Proof of Existence of Infinite Twin Primes

Nikhil Datar

nsdatar@gmail.com

ABSTRACT

This submission gives a proof of existence of infinite twin primes. The basic principle used here is of Fundamental Principle of Counting and Excel has been as a major tool for explanation.

KEYWORDS
Twin Primes

INTRODUCTION

Twin Primes are pairs of prime numbers which have a difference of 2. For instance (3,5), (5,7), (11,13), (17,19) etc. are examples of twin primes. In this paper the first number of the pair will be referred to as a twin prime.

MAIN

1) The Table of Remainders

Let's first consider a table where the columns will represent the divisors (denominators), the rows will represent the dividends (numerators) while the corresponding cell values will represent the *remainders* of its row ÷ column. Let's call this table the Table of Remainders.

Lets start by looking at the following Table of Remainders:

TABLE A: Table of Remainders

	Denominators					
Table A	2	3	5	7	9	11
Numerators: 1	1	1	1	1	1	1
2	0	2	2	2	2	2
3	1	0	3	3	3	3
4	0	1	4	4	4	4
5	1	2	0	5	5	5
6	0	0	1	6	6	6
7	1	1	2	0	7	7
8	0	2	3	1	8	8
9	1	0	4	2	0	9
10	0	1	0	3	1	10
11	1	2	1	4	2	0
12	0	0	2	5	3	1

13	1	1	3	6	4	2
14	0	2	4	0	5	3
15	1	0	0	1	6	4
16	0	1	1	2	7	5
17	1	2	2	3	8	6
18	0	0	3	4	0	7
19	1	1	4	5	1	8
20	0	2	0	6	2	9
21	1	0	1	0	3	10
22	0	1	2	1	4	0
23	1	2	3	2	5	1
24	0	0	4	3	6	2
25	1	1	0	4	7	3
26	0	2	1	5	8	4
27	1	0	2	6	0	5
28	0	1	3	0	1	6
29	1	2	4	1	2	7
30	0	0	0	2	3	8
31	1	1	1	3	4	9
32	0	2	2	4	5	10
33	1	0	3	5	6	0
34	0	1	4	6	7	1
35	1	2	0	0	8	2
36	0	0	1	1	0	3
37	1	1	2	2	1	4
38	0	2	3	3	2	5
39	1	0	4	4	3	6
40	0	1	0	5	4	7
41	1	2	1	6	5	8
42	0	0	2	0	6	9
43	1	1	3	1	7	10
44	0	2	4	2	8	0
45	1	0	0	3	0	1
46	0	1	1	4	1	2
47	1	2	2	5	2	3
48	0	0	3	6	3	4
49	1	1	4	0	4	5
50	0	2	0	1	5	6
51	1	0	1	2	6	7
52	0	1	2	3	7	8
53	1	2	3	4	8	9
54	0	0	4	5	0	10
55	1	1	0	6	1	0
56	0	2	1	0	2	1
57	1	0	2	1	3	2

58	0	1	3	2	4	3
59	1	2	4	3	5	4
60	0	0	0	4	6	5
61	1	1	1	5	7	6
62	0	2	2	6	8	7
63	1	0	3	0	0	8
64	0	1	4	1	1	9
65	1	2	0	2	2	10
66	0	0	1	3	3	0
67	1	1	2	4	4	1
68	0	2	3	5	5	2
69	1	0	4	6	6	3
70	0	1	0	0	7	4
71	1	2	1	1	8	5
72	0	0	2	2	0	6
73	1	1	3	3	1	7
74	0	2	4	4	2	8
75	1	0	0	5	3	9
76	0	1	1	6	4	10
77	1	2	2	0	5	0
78	0	0	3	1	6	1
79	1	1	4	2	7	2
80	0	2	0	3	8	3
81	1	0	1	4	0	4
82	0	1	2	5	1	5
83	1	2	3	6	2	6
84	0	0	4	0	3	7
85	1	1	0	1	4	8
86	0	2	1	2	5	9
87	1	0	2	3	6	10
88	0	1	3	4	7	0
89	1	2	4	5	8	1
90	0	0	0	6	0	2
91	1	1	1	0	1	3
92	0	2	2	1	2	4
93	1	0	3	2	3	5
94	0	1	4	3	4	6
95	1	2	0	4	5	7
96	0	0	1	5	6	8
97	1	1	2	6	7	9
98	0	2	3	0	8	10
99	1	0	4	1	0	0
100	0	1	0	2	1	1
101	1	2	1	3	2	2
102	0	0	2	4	3	3

103	1	1	3	5	4	4
104	0	2	4	6	5	5
105	1	0	0	0	6	6
106	0	1	1	1	7	7
107	1	2	2	2	8	8
108	0	0	3	3	0	9
109	1	1	4	4	1	10
110	0	2	0	5	2	0
111	1	0	1	6	3	1
112	0	1	2	0	4	2
113	1	2	3	1	5	3
114	0	0	4	2	6	4
115	1	1	0	3	7	5
116	0	2	1	4	8	6
117	1	0	2	5	0	7
118	0	1	3	6	1	8
119	1	2	4	0	2	9
120	0	0	0	1	3	10
121	1	1	1	2	4	0

In Table A, the columns will assume values from 2 and then all odd numbers# starting from 3.
Let the last column in Table A, the Table of Remainders, be represented by 'n' i.e. here let n=11. The rows will run from 1 to n^2 i.e. from 1 to 121.

<u>Terminologies:</u>
Here, Column 2 would mean either Value 2 or Column with Header Value 2, while, Column 5 would mean either Value 5 or Column with Header Value 5 depending on the context its used.
Similarly for Rows, Row 5 would mean either Value 5 or Row with Header Value 5, while, Row 10 would mean either Value 10 or Row with Header Value 10 depending on the context its used.
The Cells would be represented by Cell (row, column) i.e. Cell (7,3) would mean Cell representing Row 7 and Column 3.

It is Given that in this Table of Remainders, the values in all columns are:

i) Periodic (i.e. moving in continuous cycles)
ii) In Same Order (i.e. their order will never change)
iii) With Frequency of the Column Header Value.

<u>2) Finding Twin Prime Numbers in the Table</u>:

The idea is to eliminate all such rows (r) which are - either themselves not prime or the second row after them (r+2) is not a prime (since twin primes have a gap of 2). All the rows that remain are then twin primes. Let's see how it shall be done.

Note: We will consider rows only up to n^2= 121 for reasons that will be explained in a bit.

a) <u>Eliminate rows which are not themselves prime</u>:

If at least any one cell of a row is zero i.e. remainder is zero, then such row is not a prime and hence cannot be a twin prime. We delete such rows (here by using excel filter).
*We shall call this the '**Rule of Zero**' for easy reference later.*

b) Eliminate rows where the second row after it is not a prime:
It means if we a considering row 'r' where row 'r+2' is not a prime then we delete row 'r' since it cannot be twin prime. We determine this by checking for each column of that row individually. If in a 'column c' there is a cell with value c-2 (c being the column value), for a row 'r', then it means that the second row after 'r' (i.e. row 'r+2') is a multiple of 'c' and hence 'r' cannot be twin prime since r+2 is not a prime.

For instance, for Row 23 and Column 5 the value in the Cell (23,5) is '3' that is two less than its column value 5 (C-2 i.e. 5-2). This indicates that the next multiple of Column 5 is just two rows away from the said Row 23 i.e. Row 25 and hence Row 23 can't be a twin prime and is eliminated.

Similarly, for Row 13 and Column 3, the value in Cell (13, 3) is 1 which is two short of Column 3 (C-2) and hence the next multiple of Column 3 is just two away from the said Row 13 i.e. Row 15 and hence 13 cannot be a twin prime and hence is eliminated. Also, for Row 13, Cell (13,5) i.e. 5th column has value 3 which is 2 less than Column Value 5 and is eliminated from being a twin prime this way too.

However, for Row 17, none of its cells have values that are two less than their corresponding column value (and also none of the cells are zero) and hence 17 is a twin prime i.e. (17,19).

Thus, for each row, apart from checking if the values in any of its cells is zero or not, we will also check if any of the rows' cells has value C-2 where 'C' is the corresponding column value and delete if it is.
*We shall call this '**The Rule of C-2**' for easy reference later.*

A summary of values for 'The Rule of C-2' is given below for each column.

For Column Value 2: Eliminate rows where the cell value is 2-2 i.e. 0 (which is already excluded)
For Column Value 3: Eliminate rows where the cell value is 3-2 i.e. 1
For Column Value 5: Eliminate rows where the cell value is 5-2 i.e. 3
For Column Value 7: Eliminate rows where the cell value is 7-2 i.e. 5
For Column Value 9: Eliminate rows where the cell value is 9-2 i.e. 7
For Column Value 11: Eliminate rows where the cell value is 11-2 i.e. 9.

So if we apply the 'Rule of Zero' and 'Rule of C-2' to Table A i.e. the Table of Remainders we get the following Table C :

TABLE C: Table of Twin Primes

Table C	2	3	5	7	9	11
17	1	2	2	3	8	6
29	1	2	4	1	2	7
41	1	2	1	6	5	8
59	1	2	4	3	5	4
71	1	2	1	1	8	5
101	1	2	1	3	2	2
107	1	2	2	2	8	8

In Table C, we get 17(19), 29(31), 41(43), 59(61), 71(73), 101(103) and 107(109) as twin primes.

3) 'Desired Ratios' in Each Column and Proof of its Accuracy

We will first define what are the 'Desired Ratios'.
We have following number of possible remainders in each column:
Column 2 = 2 (0, 1)
Column 3 = 3 (0, 1, 2)
Column 5 = 5 (0,1,2,3,4)
Column 7 = 7 (0,1,2,3,4,5,6)
Column 9 = 9 (0,1,2,3,4,5,6,7,8)
Column 11 = 11 (0,1,2,3,4,5,6,7,8,9,10)
(for our purpose we are including 0 as a remainder)

By Fundamental Principle of Counting, the total number of combinations of all the remainders above would be 11*9*7*5*3*2. These combinations are nothing but total number of rows possible.

Further, after eliminating Rows by 'Rule of Zero' and 'Rule of C-2', the 'Allowed Values' i.e. Allowed Remainders (i.e. remainders apart from 0 and C-2) in each column would be as follows:

Column 2 = 1 (Value 1)
Column 3 = 1 (Value 2)
Column 5 = 3 (Values 1,2,4)
Column 7 = 5 (Values 1,2,3,4,6)
Column 9 = 7 (Values 1,2,3,4,5,6,8)
Column 11 = 9 (Values 1,2,3,4,5,6,7,8,10)
Thus, again by Fundamental Principle of Counting the combination of Allowed Values will be 1*1*3*5*7*9.

Thus,

$$\frac{\text{Total Rows After Elimination}}{\text{Total Rows Before Elimination}} = \frac{1*1*3*5*7*9}{2*3*5*7*9*11} = \frac{1}{2}*\frac{1}{3}*\frac{3}{5}*\frac{5}{7}*\frac{7}{9}*\frac{9}{11}$$

From the above table we can see that the ratios of allowed values to total values in each column =

$$\frac{1}{2},\frac{1}{3},\frac{3}{5},\frac{5}{7},\frac{7}{9},\frac{9}{11}\ respectively.$$

We shall call these ratios as the 'Desired Ratios' of each column. These 'Desired Ratios' in 'All Columns' will be the cornerstone of this paper.
Thus, if we have so many Number of Rows which will ensure that these 'Desired Ratios' exist in 'All Columns' then we can say that the Rows left after Elimination (which we shall see later comprises of 'Twin Primes' and 'Possible Twin Primes') will be as per the above formula i.e.:

$$Rows\ After\ Elimination = Total\ Rows * \frac{1}{2}*\frac{1}{3}*\frac{3}{5}*\frac{5}{7}*\frac{7}{9}*\frac{9}{11}$$

Now in order to determine and maintain these desired ratios in all respective columns we would need to know the number of occurrences of each of these values i.e. remainders in each column and we have to prove that the number of occurrences of these values is same in all respective columns. Note, that this criterion of equal number of occurrences of values in all respective columns is crucial which we shall see later ##.

As mentioned earlier, the values (i.e. remainders) in the Table of Remainders will always be periodic and in the same order though with different frequencies for each column as it runs in cycles repeatedly. These values i.e. remainders will be from 0 to C-1, where 'C' is the Column Value. For instance, for Column 7, the remainders will be 0,1,2,3,4,5,6 i.e. from 0 to 6(7-1) and will be periodic and in the same order. As mentioned above, it will be crucial for our purpose to find the number of occurrences of each of these values viz. 0,1,2,3,4,5,6 in column 7 and similarly for all other columns.

We will first prove that the number of occurrences of these values i.e. remainders in any given respective Column C will tend to be same

Now, If we take n^2 rows and populate the table cells by dividing each row with each column, then there will be certain number of full cycles completed in each column (which will be represented by Quotient (Q)). Also, except Column 'n' (and its divisors, if any) all other columns will have remainders i.e. extra values (after completing Q cycles). Lets look at the following tables for that.

Let the number of rows be n^2. Since n= 11 here, we have 11^2 i.e. 121 rows.
We will divide n^2 i.e. 121 by each of the Column and find the Quotients(Q) and Remainder

TABLE D

Column	Quotient	Formula		Remainder
Column 11	Q11 = 11	(n^2)/ n	121/11	0
Column 9	Q9 = 13	(n^2)/ (n-2)	121/9	4
Column 7	Q7 = 17	(n^2)/ (n-4)	121/7	2
Column 5	Q5 = 24	(n^2)/ (n-6)	121/5	1
Column 3	Q3 = 40	(n^2)/ (n-8)	121/3	1
Column 2	Q2 = 60	(n^2)/ (n-9) till n-(n-2)	121/2	1

This means that since only Column 11 has remainder as 0 all the occurrences of each of the values i.e. remainders (0 to 10) in Column 11 will be same.
Lets call values 0 and C-2 in each column as 'Disallowed Values' and all other values as 'Allowed Values'. Then we can say that the ratio of occurrences of 'Allowed Values' to Total Values is 9 /11 in Column 11 which is the Desired Ratio.
For other columns viz. 2, 3, 5, 7 and 9, the ratio of 'Allowed Values' to 'Total Values' will be close to

$$\frac{1}{2},\frac{1}{3},\frac{3}{5},\frac{5}{7}\ and\ \ \frac{7}{9}\ respectively.$$

These '**Desired Ratios'** will be favorable if the Allowed Values (i.e. the numerator) get more (by virtue of remainders) and will become unfavorable otherwise. Though these ratios are not exact, we will prove below that these ratios tend to become exact under certain conditions.

Just to confirm once again, we look at the extract of Table A below:

Table A	2	3	5	7	9	11
1	1	1	1	1	1	1

……

……

117	1	0	2	5	0	7
118	0	1	3	6	1	8
119	1	2	4	0	2	9
120	0	0	0	1	3	10
121	1	1	1	2	4	0

As we can see in the extract of Table A above i.e. Table of Remainders, all columns started with Remainders 1 and after finishing their respective cycles ended at 0.

Row 117 is where the cycle of Column 9 ended (corresponding cell being the last zero of column 9) and 121-117 = 4 is the remainder as we confirmed above.
Likewise, we can check for all columns to confirm.

Note that for each Column 'C' we can have at most C-1 remainders i.e. at most C-1 extra values after finishing respective column's Q full cycles..
However, If these remainders happen to be 'Disallowed Values' either 0, or C-2 for each column, then the above ratios of Allowed Values to Total Values will become Unfavorable i.e. less than the Desired Ratio., whereas if the remainders are other than 0 or C-2, then the ratios will become favorable.

Lets assume only the worst case scenario where the ratios become unfavorable such that, at n^2 rows, there are exactly two remainders viz. 0 and C-2 in each column (other than Column 'n') making the above ratios most unfavorable for each column.

Thus, now the ratios in respective columns will be:
(1*Q2) / ((2*Q2) + 2) instead of the desired ratio 1/2
(1*Q3) / ((3*Q3) + 2) instead of the desired ratio 1/3
(3*Q5) / ((5*Q5) + 2) instead of the desired ratio 3/5
(5*Q7) / ((7*Q7) + 2) instead of the desired ratio 5/7
(7*Q9) / ((9*Q9) + 2) instead of the desired ratio 7/9
(9*Q11) / ((11*Q11) + 2) instead of the desired ratio 9/11.

Now as 'n' tends to infinity, then from Table D, even Q2, Q3, Q5, Q7, Q9 and Q11 will also tend to infinity (because of the formula) which will in turn further lead to:

(1*Q2) / ((2*Q2) + 2) will tend to 1/2
(1*Q3) / ((3*Q3) + 2) will tend to 1/3
(3*Q5) / ((5*Q5) + 2) will tend to 3/5
(5*Q7) / ((7*Q7) + 2) will tend to 5/7
(7*Q9) / ((9*Q9) + 2) will tend to 7/9
(9*Q11) / ((11*Q11) + 2) will tend to 9/11.

Thus, these ratios in All columns will tend to Accuracy i.e. to Desired Ratios of 1/2, 1/3, 3/5, 5/7, 7/9, 9/ 11 (i.e. up to (n-2)/n) as 'n' tends to infinity.

4) Proof of Infinite Twin Primes

Now, given an Initial Number of Rows (R), which as seen above should be large enough, when we apply the above ratios to R to filter out rows as per 'Rule of Zero' and 'Rule of C-2' for every single column, we would get the following number of rows left: (we will be using R=121 in our case)

$$\text{No. of Rows Left in Table of Remainders after Elimination} = R * \frac{1}{2} * \frac{1}{3} * \frac{3}{5} * \frac{5}{7} * \frac{7}{9} * \frac{9}{11}$$

LHS above which we shall call "Rows Left" can be categorized as below:

a) 'Twin Primes' (up to rows n^2)* and
b) 'Possible Twin Primes' (after row n^2)

*Since we have columns only till 'n' hence beyond n^2 we could not check if the rows are prime or not and hence we can determine if a row is prime or not only till rows n^2.)

$$\text{Since n} = 11, \text{'Rows Left'} = R * \frac{1}{2} * \frac{1}{\text{n}} = \frac{R}{2n}$$

Now, in order to find 'Twin Primes' we set R = n^2, and we would get:

$$\text{Twin Primes (till } n^2) = n^2 * \frac{1}{2} * \frac{1}{\text{n}} = \frac{n^2}{2n} = \frac{n}{2}$$

Thus, up to rows n^2, the number of twin primes = n/2.
As n -> infinity, n/2 -> infinity thereby proving that there are infinite twin primes.

Notes:

 # Inclusion of all odd numbers in Columns even redundant ones like Column 9

Column 9 is a redundant column (on account of inclusion of Column 3) and hence 7/9 is also redundant however we still use it for the following reasons:

1) Including column 9 and hence 7/9, the expression becomes

1/2 * 1/3 * 3/5 * 5/7 * 7/9 * 9/11 = 0.0455

if we had excluded column 9 and hence 7/9, the equation would have been

1/2 * 1/3 * 3/5 * 5/7 * 9/11 = 0.0584.

Because, 7/9 is less than 1, the value of the first expression above (which includes 7/9) is 0.0455 and is less than the value of second expression (which excludes 7/9) 0.0584

Thus, when we multiply the Total Number of Rows (n^2) with 0.0455 (instead of 0.0584) above to calculate the number of twin primes, we would get in fact a conservative estimate of twin primes which makes this inclusion of 7/9 unbiased.

2) Further, inclusion of Column 9 and hence 7/9 make algebraic manipulation simpler and helps us get neat formulas (as it helps in cancelling numerators and denominators) while being unbiased.

'At the Least' Values

As pointed above, we have included 7/9 ratio in the expressions above. Thus our estimate of number of twin primes for given number of rows becomes conservative and the actual number of twin primes may be greater.

Hence, the number of twin primes i.e. n/2 till rows n^2 are 'at least' i.e. there are at least 'n/2' twin primes till rows n^2 (and 'n' being sufficiently large)

Diversion from Desired Ratio

The key here is that ratio of 'Allowed Values' to 'Total Values' in each column should tend to '**Desired Ratio**' in '**All Columns**'. And this as we have seen above will happen when 'n' is large.

If we are picking only a subset of rows from the total number of rows, then these desired ratios in each column may distort and become either favorable or unfavorable depending on the subset that we have chosen and its size and hence we may get no twin primes or more than expected twin primes in such selected subsets due to such distorted ratios. This may indicate that twin primes may not be evenly distributed but when large number of rows are taken the discrepancies in ratios ‘even out’ and the ratios in ‘all’ columns tend to its desired ratios as already shown above.

CONCLUSION:

Since we have proved that there are at least (n/2) twin primes between 1 and n^2 rows, we can thus say that, as ‘n’ tends to infinity, ‘(n/2)’ too tends to infinity i.e.

The number of twin primes (n/2) tends to infinity thus proving that **there are infinite twin primes**.