The Distribution of Relatively Prime Numbers

Abdulkadir Hassen
Matt Oster
Mathematics Department
Rowan University
Glassboro, NJ 08028

hassen@rowan.edu
osterm38@students.rowan.edu

1. Introduction

In this article, we aim to prove that if one picks an ordered pair of integers (x, y) at random, then the probability that these integers are relatively prime is $\frac{6}{\pi^2}$ or 60.8%. Recall that two integers x and y are called *relatively prime* or *co-prime* if their greatest common divisor is 1. We will do this in three different ways. Our first proof is longer and makes use of the Möbius Inversion Formula. We have made every effort to make this as elementary as possible. While this is longer than the other two, it gives the flavor of the beautiful mathematics behind the techniques of Analytic Number Theory. The other two methods are shorter and use the concept of probability and modulo arithmetic.

We state our result as

Theorem 1: If x and y are randomly selected integers, then $\Pr\{(\gcd(x,y)=1\} = \frac{6}{\pi^2}$,

where $\gcd(x,y)$ stands for the greatest common divisor of x and y

Note: 1. Euler proved that $\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}$

2. $\zeta(s)=\sum\frac{1}{n^s}$, s > 1 is the Riemann Zeta function. Thus our probability is $\frac{1}{\zeta(2)}$.

2. Proof of Theorem 1 using Mobius Inversion Formula

For $r>0,\ r\in\mathbb{Z}$, define $S_r=\{(x,y)\,|\,x,y\in\mathbb{Z},\ -r\le x\le r \text{ and } -r\le y\le r\}$

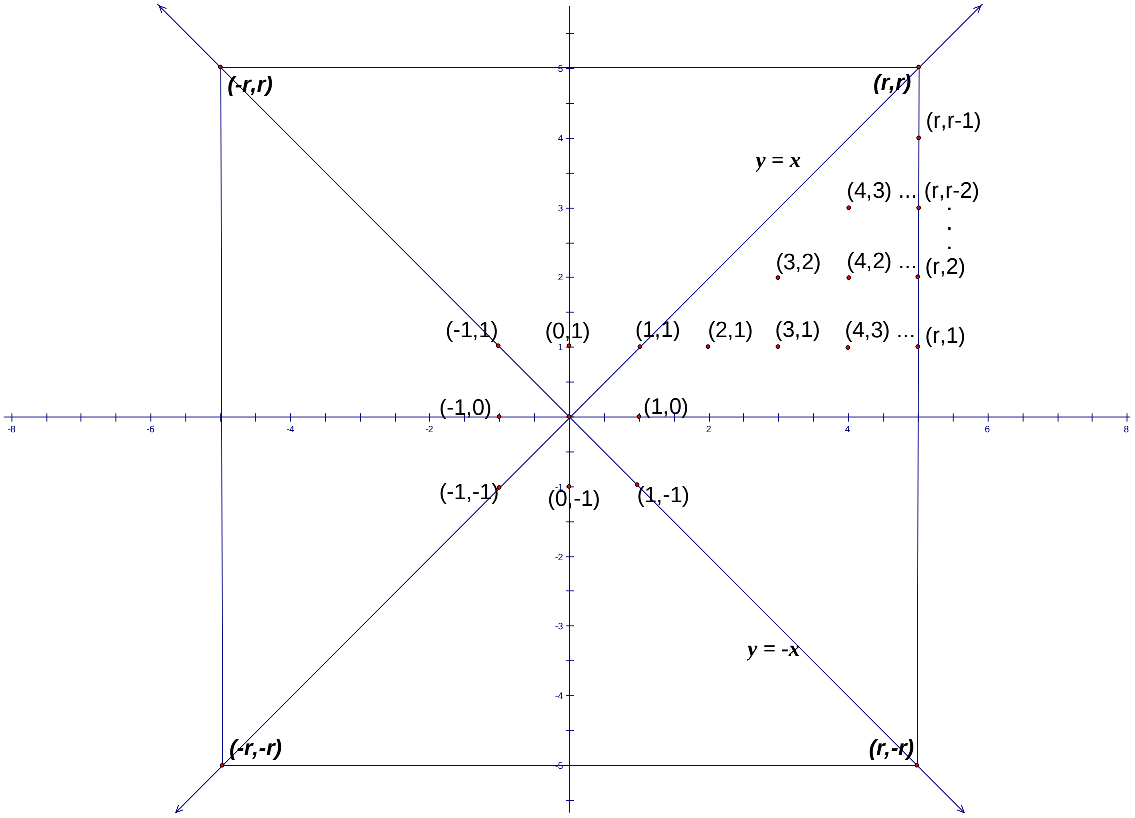
Let $N(r)$ be the number of elements in S_r and $M(r)$ be the number of elements (x,y) in S_r that are relatively prime. It is an interesting exercise to show that $\gcd(x,y)=1$ if and only if the line segment joining $(0,0)$ to (x,y) contains no point (a,b) with *a*, *b* integers. When the latter happens, we say (x,y) is visible from the origin.

Theorem 1 will follow if we prove

Theorem 2:

$$\lim_{r\to\infty}\frac{M(r)}{N(r)}=\frac{6}{\pi^2}$$

To prove Theorem 2, we consider the following diagram.



Clearly the eight points $(0,\pm 1)$, $(\pm 1, 0), (\pm 1, \pm 1)$ are relatively prime. Furthermore, if $\gcd(x,y)=1$, then $\gcd(\pm x,\pm y)=1$ and $\gcd(\pm y,\pm x)=1$. Thus, we count the relative prime pairs (that is, count (x,y) for which $\gcd(x,y)=1$) in the set $2\le x\le r, 1\le y\le x$. In other words, if $M'(r)$ is the number of ordered pairs (x, y) such that $\gcd(x,y)=1$ and $2\le x\le r$ and $1\le y\le x$, then

$$M(r)=8+8M'(r).$$

But for each *n* from *2* to *r*, the number of integers *y*, such that $\gcd(n,y)=1$, is $\varphi(n)$. Thus

$M'(r)=\sum_{n=2}^{r}\varphi(n)$. Since $\varphi(1)=1$, we conclude that

$$M(r)=8\sum_{n=1}^{r}\varphi(n).$$

Theorem 2 (and hence Theorem 1) will be a consequence of

Theorem 3: $$\sum_{n=1}^{r} \varphi(r) = \frac{3}{\pi^2} r^2 + g(r)$$

where $g(r)$ is a function with the property that $|g(r)| \le Mr \ln r$ for some constant *M.*

We will prove Theorem 3 in a moment. But first, let us assume its validity and prove Theorem 2. By the remarks preceding Theorem 3, we have

$$M(r) = 8\sum_{n=1}^{r} \varphi(n) = 8\left(\frac{3}{\pi^2} r^2 + g(r)\right) = \frac{24}{\pi^2} r^2 + 8g(r).$$

Clearly $N(r) = (2r+1)^2 = 4r^2 + 4r + 1$. Therefore,

$$\frac{M(r)}{N(r)} = \frac{\left(\frac{24r^2}{\pi^2} + 8g(r)\right)}{4r^2 + 4r + 1} = \frac{\frac{24}{\pi^2} + 8\frac{g(r)}{r^2}}{4 + \frac{4}{r} + \frac{1}{r^2}}$$

And hence,

$$\lim_{r\to\infty} \frac{M(r)}{N(r)} = \frac{\left(\frac{24}{\pi^2}\right)}{4} = \frac{6}{\pi^2}.$$

Before giving the proof of Theorem 3, let us consider two examples.

Example 1: Let n = 20 and S = {1, 2, ….., 20}. We need to partition S in to subsets whose elements have the same gcd with 20. For each divisor d of 20, let

$S_d = \{x \in S \mid \gcd(x, 20) = d\}$. Thus $S_1 = \{x \mid \gcd(x, 20) = 1\} = \{1, 3, 7, 11, 13, 17, 19\}$. These numbers are all relatively prime to 20. Hence S_1 has $\varphi(20)$ elements.

$S_2 = \{x \mid \gcd(x, 20) = 2\} = \{2, 6, 14, 18\} = \{2\cdot 1, 2\cdot 3, 2\cdot 7, 2\cdot 9\}$. If we factor 2, the remaining numbers in S_2 are those relatively prime to 10 = $\frac{20}{2}$. Hence there are $\varphi(10)$ of them.

The reader can show that S_4 has $\varphi(\frac{20}{4}) = \varphi(5)$ elements, S_5 has $\varphi(\frac{20}{5}) = \varphi(4)$, S_{10} has $\varphi(\frac{20}{10}) = \varphi(2)$, and S_{20} has $\varphi(\frac{20}{20}) = \varphi(1)$ element. The sets $S_1, S_2, S_4, S_5, S_{10}, S_{20}$ are mutually disjoint. Hence

$$20 = \varphi(20) + \varphi(10) + \varphi(5) + \varphi(4) + \varphi(1)$$

In general, we have

Lemma 1: For any positive integer *n* we have $n = \sum_{d|n} \varphi(d)$.

Here $d \mid n$ means *d* is a divisor of *n* and $\sum_{d|n}$ means that we are adding over the positive divisors of *n*.

Next, we consider the problem of recovering a function $g(n)$, if we are given that

$$f(n) = \sum_{d|n} g(d).$$

In other words, given the above sum, can we find $g(n)$ in terms of $f(d)$, $d \mid n$?

The answer is yes. It is given by the use of Mobius Inversion Formula. We show this with an example.

Example 2: Again we use *n= 20* as an example to show how this can be done. Suppose then

$$f(20) = \sum_{d|20} g(d) = g(1) + g(2) + g(4) + g(5) + g(10) + g(20).$$

We want to express $g(20)$ in terms of $f(1),\ f(2),\ f(4),\ f(5),\ f(10),\ f(20).$ We list the values of f at $d=1,2,4,5,10,20$ as follow

$$f(20)=g(1)+g(2)+g(4)+g(5)+g(10)+g(20)$$

$$f(10)=g(1)+g(2)+g(5)+g(10)$$

$$f(5)=g(1)+g(2)+g(5)$$

$$f(4)=g(1)+g(2)+g(4)$$

$$f(2)=g(1)+g(2)$$

$$f(1)=g(1)$$

Thus

$$f(20)\ \text{-}\ f(10)\ =\ g(4)\ +\ g(20).$$

To get rid of $g(4)$ let us subtract $f(4).$ This gives us

$$f(20)\ \text{-}\ f(10)\ \text{-}\ f(4)\ =\ g(20)\ \text{-}\ g(1)\ \text{-}\ g(2).$$

However, we do not want $-g(1)\ -\ g(2),$ so we add $f(2).$ Thus

$$f(20)\ \text{-}\ f(10)\ \text{-}\ f(4)\ +\ f(2)\ =\ g(20).$$

Therefore

$$g(20)\ =\ f(2)\ \text{-}\ f(4)\ \text{-}\ f(10)\ +\ f(20).$$

We are missing $f(1)$ *and* $f(5).$ Let us note that 20/1 = 20 and 20/5 = 4 have square factors; 20/4 = 5 and 20/10 = 2 have one prime factors; $20/2\ =\ 10\ =\ 2\cdot 5$ has two prime factors, and 20/20 = 1 has no prime factor. Thus if we define

$$\mu(1)=1,\ \ \mu(2)=\text{-}1,\ \ \mu(4)=0,\ \ \mu(5)=\text{-}1,\ \ \mu(10)=1,\ \text{and}\ \mu(20)=0$$

we see that

$$g(20)=\sum_{d|20}\mu\left(\frac{20}{d}\right)f(d)$$

This leads to the definition of the following:

Mobius Function

$$\mu(n)=\begin{cases}1,\ \text{if } n=1\\ 0,\ \text{if } n \text{ has a square divisor}\\ (\text{-}1)^r,\ \text{if } n \text{ is a product of } r \text{ distinct primes.}\end{cases}$$

A general statement shown by the above example is the

Mobius Inversion Formula: If $f(n)=\sum_{d|n} g(n),$ then $g(n)=\sum_{d|n}\mu(d)f\left(\frac{d}{n}\right)$

We now return to Theorem 3.

By lemma 1, $n=\sum_{d|n}\varphi(d).$ If $f(n)=n,$ and $g(n)=\varphi(n)$, then the Mobius Inversion Formula gives,

$$\varphi(n)=\sum_{d|n}\mu(d)\left(\frac{n}{d}\right).$$

Therefore

$$M(r)=\sum_{n=1}^{r}\left(\sum_{d|n}\mu(d)\frac{n}{d}\right)=\sum_{\substack{qd\le r\\ q,d}}\mu(d)q=\sum_{d\le r}\mu(d)\left(\sum_{q\le\frac{r}{d}}q\right)$$

We now use

$$\sum_{k=1}^{x}k=\frac{x(x+1)}{2}=\frac{1}{2}x^2+f(x),$$

where $f(x)=\frac{x}{2},$ and $x=\left[\frac{r}{d}\right]=$ the greatest integer less than or equal to $\frac{\mathrm{r}}{\mathrm{d}}$

Hence $\sum_{q\le\frac{r}{d}}q=\frac{1}{2}\left(\frac{r}{d}\right)^2+f\left(\frac{r}{d}\right),\qquad f\left(\frac{r}{d}\right)\le K\frac{r}{d}$. Therefore

$$\sum_{n\le r}\varphi(n)=\sum_{d\le r}\mu(d)\left[\frac{1}{2}\left(\frac{r}{d}\right)^2+f\left(\frac{r}{d}\right)\right]=\frac{1}{2}r^2\sum_{d\le r}\frac{\mu(d)}{d^2}+\sum_{d\le r}\mu(d)\,f\left(\frac{r}{d}\right)$$

$$=\frac{1}{2}r^2\left(\sum_{d=1}^{\infty}\frac{\mu(d)}{d^2}-\sum_{d>r}\frac{\mu(d)}{d^2}\right)+\sum_{d\le r}\mu(d)\,f\left(\frac{r}{d}\right)$$

$$=\frac{1}{2}r^2\sum_{d=1}^{\infty}\frac{\mu(d)}{d^2}+G(r),$$

where

$$G(r)=\sum_{d\le r}\mu(d)\,f\left(\frac{r}{d}\right)-\frac{1}{2}r^2\sum_{d>r}\frac{\mu(d)}{d^2}$$

To show that $|G(r)|\le Mr\ln r,$ we observe that

$$\left|\frac{1}{2}r^2\sum_{d>r}\frac{\mu(d)}{d^2}\right|\le\frac{1}{2}r^2\sum_{n>r}\frac{1}{n^2}<\frac{1}{2}r^2L\int_r^{\infty}\frac{1}{x^2}\,dx=\frac{1}{2}Lr.$$

and

$$\left|\sum_{d\le r}\mu(d)f\left(\frac{r}{d}\right)\right|\le Kr\left|\sum_{d\le r}\frac{\mu(d)}{d}\right|\le Kr\sum_{d\le r}\frac{1}{d}=Kr\ln r.$$

Consequently,

$$|G(r)|=\left|\sum_{d\le r}\mu(d)\,f\left(\frac{r}{d}\right)-\frac{1}{2}r^2\sum_{d>r}\frac{\mu(d)}{d^2}\right|\le\left|\sum_{d\le r}\mu(d)\,f\left(\frac{r}{d}\right)\right|+\left|\frac{1}{2}r^2\sum_{d>r}\frac{\mu(d)}{d^2}\right|\le Kr\ln r+\frac{1}{2}Lr\le Mr\ln r$$

To complete the proof of Theorem 3, all we need to show is that

$$\sum_{d=1}^{\infty}\frac{\mu(d)}{d^2}=\frac{1}{\zeta(2)}\,.$$

But this follows from

$$\begin{aligned}\prod_{p\ \text{prime}}\left(1-p^{-2}\right)&=\left(1-2^{-2}\right)\left(1-3^{-2}\right)\left(1-5^{-2}\right)\left(1-7^{-2}\right)\left(1-11^{-2}\right)\cdots\\ &=1-\left(2^{-2}+3^{-2}+5^{-2}+\cdots\right)\\ &\quad+\left((2\cdot 3)^{-2}+(2\cdot 5)^{-2}+\cdots+(3\cdot 5)^{-2}+\cdots\right)\\ &\quad-\left((2\cdot 3\cdot 5)^{-2}+(2\cdot 3\cdot 7)^{-2}+\cdots+(3\cdot 5\cdot 7)^{-2}\right)\cdots\\ &\quad=\sum\frac{\mu(n)}{n^2}.\end{aligned}$$

The sum on the first row is obtained when we multiply each of the prime factors with one form all the other factors, the sum in the second row is obtained by taking the primes multiplied by 1 taken from the rest of the factors. The sum in the second row comes from the product of two primes from two of the factors and all 1 from the other factors. And so on. The last equality is clear from the definition of the Mobius Function. A Similar argument yields

$$\begin{aligned}\prod\left(1-p^{-2}\right)^{-1}&=\frac{1}{\left(1-2^{-2}\right)}\cdot\frac{1}{\left(1-3^{-2}\right)}\cdot\frac{1}{\left(1-5^{-2}\right)}\cdots\\ &=\left(1+(2^{-2})+(2^{-2})^2+\cdots\right)\left(1+(3^{-2})+(3^{-2})^2+\cdots\right)\left(1+(5^{-2})+(5^{-2})^2+\cdots\right)\\ &=1+2^{-2}+3^{-2}+5^{-2}+\cdots+(2\cdot 2)^{-2}+(2\cdot 3)^{-2}+(2\cdot 5)^{-2}+\cdots\\ &=\sum\frac{1}{n^2}=\frac{\pi^2}{6}\end{aligned}$$

Therefore

$$\sum\frac{\mu(n)}{n^2}=\frac{1}{\sum\frac{1}{n^2}}=\frac{6}{\pi^2}.$$

This completes the proof of Theorem 3.

3. Probability Argument

We will use the fact that if S is a sample space of all possible outcomes of an experiment, then *pr(S) = 1,* where *pr(S)* denotes probability of S. Also if $S = \bigcup_n S_n$ and the collection $\{S_j\}$ is pairwise disjoint, then $\sum \Pr\left(S_j\right) = 1$.

We also need the modulo notation for easier argument. We say x is congruent to y modulo *n* and write $x \equiv y \left(\operatorname{mod} n\right)$ if *n* is a divisor of $x - y$. It is not hard to see that for any given $x \in \mathbb{Z}$ and $n \geq 1$, then

$$\Pr(\{x \mid x \equiv c(\operatorname{mod} n)\}) = \frac{1}{n}$$

If we pick *x* and *y* randomly and independently, then

$$\Pr\left(\left\{(x, y) \mid x \equiv 0(\operatorname{mod} n) \text{ and } y \equiv 0(\operatorname{mod} n)\right\}\right) = \frac{1}{n^2}$$

Let *P = Pr(gcd(x, y) = 1)* where *x, y* are randomly selected. Then, for a fixed *n > 1,*

$$\Pr\left(\gcd\left(\frac{x}{n}, \frac{y}{n}\right) = 1\right) = P.$$

A fact that can easily be established is that

$$\text{gcd(x, y) = n if and only if} \begin{cases} x \equiv 0\left(\operatorname{mod} n\right) \\ y \equiv 0\left(\operatorname{mod} n\right) \\ \gcd\left(\frac{x}{n}, \frac{y}{n}\right) = 1. \end{cases}$$

Therefore $\Pr\left(\gcd(x, y) = n\right) = \frac{1}{n} \cdot \frac{1}{n} \cdot P$

And since $\sum_{n=1}^{\infty} \Pr\left(\gcd(x, y) = n\right) = 1,$ we see that $\sum \frac{1}{n^2} P = 1$ and hence

$$P = \frac{1}{\sum \frac{1}{n^2}} = \frac{6}{\pi^2}$$

Another probabilistic argument can be given as follows:

$$\gcd(x,\ y)\ =\ 1 \text{ if and only if } \begin{cases} x \equiv 0(\text{mod } p) \\ or \\ y \equiv 0(\text{mod } p) \end{cases} \quad \text{for all prime } p$$

Since $\Pr\{x \equiv 0(\text{mod } p) \text{ and } y \equiv 0(\text{mod } p)\} = \frac{1}{p^2}$ and the negation of $x \equiv 0(\text{mod } p)$ and $y \equiv 0(\text{mod } p)$ is $x \neq 0(\text{mod}\, P)$ or $y \neq 0(\text{mod}\, P)$, we see that

$$\Pr\{x \neq 0(\text{mod } p) \text{ or } y \neq 0(\text{mod } p)\} = 1 - \frac{1}{p^2}$$

Since primes are independent, we conclude that

$$\Pr\{\gcd(x, y) = 1\} = \prod_p \left(1 - \frac{1}{p^2}\right) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^2}\ .$$

4. References

[1] Apostol, Tom, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, NY 1976.

[2] Jones, Gareth A., and Jones, Mary J., *Elementary Number*, Springer-Verlag, New York, NY, 1998.