Erdos and the Twin Prime Conjecture: Elementary Approaches to Characterizing the Differences of Primes

Jerry Li

June 2, 2010

Contents

1 Introduction

This paper will contain a fuller proof of two results derived by Erdos over the span of about 8 years. The first, given in The Difference of Consecutive Primes, Duke Math J., Vol 2, Number 6(1940), 438 – 441, was the first proof that

$$\liminf \frac{p_{n+1} - p_n}{\log_n} < A, A < \infty \tag{1}$$

where p_n is the nth prime, and in particular found that $A \leq 1 - c$, for some $c > 0$. More modern methods (indeed, published in 2009!) finally showed that $A = 0$, as long suspected.

The other efforts approaches the question from a different point of view. Rather than attempting to bound the differences of infinitely many primes, Erdos and Turan ask instead how the primes behave under different mean values, and in particular, whether there are infinitely many solutions to the inequalities

$$(\frac{p_{n-1}^t + p_{n+1}^t}{2})^{1/t} > p_n, (\frac{p_{n-1}^t + p_{n+1}^t}{2})^{1/t} < p_n, \tag{2}$$

for any t. The cases of $t = 0$ (using a limiting argument) and $t = 1$ reduce down to the question of whether there are infinitely many primes that satisfy

$$p_{n-1}p_{n+1} > p_n^2, p_{n+1}p_{n-1} < p_m^2 \tag{3}$$

and

$$\frac{p_{n-1} + p_{n+1}}{2} > p_n, \frac{p_{n-1} + p_{n+1}}{2} < p_n \tag{4}$$

respectively, which are of course just the familiar geometric and arithmetic means.

In §2 we will present a more detailed proof of (1) based on Erdos's proof in [1], with more details being filled in as necessary, and in §3 we will present two proofs that (2) has infinitely many solutions, one based on Erdos's proof in [4], the second on [2]. Later in [4] Erdos proves some generalizations of the above statement dealing with any sequence of integers bounded in their growth rates, and then some non-elementary proofs that (2) has infinitely many solutions, but we will not discuss these in this current paper. We will note here that the latter approach for showing that (2) has infinitely many solutions actually does better; in fact it demonstrates that the number of solutions for p_k such that $k \leq n$ does not drop below $(c_2/2)/n$, hence, it seems that that method is worthy of study. A fourth paper by Erdos is cited here as well; however its content was deemed more technical and less interesting and the author decided to use it merely as a reference.

Notation Throughout this paper, n will be used to denote natural numbers, p_n will be the nth prime, and c_k will be a positive constant. We will also use d_k to denote $p_{k+1} - p_k$.

2 Proof of Formula 1

At the time of Erdos's publication, it was known that

$$\limsup \frac{p_{n+1} - p_n}{\log p_n} = \infty,$$

and various lower bounds on the differences between infinitely many consecutive primes were known; however, much less was known about the quantity A in (1). As mentioned previously, Erdos's proof of (1) was the first unconditional result concerning the limit inferior. Hardy and Littlewood, proved that the A in (1) must be less than 2/3 using the Riemann Hypothesis, a few years before, and Rankin, again assuming the truth of the Riemann Hypothesis, proved that it was less than 3/5.

The article in question [1] was surprisingly terse and as noted by Dr. Morrow much more like a proof sketch than a full-blown proof, however, it was as far as the author knows Erdos's most extensive treatment on the topic. Therefore some details which Erdos omitted will be expanded upon, and afterwards there will be a brief discussion on late developments on this topic.

2.1 Proof

Erdos's proof of formula (1) relies on several lemmas.

Lemma 2.1. *For any a, the number of solutions of $p_i+p_j = a$, where $p_i, p_j < n$, is at most*

$$c \prod_{p|a} (1 + \frac{1}{p}) \frac{n}{(\log n)^2},$$

for some c, and for p ranging over prime numbers.

I cannot find a proof of 2.1 in English, unfortunately, and it is beyond my ability as of now to prove it independently, so I will have to take it on faith. Erdos cites the paper *Uber additive Eigenschaften von Zahlen* by L. Schnirelmann publishced in Math. Annalen, vol. 107 (1933), and Dr. Greenberg assures me that the product is indeed over the primes, and with that we disregard the proof of the above lemma.

Lemma 2.2. *For c_1 sufficiently small, we have that*

$$\sum_R \prod_{p|a} (1 + \frac{1}{p}) \leq \frac{\log n}{6c},$$

where the R indicates that the summation is over $(1 - c_1) \log n \leq a \leq (1 + c_1) \log n$, the ps are prime, and the c is the same as in the statement of Lemma 2.1.

Proof. By the existance of a unique prime factorization, we see that the largest denominator that can result from expanding the product is in fact $(1+c_1) \log n$, so if we correctly adjust the coefficients, we should be able to bound the entire thing by $\sum_{i=1}^{(1+c_1)\log n} k_i/i$ for some k_i. There are at most $\lceil 2c_4 \log n \rceil$ integers over which the a can range, where the brackets indicate the ceiling function. As for any a the product

$$\prod_{p|a}(1+\frac{1}{p})$$

will, once expanded, provide at most a coefficient of 1 to the $1/d$ term, for any $d \leq p$, again by the uniqueness of the prime factorization, and as in a span of n numbers a number d can only be a divisor of $\lceil n/d \rceil$ such numbers, we see that

$$\sum_R \prod_{p|a}(1+\frac{1}{p}) \leq \sum_{d\leq(1+c_1)\log n} \frac{1}{d}(\frac{2c_1 \log n}{d}+1),$$

as

$$\lceil \frac{\lceil 2c_1 \log n \rceil}{d} \rceil \leq (\frac{2c_1 \log n}{d}+1),$$

for $d \geq 1$, and so as

$$\sum_{d\leq(1+c_1)\log n} \frac{1}{d}(\frac{2c_1 \log n}{d}+1) \leq c_2 \log n + \sum_{d\leq(1+c_1)\log n} \frac{1}{d} \leq \frac{\log n}{6c}$$

for c_1 sufficiently small. □

The proof of the theorem now depends on a basic estimate on how many primes can exist between $n/2$ and n for any n. Let $p_k, p_{k+1}, \cdots, p_{k+j}$ be the primes such that $p_i \in (n/2, n)$. We now require another lemma.

Lemma 2.3. *Let $p_k, p_{k+1}, \cdots, p_{k+j}$ be defined as above. For sufficiently large n, $j > (1/2-\epsilon)n/\log n$.*

Proof. By the prime number theorem, as $\pi(n) > (1-\epsilon)n/\log n$ for large n, and $\pi(n/2) > ((1+\epsilon)n/2)/\log(n/2) > ((1+\epsilon)n/2)/\log n$, we get that the number of primes between $(n/2, n)$ is just $\pi(n) - \pi(n/2)$, which is greater than $(1/2 - 3\epsilon/2)n/\log n$, which, as ϵ is arbitrary, proves the claim. □

Now to show formula (1), it is enough to show that there is at least one i, for n large, such that $p_{k+i+1} - p_{k+i} < (1-c_4)\log n$, as then we see that

$$\liminf_{i\to\infty} \frac{p_{i+1}-p_i}{\log p_r} \leq \frac{(1-c_4)\log n}{\log n/2} \to 1-c_4.$$

The trick is to use telescoping sums. Let $d_{k+i} = p_{k+i+1} - p_{k+1}$. As $p_{k+j} - p_k \leq n/2$, we get that

$$\sum_{i=1}^{j-1} d_i \leq n/2. \tag{5}$$

Now, from Lemmas 2.1 and 2.2, we gather that the number of bs such that $(1-c_1)\log n \le b \le (1+c_1)\log n$ does not exceed

$$c\sum_U \prod_{p|a}(1+\frac{1}{p}) \le \frac{n}{6\log n}.$$

Prooceed now by contradiction. Assume that there is no d_{k+i} such that $d_{k+i} < (1-c_1)\log n$. Then that means, as at most $n/(6\log n)$ of the $(1/2 - \epsilon n)/\log n$ (by Lemma 2.3) d's fall within the range $((1-c_1)\log n, (1+c_1)\log n)$, we get that

$$\sum_{i=1}^{j-1} d_{k+i} \ge \frac{n}{6\log n}(1-c_1)\log n + (1/2 - 1/6 - \epsilon)\frac{n}{\log n}(1+c_1)\log n,$$

which is, after cancelling out redundant terms, $(1-2\epsilon)n/2 + (1/6-\epsilon)c_1 n > n/2$ if ϵ is sufficiently small, which clearly contradicts (3). Hence we must have at least one $d_{k+i} < (1-c_1)\log n$, which completes the proof.

2.2 More Recent Developments and Conjectures

This result was at the time groundbreaking, but still very unsatisfactory. Indeed, if the twin prime conjecture is true, as it likely is, then it should be trivially obvious that $\liminf(p_{k+1}-p_k)/\log p_k$ is zero. Eventually, in 2009, Goldston, Pintz, and Yildirim proved that the limit inferior is in fact zero, and furthermore produced the unconditional result that, letting q_n denote the nth number that has at most two nontrivial factors, $\liminf(q_{k+1}-q_k) < 26$. However, the techniques that they used are too advanced for the purposes of this paper, and it might be said much less elegant than Erdos's method here.

In [1], Erdos goes on to consider a similar looking sum to the one used to produce a contradiction in the above proof, which however does not telescope so neatly. He considers the k primes less than n, and conjectures that

$$\sum_{i=1}^{k-1}(p_{i+1}-p_i)^2 = O(n\log n).$$

Cramer, in 1937, provide by assuming the truth of the Riemann Hypothesis, that

$$\sum_{i=1}^{k-1}(p_{i+1}-P_i) = O(\frac{n}{\log\log n}), \qquad (p_{i+1}-P_i) > (\log q_i)^2,$$

which motivated Erdos's own conjecture about the topic. However, Erdos could not find a way to prove his conjecture, and in fact could not even prove a highly simplified version of this conjecture, and indeed, the author has not found any progress on these questions even to this day, and as Erdos himself writes, "the result, if true, must be very deep."

3 Considering the Differences of Primes

This section will fundamentally deal with the proof of the following theorem:

Theorem 3.1. *There are infinitely many solutions to (2).*

We will approach this in two different ways. First, by considering different mean values directly, and then by considering the differences between primes and putting a lower bound on the number of solutions to (2). The first was suggested in [4], and the second is dealt with thoroughly in [2]. The second is more intensive but gives us better results than the first method.

3.1 By Considering Mean Values

This section will deal directly with (2), and prove Theorem 3.1 directly by that.

This following result was claimed previously, and now will be proved.

Lemma 3.2. *For positive fixed a, b,*

$$\lim_{t\to 0}(\frac{a^t + b^t}{2})^{1/t} = \sqrt{ab}.$$

Proof. The proof is straightforward. As

$$(\frac{a^t + b^t}{2})^{1/t} = \exp(\frac{\log((a^t + b^t)/2)}{t}),$$

and

$$\lim_{t\to 0} \frac{\log((a^t + b^t)/2)}{t} = \frac{1}{2}(\log a + \log b),$$

the result follows immediately. □

Erdos first notes that the first inequality in (4) has infinitely many solutions, by the following lemma.

Lemma 3.3. $\limsup d_k = \infty$.

Proof. This follows directly from the fact that $n! + 2, n! + 3, n! + 4, \cdots, n! + n$ are all composite, and so the largest prime less than $n! + 2$ and the first prime after that must differ by at least n, hence there are infinitely many consecutive primes for which the difference grows without bound. □

We can now prove the above claim.

Theorem 3.4. *The first inequality in (4) has infinitely solutions.*

Proof. We note that the claim is the same as claiming that there are infinitely many n such that

$$p_{n+1} - p_n > p_n - p_{n-1}.$$

Assume this is not the case, that is, for some n_0 for all $n \geq n_0$ we would have that

$$p_{n+1} - p_n \leq p_n - p_{n-1},$$

which in particular would imply that $p_{n-1} - p_n$ was bounded for all large enough n, which contradicts Lemma 3.1. $\square$

The other inequalities require slightly more delicate bounds. First we note the well-known (actually well-known!) fact that for $\lambda \approx 1.25506$,

$$\pi(x) > \lambda x / \log x, \tag{6}$$

and we require a lemma.

Lemma 3.5. *Let $A > 0$ be any constant. Then there are infinitely many n such that*

$$p_n - p_{n-1} < p_{n+1} - p_n, \qquad p_k - p_{k-1} < A p_k^{1/2} \tag{7}$$

and

$$p_n - p_{n-1} > p_{n+1} - p_n, \qquad p_{k+1} - p_k < A p_k^{1/2}. \tag{8}$$

Proof. We first show that from (6), that there is a c_2 and are infinitely many m such that $p_{m+1} - p_m < c_2 \log p_m$. Assume otherwise. Then for all $k \geq m$ we have that $p_{m+1} - p_m > c_2 \log p_m$. We notice that for a suitable $c_3 > 1$, we get that for sufficiently large k,

$$\sum_{m}^{k-1} \log p_{m+1} > (m+k)\log(m+k-1), \tag{9}$$

as by the prime number therem, for large n,

$$\pi(n) \leq \frac{(1+\epsilon)p_n}{\log p_n},$$

and so plugging in $n = p_k$, we see that

$$k \leq \frac{(1+\epsilon)p_k}{\log p_k},$$

and in particular

$$\frac{p_k}{k} \geq \frac{\log p_n}{(1+\epsilon)}.$$

Furthermore, we see that by [5] on pg. 460,

$$(m+k)\log(m+k-1) > (1-\epsilon_0)(m+k)\log(m+k) > c_3 n \log p_{m+k}$$

Hence, by assumption we see that $p_{m+k} > p_m + c_2 \sum_{m}^{k-1} \log p_{m+1} > c_2 c_3 n \log p_{m+k}$, so

$$\frac{\lambda p_{m+k}}{\log p_{m+k}} \geq \frac{n \log p_{m+k}}{\log p_{m+k}} \geq n,$$

which contradicts (6). Hence there are indeed infinitely many m such that $p_{m+1} - p_m < c_2 \log p_n$. By Theorem 3.4, there are infinitely many primes such that $p_{n+1} - p_n > p_n - p_{n-1}$. Choose the first $k > m$ for m large for which $p_{n+1} - p_n > p_n - p_{n-1}$. Then clearly if m is large then both conditions of (7) are satisfied.

To prove (8) has infinitely many solutions, assume again that there are only finitely many; that is there are no solutions for $p > p_0$. We just proved above that there are infinitely many primes for which $p_{m+1} - p_m < c_2 \log p_m$, so choose such a m large. Letting p_r be the smallest prime greater than $p_m^{1/2}$, we see that

$$p_{r+1} - p_r \leq p_{r+2} - p_{r+1} \leq \cdots \leq p_{m+1} - p_m \leq c_2 \log p_m,$$

for if not then letting k such that $r < k \leq m$ be the greatest such that $p_{k+1} - p_k \leq p_k - p_{k+1}$, we get that $p_{k+1} - p_k \leq p_{m+1} - p_m < c_2 \log p_m \leq Ap_m^1/2$, which cannot be true because then p_k is a solution for (8). But now, if

$$p_{i+1} - p_i = p_{i+2} - p_2 = \cdots = p_{i+s+1} - p_{i+s} = d,$$

since that the numbers $\alpha, \alpha + \beta, \alpha + 2\beta, \cdots, \alpha + \alpha\beta = \alpha(\beta + 1)$ cannot all be prime, we get that $s \leq d$, so by the above we get that $m - r \leq c_2 \log p_m$, so

$$m = \pi(p_m) \leq r + c_2 \log p_m,$$

which contradicts (6), which completes the proof of the lemma. □

Now we have all the necessary tools to prove Theorem 3.1. We first show that there are infinitely many solutions to the first inequality of (2). It is trivial to show (by taking derivatives) that $((a^t + b^t)/2)^{1/t}$ is an increasing function of t, so it suffices to prove it for $t = -l$, where $l \geq 2$. Let p_{k-1}, p_k, p_{k+1} satisfy (7) with $A \leq 1/(2l^2)$. We will show that it also satisfies (2). Let $p_k - p_{p-1} = u$, as again it is trivial to show that $((a^t + b^t)/2)^{1/t}$ is increasing with respect to a and b, it is sufficient to show that the first inequality of (2) is satisfied for $p_{k+1} - p_k = u + 1$, which is tantamount to showing that

$$(\frac{(p_k - u)^{-l} + (p_k + u + 1)^{-l}}{2})^{-1/l} > p_k.$$

By moving everything over, this is equivalent to showing that

$$(p_k - u)^{-l} + (p_k + u + 1)^{-l} < 2p_k^{-l},$$

or

$$(p_k + u + 1)^l (2(p_k - u)^l - p_k^l) > p_k^l (p_k - u)^l,$$

by moving everything under a common denominator (These steps were done carefully by Erdos, which is surprising considering the terse tone of the rest of the paper). Now as $u = p_k - p_{k-1} < p_k^{1/2}/(2l^2)$ by assumption, we get that

$$p_k^l - ulp_k^{l-1} < (p_k - u)^l < p_k^l - ulp_k^{l-1} + \binom{l}{2} u^2 p) k^{l-2} + \cdots < p_k^l - (ul - 1/2)p_k^{l-1},$$

for large k, so it suffices to show that

$$p_k^l + (u+1)lp_k^{l-1}(p_k - 2ulp_k^{l-1}) > p_k^l(p_k^l - (ul-1/2)p_k^{l-1}),$$

which after expanding everything out, is equivalent to showing that

$$(1-1/2)p_k^{l-1} > 2l^2u(u+1)p_k^{l-2},$$

which is obviously satisfied for $u < p_k^{1/2}/2l^2$, which shows half of Theorem 3.1. To show the second half, procced as before, but put $p_{k+1} - p_k = u$, and consider the case of $p_k - p_{k-1} = u+1$ and $t \geq 2$. Then we show that

$$(p_k - u - 1)^t + (p_k + u)^t - 2p_k^t < 0,$$

so letting $u < p_k^{1/2}/2l^2$, by a symmetric argument with inequalities reversed to what we did above, we get that

$$(p_k + u)^t < p_k^t + (tu + 1/2)p_k^{t-1},$$

so again, essentially repeating what we did above, we get that

$$(p_k-(u+1))^t+(p_k+u)^t-2p_k^t < 2p_k^t-((u+1)t-1/2)p_k^{t-1}+(tu+1/2)p_K^{t-1}-2p_k^t < 0,$$

which completes the proof of Theorem 3.1.

3.2 By Considering Differences

In this section we will approach Theorem 3.1 by considering the equations

$$d_{k+1} > (1+c_1)d_k \qquad k \leq n \tag{10}$$

and

$$d_{l+1} < (1+c_1)d_l \qquad l \leq n \tag{11}$$

for $c_1 < 1$. We will eventually build up the following result:

Theorem 3.6. *There are infinitely many solutions to equations (10) and (11); in fact, for some $c_2 < 1$, the number of solutions of both is greater than $c_2 n$.*

If we assume this, then we can almost immediately obtain a stronger version of Theorem 3.1, after a few lemmas.

Lemma 3.7. *For all n,*

$$p_n < 2n \log n. \tag{12}$$

Proofs of this are widely known, for instance in [5]. Using Lemma 2.2, we can easily deduce the following:

Lemma 3.8. *Let ϵ be fixed. The number of $k \leq n$ with $p_{k+1} > (1+\epsilon)p_k$ is less than $c_1 \log n$ for some fixed c_1.*

Proof. The result follows from pigeonholing. Let $p_{j_1}, p_{j_2}, \cdots, p_{j_k}$ be the sequence of k primes with the property that $p_{j_\alpha} > (1+\epsilon)p_{(j_\alpha - 1)}$ and $j_\alpha < n$; that is, the sequence of primes that satisfy the condition in the lemma. Then clearly by applying the defining inequality and the fact that $p_{k+1} > p_k$, we get that $p_n > p_{j_k} > (1+\epsilon)^k p_{(j_1-1)}$, so if $k > c_1 \log n$ for all c_1 for n sufficiently large, we get that, by lemma 1, $2n \log n > p_{(j_1-1)} n^{c_1 \log(1+\epsilon)}$ for all c_1 for n sufficiently large, which is clearly false, hence, by contradiction, $k < c_1 \log n$. □

Using these lemmas, we can now show the following:

Theorem 3.9. *For all t and all sufficiently large n the number of solutions of*

$$(\frac{p_{k-1}^t + p_{k+1}^t}{2})^{1/t} > p_k, \qquad\qquad k \le n$$

is bounded below by $(c_2/2)n$.

Proof. From the above lemma, and Theorem 3.2, we gather that there are, for sufficiently large k and n, there are at least $c_2 n - c \log n > (c_2/2)n$ simultaneous solutions to the equations

$$p_{k+1} < (1+\epsilon)p_k, \qquad d_k > (1+c_1)d_{k-1}, \qquad k \le n. \tag{13}$$

Now we show that if p_{k-1}, p_k, and p_{k+1} satisfy (10), then they satisfy the first inequality of Theorem 3.2. We note that it immediately follows from (13) that

$$\epsilon p_k > p_{k+1} - p_k > (1+c_1)d_{k-1},$$

so as the first inequality in Theorem 3.2 is equivalent to showing the inequality that

$$p_{k+1}^t + p_{k-1}^t > 2p_k^t,$$

or that

$$p_{k+1}^t - p_k^t > p_k^t - p_{k-1}^t,$$

which as the right side is less than $p_k^t - p_{k-1}^t < (p_k - p_{k-1})^t < \epsilon^t p_k^t$, and the right side is greater than $c(p_{k+1} - p_k)^t$ for some c, we get that for sufficiently small ϵ, as $c(1+c_1)d_k > \epsilon p_k$, for small ϵ, our claim is proved. □

The proof of the opposite directions uses virtually the same lemmas but with orientations reversed, and so if Theorem 3.2 is true we get some very nice lower bounds on the number of solutions to (2), which immediately implies that there are infinitely many such solutions.

Now we proceed with the proof of Theorem 3.2 itself. We require several lemmas again.

Lemma 3.10. *For sufficiently small $c_1 > 0$ the number of solutions of the inequalities*

$$1 - c_1 < d_{k+1}/d_k < 1 + c_1, \qquad\qquad k \le n \tag{14}$$

is less than $n/4$.

Proof. The proof requires surprisingly technical sieve methods. First we construct a general bound to the number of solutions of equations relating to the differences of primes. Let $g(n;a,b)$ denote the number of solutions to the simultaneous equations

$$d_{k+1} = a, d_k = b, \qquad\qquad k \leq n.$$

Let V be the number of primes $p_0 < 2n\log n$ such that $p_0 + a$ and $p_0 + a + b$ are still prime. As $p_n < 2n\log n$, it is clear that $g(n;a,b) \leq V$. Now for small c_2, letting $q_1, q_2, \cdots, q_j$ be the primes less than n^{c_1}, it is clear that V is not greater than $n^{c_2} + U$, where U is the number of integers $m \leq 2n\log n$ which satisfy, for all i,

$$m \not\equiv 0 \mod q_i, \qquad m \not\equiv -a \mod q_i, \qquad m \not\equiv -(a+b) \mod q_i,$$

as these would be the solutions to the above simultaneous equations assuming there were only the j primes below n^{c_1} so for small c_1 this is clearly greater than V as there are less and less primes with which the numbers are filtered through, resulting in a larger and larger value. Now by elementary number theory, if $q \nmid ab(a+b)$, then none of the above equivalences follow trivially from any other; that is, it is not the case that

$$0 \equiv a, 0 \equiv (a+b) \mod q_i, \text{etc.},$$

and as it could not be the case that q, $q+a$ and $q+a+b$ could be prime and have $q \nmid ab(a+b)$ (as we can create a factorization), we may assume that the residues (the modulus) are distinct (Throughout this proof we will assume that q is prime). Now, Erdos proved in [3] that the number of solutions to the modular equivalences above with $m < r$ is less than

$$cr\prod_i (1 - 3/q_i);$$

hence we have that

$$U < c_4 n\log n \prod_i (1 - 3/q_i), q < n^{c_3}, \qquad q \nmid ab(a+b)$$

with the factor of two being absorbed into the constant. Now, by [5] on p. 349, we see that

$$\prod_{q<x} (1 - \frac{3}{q}) < \frac{c}{(\log x)^3},$$

so writing that product as the product of all the terms for which $q < n^{c_3}$ and $q \nmid ab(a+b)$ and the others, or those for which $q \mid ab(a+b)$, as we discussed before that if $q \mid ab(a+b)$ then $q > n^{c_3}$, as otherwise contradictions arise. Hence, we get that

$$U < c_5 \frac{n}{(\log n)^2} \prod (1 + \frac{3}{q-3}), \qquad\qquad q \mid ab(a+b),$$

so reindexing the product slightly simplifies the above to the assertion that

$$U < c_5 \frac{n}{(\log n)^2} \prod (1 + \frac{3}{q}), \qquad q \mid ab(a+b),$$

and so we get, that as $V \leq U + n^{c_1}$, that

$$g(n; a, b) < c_5' \frac{n}{(\log n)^2} \prod (1 + \frac{3}{q}), \qquad q \mid ab(a+b).$$

as for small c_1 $n^{c_1} \leq n/(\log x)^2$.

Now we can proceed as follows. Split the $k's$ that solve (14) into two categories. First consider all such k for which $d_k > 20 \log n$. As $p_n < 2n \log n$, we get that the number of k's in that catagory cannot exceed $n/10$, as otherwise the primes would have to be larger. The second group requires subtler consideration. As $d_{k+1} = b$ and $d_k = a$, and so by (14) $1 + c_1 > b/a > 1 - c_1$, the number of k's in the second class is not greater than

$$\sum_R g(n; a, b) < c_6 \frac{n}{(\log n)^2} \sum_R \prod_q (1 + \frac{3}{q}), \qquad q \mid ab(a+b), \tag{15}$$

where the R indicates that the summation is over such a and b for which $a < 20 \log n$, as this is the second category, and $1 + c_1 > b/a > 1 - c_1$, (in the second sum we consider a to be fixed) as explained above. Now as q is prime, as a prime only divides $ab(a+b)$ if it divides a, b, and/or $a+b$, we get that

$$\sum_R \prod_q (1 + \frac{3}{q}) \leq \sum_1 (\prod_{q|a} (1 + \frac{3}{q}) \sum_2 \prod_{q|b(a+b)} (1 + \frac{3}{q})),$$

where $\sum_1$ indicates that the sum is over $a < 20 \log n$, and the $\sum_2$ indicates that the sum is over $1 + c_1 > b/a > 1 - c_1$, as the latter sum includes a few possible repeated terms but no omitted terms. By that same logic,

$$\prod_{q|b(a+b)} (1 + \frac{3}{q}) \leq \prod_{q|b} (1 + \frac{3}{q}) \prod_{q|(a+b)} (1 + \frac{3}{q}),$$

and so as

$$\prod_{q|b} (1+\frac{3}{q}) \prod_{q|(a+b)} (1+\frac{3}{q}) < 2 \prod_{q|b} (1+\frac{3}{q}) \prod_{q|(a+b)} (1+\frac{3}{q}) \leq \prod_{q|b} (1+\frac{3}{q})^2 + \prod_{q|(a+b)} (1+\frac{3}{q})^2,$$

as $(\alpha - \beta)^2 \geq 0$, for any α, β, and so $\alpha^2 + \beta^2 \geq 2\alpha\beta$, and so as

$$(1 + \frac{3}{q})^2 = 1 + \frac{6q + 9}{q^2} < 1 + \frac{15}{q}$$

for all prime q, we get that

$$\sum_2 (\prod_{q|b(a+b)} (1 + \frac{3}{q})) < \sum_2 (\prod_{q|b} (1 + \frac{15}{q}) + (\prod_{q|(a+b)} 1 + \frac{15}{q})).$$

Now we use a similar technique to one used in the proof of Formula (1). Fixing a, we notice that for b such that $(1+c_1)a > b > (1-c_1)a$ for any number m there are at most $1 + 2c_1a/m$ solutions to the equation $b \equiv 0 \mod m$. By replacing b with $a+b$ at every instance, we see that there are at most $1 + 2c_1a/m$ to the equation $a + b \equiv 0 \mod m$, and so letting $v(m)$ denote the number of prime divisors of m, we get that (as a factor of 15 gets multiplied to the coefficient of $1/m$ for every prime factor m has, which is evident just by expanding) that the above sum of products is less than

$$\sum_{m<3a} 2(1+\frac{2c_1a}{m})\frac{15^{v(m)}}{m},$$

as q ranges over primes and so the largest number d such that $1/d$ could appear in the sum is $a+b < 3a$ as we are summing over that restricted range. As after some value the number of unique prime factors of a number n grows slower than $\log_{15} n$, this grows at $O(n)$, which can be shown by splitting the sum up into those two groups, so as c_1 is constant, the above is bounded by c_6c_1n for some c_6. Notice also that c_6 should intuitively be quite large to cover the constant factors (and increasing as c_1 decreases). We will not provide the full argument here; it is rather tedious and only requires very basic counting techniques, and so we leave it to the reader.

Given what was stated above, we see then that, by putting the inequality in the expression for the original sum, that, by similar arguments to the above

$$\sum_R \prod_{q|ab(a+b)} (1+\frac{3}{q}) < c_6c_1 \sum_1 a \prod_{q|a}(1+\frac{3}{q}) < 20c_6c_1 \log n \sum_1 \prod_{q|a}(1+\frac{3}{q}), \quad (16)$$

which using similar logic as above, is bounded by

$$20c_6c_1 \log n \sum_{m=1}^{\infty} \frac{20 \log n 3^{v(m)}}{m^2},$$

and as $v(m)$ actually grows slower than logarithmically with base 3 (we mentioned this above; again, it is not difficult to prove) and so the infinite sum converges to some finite value, and so that above is less than

$$c_7c_1(\log n)^2 < \frac{1}{10c_5'}(\log n)^2$$

for c_1 sufficiently small, so the number of solutions in the second category is also less than $n/10$, so the whole thing is less than $n/4$, which finally proves the lemma. □

The above lemma lets us set a bound on the number of differences between two numbers, and now we set our sights on bounding the number of d_ks such that they are unbounded from above and unbounded from below.

Lemma 3.11. *For some constant c_8, the number of integers $k \le n$ such that*

$$d_{k+1}/d_k > t \quad or \quad d_{k+1}/d_k < 1/t \tag{17}$$

is less than $c_8 n/t^{1/2}$.

Proof. The proof is similiar to the proof of the previous lemma, but we will only consider this for large t; this is sufficient as for an appropriate choice of c_8 we can adjust it for smaller t's. We first split the integers satisfying (16) into two classes. The first class contains all the k's such that either $d_k \ge t^{1/2}\log n$ or $d_{k+1} \ge t^{1/2}\log n$. The second class contains all the k's such that $d_k \le (\log n)/t^{1/2}, or d_{k+1} \le \log n/t^{1/2}$. Clearly these two classes contain all such k's, as if none of these inequalities held an evident contradiction arises from considering the ratio of the successive d_k's. By Lemma 3.7, the number of k's in the first group must be less than $4n/t^{1/2}$. Again, the second group requires more delicate handling. Erdos uses another result from [3] that states that the number Z_a of solutions of the equation $d_u = a, u \le n$ is less than

$$Z_a < c_9 \log n \prod_q (1 - \frac{2}{q}), \qquad q|a, q < n^{c_9}.$$

We do as we did in the previous proof, and assert that

$$Z_a < c_{10}\frac{n}{\log n}\prod p \mid a(1 + \frac{2}{q}),$$

and so from almost mirror reasoning as in the previous lemma we find that

$$Z_\alpha < 2c_{10}\frac{n}{\log n}\sum_{a<\log n/t^{1/2}}\prod_{q|a}(1+\frac{2}{q}) < 2c_{10}\frac{n}{\log n}\sum_{k=1}^{\infty}\frac{\log n 2^{v(k)}}{t^{1/2}k^2} < 2\frac{c_{11}n}{t^{1/2}}, \tag{18}$$

where the factor of two in those first inequalities arises just because either d_k or d_{k+1} could satisfy the conditions. The proof is now complete; combining the two parts gives use the necessary estimates.

After these technically challenging lemmas, however, the proof of Theorem 3.6 follows almost immediately. We will only prove the first claim; the second follows symmetrically. We proceed by contradiction; suppose that for all $c_1 > 0$ and $\epsilon > 0$ there is an arbitrarily large n such that the number of solutions of

$$d_{k+1} > (1 - c_1)d_k, \qquad k \le n$$

is less than ϵn. Consider the telescoping product

$$\frac{d_n}{d_1} = \frac{d_2}{d_1}\frac{d_3}{d_2}\cdots\frac{d_n}{d_{n-1}}.$$

We now consider the range of all the possible d_{k+1}/d_k's and bound them all:

1. By Lemma 3.11 the number of $k \leq n$ satisfying $d_{k+1}/d_k > 2^{2l}$ is less than $c_8 n/2^l$.

2. By the contradiction hypothesis, the number of d_{k+1}/d_k which are bounded by 2^{2u} for all u does not exceed ϵn

3. By Lemma 3.10 the number the number of d_{k+1}/d_k which are bounded by $1 + c_1$ and greater than $1 - c_1$ is less than $n/4$

4. The remaining ratios must not exceed $1 - c_1$, and as the bounds above do not sum up to $n/2$ for large l and n, the number of such ratios does not exceed $n/2$.

Hence their product cannot exceed

$$\frac{d_n}{d_n} < 2^{2u\epsilon n}(\prod_{l \geq 2^{2u}} (2^{2l})^{c_8 n/2^l})(1+c_1)^{n/4}(1-c_1)^{n/2},$$

(the infinite product converges very quickly and its convergence is quite trivial to show) and so

$$\frac{d_n}{d_1} < 2^{2u\epsilon n} \exp \sum_{l \geq u} \frac{c_9 n l \log 4}{2^l}(1-c_1)^{n/4}.$$

Now choosing ϵ small then for some u we can get that $2^{2u\epsilon n} < (1+c_1)^{n/8}$ and the exponentiated sum to be less than that same amount, and so we evidently get that $d_n/d_1 = d_n < (1+c_1)^n/4(1-c_1)^n/4 = (1-c_1^2)^n/4 < 1/n$, which does not make sense, hence there is a contradiction, and so the first proposition of Theorem 3.6 is proved. The other direction uses similar estimates, in fact uses slightly simpler bounds, and so we will omit its proof, which follows almost identically as above.

3.3 Conjectures

Erdos's derived result here has not seen much application as far as the author knows, and so these theorems seem to remain merely a mathematical curiosity. These results however do dispel any notion of the "convexity" or "concavity" of prime numbers, and again reinforce then notion of the utter unpredictability of the abundance of prime numbers. However, in the course of those two papers Erdos does conjecture several interesting and to my knowledge unsolved formulas, which if resolved could provide further insight into the abundance of primes. His primary concern throughout [2] and [4] was to try to find bounds on the number of solutions to $d_k > d_{k-1}$ and $d_k < d_{k-1}$; both papers offer lower bounds, but not much other detail into how d_k acted as a function of k. In [2] Erdos out of curiosity defines the function

$$w(k) = \begin{cases} 1 & \text{if } d_k > d_{k-1}; \\ 0 & \text{if } d_k \leq d_{k-1}. \end{cases}$$

It would not be unreasonable to consider the behavior of the number $W = \sum_{i=0}^{\infty} w(k)/2^i$; and it would seem that it should (or at least the author would think so) be true that W is at least irrational; however, as there is in fact no proof known today even of the much simpler proposition that $w(k)$ does not merely alternate between 1 and 0 in a predictable fashion for k sufficiently large, which is rather shocking, actually, we cannot appraoch this problem at all. This perhaps only serves to show the number of open questions and general unpredictability that the prime numbers continue to exhibit even after thousands of years of study.

References

[1] Erdos, Paul, The Difference of Consecutive Primes, Duke Math J., Vol 2, Number 6(1940), 438 – 441.

[2] Erdos, Paul, Some new questions on the distribution of primes. Bull. Amer. Math. Soc. vol 54 (1948) pp 371-378.

[3] Erdos, Paul, On the difference of consecutive primes. Bull. Amer. Math. Soc., Volume 54, Number 10 (1948), 885–889.

[4] Erdos, Paul & Turan, Paul, On Some New Questions on the Distribution of PrimeNumbers, Bull. Amer. Math. Soc. Vol 54, Number 4 (1948), 371–378.

[5] Hardy, G. H., and E. M. Wright. An Introduction to the Theory of Numbers. Oxford: Clarendon Press, 1979.