On the Infinite Series Characterizing the Elimination of Twin Prime Candidates

by

Dennis R. Martin
DP Technology Corp., Camarillo, CA
dennis.martin@dptechnology.com
BSME, Michigan Technological University, Houghton, MI
Department of Mathematics, University of California, Santa Barbara
drmartin@umail.ucsb.edu

This document is available in other formats. See www.primenace.com.

Abstract

The pattern of the composite numbers that have a particular lowest prime factor repeats over intervals equal to the primorial of that lowest prime factor. The number of composites having that prime for their lowest factor is constant over those primorial intervals, and the value of that constant for each prime is directly related to the value of the previous prime and its constant composite to primorial ratio.

Those primorial patterns apply to twin primes. The composite numbers that eliminate twin prime candidates can be counted in terms of their lowest prime factors. Those twin prime eliminations repeat over primorial intervals such that the rate of elimination can be represented by an infinite series over all primes 5 and greater. We calculate the partial sums for that series for the first 15 million primes and examine the implications of the convergence of that infinite series.

1 Primorial Soup

It is known that the pattern of composite numbers repeats with a period equal to the primorial of each prime factor. Dickson [1] refers to remarks by H. J. S. Smith and theorems by J. DesChamps regarding this property, and Weisstein [2] makes note of it. The **primorial** is analogous to a factorial applied to the sequence of prime numbers. The primorial for the prime p_n is the product of all primes up to and including p_n, and it is denoted as $p_n\#$. By the definition of factorial, $p_1\# = 2$, and then for every prime greater than 2:

$$p_n\# = p_n \bullet p_{n-1}\# \qquad (1)$$

Since p_n itself is prime and there can be no composites less than p_n having p_n for a factor, it is intuitive to start the first primorial interval at $p_n + 1$, such that it ends at $p_n + p_n\#$. Because in some ways we can think of composite numbers as molecules composed of their constituent prime factor atoms and arranged in a lattice across these primorial intervals, the value $p_n + p_n\#$ will be called the **first atomic boundary** and labeled α_n.

If we make a function $T_n(x)$ to count the total number of composites that have p_n as their lowest prime factor, then we find that the count in the first and all subsequent primorial intervals is a constant. This constant **composite to primorial ratio** shall be labeled ρ_n, where $\rho_n = T_n(\alpha_n)$.

Table 1: Composite to Primorial Ratio and Ratio Summation for the First Twelve Primes

n	p_n	$p_n\#$	$\alpha_n = p_n + p_n\#$	$\rho_n = T_n(\alpha_n)$	Σ (Numerator)
1	2	2	4	1	1
2	3	6	9	1	4
3	5	30	35	2	22
4	7	210	217	8	162
5	11	2310	2321	48	1830
6	13	30030	30043	480	24270
7	17	510510	510527	5760	418350
8	19	9699690	9699709	92160	8040810
9	23	223092870	223092893	1658880	186597510
10	29	6469693230	6469693259	36495360	5447823150
11	31	200560490130	200560490161	1021870080	169904387730
12	37	7420738134810	7420738134847	30656102400	6317118448410

It is readily apparent in Table 1 that each composite to primorial ratio is related to the previous prime and its ratio by the following:

$$\rho_n = \rho_{n-1} \bullet (p_{n-1} - 1) \qquad (2)$$

This relationship can be rigorously proven by observing the number of complete residue systems modulus the primorial of each of the previous primes up to p_n that exist within a $p_n\#$ primorial interval. Note also that Sloane [3] gives this as sequence A005867 in the On-Line Encyclopedia of Integer Sequences (OEIS) with the comment that it corresponds to the local minima of Euler's phi (totient) function.

2 To Be, Or Not To Be

Twin primes are primes of the form (p, $p + 2$), such as (3, 5), (5, 7), (11, 13), (17, 19), and so on. The **twin prime conjecture** is a currently unproven conjecture stating that there are infinitely many twin primes. It is well established that all twin primes after the first are of the form $6n \pm 1$.

Every pair of positive integers ($6n - 1$, $6n + 1$) can be considered to be a **twin prime candidate**, where the value $6n$ is the **center post**. In order to determine when both members of a twin prime candidate are indeed prime, it is important then to understand when and how one or both of them are not. A candidate is eliminated when either or both of the numbers adjacent to its center post are composite. Let us call such a composite an **eliminating composite**.

It is fairly easy to establish that the primes $p_1 = 2$ and $p_2 = 3$ cannot produce any eliminating composites, and that all composites having $p_3 = 5$ or greater for their lowest prime factor are eliminating composites. See Table 2, for example, and notice that all of the lowest prime factorizations are of the form $5 \bullet (6n \pm 1)$ and that starting above 5 there are two such composites in each interval of $p_3\# = 30$, which illustrates the relationship presented in the previous section.

Table 2: Eliminations of Twin Prime Candidates by $p_3 = 5$ in the Interval up to 217

Candidate Center Post	Eliminating Composite	Prime Factorization	Lowest Prime Factorization	Residues	
				(mod 30)	(mod 6)
24	25	5 • 5	5 • (6 – 1)	25	1
36	35	5 • 7	5 • (6 + 1)	5	5
54	55	5 • 11	5 • (12 – 1)	25	1
66	65	5 • 13	5 • (12 + 1)	5	5
84	85	5 • 17	5 • (18 – 1)	25	1
96	95	5 • 19	5 • (18 + 1)	5	5
114	115	5 • 23	5 • (24 – 1)	25	1
126	125	5 • 5 • 5	5 • (24 + 1)	5	5
144	145	5 • 29	5 • (30 – 1)	25	1
156	155	5 • 31	5 • (30 + 1)	5	5
174	175	5 • 5 • 7	5 • (36 – 1)	25	1
186	185	5 • 37	5 • (36 + 1)	5	5
204	205	5 • 41	5 • (42 – 1)	25	1
216	215	5 • 43	5 • (42 + 1)	5	5

There are 30 / 6 = 5 candidates out of every interval of 30, so that means that three candidates out of every interval of 30 are not eliminated by $p_3 = 5$. If those candidates are to be eliminated, they must be eliminated by a composite having $p_4 = 7$ or higher as its lowest prime factor.

The composites having 7 as their lowest prime factor have a pattern that repeats over intervals of $p_4\# = 210$, and there are $\rho_4 = 8$ of them within each such interval. Up to $\alpha_4 = 217$, for instance, those eight composites are (49, 77, 91, 119, 133, 161, 203, 217) while in the next interval to 427 they are (259, 287, 301, 329, 343, 371, 413, 427). This pattern is summarized in Table 3.

Table 3: Factorization of Any $p_4\# = 210$ Primorial Interval for the Lowest Prime Factor $p_4 = 7$

Primorial Interval Member	Lowest Prime Factorization	Residues		
		Modulo $p_4\#$	Modulo $p_3\#$	Modulo $p_2\#$
$210n + 49$	$7 \cdot (30n + 7)$	49 (mod 210)	19 (mod 30)	1 (mod 6)
$210n + 77$	$7 \cdot (30n + 11)$	77 (mod 210)	17 (mod 30)	5 (mod 6)
$210n + 91$	$7 \cdot (30n + 13)$	91 (mod 210)	1 (mod 30)	1 (mod 6)
$210n + 119$	$7 \cdot (30n + 17)$	119 (mod 210)	29 (mod 30)	5 (mod 6)
$210n + 133$	$7 \cdot (30n + 19)$	133 (mod 210)	13 (mod 30)	1 (mod 6)
$210n + 161$	$7 \cdot (30n + 23)$	161 (mod 210)	11 (mod 30)	5 (mod 6)
$210n + 203$	$7 \cdot (30n + 29)$	203 (mod 210)	23 (mod 30)	5 (mod 6)
$210(n + 1) + 7$	$7 \cdot (30(n + 1) + 1)$	7 (mod 210)	7 (mod 30)	1 (mod 6)

There are 210 / 6 = 35 candidates out of every interval of 210. With an interval starting above 7, there would be 2 • (210 / 30) = 14 eliminating composites produced by $p_3 = 5$. That matches the count from Table 2, which conveniently listed them up to 217. Add those 14 to the 8 eliminating composites produced by $p_4 = 7$ and that is still only 22. At least 35 – 22 = 13 candidates are not eliminated by p_3 or p_4 in each interval of 210. If those twin prime candidates are to be eliminated, it has to be by a composite with a lowest prime factor of $p_5 = 11$ or higher.

But there would only be 13 candidates left if all 22 of those eliminating composites corresponded to exactly one unique elimination of a twin prime. They do not. There are cases of **double eliminations** where both of the numbers in the candidate pair are composite. Two such double eliminations can be seen in the preceding tables. The value 205 in Table 2 and $210n + 203$ from Table 3, when $n = 0$, will both be adjacent to the center post 204, while the 215 from Table 2 and $210(n + 1) + 7$ from Table 3, when $n = 0$, will doubly eliminate the center post 216.

3 The Double Elimination Round

Let $E''_{m,n}$ represent the double eliminations made by a lower prime p_m in combination with the prime p_n. We can calculate the twin prime count $\pi_2(x)$ by subtracting the total composite count contributed by each lowest prime factor and then adding back in the double eliminations within that range. For $n \geq 3$ such that $p_n \leq x^{1/2}$ and $3 \leq m \leq (n - 1)$ that count is as follows, where the initial 1 is to account for the first twin prime (3, 5) and the \ in (x \ 6) indicates integer division.

$$\pi_2(x) = 1 + (x \setminus 6) - \Sigma T_n(x) + \Sigma E''_{m,n}(x) \qquad (3)$$

If we account for the eliminations of twin prime candidates by lower prime factors $p_m < p_n$ in this way, then p_n^2 is the earliest that a prime factor p_n can generate a previously unaccounted for eliminating composite. For example, Table 2 lists the composites 35 and 175, which have both 5 and 7 as factors, as being eliminations by $p_3 = 5$. Those composites are not included in the factorizations in Table 3. Table 1 also includes 55, while Table 2 includes the factorization representing 77, both of which have 11 as a factor. Since those composites are counted there, the first composite that has 11 as a factor that needs to be counted in one of the $E''_{m,5}$ components assigned to the prime factor p_5 is its square 121.

To give an example using (3), let us evaluate the twin prime count up to $x = 217$. For this we need to consider eliminations by $p_5 = 11$ and $p_6 = 13$, because 11^2 and 13^2 are less than 217. The square of 17 is 289, though, so we do not need to consider p_7, and since 13 • 17 = 221 we know that 169 is the only eliminating composite that can be attributed to p_6 in this interval. On the other side of its center post is the prime 167, so it can be thought of as a **single elimination**.

The eliminating composites less than 217 having 11 for their lowest prime factor are 121, 143, 187, and 209. The corresponding candidate center posts are 120, 144, 186, and 210. The factorization for 119 is listed in Table 3, and that is adjacent to 120, so it is a double elimination with p_4, and 144 and 186 are adjacent to 145 and 185, which would be counted as eliminations by p_3. The only one of the four candidates that is singly eliminated by p_5 is the center post at 210.

For $x = 217$, as can be deduced from the discussion above, $E''_{3,4}(217) = 2$; $E''_{3,5}(217) = 2$; and $E''_{4,5}(217) = 1$. There were no double eliminations involving p_6 so $E''_{m,6}(217) = 0$.

$$\pi_2(217) = 1 + (217 \setminus 6) - (14 + 8 + 4 + 1) + (2 + 2 + 1) = 1 + 36 - 27 + 5 = 15$$

Those fifteen, along with the next twin prime after 217, are listed in Table 4.

Table 4: The First Fifteen Twin Primes Up to 217, Plus One More

π_2	Center Post	Twin Prime
1	4	(3, 5)
2	6	(5, 7)
3	12	(11, 13)
4	18	(17, 19)
5	30	(29, 31)
6	42	(41, 43)
7	60	(59, 61)
8	72	(71, 73)

π_2	Center Post	Twin Prime
9	102	(101, 103)
10	108	(107, 109)
11	138	(137, 139)
12	150	(149, 151)
13	180	(179, 181)
14	192	(191, 193)
15	198	(197, 199)
16	228	(227, 229)

From the periodic primorial property presented in part 1, it follows that the pattern of double eliminations of twin prime candidates will repeat over intervals of the primorial of the higher of the two lowest prime factors involved in each of the eliminating composites.

4 Double Jeopardy: The Primorial Patterns of Double Eliminations

We previously noted that p_3 and p_4 both eliminate the twin prime candidates centered at 204 and 216. If we check every multiple of $p_4\# = 210$ thereafter we find a similar double elimination. 413 is 7 • 59 and 415 is 5 • 83, while $425 = 5^2 \cdot 13$ and 427 = 7 • 61, thus the center posts at 414 and 426 are doubly eliminated by p_3 and p_4. 623 = 7 • 89 and $625 = 5^4$, while 635 is 5 • 127 and 637 is $7^2 \cdot 13$, thus the center posts at 624 and 636 are doubly eliminated by p_3 and p_4 also. These results are summarized in Table 5.

Table 5: Double Eliminations by $p_3 = 5$ and $p_4 = 7$ within Intervals of $p_4\# = 210$

Center Post	Interval Member	Factorization	Residues (modulus)		
			(210)	(30)	(6)
210*n* + 204	210*n* + 203	7 • (30*n* + 29)	203		5
	210*n* + 205	5 • (42*n* + 41)		25	1
210(*n* + 1) + 6	210(*n* + 1) + 5	5 • (42(*n* + 1) + 1)		5	5
	210(*n* + 1) + 7	7 • (30(*n* + 1) + 1)	7		1

120 was a double elimination by p_4 and p_5, and 144 and 186 were double eliminations by p_3 and p_5. The primorial $p_5\# = 2310$. If we check by 2430, we see that 2429 = 7 • 347 while 2431 = 11 • 13 • 17. For 2454 and 2496, obviously to one side is a number divisible by 5, while the values on the other side, 2453 and 2497, both have 11 for their lowest prime factor. There are many more double eliminations for p_5 across intervals of its primorial. They are listed in Tables 6 and 7.

Table 6: Double Eliminations by $p_3 = 5$ and $p_5 = 11$ within Intervals of $p_5\# = 2310$

Center Post	Interval Member	Factorization	Residues (modulus)		
			(2310)	(30)	(6)
2310*n* + 144	2310*n* + 143	11 • (210*n* + 13)	143		5
	2310*n* + 145	5 • (462*n* + 29)		25	1
2310*n* + 186	2310*n* + 185	5 • (462*n* + 31)		5	5
	2310*n* + 187	11 • (210*n* + 17)	187		1
2310*n* + 474	2310*n* + 473	11 • (210*n* + 43)	473		5
	2310*n* + 475	5 • (462*n* + 95)		25	1
2310*n* + 516	2310*n* + 515	5 • (462*n* + 103)		5	5
	2310*n* + 517	11 • (210*n* + 47)	517		1
2310*n* + 804	2310*n* + 803	11 • (210*n* + 73)	803		5
	2310*n* + 805	5 • (462*n* + 161)		25	1
2310*n* + 1134	2310*n* + 1133	11 • (210*n* + 103)	1133		5
	2310*n* + 1135	5 • (462*n* + 227)		25	1
2310*n* + 1176	2310*n* + 1175	5 • (462*n* + 235)		5	5
	2310*n* + 1177	11 • (210*n* + 107)	1177		1
2310*n* + 1506	2310*n* + 1505	5 • (462*n* + 301)		5	5
	2310*n* + 1507	11 • (210*n* + 137)	1507		1
2310*n* + 1794	2310*n* + 1793	11 • (210*n* + 163)	1793		5
	2310*n* + 1795	5 • (462*n* + 359)		25	1
2310*n* + 1836	2310*n* + 1835	5 • (462*n* + 367)		5	5
	2310*n* + 1837	11 • (210*n* + 167)	1837		1
2310*n* + 2124	2310*n* + 2123	11 • (210*n* + 193)	2123		5
	2310*n* + 2125	5 • (462*n* + 425)		25	1
2310*n* + 2166	2310*n* + 2165	5 • (462*n* + 433)		5	5
	2310*n* + 2167	11 • (210*n* + 197)	2167		1

Table 7: Double Eliminations by $p_4 = 7$ and $p_5 = 11$ within Intervals of $p_5\# = 2310$

Center Post	Interval Member	Factorization	Residues (modulus)		
			(2310)	(210)	(6)
2310*n* + 120	2310*n* + 119	7 • (330*n* + 17)		119	5
	2310*n* + 121	11 • (210*n* + 11)	121		1
2310*n* + 342	2310*n* + 341	11 • (210*n* + 31)	341		5
	2310*n* + 343	7 • (330*n* + 49)		133	1
2310*n* + 582	2310*n* + 581	7 • (330*n* + 83)		161	5
	2310*n* + 583	11 • (210*n* + 53)	583		1
2310*n* + 1728	2310*n* + 1727	11 • (210*n* + 157)	1727		5
	2310*n* + 1729	7 • (330*n* + 247)		49	1
2310*n* + 1968	2310*n* + 1967	7 • (330*n* + 281)		77	5
	2310*n* + 1969	11 • (210*n* + 179)	1969		1
2310*n* + 2190	2310*n* + 2189	11 • (210*n* + 199)	2189		5
	2310*n* + 2191	7 • (330*n* + 313)		91	1

If we count up the entries in the three previous tables, we have 2 double eliminations by p_3 and p_4 within Intervals of $p_4\#$; 12 double eliminations by p_3 and p_5 within Intervals of $p_5\#$; and 6 double eliminations by p_4 and p_5 within intervals of $p_5\#$. Is there a pattern to those values? Three counts are not many to work with. Fortunately, since we are only concerned with tabulating the counts for a certain number of lowest prime factors, and not with completely factorizing every candidate pair, it is easy to construct a program that can quickly give us more such counts.

Table 8 shows the numerical results of just such a program which considered prime factors up to $p_{11} = 31$ within the interval to $\alpha_{11} + 2 = 200560490163$. It was necessary to go to $\alpha_n + 2$ for each count in case $\alpha_n + 1$ was a potential center post that was eliminated by $\alpha_n + 2$ being a composite of one of the factors under consideration. To shorten the notation, let $\alpha''_n = \alpha_n + 2$.

Table 8: Double Elimination Counts within Primorials, $E''_{m,n}(\alpha''_n)$,
With Lower Prime p_m Along the Left and Higher Prime p_n Across the Top

m,n		4	5	6	7	8	9	10	11
	p	7	11	13	17	19	23	29	31
3	5	2	12	120	1440	23040	414720	9123840	255467520
4	7		6	60	720	11520	207360	4561920	127733760
5	11			30	360	5760	103680	2280960	63866880
6	13				270	4320	77760	1710720	47900160
7	17					2970	53460	1176120	32931360
8	19						44550	980100	27442800
9	23							757350	21205800
10	31								15904350
	Total	2	18	210	2790	47610	901530	20591010	592452630

For $p_m = 5$ and $p_n = 23$, double elimination number 414720 occurs at $\alpha_9 + 1 = 223092894$.

For all entries in the Table 8 that have an entry to their left such that $3 < m < (n - 1)$, the following relationship holds:

$$E''_{m,n}(\alpha''_n) = (p_{n-1} - 1) \bullet E''_{m,n-1}(\alpha''_{n-1}) \qquad (4)$$

The logic for proving equation (4) would be the same as that used to prove equation (2), namely that only certain residues in the complete residue system of previous primorials are lined up in the correct positions necessary to produce an eliminating composite when the members of those previous primorials are multiplied by the next prime to initially generate the primorial interval corresponding to that next prime.

What about the entries in Table 8 that do not have a value to their left on which to base the next double elimination count? To find out, let us use E''_n to represent the total number of double eliminations that are already attributed to a lower prime $p_m < p_n$ within each $p_n\#$ primorial interval, and then E'_n to represent those eliminations (single or double) that can be attributed to p_n itself within each $p_n\#$ primorial interval.

$$E''_n = \Sigma E''_{m,n}(\alpha''_n) \qquad \{\text{for } m = 3 \text{ to } (n - 1) \qquad (5)$$

$$E'_n = \rho_n - E''_n \qquad (6)$$

For example, $E''_5 = E''_{3,5}(\alpha''_5) + E''_{4,5}(\alpha''_5) = 12 + 6 = 18$, which means that $E'_5 = 48 - 18 = 30$.

Once again a word of caution; E'_n does *not* represent the number of single eliminations. Some of the eliminations counted by E'_n may be double eliminations that are later attributed to a higher prime. That is, they may be included in a subsequent E''_q term for a prime $p_q > p_n$. For instance, 168 was a single elimination by p_6, as is $168 + p_6\#$, since 30197 is prime and $30199 = 13 \bullet 2323$. But $168 + 2p_6\#$ is a double elimination because $60227 = 229 \bullet 263$.

What E'_n does represent is the **unique twin prime elimination count**. It is the number of twin prime eliminations over intervals of the primorial $p_n\#$ that we get when we take out those that were double counted because the other half of the pair was eliminated by an earlier prime. For example, $E'_4 = 8 - 2 = 6$. The two that are subtracted out initially are the eliminations of 204 and 216, which are counted in the E'_3 component for p_3. One of those six that remain will be counted again in E''_5 (that being the elimination of 120) but then that over-count is corrected in the computation of E'_5. Table 9 shows the E'_n value for the first eleven primes. Since there are no eliminations by p_2 whatsoever we can say that $E''_2 = 0$ so that $E'_3 = \rho_3 = 2$.

Table 9: Unique Twin Prime Elimination Counts Over Primorial Intervals

n	3	4	5	6	7	8	9	10	11
p_n	5	7	11	13	17	19	23	29	31
ρ_n	2	8	48	480	5760	92160	1658880	36495360	1021870080
E''_n	0	2	18	210	2790	47610	901530	20591010	592452630
E'_n	2	6	30	270	2970	44550	757350	15904350	429417450

Notice that the E'_n values in Table 9 all match the values that appear in Table 8 as the double elimination count for that prime when paired with the next higher prime.

$$E''_{n-1,n}(\alpha''_n) = E'_{n-1} \quad (7)$$

When equation (7) is combined with those presented earlier, it is straightforward to derive that:

$$E'_n = (p_{n-1} - 2) \bullet E'_{n-1} \quad (8)$$

This relationship is clearly demonstrated in the last row in Table 9. The integer sequences formed by E''_n and by E'_n have been submitted to the On-Line Encyclopedia of Integer Sequences as sequences A121407 and A121406, respectively.

4 Double Vision: The Implications of Primorial Patterns on Twin Primes

Equation (8) allows us to calculate a **twin prime elimination ratio**, $(E'_n / p_n\#)$, for each prime factor. Starting with $(E'_3 / p_3\#) = 2 / 30$, we can easily calculate subsequent ratios.

$$(E'_n / p_n\#) = (E'_{n-1} / p_{n-1}\#) \bullet (p_{n-1} - 2) / p_n \quad (9)$$

The first several values of the ratios formed by equation (9) are listed in Table 10.

Table 10: Twin Prime Elimination to Primorial Ratios and Ratio Summation

n	p_n	E'_n	$p_n\#$	$\approx (E'_n / p_n\#)$	$\approx \Sigma (E'_n / p_n\#)$
3	5	2	30	0.06666667	0.06666667
4	7	6	210	0.02857143	0.09523810
5	11	30	2310	0.01298701	0.10822511
6	13	270	30030	0.00899101	0.11721612
7	17	2970	510510	0.00581771	0.12303383
8	19	44550	9699690	0.00459293	0.12762676
9	23	757350	223092870	0.00339477	0.13102153
10	29	15904350	6469693230	0.00245829	0.13347982
11	31	429417450	200560490130	0.00214109	0.13562091
12	37	12453106050	7420738134810	0.00167815	0.13729905
13	41	435858711750	304250263527210	0.00143257	0.13873162
14	43	16998489758250	13082761331670030	0.00129930	0.14003093
15	47	696938080088250	614889782588491410	0.00113344	0.14116436
16	53	31362213603971250	32589158477190044730	0.00096235	0.14212671

One application for these ratios is to estimate values of the twin prime counting function, $\pi_2(x)$. For $x \geq 4$ and $n \geq 3$ such that $p_n < x^{1/2}$ the estimation is given by:

$$\pi_2(x) \approx 1 + (x / 6) - \Sigma ((x - p_n) \bullet (E'_n / p_n\#)) \quad (10)$$

For $\pi_2(217)$, that value would be would be calculated as:

$$1 + (217 / 6) - 212 \cdot 2 / 30 - 210 \cdot 6 / 210 - 206 \cdot 30 / 2310 - 204 \cdot 270 / 30030$$

The result is $\pi_2(217) \approx 12.5$, which is about a 16.7% error from the correct count of 15 that we found earlier. Table 11 shows the results from equation (10) for powers of ten up to 10^{16}. The values were calculated using the primes up to 10^8 as published by Caldwell [4].

Table 11: Twin Prime Counting Function Estimates Up to 10^{16}

x	$\pi_2(x)$ Estimated	$\pi_2(x)$ Actual[a]	% Error
10^1	2.7	2	35.0
10^2	8.7	8	8.8
10^3	33.2	35	–5.1
10^4	194.4	205	–5.2
10^5	1234.6	1224	0.9
10^6	8662.4	8169	6.0
10^7	63973.6	58980	8.5
10^8	489460.4	440312	11.2
10^9	3873936.6	3424506	13.1
10^{10}	31382177.5	27412679	14.5
10^{11}	259468905.1	224376048	15.6
10^{12}	2180467679.1	1870585220	16.6
10^{13}	18579615197.2	15834664872	17.3
10^{14}	160206996577.1	135780321665	18.0
10^{15}	1395577415965.0	1177209242304	18.5
10^{16}	12265983873368.7	10304195697298	19.0

a. Source: Sloane's A007508; Ribenboim 1996, p. 263; Nicely 1998, 1999; Sebah 2002; as referenced in Weisstein, Eric W. "Twin Primes." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/TwinPrimes.html

Notice the estimation undercounts for 10^3 and 10^4 and over-counts for all of the other entries, with the percent over-count increasing for the estimates after 10^4. The estimates vary because the distribution of the double eliminations is not regular within the primorial intervals. In the case of the double eliminations by p_3 and p_4, for instance, both double eliminations come near the end of each $p_4\#$ primorial interval. However, there is no reason to believe that the percent error will always continue to increase past 10^{16}. The rate of increase in the error seems to be stabilizing towards the end of Table 11, and it may be the case that it will decrease again at some point.

That topic will be left for later research, for what is even more significant than using the individual ratios to estimate $\pi_2(x)$ is that we can take a summation of the $(E'_n / p_n\#)$ ratios. Let us call this value E_2, for the **twin prime elimination constant**, where for $n = 3$ to ∞ its limit is:

$$E_2 = \Sigma\, (E'_n / p_n\#) \qquad (11)$$

Expansion of equation (11) gives the following:

$$E_2 = (2 / 30) + (6 / 210) + (30 / 2310) + (270 / 30030) + (2970 / 510510) + \ldots$$

The first few of those partial sums were also shown in Table 10. The individual ratios are getting smaller, so it appears that the summation will converge to a definite value. But what value?

Thoerem 1: The value of the twin prime elimination constant E_2 as given by equation (11) must converge to a value of 1 / 6 from below, such that it never exceeds 1 / 6.

*Proof***:** The numerators of each term in E_2 are the unique twin prime elimination counts over primorial intervals, such that all double eliminations of twin prime candidates are counted as one instance of a twin prime elimination. Since the ratio of the total number of candidates is 1 / 6, their total unique elimination ratio could never exceed 1 / 6, because that would imply that more candidates had been eliminated than actually existed.

But by the prime number theorem, the density of the prime numbers tends to thin out as they go off to infinity. This means that the proportion of composites to primes is always increasing. This increase can be seen back in Table 1, where the summation of the composite counts will continue to grow such that the quotient of the summation over the primorial will forever keep getting closer and closer to 1.0. Likewise the proportion of composites that eliminate twin primes will forever continue to increase and approach 1 / 6 as the primes go off to infinity.

Q.E.D.

The numerical evidence presented in Table 12 supports the assertion of Theorem 1.

Derbyshire [5] discusses how Euler showed that $\zeta(2)$ converges to $\pi^2 / 6$, where $\zeta(x)$ is the Riemann Zeta Function and the π in the numerator is the familiar constant ≈ 3.1415927 and does not represent a prime number counting function.

$$\zeta(2) = 1 + (1/2^2) + (1/3^2) + (1/4^2) + (1/5^2) + (1/6^2) + (1/7^2) + \ldots \qquad (12)$$

Since $\zeta(2) = \pi^2 / 6$ it follows that $\zeta(2) / \pi^2 = 1 / 6$. Table 12 divides each term in the $\zeta(2)$ summation by π^2 in order to compare them to the individual terms for E_2 and it shows a comparison of their partial sums for a select group of terms.

The first term in $\zeta(2) / \pi^2$, for example, is $\pi^{-2} \approx 0.101321183642$. That is much larger than the first term in E_2, which is (2 / 30). But by the second term the former is already smaller than the latter: $(1 / 4\pi^2) \approx 0.025330295911$ which is less than $(6 / 210) \approx 0.028571428571$. Each of the successive terms shown for E_2 is larger than its counterpart in $\zeta(2) / \pi^2$ as well.

However, the head start given by that larger first term has $\zeta(2) / \pi^2$ converging towards 1 / 6 much more rapidly than E_2 is. As shown at the end of Table 12, the partial sums for latter have not caught up to the former even as far out as nearly 15 million terms. These values were again computed using the first 15 million primes up to 275604541 as given by Caldwell [4].

Table 12: Comparison of the Convergence of E_2 to $\zeta(2) / \pi^2$

$n+2$	$p_{(n+2)}$	E_2		$\zeta(2) / \pi^2$	
		n-th Term	*n*-th Partial Sum	*n*-th Term	*n*-th Partial Sum
3	5	0.066666666667	0.066666666667	0.101321183642	0.101321183642
4	7	0.028571428571	0.095238095238	0.025330295911	0.126651479553
5	11	0.012987012987	0.108225108225	0.011257909294	0.137909388847
6	13	0.008991008991	0.117216117216	0.006332573978	0.144241962824
7	17	0.005817711700	0.123033828916	0.004052847346	0.148294810170
8	19	0.004592930290	0.127626759206	0.002814477323	0.151109287493
9	23	0.003394774562	0.131021533768	0.002067779258	0.153177066751
10	29	0.002458285028	0.133479818795	0.001583143494	0.154760210246
11	31	0.002141086959	0.135620905755	0.001250878810	0.156011089056
12	37	0.001678149238	0.137299054993	0.001013211836	0.157024300892
13	41	0.001432566423	0.138731621416	0.000837365154	0.157861666047
14	43	0.001299304430	0.140030925846	0.000703619331	0.158565285377
15	47	0.001133435780	0.141164361626	0.000599533631	0.159164819008
16	53	0.000962351134	0.142126712760	0.000516944815	0.159681763823
…					
25	97	0.000403126748	0.147518146114	0.000191533428	0.162355777128
26	101	0.000379178625	0.147897324739	0.000175904833	0.162531681960
…					
168	997	0.000017399629	0.158010351238	0.000003676919	0.166058132858
169	1009	0.000017158207	0.158027509445	0.000003633016	0.166061765874
…					
1229	9973	0.000000981730	0.161772250511	0.000000067299	0.166584123955
1230	10007	0.000000978198	0.161773228710	0.000000067190	0.166584191145
…					
9592	99991	0.000000062771	0.163528456887	0.000000001102	0.166656101922
9593	100003	0.000000062762	0.163528519650	0.000000001101	0.166656103023
…					
78498	999983	0.000000004361	0.164486199406	0.000000000016	0.166665375893
78499	1000003	0.000000004361	0.164486203767	0.000000000016	0.166665375909

$n+2$	$p_{(n+2)}$	E_2		$\zeta(2) / \pi^2$	
		n-th Term	*n*-th Partial Sum	*n*-th Term	*n*-th Partial Sum
664579	9999991	3.2041434E–10	0.165064596726	2.2940857E–13	0.166666514207
664580	10000019	3.2041338E–10	0.165064597046	2.2940788E–13	0.166666514207
…					
5761455	99999989	2.4531971E–11	0.165440068281	3.0523632E–15	0.166666649081
5761456	100000007	2.4531966E–11	0.165440068305	3.0523621E–15	0.166666649081
…					
14999999	275604533	7.9967276E–12	0.165564699483	4.5031655E–16	0.166666659912
15000000	275604541	7.9967273E–12	0.165564699491	4.5031649E–16	0.166666659912

5 Double the Pleasure, Double the Fun: Locating Twin Primes

How can these primorial patterns help in locating twin primes? A good place to look for twin primes is in the interval between the first atomic boundary for that prime and the next composite that corresponds to its square, that is, between $p_n\# + p_n + 2$ and $p_n\# + p_n^2 – 2$. The +/– 2 is specified so as to leave out the adjacent values, since p_n would eliminate them. But p_n is not capable of eliminating any other candidates in that interval.

The numerical results for this interval look promising, at least in the beginning. The twin prime (41, 43) falls between 37 and 53 while (227, 229) and (239, 241) are both between 219 and 257. The twin primes (2339, 2341) and (2381, 2383) are between 2323 and 2429, and there are two twins between 30045 and 30197 as well: (30089, 30091) and (30137, 30139). And if we check the interval from 510529 to 510797, we find a total of five twin primes: (510551, 510553), (510581, 510583), (510611, 510613), (510617, 510619), and (510707, 510709).

Because the pattern of the composites having prime factors lower than p_n repeats, what happens is that all of the twin primes that exist in the interval from $p_n + 2$ to $p_n^2 – 2$ get through and their corresponding candidates survive past p_n in the interval from $p_n\# + p_n + 2$ to $p_n\# + p_n^2 – 2$. If those corresponding candidates are to be eliminated at all, it must be by a prime factor greater than p_n.

There are three twin primes (9699731, 9699733), (9699887, 9699889), and (9699917, 9699919) between 9699711 and 9700049. They correspond to 19# + (41, 43); 19# + (197, 199); and 19# + (227, 229). Unfortunately, however, there are no twin primes between 223092895 and 223093397. Of the 83 candidates total in that interval, 62 are eliminated by primes less than p_9 = 23. The other 21 are eliminated by various primes greater than 23.

But just because there are no twin primes in a particular interval does not mean that the trend completely stops there. If we look in the interval from 6469693261 to 6469694069 above 29#, we find five twin primes again: (6469693331, 6469693333), (6469693511, 6469693513), (6469693661, 6469693663), (6469694039, 6469694041), and (6469694057, 6469694059).

Higher multiples of those high density base primorials do well, too. There are four twin primes between $2 \bullet 29\# + 29 + 2 = 12939386491$ and $2 \bullet 29\# + 29^2 – 2 = 12939387299$. Those four correspond to 2 • 29# plus the pairs (227, 229); (311, 313); (599, 601); and (641, 643).

So while no theorem will be offered here as to where twin primes might always be found, the evidence points to the intervals between $N \bullet p_n\# + p_n + 2$ and $N \bullet p_n\# + p_n^2 – 2$ as being favorable.

Finally, consider that there is one twin prime between $37\# + 37 + 2$ and $37\# + 37^2 – 2$, that being (7420738134911, 7420738134913). It corresponds to 37# + (101, 103). The other corresponding candidates in this and similar such intervals are eliminated by primes larger than the prime of the base primorial, and in some cases much larger.

For example, in the twin prime candidate that corresponds to 37# + (599, 601), the first half of the pair, the value 7420738135409, is prime, while the other, 7420738135411, is a product of

2459383 • 3017317. That such eliminations exist further validates our infinite series. While the elimination ratio for the prime 2459383 is only about $1.563 \cdot 10^{-9}$, that is still not zero, and because that ratio can never be zero, it leads us to another theorem.

Theorem 2: It would take infinitely many primes to eliminate all of the twin prime candidates greater than any particular value.

Proof: As given by equation (11), we have an infinite series over the primes p_3 and greater that represents the rate of elimination of twin prime candidates. As given by equation (9), each term in that infinite series is generated by multiplying the previous term by $(p_{n-1} - 2) / p_n$. Since the previous prime p_{n-1} is greater than 2 for all $n > 3$, the multiplier $((p_{n-1} - 2) / p_n)$ is always greater than zero, and thus each and every term in the infinite series in non-zero.

It can never be the case that all of the twin prime candidates past a particular value are entirely eliminated by only prime factors lower than a certain prime. If that was the case, then those eliminating composites would repeat in subsequent primorial intervals such that the elimination ratio for larger primes would have to be zero, and that would be a contradiction.

Not only would it take infinitely many primes to eliminate all twin prime candidates from a particular value onward, but it would take all of the primes 5 and greater to eliminate all of the candidates from any point out to infinity.

Q.E.D.

5 Open Questions and Conclusions

What does theorem 2 imply about the infinitude of the twin primes themselves? Let us assume that the number of twin primes is finite. That would mean that there is a last or maximum twin prime, $(p_{\max} - 2, p_{\max})$. If that is true, then there must be a prime number p_e whose square, the first eliminating composite that it can generate, is greater than that last twin prime, i.e., $p_e^2 > p_{\max}$. There cannot be any twin primes in the primorial interval from p_e^2 to $p_e^2 + p_e\#$.

There would be an integer number of primorial intervals for each of the primes up to p_e within an interval of $p_e\#$. Therefore the elimination rate for the primes less than or equal to p_e within that interval would precisely match the summation of E_2 up to and including the prime p_e. The elimination rate for the primes greater than p_e within that interval would have to be (1/6) minus the elimination rate of the primes up to and including the prime p_e.

But that would match the elimination rate for all of the primes greater than p_e out to infinity. There are only a finite number of primes greater than p_e, though, which could be the lowest prime factor of a composite within that interval. That is, only the primes up to a $p_i \leq (p_e^2 + p_e\#)^{1/2}$ could account for eliminations of twin prime candidates between p_e^2 and $p_e^2 + p_e\#$. Now, the pattern for those composites would not repeat within intervals of $p_e\#$; it would repeat within intervals of the larger primorials within $p_i\# > p_e\#$. So while the elimination rate for those primes would tend toward the summation of their individual $(E'_n / p_n\#)$ ratios, the actual rate could increase in more narrow intervals. But then by the time a primorial interval from p_i^2 to $p_i^2 + p_i\#$ was completed, the elimination rate for them within that $p_i\#$ primorial interval would precisely match the summation of their individual ratios, and now the primes from p_i up to $(p_i^2 + p_i\#)^{1/2}$ would have to exceed their primorial ratios to eliminate the remaining candidates in this larger interval.

Is it possible for the twin prime elimination rate of a finite number of lowest prime factors to forever exceed their primorial elimination ratios within earlier and smaller primorials in this way? It also seems interesting to note that the number of unique twin prime eliminations for each prime is always an even number over its primorial, whereas the total number of twin prime candidates in any primorial of $p_3\#$ or larger is always odd. Could a finite number of prime factors always produce an odd number of twin prime eliminations exactly as needed from a certain point forever onward out to infinity so as to make the number of twin primes finite?

While it seems highly unlikely that could be the case, an actual proof has been elusive. Shanks [6] stated that the evidence for the twin prime conjecture was overwhelming. At the very least, the infinite series introduced here adds a strong argument to that arsenal. But the primorial patterns presented here also have the potential to be applied to other prime gaps, and to sieving in general. Perhaps they may eventually lead to other useful insights regarding the distribution of prime numbers.

7 Acknowledgments

I am grateful to the following individuals for their assistance and advice: Dr. Peter Gibson, The University of Alabama in Huntsville; Dr. Charles Ryavec and Dr. Jeffrey Stopple, University of California, Santa Barbara; Kaustuv M. Das and Joseph Paunovich of DWT. And finally a special thanks to DP Technology Corp. and its management for their support and encouragement.

References:

1. Dickson, Leonard E. “History of the Theory of Numbers Volume I: Divisibility and Primality”. Mineola, NY: Dover Publications, Inc., 2005.

2. Weisstein, Eric W. "Twin Peaks." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/TwinPeaks.html

3. Sloane, Neil J. A. Sequence A005867 in “The On-Line Encyclopedia of Integer Sequences.” http://www.research.att.com/~njas/sequences/A005867

4. Caldwell, Chris K. “The Prime Pages”, http://primes.utm.edu/.

5. Derbyshire, J. *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics.* New York: Penguin, 2004, p. 64.

6. Shanks, D. *Solved and Unsolved Problems in Number Theory, 4th ed.* New York: Chelsea, 1993, p. 219.