On the Infinitude of Primes*

An Elementary Approach Through an Inequality

Sharv Laad

There have been many proofs of the infinitude of primes since Euclid (circa 300 BC) first gave his proof. In this article, I present my contribution to this collection of proofs.

Sharv Laad is a high school senior, pursuing mathematics. He has been involved in the Mathematical Olympiad Programme and has cleared the regional level.

Introduction

Euclid presented the world with the first-ever proof of the infinitude of primes. Since then, there have been numerous proofs of the same result using different ideas. Euclid's idea was very simple. He assumed that there were finitely many primes, say $p_1, p_2, ..., p_n$. He then considered the number $N = p_1 p_2 ... p_n + 1$. Clearly, N is not divisible by any of the n primes, and thus we arrive at a contradiction.

My proof, however, makes use of the number of integers less than or equal to n that are indivisible by any of the primes out of a given set. I use this to arrive at an inequality which proves the result.

1. The Principle of Inclusion And Exclusion

This section will help you understand the principle of inclusion and exclusion that I make use of in the sections that follow.

One will find *Figure* 1 in almost any standard combinatorics book, and it is the best way to understand this principle. The diagram shows two sets—A and B. We want to count the number of elements in $A \cup B$. To do this, notice that the area denoted by $A \cup B$ is nothing but the area in set A plus that in B minus the orange area, i.e. $A \cap B$ because we count this area twice. Thus,

Keywords
Number theory, primes, infinitude.

*Vol.25, No.10, DOI: https://doi.org/10.1007/s12045-020-1060-5

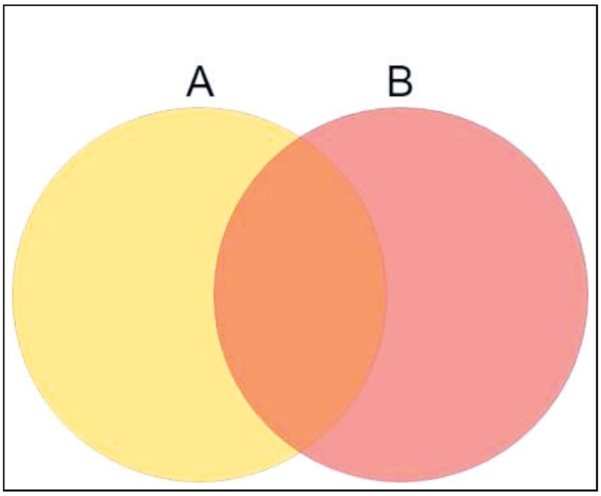

Figure 1. Venn diagram showing two sets A and B having common elements.

$$|A \cup B| = |A| + |B| - |A \cap B| .$$

Similarly, one can show for three sets A, B and C that

$$|A \cup B \cup C| = |A|+|B|+|C|-|A \cap B|-|B \cap C|-|C \cap A|+|A \cap B \cap C| .$$

And thus, for n sets.

2. A Counting Problem

I will first prove a simple counting problem stated below. The result of this problem will be used in the proof.

Consider k arbitrary primes $p_1, p_2, \ldots, p_k$. Let the number of positive integers less than or equal to n which are not divisible by any of the k primes be denoted by $\lambda(n; p_1, p_2, \ldots, p_k)$. We have [1]:

$$\lambda(n; p_1, p_2, \ldots, p_k) =$$
$$n - \sum_{sym}\left[\frac{n}{p_1}\right] + \sum_{sym}\left[\frac{n}{p_1 p_2}\right] - \sum_{sym}\left[\frac{n}{p_1 p_2 p_3}\right] \ldots (-1)^k \left[\frac{n}{p_1 p_2 ... p_k}\right],$$

[1] $[x]$ represents the greatest integer function.

where,

$\sum_{sym}\left[\frac{n}{p_1 p_2 \cdots p_i}\right]$ denotes the symmetric sum $\sum_{1\leq i_1<i_2<\ldots<i_s\leq n}\left[\frac{n}{p_{i_1} p_{i_2} \cdots p_{i_s}}\right]$.

For example, if we take $p_1 = 2$, $p_2 = 5$ and $n = 10$, the problem states that the number of integers less than or equal to 10 which are not divisible by 5 or 2, i.e. $\lambda(10; 2, 5)$ is given by,

$$10 - \left[\frac{10}{2}\right] - \left[\frac{10}{5}\right] + \left[\frac{10}{2\times 5}\right] = 4.$$

I will prove this using the inclusion-exclusion principle. The number of multiples of an integer t less than or equal to an integer m is given by $\left[\frac{m}{t}\right]$. If we denote by A_i, the set of multiples of the prime p_i less than or equal to n, we know that $|A_i| = \left[\frac{n}{p_i}\right]$. Also, $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}$ where $1 \leq i_1 < i_2 \cdots < i_s \leq k$, denotes the number of integers less than or equal to n which are multiples of each of the s primes, i.e. multiples of $p_{i_1} p_{i_2} \ldots p_{i_s}$. Thus we have

$$\left|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}\right| = \left[\frac{n}{p_{i_1} p_{i_2} \cdots p_{i_s}}\right].$$

The number of multiples of an integer t less than or equal to an integer m is given by $\left[\frac{m}{t}\right]$.

Since $A_1 \cup A_2 \cup \cdots \cup A_k$ will denote the number of integers less than or equal to n which are multiples of at least one of the k primes, we have

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = n - \lambda(n; p_1, p_2, \ldots, p_k).$$

Thus, by the principle of inclusion-exclusion, we have:

$$n - \lambda(n; p_1, p_2, \ldots, p_k) =$$
$$\sum_{sym}\left[\frac{n}{p_1}\right] - \sum_{sym}\left[\frac{n}{p_1 p_2}\right] + \sum_{sym}\left[\frac{n}{p_1 p_2 p_3}\right] \ldots (-1)^{k+1}\left[\frac{n}{p_1 p_2 \ldots p_k}\right].$$

And the result is immediate from here.

3. An Inequality

Using the result of the problem described in the previous section, I will now prove the following inequality.

If $p_1, p_2, ..., p_{\pi(n)}$ denote the $\pi(n)$ primes less than or equal to n. Then the following inequality holds[2]:

[2] $\pi(n)$ denotes the prime counting function, i.e. $\pi(n)$ is the number of primes less than or equal to n.

$$n \prod_{i=1}^{\pi(n)} \left(1 - \frac{1}{p_i}\right) < 1 + 2^{\pi(n)-1}$$

As done above, let us denote by $\lambda(n; p_1, p_2, \ldots, p_k)$ the number of integers less than or equal to n which are not multiples of any of these k primes.

We know that,

$$\lambda(n; p_1, p_2, \ldots, p_k) =$$
$$n - \sum_{sym} \left[\frac{n}{p_1}\right] + \sum_{sym} \left[\frac{n}{p_1 p_2}\right] - \sum_{sym} \left[\frac{n}{p_1 p_2 p_3}\right] \ldots (-1)^k \left[\frac{n}{p_1 p_2 ... p_k}\right].$$

We have,

$$x - 1 < [x] \leq x.$$

Multiplying throughout by -1,

$$-x \leq -[x] < -(x - 1).$$

Thus,

$$\begin{array}{lllll}
-\displaystyle\sum_{sym}\frac{n}{p_1} & < & -\displaystyle\sum_{sym}\left[\frac{n}{p_1}\right] & \leq & -\displaystyle\sum_{sym}\left(\frac{n}{p_1}-1\right), \\
\displaystyle\sum_{sym}\left(\frac{n}{p_1p_2}-1\right) & \leq & \displaystyle\sum_{sym}\left[\frac{n}{p_1p_2}\right] & < & \displaystyle\sum_{sym}\frac{n}{p_1p_2}, \\
-\displaystyle\sum_{sym}\frac{n}{p_1p_2p_3} & < & -\displaystyle\sum_{sym}\left[\frac{n}{p_1p_2p_3}\right] & \leq & -\displaystyle\sum_{sym}\left(\frac{n}{p_1p_2p_3}-1\right), \\
\vdots & & \vdots & & \vdots
\end{array}$$

Adding all these and adding n throughout, we get,

$$n-\sum_{sym}\frac{n}{p_1}+\sum_{sym}\left(\frac{n}{p_1p_2}-1\right)-\sum_{sym}\frac{n}{p_1p_2p_3}\cdots<\lambda(n;p_1,p_2,\ldots,p_k)$$

$$\implies n-\sum_{sym}\frac{n}{p_1}+\sum_{sym}\frac{n}{p_1p_2}-\binom{k}{2} -\sum_{sym}\frac{n}{p_1p_2p_3}+\sum_{sym}\frac{n}{p_1p_2p_3p_4}-\binom{k}{4}\ldots <\lambda(n;p_1,p_2,\ldots,p_k)$$

$$\implies \left(n-\sum_{sym}\frac{n}{p_1}+\sum_{sym}\frac{n}{p_1p_2}-\sum_{sym}\frac{n}{p_1p_2p_3}\ldots\right)-\sum_{i=1}\binom{k}{2i} <\lambda(n;p_1,p_2,\ldots,p_k)$$

$$\implies n\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\ldots\left(1-\frac{1}{p_k}\right)-2^{k-1} <\lambda(n;p_1,p_2,\ldots,p_k)$$

$$\implies n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)-2^{k-1} <\lambda(n;p_1,p_2,\ldots,p_k).$$

Similarly we can get the upper bound,

$$n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)-2^{k-1} < \lambda(n; p_1, p_2, \ldots, p_k) < n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)+2^{k-1}.$$

How is this inequality useful? It becomes useful when we consider the first $\pi(n)$ primes $2 = p_1, p_2, \ldots, p_{\pi(n)}$ less than or equal to n. Then by the definition of λ, $\lambda(n; p_1, p_2, \ldots, p_{\pi(n)}) = 1$, because each of the integers less than or equal to n, except 1, is a multiple of at least one of these primes. The above inequality then gives

$$n\prod_{i=1}^{\pi(n)}\left(1-\frac{1}{p_i}\right) - 2^{\pi(n)-1} < 1.$$

And the result follows.

4. The Infinitude Of Primes

Now that I have proved the inequality in the previous section, the proof of the infinitude of primes becomes a matter of few steps. We will do this by contradiction.

Let us assume to the contrary that there are only k primes, $p_1, p_2, \ldots, p_k$. For all $n > p_k$, $\pi(n) = k$ and $\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)$ is constant. We have proved that

$$n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right) < 1 + 2^{k-1}.$$

The left side of this inequality is clearly unbounded, while the right side is constant. This is a contradiction. Hence, there are infinitely many primes.

5. Similar Proofs

My proof involved an inequality related to primes. In this section, I will present two more proofs which are based on similar ideas.

5.1 *Chebyshev's Proof*

Chebyshev's proof makes use of De Polignac's formula, that gives the power of a prime dividing $n!$.

Let k_i denote the largest power of p_i in $n!$ for $i = 1, 2, ..., \pi(n)$. De Polignac's formula states,

$$k_i = [\frac{n}{p_i}] + [\frac{n}{p_i^2}] + [\frac{n}{p_i^3}]...$$

But we have,

$$[\frac{n}{p_i^s}] \leq \frac{n}{p_i^s}.$$

This gives,

$$\begin{aligned} k_i &= [\frac{n}{p_i}] + [\frac{n}{p_i^2}] + [\frac{n}{p_i^3}]... \\ &< \frac{n}{p_i} + \frac{n}{p_i^2} + \frac{n}{p_i^3}... \\ &= \frac{n}{p_i}\frac{1}{1-\frac{1}{p_i}}, \\ &= \frac{n}{p_i - 1}. \end{aligned}$$

Notice that we will not have equality because when $p_i^s > n, 0 < \frac{n}{p_i^s} < 1$. So, $[\frac{n}{p_i}] = 0$. Thus, the RHS of the inequality will still have a positive increment, while, the LHS will not change. Now, Since $p_i > 1$, we have:

$$p_i^{k_i} < p_i^{\frac{n}{p_i - 1}},$$

$$\implies \prod_{i=1}^{\pi(n)} p_i^{k_i} < \prod_{i=1}^{\pi(n)} p_i^{\frac{n}{p_i - 1}}.$$

Chebyshev's proof makes use of De Polignac's formula, that gives the power of a prime dividing $n!$.

But $\prod_{i=1}^{\pi(n)} p_i^{k_i}$ is nothing but $n!$ because the prime factorisation of $n!$ would contain all the $\pi(n)$ primes less than n raised to their respective highest powers in $n!$. Thus,

$$n! < \prod_{i=1}^{\pi(n)} p_i^{\frac{n}{p_i - 1}}.$$

Taking log on both sides and dividing by n, we get:

$$\frac{\sum_{i=1}^{n} \log i}{n} < \sum_{i=1}^{\pi(n)} \frac{\log p_i}{p_i - 1}.$$

It can be shown that the left side of this inequality is unbounded. Thus, there must be infinitely many primes.

5.2 *Erdos' Proof*

Erdos used the fact that any number n can be expressed as $n = rs^2$ where r is square-free.

In his proof, Erdos used the fact that any number n can be expressed as $n = rs^2$ where r is square-free. To see this, choose s to be the largest integer such that s^2 divides n. Next he overestimated how many factorisations there are $\leq n$ as follows.

The number of squares $\leq n$ are not more than $\sqrt{n}$. Thus, there are at most $\sqrt{n}$ possibilities for s. r will simply be a product of some of the distinct primes less than n, since it cannot contain a perfect square. Thus, the maximum number of possibilities for r are given by the number of subsets of $\{p_1, p_2, \ldots, p_{\pi(n)}\}$, i.e., $2^{\pi(n)}$.

Thus, we have:

$$n \leq 2^{\pi(n)} \sqrt{n}.$$

Dividing by $\sqrt{n}$ and taking log on both sides we get,

$$\frac{\log n}{\log 4} \leq \pi(n).$$

And clearly the LHS is unbounded. Hence there are infinitely many primes.

6. Comparison

All the three proofs described above are based on the same idea. Chebyshev's proof used the power of a prime in $n!$ and Erdos' proof used an intelligent way to express any integer. Whereas, my proof used the number of integers less than n which are not divisible by a given set of primes. Each approach involved the derivation of an inequality, which lead to the proof.

It should be noted however, the inequality derived in my proof is stronger than the one derived in Erdos' proof for $n \geq 146$. This can be shown as follows.

The RHS is clearly a stronger bound since $2^{\pi(n)-1} + 1 < 2^{\pi(n)}$. We need to prove that

$$\sqrt{n} < n \prod_{i=1}^{\pi(n)} \left(1 - \frac{1}{p_i}\right),$$

Or,

$$\frac{1}{\sqrt{n}} < \prod_{i=1}^{\pi(n)} \left(1 - \frac{1}{p_i}\right). \tag{*}$$

Using computer programs, I have confirmed that this inequality is valid for $n = 146$. I will prove that the inequality is true for $n > 146$ by induction.

Let the inequality be true for some $n > 146$. Consider the inequality

$$\frac{\sqrt{n}}{\sqrt{n+1}} < 1 - \frac{1}{p_{n+1}}. \tag{**}$$

This reduces to

$$\frac{1}{p_{n+1}} < \frac{\sqrt{n+1} - \sqrt{n}}{\sqrt{n+1}},$$

Or,

$$p_{n+1} > \frac{\sqrt{n+1}}{\sqrt{n+1} - \sqrt{n}} = \sqrt{n+1}(\sqrt{n+1} + \sqrt{n}) = n+1 + \sqrt{n(n+1)},$$

i.e.,

$$p_n > n + \sqrt{n(n-1)}.$$

However, we know that $p_n > 2n > n + \sqrt{n(n-1)}$ for $n >= 5$ [3]. Thus, inequality (**) is true $n \geq 146$. Multiplying (**) with (*), the induction step (and thus the proof) is complete.

[3]This can also be done via induction. Assuming it is true for some n, notice that $p_{n+1} \geq p_n + 2 > 2n + 2$

Thus, the inequality described in my proof becomes stronger for $n \geq 146$.

7. Conclusion

I began with a basic combinatorics problem that used the principle of inclusion and exclusion and arrived at a formula. As you have seen, this formula made use of the greatest integer function, which provided me with an opportunity to introduce bounds. By doing so, I arrived at an inequality that, in a way, related the primes less than n to $\pi(n)$. This inequality helped me prove that there are infinitely many primes. Finally, I have described a few other proofs which were based on similar ideas and have compared them to my proof.

Acknowledgement

I would like to thank Dr S A Katre for assistance with proof reading and Mr Siddharth Bhalerao for his ideas that greatly improved the manuscript. I would also like to thank Mr Kiran Barve for his immense support in helping me conduct my research.

Suggested Reading

[1] Bibekananda Maji, A new proof of the infinitude of primes, *Resonance: Journal of Science Education*, Vol.20, No.12, pp.1128–1135, 2015.

[2] Shailesh A Shirali, On the infinitude of prime numbers: Euler's proof, *Resonance: Journal of Science Education*, Vol.1, No.3, pp.78–95, 1996.

[3] P Ribenboim, *The Little Book Of Bigger Primes*, Springer-Verlag, New York, 1996.

[4] Ivan Niven, Herbert S Zuckerman, Hugh L Montgomery, *An Introduction To The Theory Of Numbers, 5th edition*, John Wiley and Sons, 2013.

Address for Correspondence
Sharv Laad
A/8 Hill View Residency
Baner road
Pune 411 045, India.
Email: sharvlaad@gmail.com