A Proof of the Twin Prime Conjecture

A W. Draut
Embry Riddle Aeronautical University
Prescott, AZ

draut129@erau.edu

20 August 2020

Abstract

The traditional definition of the twin prime conjecture is that there is an infinite number of twin primes. The traditional definition of a twin prime is a pair of primes separated by one even number, e.g., 29 and 31. We expand this definition and prove the infinitude of two types of twin primes.

Our primary vehicle for proving the twin prime conjecture is a structure that we call Eratosthenes' Patterns, which are created by Eratosthenes Sieve. First, we describe Eratosthenes' Sieve, then we describe Eratosthenes' Patterns, then we give the proof.

The essence of our proof is to show that the number of prime twins between p_n and p_n^2 approaches infinity as n approaches infinity.

Before we begin, we cover two nomenclature topics: a definition and mathematical notation.

Definition of prime number We restrict our definition to the natural numbers, $\mathbb{N}$.
Any number that has exactly two divisors is prime.
Any number that has three or more divisors is composite.
1 is the only number that has exactly one divisor and is neither prime nor composite. □

We have a possible confusion in our notation. We use (a,b) for the greatest common divisor of a and b. This notation is commonly used in Number Theory. We also use (a,b) for an open set of numbers, bounded by a and b. This notation is commonly used in Set Theory. The reader should be aware of the context.

The Sieve of Eratosthenes

Eratosthenes sieve is a method for finding prime numbers. Eratosthenes lived circa 200 BC.

In the Sieve of Eratosthenes, primes and multiples of primes are removed from the natural numbers. One starts with the natural numbers and creates a second set by removing all multiples of 2. Then one creates the third set by removing all multiples of 3 from the second set. This process is continued as long as desired. We name these sets A_1, A_2, A_3, ..., A_n with A_1 being the natural numbers.

We show some of these sets here.

A_1	1 2 3 4 5 6 7 8 ...	the natural numbers
A_2	1 3 5 7 9 11 13 15 ...	multiples of 2 removed (odd numbers)
A_3	1 5 7 11 13 17 19 23 25 29 31 35 37 ...	multiples of 3 removed from A_2
A_4	1 7 11 13 17 19 23 29 31 37 41 43 47 49 ...	multiples of 5 removed from A_3
$\vdots$		

All of the primes, however many there are, are in A_1
All of the primes greater than 2 are in A_2, in order.
All of the primes greater than 3 are in A_3, in order.
All of the primes greater than 5 are in A_4, in order.
$\vdots$

These statements seem obvious and trivial, but they will prove necessary below.

In each set, the first member is 1 and the second member is the nth prime.
In each set, the first composite is p_n^2.
In each set, apart from the 1, all of the numbers less than p_n^2 are prime.
When multiples of p_n are removed from A_n in order to create A_{n+1}, an infinity of composites is removed but only one prime is removed.
By choosing the 2nd member of each set, A_1 through A_n, we construct a list of the first n primes.

All of the information given above is thousands of years old, except for the names of the sets, which we have chosen.

Instead of studying the primes that have been found, we have studied the numbers that are left behind after primes and their multiples have been removed from the natural numbers and we have found interesting and useful patterns. We call them Eratosthenes' Patterns. Next, we give some of the many useful features of Eratosthenes' Patterns. We have not found these structures in the literature.

Eratosthenes' Patterns

We show A_3 and A_4 rearranged so as to demonstrate the patterns that we have found.

A_3 original multiples of 2 and 3 have been removed from A_1

1 5 7 11 13 17 19 23 25 29 31 35 37 ...

A_3 rearranged

1	7	13	19	25	31	37	43	49	55	61 ...
5	11	17	23	29	35	41	47	53	59	65 ...

A_4 original multiples of 5 have been removed from A_3

1 7 11 13 17 19 23 29 31 37 41 43 47 49 53 59 61 67 ...

A_4 rearranged

1	31	61	91	121	151	181	211	241	271	301	331	361	391	421 ...
7	37	67	97	127	157	187	217	247	277	307	337	367	397	427 ...
11	41	71	101	131	161	191	221	251	281	311	341	371	401	431 ...
13	43	73	103	133	163	193	223	253	283	313	343	373	403	433 ...
17	47	77	107	137	167	197	227	257	287	317	347	377	407	437 ...
19	49	79	109	139	169	199	229	259	289	319	349	379	409	439 ...
23	53	83	113	143	173	203	233	263	293	323	353	383	413	443 ...
29	59	89	119	149	179	209	239	269	299	329	359	389	419	449 ...

As stated above, the first member in the first column is 1 and the second number is the nth prime. We call the columns patterns. The first pattern is the 'fundamental pattern'. Note the spaces after every p_n patterns. These delineate what we call extended patterns (to be explained below). The first extended pattern is the 'fundamental extended pattern'.

We use three parameters to describe the patterns, λ, ν, and τ. λ is used for the length of a pattern. ν is used for the number of members in a pattern, and τ is used for the number of twins in a pattern. (We explain twins below.) p is used for prime numbers. Note that the boundaries of the patterns are not members of the set. For example, in A_4, the boundaries are at 0, 30, 60, 90, ... Note that $\lambda_4 = 30$. The length of an extended pattern in A_4 is 210. $\lambda_4 \times p_4 = 210$.

The creation of A_{n+1} from A_n involves removing all multiples of p_n from A_n. The extended patterns in A_n become the patterns of A_{n+1}. When an extended pattern of A_n becomes a pattern of A_{n+1}, the patterns of A_n become subpatterns of the patterns of A_{n+1}. For example, in the 2nd pattern of A_5, which is bounded by 210 and 420, the subpatterns are bounded by 210, 240, 270, 300, ... (A_5 is given below.) In A_4, the patterns all contain 8 members. In A_5, the subpatterns have less than 8 members.

We speak of of the first and second halves of an extended pattern. In the case of A_4, the first half of the fundamental extended pattern is from 1 to 103 and the second half is from 107 to 209.

We speak of 'stepping' from A_n to A_{n+1}.

Next, we show the sets A_1 through A_5 with the parameters added and certain numbers in boldface. The numbers in boldface are multiples of p_n. The distribution of the multiples of p_n is critical.

A_1 $\quad p_1 = 2$, $\lambda_1 = 1$, $\nu_1 = 1$

1 **2** 3 **4** 5 **6** 7 **8** 9 ...

A_2 $\quad p_2 = 3$, $\lambda_2 = 2$, $\nu_2 = 1$

1 **3** 5 7 **9** 11 13 **15** 17 19 ...

A_3 $\quad p_3 = 5$, $\lambda_3 = 6$, $\nu_3 = 2$, $\tau_3 = 1$

1	7	13	19	**25**	31	37	43	49	**55**	61 ...
5	11	17	23	29	**35**	41	47	53	59	**65** ...

A_4 $\quad p_4 = 7$, $\lambda_4 = 30$, $\nu_4 = 8$, $\tau_4 = 3$

1	31	61	**91**	121	151	181	211	241	271	**301**	331	361	391	421 ...
7	37	67	97	127	157	187	**217**	247	277	307	337	367	397	**427** ...
11	41	71	101	131	**161**	191	221	251	281	311	341	**371**	401	431 ...
13	43	73	103	**133**	163	193	223	253	283	313	**343**	373	403	433 ...
17	47	**77**	107	137	167	197	227	257	**287**	317	347	377	407	437 ...
19	**49**	79	109	139	169	199	229	**259**	289	319	349	379	409	439 ...
23	53	83	113	143	173	**203**	233	263	293	323	353	383	**413**	443 ...
29	59	89	**119**	149	179	209	239	269	299	**329**	359	389	419	449 ...

Note the repetition in the distribution of the numbers in boldface (multiples of p_n) across the extended patterns. This is due to the fact that a member of an extended pattern plus the length of an extended pattern ($\lambda_n \times p_n$) gives a corresponding member in the next extended pattern. If one of these is a multiple of p_n then the other must also be a multiple. Also, we find that in any row of an extended pattern, there is exactly one multiple of p_n. We explain this below.

We continue with A_5.

A_5 $\qquad p_5 = 11,\ \lambda_5 = 210,\ \nu_5 = 48,\ \tau_5 = 15$

1	211	421	631	841	1051	1261	1471	1681	1891	**2101**	2311	2521 ...
11	221	431	641	851	1061	1271	1481	1691	1901	2111	**2321**	2531 ...
13	223	433	643	853	1063	1273	1483	1693	**1903**	2113	2323	2533 ...
17	227	437	647	857	**1067**	1277	1487	1697	1907	2117	2327	2537 ...
19	229	439	**649**	859	1069	1279	1489	1699	1909	2119	2329	2539 ...
23	233	443	653	863	1073	1283	1493	1703	1913	**2123**	2333	2543 ...
29	239	449	659	**869**	1079	1289	1499	1709	1919	2129	2339	2549 ...
31	241	**451**	661	871	1081	1291	1501	1711	1921	2131	2341	2551 ...
37	247	457	667	877	1087	1297	**1507**	1717	1927	2137	2347	2557 ...
41	251	461	**671**	881	1091	1301	1511	1721	1931	2141	2351	2561 ...
43	**253**	463	673	883	1093	1303	1513	1723	1933	2143	2353	**2563** ...
47	257	467	677	887	1097	1307	1517	**1727**	1937	2147	2357	2567 ...
53	263	**473**	683	893	1103	1313	1523	1733	1943	2153	2363	2573 ...
59	269	479	689	899	1109	1319	**1529**	1739	1949	2159	2369	2579 ...
61	271	481	691	901	**1111**	1321	1531	1741	1951	2161	2371	2581 ...
67	277	487	697	907	1117	1327	1537	1747	1957	**2167**	2377	2587 ...
71	281	491	701	911	1121	**1331**	1541	1751	1961	2171	2381	2591 ...
73	283	493	703	**913**	1123	1333	1543	1753	1963	2173	2383	2593 ...
79	289	499	709	919	1129	1339	1549	1759	**1969**	2179	2389	2599 ...
83	293	503	713	923	**1133**	1343	1553	1763	1973	2183	2393	2603 ...
89	299	509	719	929	1139	1349	1559	1769	1979	**2189**	2399	2609 ...
97	307	**517**	727	937	1147	1357	1567	1777	1987	2197	2407	2617 ...
101	311	521	731	941	1151	1361	1571	1781	**1991**	2201	2411	2621 ...
103	313	523	733	943	1153	1363	**1573**	1783	1993	2203	2413	2623 ...
107	317	527	**737**	947	1157	1367	1577	1787	1997	2207	2417	2627 ...
109	**319**	529	739	949	1159	1369	1579	1789	1999	2209	2419	**2629** ...
113	323	533	743	953	1163	1373	1583	**1793**	2003	2213	2423	2633 ...
121	331	541	751	961	1171	1381	1591	1801	2011	2221	**2431**	2641 ...
127	337	547	757	967	**1177**	1387	1597	1807	2017	2227	2437	2647 ...
131	**341**	551	761	971	1181	1391	1601	1811	2021	2231	2441	**2651** ...
137	347	557	767	977	1187	**1397**	1607	1817	2027	2237	2447	2657 ...
139	349	559	769	**979**	1189	1399	1609	1819	2029	2239	2449	2659 ...
143	353	563	773	983	1193	1403	1613	1823	2033	2243	**2453**	2663 ...
149	359	569	779	989	**1199**	1409	1619	1829	2039	2249	2459	2669 ...
151	361	571	**781**	991	1201	1411	1621	1831	2041	2251	2461	2671 ...
157	367	577	787	997	1207	1417	1627	**1837**	2047	2257	2467	2677 ...
163	373	**583**	793	1003	1213	1423	1633	1843	2053	2263	2473	2683 ...
167	377	587	797	1007	1217	1427	1637	1847	**2057**	2267	2477	2687 ...
169	379	589	799	1009	1219	1429	**1639**	1849	2059	2269	2479	2689 ...
173	383	593	**803**	1013	1223	1433	1643	1853	2063	2273	2483	2693 ...
179	389	599	809	1019	1229	1439	1649	**1859**	2069	2279	2489	2699 ...
181	391	601	811	1021	1231	**1441**	1651	1861	2071	2281	2491	2701 ...
187	397	607	817	1027	1237	1447	1657	1867	2077	2287	**2497**	2707 ...
191	401	611	821	1031	1241	1451	**1661**	1871	2081	2291	2501	2711 ...
193	403	613	823	1033	**1243**	1453	1663	1873	2083	2293	2503	2713 ...
197	**407**	617	827	1037	1347	1457	1667	1877	2087	2297	2507	**2717** ...
199	409	619	829	1039	1249	1459	1669	1879	2089	**2299**	2509	2719 ...
209	419	629	839	1049	1259	1469	1679	1889	2099	2309	**2519**	2729 ...

Note that the distribution of multiples of 11 in A_5 is similar to what we saw in A_4: the distribution is the same in all extended patterns and there is exactly one multiple of 11 in each row of an extended pattern.

We review the structure of these sets. We depict them with the patterns as vertical columns. The vertical length of an extended pattern is determined by the patterns. The horizontal width of an extended pattern is determined by p_n. As we step through the sets, A_n, the extended patterns develop a very large aspect ratio. i.e., the vertical size vs the horizontal width. The extended patterns are placed left to right and extend to infinity.

The set is shown as a collection of extended patterns, left to right, with finite vertical height and an infinite horizontal width.

The members of the fundamental pattern in A_n are a reduced residue system[1] of λ_n since every member of A_n is relatively prime to λ_n.

The members of the fundamental pattern in A_n are a multiplicative group modulo λ_n[2]

The number of members in the fundamental pattern in A_n, ν_n, is Euler's totient[3] for λ_n.

The members of a pattern are symmetrically arranged about $\lambda_n/2$. For example, in A_4, 7 is a member and 30 - 7 = 23 is a member. ($\lambda_4 = 30$). This symmetry occurs because $(\lambda_n, a) = (\lambda_n - a, a) = 1$ for any member, a, of A_n. This is a specific example of a more general case involving any reduced residue system. In a reduced residue system of m, in which all members are positive and less than m, the members are symmetrically arranged about m/2. We show this with the following theorem.

Theorem 1 $\quad (m, a) = (m - a, a), \quad$ where $m > a$

Proof Let $g_1 = (m, a)$ and $g_2 = (m - a, a)$. g_1 also divides $m - a$ and therefore, $g_1 | g_2$. g_2 also divides $m - a + a = m$ and therefore, $g_2 | g_1$. Thus $g_1 = g_2$. □

There is also a symmetry about the center of an extended pattern of A_n. If a is a member, so is $(\lambda_n \times p_n) - a$. For example, look at the fundamental extended pattern of A_4 in the table above. 53 and 157 are images of each other in this specific type of symmetry. ($53 + 157 = 210$)

As we search through the extended patterns of A_n we have seen that there is a multiple of p_n in the same position of a every extended pattern. These particular multiples are separated by $p_n \times \lambda_n$. Thus, they are a residue class[4] in A_n.

These are the reasons for our choice of p_n patterns as the size of the extended patterns.

When one studies the patterns, in any set, and notes the infinitude of numbers that have been removed from the natural numbers, one should keep in mind the fact that only composite num-

[1]See, for example, Apostol, section 5.2
[2]See, for example, Niven, section 2.11
[3]See, for example, Niven, section 2.1
[4]See, for example, Apostol section 5.2

bers have been removed from the natural numbers, with one exception: the gap between 1 and p_n (the first two numbers in a fundamental pattern) is the only place where primes have been removed.

When stepping from one set to the next, apart from the fundamental extended pattern, when an extended pattern in A_n becomes a pattern in A_{n+1}, the ratio of primes to composites increases.

The Twins

Here, we define twins. We define two types of twins: two-twins and four-twins. Historically, twins have referred to what we call two-twins. Two-twins are two odd numbers which differ by two, such as 17 and 19. Four-twins differ by four, such as 19 and 23. We shall show that the four-twins are as significant as the two-twins. The two-twins and four-twins are equal in number in any pattern. Their distributions in a pattern are similar and their distributions in extended patterns are similar.

We start by showing different arrangements of A_3.

A_3 Four-Twins

1	7	13	19	25	31	37	43	49	55	61 ...
5	11	17	23	29	35	41	47	53	59	65 ...

A_3 (modified) Two-Twins

	5	11	17	23	29	35	41	47	53 ...
1	7	13	19	25	31	37	43	49	55 ...

	5	11	17	23	29		35	41	47	53	59 ...	(2nd version)
1	7	13	19	25		31	37	43	49	55 ...		

In the first case given above, the patterns are four-twins and, in the second and third cases, the patterns are two-twins.

All the members of A_3 are of the form $6n \pm 1$. Thus, all members of any set, A_k, where $k \geq 3$, are of the form $6n \pm 1$. Since all primes greater than or equal to 5 are in the set, A_3, we can justify the following theorem.

Theorem 2 All primes ≥ 5 are of the form, $6n \pm 1$ □

To designate a 2-twin, we use the multiple of 6 (even multiple of 3) on which the twin is centered. For example, the 2-twin, 17,19, is designated as 18. To designate a 4-twin, we use the odd multiple of 3 on which the twin is centered. For example, the four-twin, 19,23 is designated as 21.

A twin with two prime members is called a prime twin. A twin with one prime member is called a combination twin, and a twin with two composite members is called a composite twin.

We have adopted the convention that 3,5 is not a two-twin, since 3 is not a member of A_3. We note that no primes fall between the two members of a twin. However, looking at the set A_3 above we see that we have to make an exception in the case of the four-twin, 1,5, since there are two primes, 2 and 3, between 1 and 5. We explain the significance of A_3 next.

Another way to describe these features is to note that the number that falls between the members of a two-twin is always a multiple of 6 and, of the three numbers that fall between the members of a four-twin, two are multiples of 2 and one is an odd multiple of 3. A number can only be a member of a unique two-twin. A number can only be a member of a unique four-twin. However, a number can simultaneously be a member of both a two-twin and a four-twin. For example, 19 is a member of the two-twin, 18, and is also a member of the four twin, 21. But 19 is not a member of any other two-twin nor any other four-twin. In the steps to higher numbered sets, two-twins and four-twins can only be destroyed and cannot be created. Thus, the density of either type of these twins in A_{n+1} is always less than that of A_n. However, six-twins, eight-twins, and others, are created in almost all steps from one set to the next. The set A_3 is an infinite sequence of alternating 2-twins and 4-twins.

All twins, 2-twins and 4-twins, are created in A_3.

According to our definitions there are no twins in A_1 nor in A_2.

Also, it is clear, from the structure of A_3, that the spacing between any two twins, either 2-twins or 4-twins, is a multiple of 6.

In the remainder of this paper, unless otherwise stated, when we speak of stepping from A_n to A_{n+1}, $n \geq 3$. Also, our use of the word 'twins' will refer to 2-twins and 4-twins only.

Theorem 3 In any pattern, the number of two-twins is exactly equal to the number of four-twins.

Proof by induction.

The basis case: A_4 has exactly 3 two-twins and 3 four-twins in each pattern. Also, there are 21 two-twins and 21 four-twins in each extended pattern. Recall that there is one half of a two-twin at the beginning and ending of each pattern.

The induction case: There are τ_n two-twins and τ_n four-twins in the patterns of A_n. There are $p_n \tau_n$ twins of each type in each extended pattern. Each twin in the fundamental pattern is at the head of two rows of twins and these two rows have two multiples of p_n. There will be $2\tau_n$ two-twins and $2\tau_n$ four twins removed from the fundamental extended pattern, and from each extended pattern, in the step to A_{n+1}. This will leave $\tau_n(p_n - 2)$ two-twins and $\tau_n(p_n - 2)$ four-twins in the extended patterns of A_{n+1}. $\square$

Next we show A_5 with a count of the twins added.

A_5 $p_5 = 11$, $\lambda_5 = 210$, $\nu_5 = 48$, $\tau_5 = 15$ (count of twins added)

2-twin													4-twin
15	1	211	421	631	841	1051	1261	1471	1681	1891	**2101**	2311 ...	
1	**11**	221	431	641	851	1061	1271	1481	1691	1901	2111	**2321** ...	
1	13	223	433	643	853	1063	1273	1483	1693	**19.5**	2113	2323 ...	1
2	17	227	437	647	857	**1067**	1277	1487	1697	1907	2117	2327 ...	1
2	19	229	439	**649**	859	1069	1279	1489	1699	1909	2119	2329 ...	2
	23	233	443	653	863	1073	1283	1493	1703	1913	**2123**	2333 ...	2
3	29	239	449	659	**869**	1079	1289	1499	1709	1919	2129	2339 ...	
3	31	241	**451**	661	871	1081	1291	1501	1711	1921	2131	2341 ...	
	37	247	457	667	877	1087	1297	**1507**	1717	1927	2137	2347 ...	3
4	41	251	461	**671**	881	1091	1301	1511	1721	1931	2141	2351 ...	3
4	43	**253**	463	673	883	1093	1303	1513	1723	1933	2143	2353 ...	4
	47	257	467	677	887	1097	1307	1517	**1727**	1937	2147	2357 ...	4
	53	263	**473**	683	893	1103	1313	1523	1733	1943	2153	2363 ...	
5	59	269	479	689	899	1109	1319	**1529**	1739	1949	2159	2369 ...	
5	61	271	481	691	901	**1111**	1321	1531	1741	1951	2161	2371 ...	
	67	277	487	697	907	1117	1327	1537	1747	1957	**2167**	2377 ...	5
6	71	281	491	701	911	1121	**1331**	1541	1751	1961	2171	2381 ...	5
6	73	283	493	703	**913**	1123	1333	1543	1753	1963	2173	2383 ...	
	79	289	499	709	919	1129	1339	1549	1759	**1969**	2179	2389 ...	6
	83	293	503	713	923	**1133**	1343	1553	1763	1973	2183	2393 ...	6
	89	299	509	719	929	1139	1349	1559	1769	1979	**2189**	2399 ...	
	97	307	**517**	727	937	1147	1357	1567	1777	1987	2197	2407 ...	7
7	101	311	521	731	941	1151	1361	1571	1781	**1991**	2201	2411 ...	7
7	103	313	523	733	943	1153	1363	**1573**	1783	1993	2203	2413 ...	8
8	107	317	527	**737**	947	1157	1367	1577	1787	1997	2207	2417 ...	8
8	109	**319**	529	739	949	1159	1369	1579	1789	1999	2209	2419 ...	9
	113	323	533	743	953	1163	1373	1583	**1793**	2003	2213	2423 ...	9
	121	331	541	751	961	1171	1381	1591	1801	2011	2221	**2431** ...	
	127	337	547	757	967	**1177**	1387	1597	1807	2017	2227	2437 ...	10
	131	**341**	551	761	971	1181	1391	1601	1811	2021	2231	2441 ...	10
9	137	347	557	767	977	1187	**1397**	1607	1817	2027	2237	2447 ...	
9	139	349	559	769	**979**	1189	1399	1609	1819	2029	2239	2449 ...	11
	143	353	563	773	983	1193	1403	1613	1823	2033	2243	**2453** ...	11
10	149	359	569	779	989	**1199**	1409	1619	1829	2039	2249	2459 ...	
10	151	361	571	**781**	991	1201	1411	1621	1831	2041	2251	2461 ...	
	157	367	577	787	997	1207	1417	1627	**1837**	2047	2257	2467 ...	
	163	373	**583**	793	1003	1213	1423	1633	1843	2053	2263	2473 ...	12
11	167	377	587	797	1007	1217	1427	1637	1847	**2057**	2267	2477 ...	12
11	169	379	589	799	1009	1219	1429	**1639**	1849	2059	2269	2479 ...	13
	173	383	593	**803**	1013	1223	1433	1643	1853	2063	2273	2483 ...	13
12	179	389	599	809	1019	1229	1439	1649	**1859**	2069	2279	2489 ...	
12	181	391	601	811	1021	1231	**1441**	1651	1861	2071	2281	2491 ...	
	187	397	607	817	1027	1237	1447	1657	1867	2077	2287	**2497** ...	14
13	191	401	611	821	1031	1241	1451	**1661**	1871	2081	2291	2501 ...	14
13	193	403	613	823	1033	**1243**	1453	1663	1873	2083	2293	2503 ...	15
14	197	**407**	617	827	1037	1347	1457	1667	1877	2087	2297	2507 ...	15
14	199	409	619	829	1039	1249	1459	1669	1879	2089	**2299**	2509 ...	
15	**209**	419	629	839	1049	1259	1469	1679	1889	2099	2309	**2519** ...	

Here we give formulas for calculating λ, ν, and τ.

The length of a pattern in A_n is: $\lambda_n = \prod_{i=1}^{n-1} p_i, \quad n \geq 2$

$$\lambda_{n+1} = \lambda_n p_n$$

The number of members in a pattern in A_n is: $\nu_n = \prod_{i=1}^{n-1} (p_i - 1), \quad n \geq 2$

$$\nu_{n+1} = \nu_n(p_n - 1)$$

The number of twins in a pattern in A_n is: $\tau_n = \prod_{i=2}^{n-1} (p_i - 2), \quad n \geq 3$

$$\tau_{n+1} = \tau_n(p_n - 2)$$

Here are the parameters for the first 10 sets.

n	p_n	p_n^2	λ_n	ν_n	τ_n
1	2	4	1	1	-
2	3	9	2	1	-
3	5	25	6	2	1
4	7	49	30	8	3
5	11	121	210	48	15
6	13	169	2310	480	135
7	17	289	30030	5760	1485
8	19	361	510510	92160	22275
9	23	529	9699690	1658880	378675
10	29	841	223092870	36495360	7952175

Here we show a formula for any member of any set. This allows one to build the set A_n without having any of the previous sets available.

Theorem 4 A formula for any member of any set.

Every member of A_n, greater than 2, can be represented in the following form.

$$a = (\lambda_n/2) \pm 2^j p_{k_1}^{b_1} p_{k_2}^{b_2} \cdots , \qquad j > 0,\ k_i \geq n,\ b_i \geq 0,\ n \geq 2,\ \lambda_n/2 = \prod_{i=2}^{n-1} p_i$$

Proof. Let a be a member of A_n. a, by the definition of membership in A_n, cannot be divided by any prime less than p_n. The first of the two terms above contains the primes from 3 to p_{n-1}. The second contains multiples of 2, and possible multiples of primes that are greater than or equal to p_n. Next, we only need to show that every member of A_n can be represented in the above form.

Let z be any member of A_n, greater than 2, and $z = x + y$, with x being odd and y being even. x can be any odd number and, when chosen, y is determined. Let $x = \lambda_n/2 = \prod_{i=2}^{n-1} p_i$. Therefore, y must contain a power of 2 as a factor. If there are other factors of y, they must be divisible by powers of primes greater than or equal to p_n. Thus, x and y are of the forms of the first and second terms, right of the equal sign, in the statement of the theorem given above. □

Next, we define four new parameters that we will use in various calculations:
'Vulnerable twins'; 'Singles'; ' Blocks'; and 'g_n'.

Vulnerable twins

In the step to A_{n+1}, the number of twins removed from the extended pattern of A_n is $2\tau_n$. Recall that for each twin in a pattern, two rows are occupied in an extended pattern and each of these rows has one multiple of p_n. The two twins that contain these multiples are the twins that will be removed in the step to A_{n+1}. We call these 'vulnerable' twins, since they will not be members of A_{n+1}.

Simply put, the vulnerable twins are those twins that are removed from A_n in the step to A_{n+1}.

Singles

A single is a member of A_n that is not a member of a twin. We consider singles in a pattern that are not members of two-twins. We also consider singles in a pattern that are not members of a four-twin. The number of singles among the two-twins is the same as the number of singles among the four-twins. We have not considered singles in a pattern that not members of both two-twins and four-twins. That would be an interesting calculation for the future.

There are ν_n members in a pattern and τ_n twins in a pattern. This gives $\nu_n - 2\tau_n$ singles in a pattern. Thus, the average number of singles between twins is $(\nu_n - 2\tau_n)/\tau_n = (\nu_n/\tau_n) - 2$. As we step through the sets this average grows monotonically without limit, but the increase, per step, in this average approaches zero.

Blocks

We separate the members of the various sets into sequences that we call blocks. A block is the subset of members between p_n^2 and p_{n+1}^2. We have two schemes for naming the blocks, one for A_1, the natural numbers, and another for the other sets.

We divide A_1 into 'blocks' as follows, using cardinal numbers.

B1 = 4 to 8 $\quad p_1^2 = 4$, $p_2^2 = 9$
B2 = 9 to 24
B3 = 25 to 48 $\quad$ Bn begins with p_n^2 and ends with $p_{n+1}^2 - 1$
B4 = 49 to 120
B5 = 121 to 168
⋮
B10 = 841 to 960
⋮
B20 = 5041 to 5328
⋮

Next, we use ordinal numbers for these same blocks when they appear in the various sets.

set	primary	2nd	3rd	4th
A_1	2 to 3	4 to 8	9 to 24	25 to 48
A_2	3 to 8	9 to 24	25 to 48	49 to 120
A_3	5 to 24	25 to 48	49 to 120	121 to 168
A_4	7 to 48	49 to 120	121 to 168	169 to 288
A_5	11 to 120	121 to 168	169 to 288	289 to 360
⋮				
A_{10}	29 to 840	841 to 960	961 to 1368	1369 to 1680
⋮				
A_{20}	71 to 5040	5041 to 5328	5329 to 6240	6241 to 6888
⋮				

We use set theory notation to define a block. For example, in A_3, the third block is [49,121).

Note that the block from p_n to $p_n^2 - 1$ in a set is called the primary block for that set.

Note that the second block in A_n is the same as Bn. Also, the jth block in A_n is the same as the jth-1 block in A_{n+1}, $j \geq 3$. Also note that the primary block in A_n ($n \geq 2$) does not have a corresponding block in A_1.

Note that the primary block in A_n consists of a merger of the primary block in A_{n-1} with the second block in A_{n-1}. In this merger, p_{n-1} and its multiples are removed.

On average, in A_n, the blocks in the jth+1 pattern are larger that those in the jth pattern.

As we step through the sets, the size of a block, say Bk, does not change (except for the primary blocks). The lower and upper boundaries of Bk do not change (except for the primary blocks). The primary blocks increase in size and both their lower and upper boundaries move forward. The number of members in the blocks decrease except that, in the primary blocks, the numbers increase. We prove these statements below.

Here are some examples.
Let Bk be the jth block in A_n. Bk = $[p_k^2, p_{k+1}^2)$. We find: k = n + j - 2. k is the cardinal number associated with Bk. j is the ordinal number associated with the jth block in the set A_n

Consider B7, [289,361). It is:
4th block in A_5 k = 7 , n = 5 , j = 4 16 members
3rd block in A_6 k = 7 , n = 6 , j = 3 14 members
2nd block in A_7 k = 7 , n = 7 , j = 2 12 members
In all cases B7 spans the numbers from 289 to 360, and its size is 72.

The primary block in A_5 is [11,121) 16 members size = 110
The primary block in A_6 is [13,169) 34 members. size = 156
The primary block in A_7 is [17,289) 55 members. size = 272

We have not found these structures that we call blocks in the literature.

g_n

g is the gap between primes. $g_n = p_{n+1} - p_n$.

$\pi(x)$ is commonly used to designate the number of primes that are less than or equal to x. According to the prime number theorem:[5]

$$\lim_{x\to\infty} \pi(x)\frac{log(x)}{x} = 1$$

If we let $\pi^*(x) = \frac{x}{logx}$, the derivative of $\pi^*(x) = \frac{logx-1}{log^2 x}$. This approaches $\frac{1}{logx}$ as x approaches infinity.

$$\lim_{x\to\infty} \frac{d\pi(x)}{dx} = \lim_{x\to\infty} \frac{d\pi^*(x)}{dx} = \frac{1}{logx}$$

As n approaches infinity, the average gap between two consecutive primes approaches log p_n, which approaches infinity.

The Distribution of Twins in a Set

We are currently studying many aspects of the distribution of the twins. We cite three of them here since we feel these will help the reader understand this proof. In a subsequent paper we will give more aspects, including constellations of primes other than twins.

This discussion applies to all twins, in all blocks, including the primary blocks.

1. Vulnerables among the twins:
There are $\tau_n\, p_n$ twins and $2\tau_n$ vulnerable twins in an extended pattern. This gives a ratio of $2/p_n$ vulnerable twins to twins. Another way to look at it is that the average number of twins

[5]references 3,4,5,6

that fall between two vulnerable twins is $p_n/2$. This number grows to infinity as n approaches infinity. As n increases, the vulnerable twins in A_n become sparse among the twins. For large n, as one steps through the sets, one finds almost no change in the distribution of twins at each step. The sparsity of the vulnerable twins is critical in the discussions below.

These calculations apply to both 2-twins and 4-twins.

2. Pattern boundary twins and pattern center twins:
The twins appear to be randomly distributed. However, there are uniform sequences of twins that are superimposed on the distribution.

We start with what we call the pattern boundary twins. They are created in A_3 where the boundaries are multiples of 6. However, we find it easier to study the sequence in A_4 where the boundaries of the patterns are 30, 60, 90, 120, ... Each of these is the center of a two-twin. There is an infinite number of these two-twins separated by a distance of 30 each. In an extended pattern (length = 210) there are 7 of these pattern boundary twins. They occupy two rows and one from each row will be eliminated in the step to A_5, leaving 5 of these twins in each pattern of A_5. In A_4, we look at the twin, 210, at the upper boundary of the fundamental extended pattern. Neither of the members of 210 is a multiple of 7, preventing it from being eliminated in the step to A_5. This leads to a pattern boundary twin in each pattern of A_5. This process continues indefinitely.

In any set, among other twins, there is a uniform distribution of pattern boundary two-twins throughout the set. They are separated by a distance of λ.

Next we look at the pattern center twins. In A_4, at the center of the fundamental pattern there is a four-twin, 15. This is explained by Theorem 4 above. The center of the fundamental pattern is $\lambda_n/2$ and the twin is generated by $(\lambda_n/2) \pm 2$. In A_4, 13 and 17 are at the heads of two rows of four-twins in the fundamental extended pattern. Of the 7 twins in these two rows, 2 will be eliminated in the step to A_5, leaving 5 four-twins in every pattern.
At the center of the fundamental extended pattern, neither of the two members of the four-twin, 105, are multiples of 7. The twin, 105, will not be eliminated in the step to A_5, leading to a four-twin at the center of every pattern.

In any set, among other twins, there is a uniform distribution of pattern center four-twins throughout the set. They are separated by a distance of λ.

There are many other constellations of twins that give a uniform distribution superimposed on the random distribution of other twins. For example, there is a constellation that we call 'hextuples'. They are created in A_4 by using Theorem 4. Their structure is $(\lambda_n/2) \pm 2$, $(\lambda_n/2) \pm 4$, and $(\lambda_n/2) \pm 8$. For example, at the center of the fundamental pattern of A_5, we find 97, 101, 103, 107, 109, 113. They include 2 two-twins and 3 four-twins. Notice that the pattern center twins are embedded in the hextuples. There is a hextuple at the center of every pattern in every set, A_n where $n \geq 4$, and they are uniformally distributed throughout the sets with a spacing of λ_n.

We will cover other constellations in a subsequent paper.

3. Growth of gaps:
Here we look at the growth of a gap between twins in the steps to subsequent sets. We shall show that this growth is limited.

First, we introduce new terminology. We borrow from the field of aeronautical engineering and speak of the leading edge of a primary block. For example, in stepping through the sets, A_6, A_7, and A_8, the leading edge of the primary blocks advances from 168 to 288 to 360. We speak of the advance of the leading edge.

In addition to the leading edge of the primary blocks we speak of the leading edge of a gap between twins and note its advance. At each step the average size of the advance of the leading edge of a gap is equal to the average distance between twins.

For a demonstration of the growth of a gap we choose four consecutive 2-twins, 462, 480, 492, and 522 which are in the fundamental pattern of A_6 and have gaps of 18, 12, and 30 between them. (A partial listing of A_6 is given in the appendix.) 481 is a multiple of 13 and, in the step to A_7, the twin, 480, will be eliminated. (479 will remain as a single.) Thus, we find in the fundamental pattern of A_7 a gap of 30 between the twins 462 and 492. We have seen a simple example of a gap growing from 18 to 30 in one step.

We now have three of the four original consecutive 2-twins in A_7, 462, 492, and 522 with gaps of 30 and 30. 493 is a multiple of 17 and the twin, 492, will be eliminated in the step to A_8. In A_8 we have two of the twins left, 462 and 522, with a gap of 60.

The original sequence of four 2-twins is in B8, [361,529), which is the 4th block of A_6 and the 2nd block of A_8. The primary block of A_9 is [23,529). The leading edge of the A_9 primary block is greater than 523 and the two twins, 462 and 522, neither of which contains a multiple of 19, in the step to A_9, will be merged into the primary block and must be prime twins.

There are two methods by which the growth of a gap can be terminated. First, in stepping through the sets, a gap can become bounded on both ends by prime twins, even though it is outside of a primary block. Second, as we have seen above, the leading edge of the primary blocks advances beyond the leading edge of the gap.

Let's look at the advance of the leading edge of the gap in the example above. The original gap was 462 to 480, with 480 being the leading edge, with a gap of 18. In the step to A_7, the leading edge advances to 492 and 12 is added to the gap. In the step to A_8, the leading edge advances to 522 and 30 is added to the gap.
Note that in A_6 the average gap size between twins (λ_6/τ_6) is 17.1 and in A_8 it's 22.9.

Let's look at the advance of the leading edge of the primary blocks during these same steps. It advanced from 168 to 288 to 360. The leading edge of the gap between the twins advanced by 40; the leading edge of the primary blocks advanced by 192.

The leading edge of the primary blocks moves forward at the same rate as the advance of p_n^2. We show below that the leading edge of a gap moves forward at a near constant rate which leads to an ever increasing ratio of block size to gap size. The speeds of the advances have the same increasing ratio.

The crucial point here is:

In one step, the leading edge of a primary block moves forward one block, and the leading edge of a gap between twins moves forward one twin.

The growth of the gap between twins is limited. The growth of any gap will be terminated in a finite number of steps by either the appearance of a prime twin or by being overtaken by the advancing primary blocks.

We have shown three aspects of the distribution of twins. There is a sparsity of vulnerable twins. There is a uniformity in the distribution of some of the twins that is superimposed on the random distribution of the other twins. The growth of the gaps between the twins is limited. We are ready to state the proof of the infinitude of the twins.

The Proof

We shall show that the number of twins in the primary blocks increases to infinity as one steps through the sets. All members of a primary block are prime. We start with a count of the twins in any block, then a count of the twins in the primary blocks.

The Number of Twins in a Block

The number of twins in a block is the product of two factors: the size (length) of the block; and the density of the twins. However, we prefer to use the gap between twins, which is the reciprocal of the density, giving the number of twins as the block size divided by the average gap between twins.

In A_n, the average gap between twins is (λ_n/τ_n) and in A_{n+1} it is $(\lambda_{n+1}/\tau_{n+1})$.

$$(\lambda_{n+1}/\tau_{n+1}) = (\lambda_n/\tau_n)(\frac{p_n}{p_n - 2})$$

$$\lim_{n\to\infty} \frac{\lambda_{n+1}/\tau_{n+1}}{\lambda_n/\tau_n} = 1$$

This implies that, for large n, in each step, the change in λ_n/τ_n is negligible. Therefore, the average number of twins in a block depends almost entirely on the size of the block. Next, we look at the sizes of the blocks and calculate the number twins in a block.

Recall the letters that we use for specifying the blocks: n, k, and j, and the equation relating them: $k = n + j - 2$. n is the set number, k is the block number in A_1, and j is the position number of the block in a set other than A_1.

There are three ways to look at the number of twins in the various blocks. First we look at a particular set, i.e., n is constant. The average gap between twins, λ_n/τ_n, is the same throughout the set. Thus, the average number twins in each block is simply the size of the block divided by the average gap between twins.

Secondly, we let k be constant. As we step through the sets, the block Bk does not change in size or location among the natural numbers, only its contents change. When stepping forward, n increases and j decreases and the number of members and the number of twins in Bk decrease. We show below that we are interested in 2nd blocks and we step forward through the sets until Bk is the 2nd block of a set. When this occurs, $j = 2$, $n = k$, and the average number twins in the block is $(p_{n+1}^2 - p_n^2)/(\lambda_n/\tau_n)$.

We are not interested in these first two ways and place our interest on the third.

Let j be constant. As an example, we let $j = 2$. Below we explain the significance of 2nd blocks. As we step through the sets, $j = 2$ and $n = k$. The size of the second block is $p_{n+1}^2 - p_n^2 = 2p_n g_n + g_n^2$ which approaches infinity as n approaches infinity. Thus, the average number of twins in a 2nd block increases without bound as n increases.

One could calculate block sizes and numbers of twins for blocks other than the 2nd. Just change the value of j and follow the same procedures. However, in calculating the numbers of twins in the primary blocks below, we need to use the number twins in the 2nd blocks.

The Number of Twins in a Primary Block

One might argue that the ratio of the size of the primary block to the size of the fundamental pattern approaches zero as n approaches infinity. It is true that this ratio approaches zero, but the size of the primary block approaches infinity as n approaches infinity. Here we have a case where each of the lengths of two sequences approaches infinity while the ratio of their lengths approaches zero.

Next we show that the number of twins in the primary block of A_{n+1} is always greater than the number of twins in the primary block of A_n.

This is the essence of the proof of the infinitude of the prime twins. As n approaches infinity the lengths of the primary blocks approach infinity and the number of twins (2-twins and 4-twins) in a primary block approaches infinity. All members of a primary block are primes.

First we show that, in some cases, the second member of a set, p_n, is the first member of a twin and this causes a twin to be eliminated from the primary block in the step to the next set. However, for large n, this occurs so infrequently that we can ignore this in our count of the number of twins in a primary block. We show this by recalling from above that the average number of singles between consecutive twins, $(\nu_n/\tau_n) - 2$, grows without bound.

Next, we compare the length of a primary block to the length of a 2nd block. The length of a primary block is $p_n^2 - p_n = p_n(p_n - 1)$.

The length of a second block is $p_{n+1}^2 - p_n^2 = 2p_n g_n + g_n^2$.
When n is large, the length of the second block is approximately $2\, p_n \, log\, p_n + log^2 p_n$

When we compare the lengths for large n, we find p_n^2 compared to $p_n \times 2\,log\,p_n$, or equivalently, we compare p_n to $2\,log\,p_n$. We see that the ratio of the length of the primary block to the length of the second block grows to infinity as n increases.
Next, we count the number of twins in the primary and 2nd blocks.

Let the numbers of twins in the primary and 2nd block of A_n be a_1 and a_2. These two blocks merge to become the primary block of A_{n+1}.

We look at the two blocks before the merger and find that the total number of twins is $a_1 + a_2$. After the merger, the total number of twins is $a_1 + ra_2$, where r is the fraction of twins remaining in what was the 2nd block of A_n. We calculate r as follows.

Recall from above that the average number of twins per vulnerable twin is $p_n/2$. The average fraction of twins removed from a 2nd block in the step from A_n to A_{n+1} is $2/p_n$, which gives $r = 1 - 2/p_n$.

It is possible that in some sets no twins are eliminated from the 2nd block in the step to the next set, i.e., all the multiples of p_n in the 2nd block are singles and $r = 1$.

The number of twins in the primary block of A_{n+1} is greater than the number in A_n. The number of twins in the primary block grows without bound as we step through the sets.

This completes our proof of the infinitude of the twin primes. We have shown that the number twins (two-twins or four-twins) in the primary blocks increases to infinity as we step through the sets, A_n, with n approaching infinity. All members of a primary block are prime.

References

1. I. Niven and H.S. Zuckerman
An Introduction to the Theory of Numbers, 4th Ed., Wiley, 1980

2. T. Apostol *Introduction to Analytic Number Theory*, Springer-Verlag, 1976

Comment on $\pi(n)$. References 2 and 3 were written almost simultaneously by different researchers who were not in contact with each other. Both used analytical methods. References 3 and 4 were similar discoveries and used elementary methods.

3. J. Hadamard *Sur la distribution des zeros de la fonction $\zeta(s)$ et ses consequences arithmetiques*
Bull. Soc. Math. France, 24, pp. 199-220, 1896

4. Ch.-J. de la vallee poussin
La Fonction $\zeta(s)$ de Riemann et les Nombres Premiers en General
Annales de la Societe Scienifique de Bruxelles, 20, pp. 185-256, 1896

5. P. Erdos *On a New Method in Elementary Number Theory Which Leads to an Elementary Proof of the Prime Number Theory*
Proceedings of the National Academy of Sciences, Vol 35, No. 7, pp. 374-384, July 1949

6. A. Selberg *An Elementary Proof of the Prime-Number Theorem*
The Annals of Mathematics, Vol. 50, N0. 2, pp. 303-313, April, 1949

Appendix

A partial listing of A_6

$A_6 \quad p_6 = 13,\ \lambda_6 = 2310,\ \nu_6 = 480,\ \tau_6 = 135$

1	2311	4621...
13	2323	4633...
17	**2327**	4637...
⋮		
461	2771	5081...
463	2773	**5083**...
467	2777	5087...
479	2789	5099...
481	2791	5101...
487	2797	5107...
491	2801	5111...
493	2803	5113...
499	2809	5119...
503	2813	5123...
509	2819	5129...
521	2831	5141...
523	2833	5143...
⋮		
2309	4619	6929...