 holds and we need only to simplify this expression to get a product which converges to the twin primes constant.

Lemma 9.

$$\sum_{k=1}^{\infty} \rho_{p^k} = \rho_p$$

Proof. The lemma is equivalent to saying $\rho_{p^k} = 0$, for all $k \geq 2$. It follows because $c_q(a) = \mu(q)$ for $(a, q) = 1$, where μ is the moebius function. This can be shown using elementary properties of the moebius function. Since $\mu(p^k) = 0$ for all $k \geq 2$, the lemma follows. □

$$\begin{aligned}
\rho_2 &= \sum_{\substack{a=1 \\ (a,2)=1}}^{2} C_2(a)e(-a/2) \\
&= \frac{c_2(1)c_2(-2)}{\phi^2(2)}e(-1/2) \\
&= \frac{e(1/2)e(-1)}{1}e(-1/2) \\
&= 1.
\end{aligned}$$

Now if $p > 2$, then $c_p(a) = c_p(-2a) = \mu(p)$ since $(a, p) = 1$. For any prime p, $\mu^2(p) = 1$, and $\phi(p) = p - 1$, so

$$\begin{aligned}
\rho_p &= \sum_{a=1}^{p-1} \frac{1}{(p-1)^2} e(-a/p), \\
&= \frac{1}{(p-1)^2}\left[-e(-0/p) + \sum_{a=0}^{p-1} e(-a/p)\right], \\
&= \frac{-1}{(p-1)^2}.
\end{aligned}$$

Lemma 10.

$$\mathfrak{S}_N = 2T_2$$

Proof. As N goes to infinity, so does $Q = \log^B N$, and thus

$$|\mathfrak{S}_N - \sum_q \rho_q|$$

goes to zero as N goes to infinity. Thus,

$$\begin{aligned}
\mathfrak{S}_N &= \prod_p (1 + \rho_p), \\
&= (1 + \rho_2) \prod_{p>2} (1 + \rho_p), \\
&= 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right), \\
&= 4T_2.
\end{aligned}$$

□

Thus, assuming Conjecture 2 we have reformulated the local Hardy-Littlewood conjecture,

$$\pi_G(N, N+l) \sim 2T_2 \left(\frac{N+l}{\log^2(N+l)} - \frac{N}{\log^2 N}\right),$$

where $T_2 = 0.6601618158 \cdots$ is the twin primes constant.

1.7 Numerical observations

In order to observe the local Hardy-Littlewood conjecture numerically, we must first chose an appropriate intervals to examine. Since the local Hardy-Littlewood conjecture relies on an asymptotic formula, the optimal place to examine this conjecture is as far out on the number line as is computationally possible. Computationally, there isn't much difficulty going out to extremely high numbers for the starting point of an interval, like say 10^{50}. In the end, the biggest computational limitation is the interval size. No matter where we start, whether it be 10 or 10^{50} we must look at as many numbers as our interval size is. All of the computations presented in this paper were performed on a Sun Ultra-80 workstation. Testing for Germain primes was done in C using the GNU Multiple Precision package. To test all the Germain primes in an interval of length 10^8 took approximately 5 hours, and an interval of length 10^9 took over a third of a day. Thus at this rate it is possible to test intervals of length 10^{10}, although it would take a great amount of time. Any larger interval size however is quite insurmountable with the computing considerations of this paper.

As mentioned earlier, we now have quite a bit of freedom concerning where we look at these intervals of size 10^8 and 10^9. Since there is an error of approximately $\sqrt{N}$ in the prime number theorem, it seems reasonable to look at intervals starting at no greater than 10^{20}. If there were the same error in our derivation, this may lead to completely unpredictable results with our chosen interval size. However, considerations from the following section will force us to chose intervals which begin after 10^{20}.

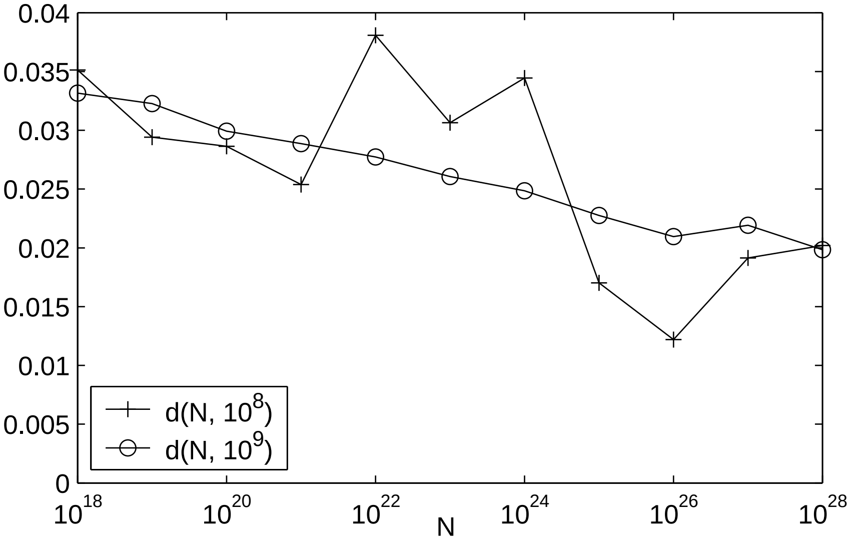

Figure 1: Convergence of $c(N, N + l)$ to $\pi_G(N, N + l)$

The first property we observed numerically was the convergence of the asymptotic formula to the actual number of Germain primes. Let

$$C(N, N + l) = \frac{N + l}{\log^2(N + l)} - \frac{N}{\log^2 N},$$

then the distance between $C(N, N + l)$ and $\pi_G(N, N + l)$, denoted $d(N, l)$ is

$$d(N, l) = \log \left| \frac{\pi_G(N, N + l)}{2T_2 C(N, N + l)} - 1 \right|.$$

Fig. 1 is a measure of $d(N, l)$ for N from 10^{18} to 10^{28} and l of 10^8 and 10^9.

Although we hoped that the plot for 10^8 would have somewhat less jitter, these numerics support our conjecture that $C(N, N+l)$ is an asymptotic formula for $\pi_G(N, N+l)$. If we were to increase the interval size to 10^{10}, one would expect that convergence plot would have even less jitter than the 10^9 plot, and the jitter would slowly decrease as interval size increases until it was undetectable.

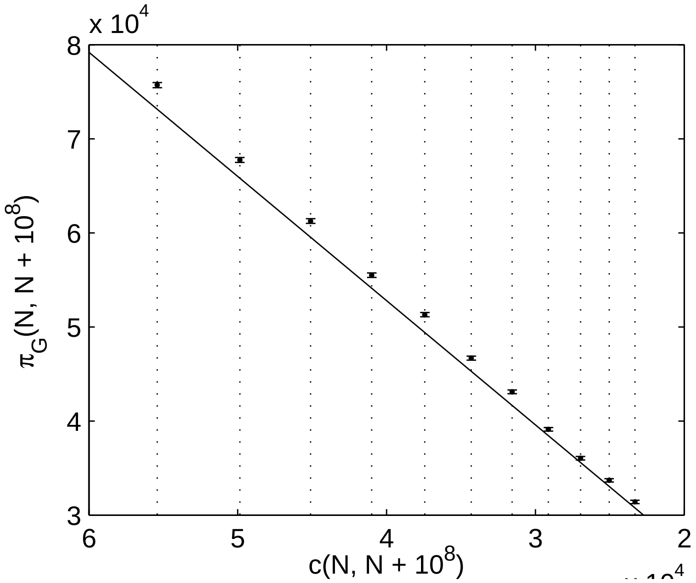

Figure 2: Measured $\pi_G(N, N+10^8)$ against asymptotic formula

Figs. 2 and 3 are plots of $\pi_G(N, N+l)$ versus $C(N, N+l)$. If our conjecture is correct the data should follow a straight line of slope $2T_2$. Notice that the bottom axis has been reversed, since as N increases with l constant, the value of $C(N, N+l)$ decreases. We wanted to show how the data behaved as N increased, so the reversed plot is more fitting. Each dotted line represents a value of N starting from 10^{18} and ending at 10^{28}. These figures reaffirm what we saw in Fig. 1.

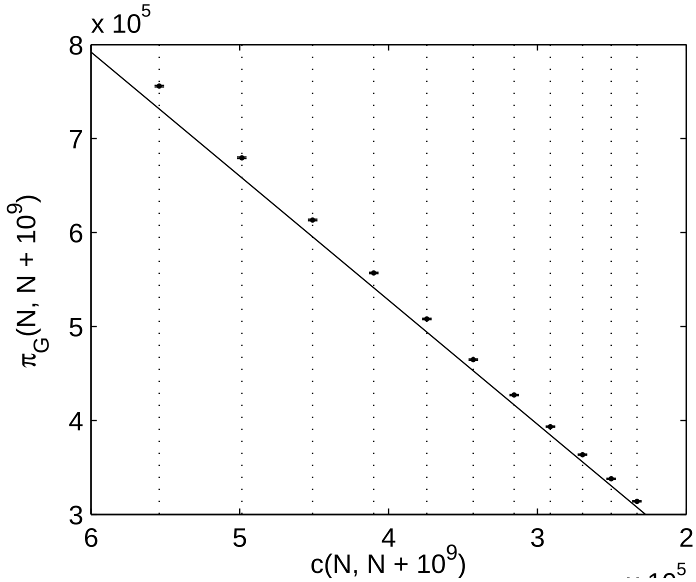

Figure 3: Measured $\pi_G(N, N + 10^9)$ against asymptotic formula

2 Distances between Germain primes

Now that we have a general understanding of the average spacing, of Germain primes, we examine how these spacings are distributed. Notice that these are two mutually exclusive questions. It could be that the spacings between every two Germain primes are exactly the average spacing. We conjecture that the Germain primes are distributed not at exactly the average spacing, but like the prime numbers. Thus, the space between two Germain primes should not affect the space between their adjacent Germain primes.

2.1 Poisson Distributions

Elements distributed such that spacing between two elements does not effect their adjacent spacings are said to follow a Poisson distribution.

Conjecture 11. *The Germain primes are a Poisson distribution.*

Assuming this Conjecture, we can describe how Germain primes should be distributed in an interval.

Consider the interval $[N, N + l]$, and let ϵ denote the probability that an integer x in the interval is a Germain prime. The probability of having exactly r Germain primes in the interval $[N, N+l]$ is. So as our x we can take the distance between two Germain primes, and our r is just $k-1$, or how many Germain primes we expect to find in between.

$$\begin{aligned}
P_k &= \binom{l}{k-1}\epsilon^{k-1}(1-\epsilon)^{l-k-1} \\
&= \frac{l(l-1)\cdots(l-(k-2))}{(k-1)!}\epsilon^{k-1}(1-\epsilon)^{l-k-1} \\
&\approx \frac{l^{k-1}}{(k-1)!}\epsilon^{k-1}(1-\epsilon)^{l-k-1} \\
&\approx \frac{l^{k-1}}{(k-1)!}\epsilon^{k-1}(1-\epsilon)^{l} \\
&= \frac{x^{k-1}}{(k-1)!}(1-\frac{x}{l})^{l}
\end{aligned}$$

Now for $l \gg r$ we can make the following two approximations Using these approximations and noticing $\epsilon = x/l$. Taking the limit as l goes to infinity gives us the probability distribution of the k-th spacings of Germain primes is

$$P_k(x) = \frac{x^{k-1}}{(k-1)!}e^{-x}\,dx.$$

Notice that for any positive integer k this is a probability distribution, because we can repeatedly reduce the power of x in the integral of this probability using integration by parts until we are left with integral whose value is 1.

$$\begin{aligned}
\int_0^\infty P_k(x) &= \int_0^\infty \frac{x^{k-1}}{(k-1)!}e^{-x}\,dx \\
&= \left.\frac{-x^{k-1}}{(k-1)!}e^{-x}\right|_0^\infty + \int_0^\infty \frac{x^{k-2}}{(k-2)!}e^{-x}\,dx \\
&= \int_0^\infty \frac{x^{k-2}}{(k-2)!}e^{-x}\,dx \\
&\vdots \\
&= \int_0^\infty e^{-x}\,dx \\
&= 1
\end{aligned}$$

$$\begin{aligned}\sum_{k=1}^{\infty} P_k(x) &= dxe^{-x}\sum_{k=0}^{\infty}\frac{x^k}{k!} \\ &= dxe^{-x}e^{x} \\ &= dx\end{aligned}$$

So for any number x, the sum of the probabilities for each spacing k is 1.

2.2 Numerical observations

In order to observe the distribution of the Germain primes with as best precision as possible, it is necessary to consider intervals $[N, N+l]$, where $N \gg l$. If the Hardy-Littlewood conjecture is correct, then for intervals where $N+l$ is significantly larger than N, the density of the Germain primes at N is considerably larger than $N+l$. Since we are concerned with a distribution whose density is essentially constant, we cannot have an interval length l that is comparable to N. Thus, we are forced to deal with the contradictory demand of having interval lengths no smaller than $\sqrt{N}$, yet much less than N. We thus chose to examine the intervals 10^{18} to 10^{28}, because it was the closest set of intervals meeting this demand.

We can see in Figures 4–6 that the distribution of Germain primes is very close to the distribution of a Poissonian process. The actual distribution was plotted using a histogram of values. Since the area under the curve of the actual distribution is equal to the bin size used in creating the histogram multiplied by the number of Germain primes in the interval, and the Poissonian curves have unit area, we multiply the Poissonian curves in each Figure by $C = 755774/12$ which is the product of the bin size and number of Germain primes, respectively.

Finally, the error bars in Figures 2 and 3 represent the fact that if the Germain primes were placed on the number line as a Poissonian process, then this amount of error would be present if they were plotted like the values in Figures 2 and 3.

References

[1] G. H. Hardy. *Collected Papers of G. H. Hardy; Including Joint Papers with J. E. Littlewood and Others.* Oxford University Press, 1966.

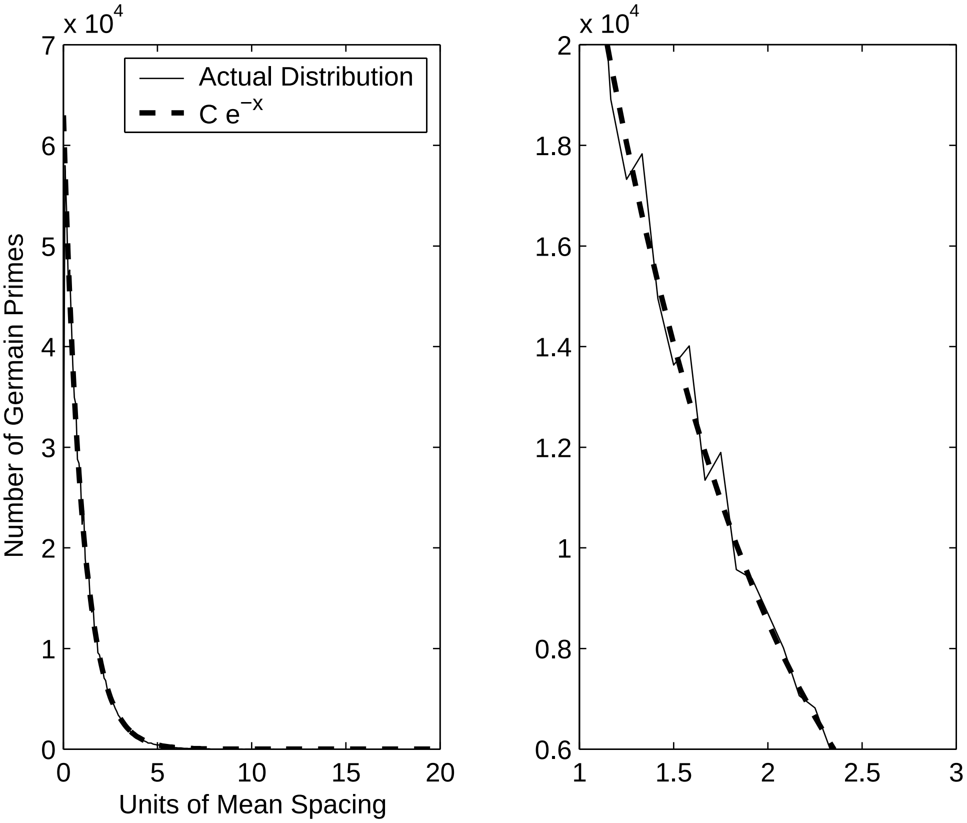

Figure 4: Distribution of the 11st spacings of Germain primes

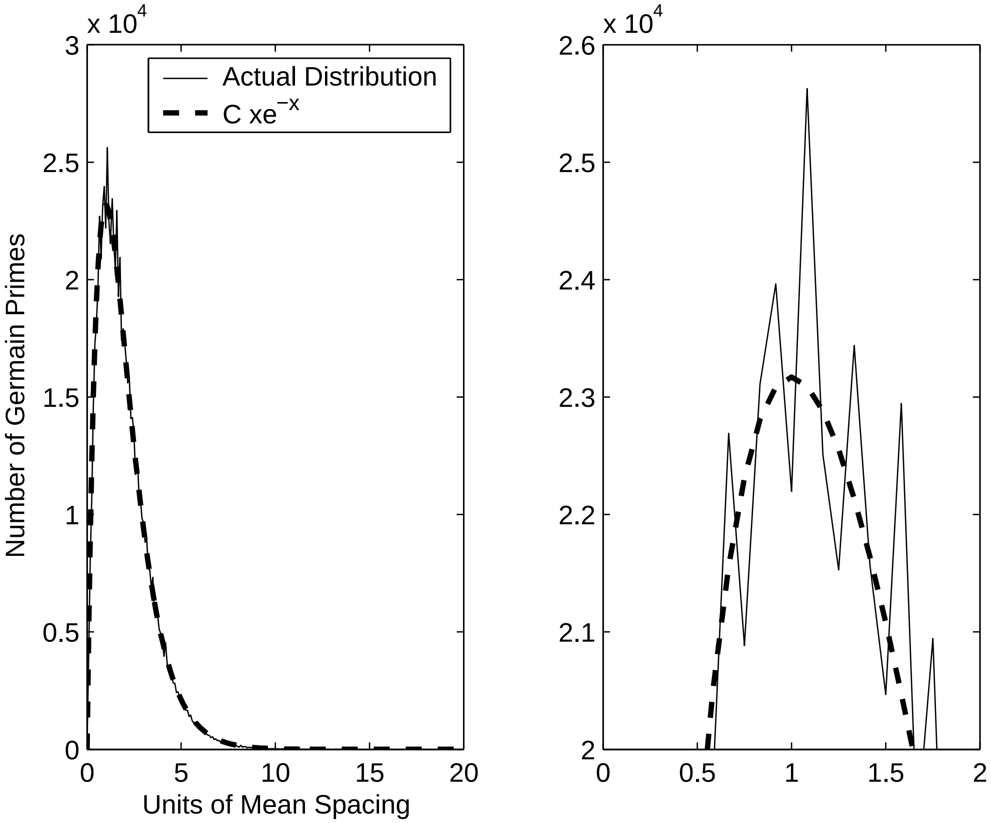

Figure 5: Distribution of the 2nd spacings of Germain primes

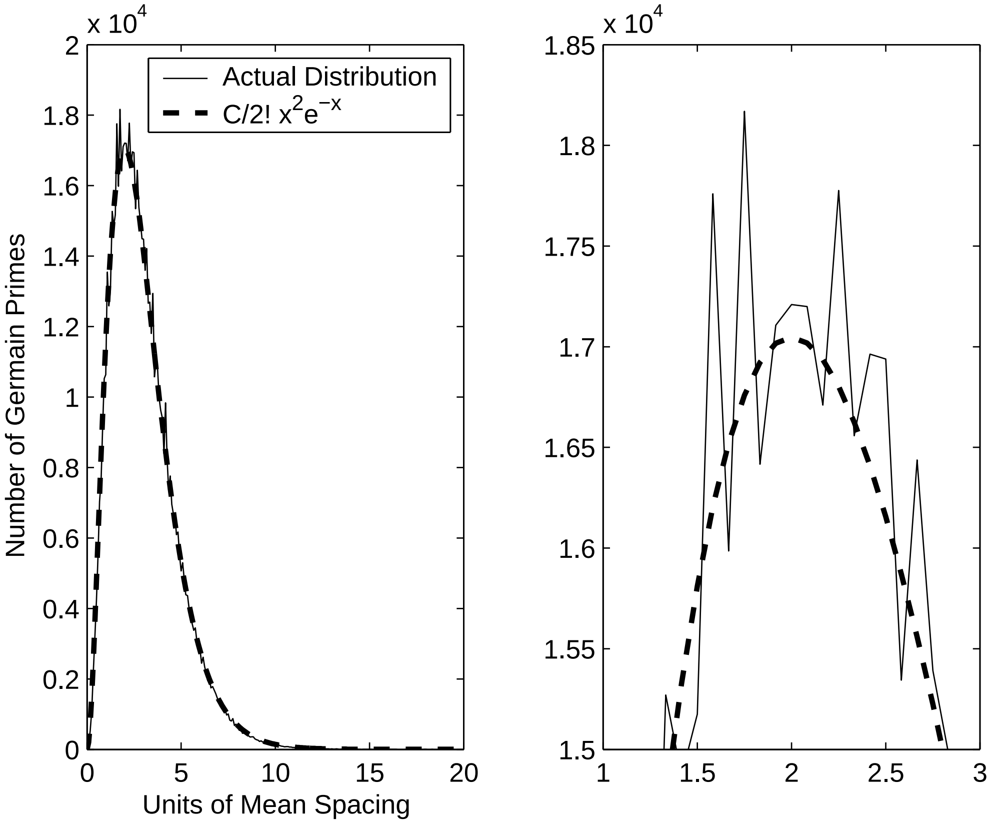

Figure 6: Distribution of the 3rd spacings of Germain primes

[2] Nitin Saxena Manindra Agrawal, Neeraj Kayal. Primes in p, 2002. http://www.cse.iitk.ac.in/users/manindra/primality.ps.

[3] Steven J. Miller. The circle method and germain primes, 2002. http://math.nyu.edu/Courses/V63.0393/classnotes/nyuGermainSJM.ps.

[4] Steven J. Miller. Notes on constructing transcendentals and poissonian behavior, 2002. http://math.nyu.edu/Curses/V63.0393/classnotes/notess.ps.

[5] Melvyn B. Nathanson. *Additive Number Theory : the Classical Bases.* Springer, 1996.

[6] R. C. Vaughn. *The Hardy-Littlewood Method.* Cambridge University Press, 1997.