Mathematics Extended Essay

Investigating the Probability of Pairwise Coprimality

What is the probability that a set of randomly chosen integers is pairwise coprime and how can it be determined?

Personal Code: gnm114

Word Count: 3304 Words

Table of Contents

Symbols and Notation

Symbols

$\zeta(s)$	Riemann zeta function
k	Size of a set
p	Denotes a prime number
ϕ	A simple probability value
$C_{pair}(k)$	Probability that a set of k randomly chosen positive integers is pairwise coprime
$\forall$	For all
iff	If and only if
$\mathbb{Z}^+$	The set of positive integers
$\rightarrow$	Tends to

Notation

(a, b) or $\gcd(a, b)$	Greatest common divisor of a and b
$\binom{n}{r}$	Binomial coefficient (Choose function)
$a \mid b$	a divides b
$a \nmid b$	a does not divide b
$a_1, a_2, \dots, a_k$	A set of integers with k members
$\prod_{i}^{\infty} f(i)$	Infinite product of $f(i)$

1. Introduction

A detailed definition of coprimality is shown in Section 2, however a simple definition is that two positive integers are considered coprime if they share no common factor, except 1. I was initially exposed to the concept of coprimality through the probability of coprimality being used to approximate the value of π (Parker). Parker generated pairs of random numbers using large dice (120 sides) and counted what percentage of those pairs were coprime. From this, Parker was able to approximate π, using the probability of coprimality, as described in Section 2. This concept instantly piqued my curiosity, as although I had never learned about number theory directly, I was very interested in its related concepts. The appearance of π in this context also appealed to me, as I am fascinated by the way π appears in so many seemingly unrelated areas of mathematics, such as Buffin's Needle, Heisenberg's Uncertainty Principle and many more.

The video that sparked my interest only involved the coprimality of two integers (Parker). This led me to think about how this concept can be expanded and extended, for example can this concept be applied to other areas or can other approaches be used. For me, the obvious next step was coprimality of a set of more than two integers. After some preliminary research, I learned that for sets, there exists two forms of coprimality, which are explained in Section 2. Further research into prior work that has been done in this area lead me to my research question. The probability of coprimality of two integers is known, but I was curious about more than two integers. As mentioned, there are two types of coprimality of sets, the probability of one of which has already been determined, however I found no research answering the question that became my research question: *What is the probability* *that a set of randomly chosen integers is pairwise coprime and how can it be determined through experimental and theoretical methods?*

2. Background Information

2.1 Introduction to Coprimality

To understand the concept of coprimality and its variations, it is necessary to understand the concept of greatest common divisor (GCD). The GCD of b and c is defined as the largest integer a such that $a|b$ and $a|c$.

The above notation, $a|b$, denotes that a divides b, or in other words, a is a factor of b.

The greatest common divisor of a & b can be notated as (a, b).

Two positive integers are considered coprime if they share no common divisor, excluding 1. Written mathematically:

$(a, b) = 1 \therefore a$ & b are coprime. (Niven et al. 9: Definition 1.3).

This can also be stated as a and b are relatively prime, or prime to one another (Niven et al. 9: Definition 1.3). Some mathematicians prefer the term relatively prime, but in this essay, I will refer to integers being coprime.

This definition applies to pairs of integers, however when addressing more than 2 integers, two separate cases of coprimality must be defined.

The first and simpler case is called setwise coprime, or mutually relatively prime. A set of integers $a_1, a_2, \dots, a_k$ is setwise coprime in case $(a_1, a_2, \dots, a_k) = 1$. Namely there is no divisor common to all members of the set (Niven et al. 9: Definition 1.3; Rosen 98-9).

The second case is called pairwise coprime, or relatively prime in pairs. Consider a set of integers $a_1, a_2, \dots, a_k$. Consider a_i and a_j to be any two members of set. The entire set is pairwise coprime if $(a_i, a_j) = 1$, for all possible pairs of integers in the set. Namely every member of the set is coprime to every other member of the set, individually (Niven et al. 9: Definition 1.3; Rosen 98-9).

A helpful way to think of coprimality is by noticing that for a fraction to be in its simplest form, the numerator and the denominator must be coprime. Namely two positive integers p and q, are coprime iff $\frac{p}{q}$ is a fraction that cannot be simplified any further (Hardy et al. 353-4). This fact is used later, in the proof in Section 2.2, and in my calculations is Section 3.4.

2.2 Probability of Coprimality

It is possible to consider the probability that two randomly chosen positive integers are coprime. This problem was solved by Ernesto Cesàro in 1881 (A. Gambinia et al.)

The details of the proof are beyond the scope of this essay, but I will include a brief explanation of the proof.

The proof consists of, for a given n, considering all possible fractions $\frac{p}{q}$ for which

$$1 \leq p \leq q \leq n, \quad p, q, n \in \mathbb{Z}^+$$

Let β equal the percentage of those fractions that are in their simplest form. (This proof can be found in Hardy et al. Theorem 332). As $n \to \infty$, β approaches $\frac{1}{\zeta(2)}$. This ζ represents the Riemann Zeta Function, which is defined as follows

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

$\zeta(2)$ was solved by Leonhard Euler in 1734. It was found to be exactly $\frac{\pi^2}{6}$. (Ayoub)

Thus, the probability that two randomly selected integers are coprime is $\frac{6}{\pi^2} \approx 0.608$. Parker used this fact to approximate π. As mentioned previously, he generated pairs of random numbers and counted what proportion of those pairs were coprime. Through simple calculation, he was able to approximate a value for π.

This can be extended to state that the probability that a set of k randomly chosen positive integers is setwise coprime is $\frac{1}{\zeta(k)}$. It is important to note that this applies only to setwise coprimality. My work is focussed on pairwise coprimality.

3. Body of Work

3.1 Possible approaches

When I initially began thinking about the problem of the probability of pairwise coprimality, I thought of two methods, namely an experimental method, similar to Parker's, and a theoretical method.

The experimental method would consist of generating random sets of integers and counting what percentage of them are coprime. This method seemed very practical, however not extremely accurate, as no constant exact value would be determined, due to the use of random numbers. However, I decided to use this method, knowing it would not provide an exact answer, but could be used to verify other methods.

My original thoughts on a theoretical method involved looking at the probability that a single pair of integers from the set are coprime and using simple probability equations to extend this to all possible pairings of integers. From there it would be possible to find the probability that a set of randomly chosen integers is pairwise coprime. However, I found flaws in this method, and then decided to approach the problem from a different angle. The theoretical method considers that every prime number, can only divide 0 or 1 integer in the set. If a prime number could divide 2 or more, the set would not be pairwise coprime. This is explained fully in Section 3.4.

3.2 Experimental Methodology

To investigate the probability of pairwise coprimality experimentally, I decided to use Microsoft Office Excel, as I knew it could generate random integers, and check the lowest common denominator of sets of integers, as well as handle very large quantities of data, which were all things I would need. This method is based on a method used by Parker in his work. Due to the computational limitations of the computers available to me, I was limited to experimenting with sets of 2 to 7 integers, the integers being between 1 and 1 000 000, and 1 000 000 trial sets per value of k, where k is how many integers are in each set. Those were the largest parameters possible, before the computer started to struggle with the large amounts of data.

In order to explain my experimental procedure, I will describe and explain the calculations for $k = 3$, however the process is much the same for all other values of k.

	A	B	C	F	G	H	I	J
1	a_1	a_2	a_3		(a_1, a_2)	(a_1, a_3)	(a_2, a_3)	Pairwise Coprime
2	71156	112591	531224		1	4	1	FALSE
3	750341	330649	240856		1	1	11	FALSE
4	644437	933833	896984		1	1	1	TRUE
5	753640	656283	338521		1	1	1	TRUE

Figure 1 Experimental Table (Source: Author's own)

Figure 1 shows the Excel Spreadsheet I created. In this spreadsheet, columns A, B and C are showing randomly generated integers between 1 and 1 000 000. Columns G, H and I are showing the GCD of a_1 & a_2, a_1 & a_3 and a_2 & a_3, respectively. Lastly column J is showing "TRUE" if all three values in columns G, H and I are equal to one, meaning that each pair of integers is coprime, and therefore the set is pairwise coprime. Otherwise column J shows "FALSE". The data collection of my experimentation consists of the program counting all the "TRUE" values in column J, and dividing that by the total number of sets, to calculate the proportion of sets that are pairwise coprime. This method is an extension of the method used by Parker. Parker's method however only used pairs of integers, and I had to create many more steps, namely comparing the numbers in each column with the numbers in every other column, rather than comparing only 2 columns, as Parker had done.

3.3 Experimental Results

The results from the above-mentioned experiment appear in Table 1, with k denoting the size of the set. Experimental Probability is given as the percentage, shown as a decimal. i.e. 0.609018 denotes that 60.9018% of the sets were coprime.

Table 1 Experimental Results (Source: Author's own)

k	Experimental Probability
2	0.609018
3	0.287004
4	0.114804
5	0.041083
6	0.013238
7	0.004064

As seen in Table 1, the experimental probability decreases as k increases. This is coherent with a logical qualitative explanation. As a set becomes larger, the probability that each and every pair is coprime decreases, as there are simply more pairs that could be not coprime.

3.4 Theoretical Calculations

To calculate the probability of pairwise coprimality theoretically, I devised the following method. As stated in section 2.1, for a set of integers to be pairwise coprime, each possible pair of integers from the set must be coprime. This can also be looked at from the perspective of divisibility by primes; a set is pairwise coprime if for each prime, p, $p \mid a_i$ is true for no more than one value of i. Namely no prime can divide multiple members of the set, as if this were the case, those members of the set would not be coprime because this prime would be a common factor.

For each prime, this can be split into two cases:

- $p \nmid a_i \; \forall \; i \in \; \mathbb{Z}^+$ i.e. the given prime cannot divide any member of the set
- $p \mid a_i$ for only one value of i, i.e. the given prime can divide only one member of the set

The probability that a randomly chosen integer is divisible by prime p is equal to $\frac{1}{p}$. This can be deduced as follows; $\frac{1}{2}$ of all integers are divisible by 2, $\frac{1}{3}$ of all integers are divisible by 3 and so on.

Using the binomial distribution, the probability of each of the above-mentioned cases can be calculated. The binomial distribution is used to calculate the probability of a certain number of "successes" occurring, out of a given number of "trials", when each event is completely independent and has equal probability. The binomial distribution, B, is given by

$$P(X = r) = \binom{n}{r} \times p^r \times (1 - p)^{n-r}$$

Where

$$X \sim B(r, n, p)$$

where r is the number of successes, n is the number of trials and p is the probability of success in a single trial.

It is convention to denote the entire probability function with P and an individual simple probability with p, as is used in the definition of the binomial distribution above. In this essay, I will deviate slightly from this convention. I will still use P to denote the entire probabilistic function, however p will denote a prime number. In places where p would normally be used to denote a simple probability I will use ϕ instead.

For the case where no members of the set are divisible by a prime, p, the binomial distribution can be used as follows: n will be k, the size of the set; r, will be 0, as no members will be divisible by p; and ϕ will be $\frac{1}{p}$, the probability that a random integer is divisible by p.

Let $X =$ the number of integers in the set divisible by p.

$$P(X = 0) = \binom{k}{0} \times \left(\frac{1}{p}\right)^{0} \times \left(1 - \frac{1}{p}\right)^{k-0}$$

$$P(X = 0) = 1 \times 1 \times \left(1 - \frac{1}{p}\right)^{k}$$

$$P(X = 0) = \left(1 - \frac{1}{p}\right)^{k}$$

For the second case where one member is divisible by p, the binomial distribution can be used as follows:

as before n will be the size of the set, k, and ϕ will be the probability of a random integer being divisible by p, which is $\frac{1}{p}$, and lastly r, will be 1 as only a single member of the set must be divisible by p. Thus

$$P(X = 1) = \binom{k}{1} \times \left(\frac{1}{p}\right)^{1} \times \left(1 - \frac{1}{p}\right)^{k-1}$$

$$P(X = 1) = k \times \frac{1}{p} \times \left(1 - \frac{1}{p}\right)^{k-1}$$

The total probability that a given prime p, divides no more than one member of a set of k randomly chosen positive integers is the sum of these two probabilities; the probabilities of the two aforementioned cases. Thus

$$P(X \leq 1) = \left(1 - \frac{1}{p}\right)^{k} + k \times \frac{1}{p} \times \left(1 - \frac{1}{p}\right)^{k-1}$$

$$P(X \leq 1) = \left(1 - \frac{1}{p}\right)^{k-1} \times \left(1 - \frac{1}{p}\right) + k \times \frac{1}{p} \times \left(1 - \frac{1}{p}\right)^{k-1}$$

$$P(X \leq 1) = \left(1 - \frac{1}{p}\right)^{k-1} \left(\left(1 - \frac{1}{p}\right) + \frac{k}{p}\right)$$

$$P(X \leq 1) = \left(1 - \frac{1}{p}\right)^{k-1} \left(1 + \frac{k-1}{p}\right)$$

This expression shows the probability that a given prime p divides no more than one member of the set. It follows that to calculate the probability that a randomly selected set of integers is pairwise coprime, the above probability equation must be calculated for all primes, and the resulting probabilities multiplied. For this, an infinite product, indexed by primes, can be used. Thus follows

$$C_{pair}(k) = \prod_{primes\ p} \left(1 - \frac{1}{p}\right)^{k-1} \left(1 + \frac{k-1}{p}\right)$$

$C_{pair}(k)$ denotes the probability that a set of k integers is pairwise coprime. The Π symbol is used to show that the result of the above expression for $P(X \leq 1)$ for each prime p must be multiplied. This is similar to Σ used to show an infinite sum.

This type of infinite product is known as an Euler product, and it is the answer to my original research question. This expression communicates how to calculate it, however in this form it is not particularly useful or understandable.

I do not have the ability to manipulate this Euler product into a more understandable form, and one obviously cannot simply calculate an infinite product. However, if each term of the infinite product converges to 1, an approximate value of the infinite product to be determined, by calculating a finite number of terms of the product. The limit of the expression is calculated below:

$$\lim_{p\to\infty}\left(\left(1-\frac{1}{p}\right)^{k-1}\left(1+\frac{k-1}{p}\right)\right)$$

$$=\lim_{p\to\infty}\left(1-\frac{1}{p}\right)^{k-1}\times\lim_{p\to\infty}\left(1+\frac{k-1}{p}\right)$$

$$=\left(\lim_{p\to\infty}(1)-\lim_{p\to\infty}\left(\frac{1}{p}\right)\right)^{k-1}\times\left(\lim_{p\to\infty}(1)+\lim_{p\to\infty}\left(\frac{k-1}{p}\right)\right)$$

$$=(1-0)^{k-1}\times(1+0)$$

$$=1\times 1$$

$$=1$$

This convergence is intuitive, because as primes get larger, it is very unlikely that they will be a factor of a random integer.

As $P(X \leq 1)$ converges to 1, $\prod_{primes\ p}\left(1-\frac{1}{p}\right)^{k-1}\left(1+\frac{k-1}{p}\right)$ is convergent.

The expression also converges very quickly, and therefore the value of the whole expression for $C_{pair}(k)$ will converge towards the final answer after very few iterations of the infinite product. For example, the term of the 40th prime, 173, is already at 0.9999000 for $k = 3$. Therefore, it is clear that the terms after the 5000th prime will have a very small effect on the probability of pairwise coprimality.

Using the first 5000 primes, the probabilities of pairwise coprimality I calculated appear in Table 2.

Table 2 Approximation of Calculated Results Using 5000 Terms of Infinite Product (Source: Author's own)

k	$C_{pair}(k)$
2	0.60792817
3	0.28674894
4	0.11488525
5	0.04093101
6	0.01332500
7	0.00403473

This values clearly decrease as the size of the set increases, which is intuitive, as larger sets, are more likely to contain a coprime pair.

4. Conclusion

4.1 Final Results

I have found, theoretically, that the probability that a set of k randomly chosen positive integers is pairwise coprime is given by the following expression

$$C_{pair}(k) = \prod_{primes\ p} \left(1 - \frac{1}{p}\right)^{k-1} \left(1 + \frac{k-1}{p}\right)$$

The results of my experimental method are shown in Table 1.

4.2 Evaluating Results

In order to assess the validity of my two differing methods, experimental and theoretical, I can compare each method's results, and assess the difference between them. If the probability function is valid, there should be very little difference between the experimentally observed probabilities and expected probabilities generated by the expression I derived. It is clearly not possible to compare the equation with the table of experimental results. Rather I will compare the approximations I created using the first 5000 terms of the infinite product, with my experimental results.

See Table 3 for a comparison of the results from both methods.

Table 3 Summary of Results (Source: Author's own)

k	Theoretical	$Experimental$
2	0.60792817	0.609018
3	0.28674894	0.287004
4	0.11488525	0.114804
5	0.04093101	0.041083
6	0.01332500	0.013238
7	0.00403473	0.004064

From Table 3, it is clear that the results appear to be very close, but for a more accurate measure of how close they are, detailed statistical analysis is required. For this analysis, I will conduct a goodness of fit test using the Chi-Squared distribution to assess the probability associated with the test statistic. This calculates the probability that the variables obtained are due to a correlation or by chance. The Chi-Squared Statistic is calculated by the following equation

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where n is the number of data points, O_i is the observed value and E_i is the expected value. In this case, the experimental results can be treated as the observed value, the theoretical results can be treated as the expected results and $n = 6$ as $2 \leq k \leq 7$.

Calculating the term for each value of k gives the results shown in Table 4.

Table 4 Calculation of Chi-Squared Statistic (Source: Author's own)

k	$\frac{(O_i - E_i)^2}{E_i}$
2	1.95406×10^{-6}
3	2.26671×10^{-7}
4	5.75029×10^{-8}
5	5.62300×10^{-7}
6	5.71750×10^{-7}
7	2.10865×10^{-7}

This gives a chi squared statistic of 3.58315×10^{-6}. From this a p-value can be calculated, using a distribution table. This p-value gives the probability that the correlation observed is due to pure coincidence. I calculated a p-value of

1.2927×10^{-15}. In most cases statisticians require a confidence interval (the p-value required to state that there is correlation) , and this is often around 0.05. The p-value from my investigation is far within any reasonable confidence interval, thus showing an extremely high and almost exact correlation between my experimental and theoretical results.

The validity of my theoretical calculations is further verified by the fact that my value for the probability of pairwise coprimality for two integers is the same as the probability of two integers being setwise coprime. This is because setwise and pairwise coprimality are exactly the same when there are only two integers in the set.

Therefore, I can conclude that my methods of calculation and experimentation are valid, and mathematically sound.

4.3 Limitations of my methods

There are some limitations to my method. Firstly, addressing my experimental methodology, due to the limitations of the technology available to me, I was limited to experimenting with sets of 2 to 7 integers, the integers being between 1 and 1 000 000, and 1 000 000 trial sets per value of k. A more rigorous investigation would require far more trials, sets with more integers and much higher numbers. As this experiment deals with random numbers, if there are too few trials, large variation could be present. However, my use of 1 000 000 trials significantly reduces these variations.

Since I only calculated the Euler product for the first 5000 terms, there are limitations to my calculations. This is only an approximation and will not yield an exact answer.

This work is by no means a proof of my equation for the probability of pairwise coprimality, therefore I cannot say with certainty that it is accurate. However, the extremely low p-value of my Chi-Squared goodness of fit test shows a distinct correlation and serves as confirmation of the validity of this work.

4.4 Unresolved and New Questions

This work prompts several new questions and possible areas of further investigation. For example, as I have investigated coprimality in pairs, it could be possible to investigate *triplewise*, or *quadruplewise* coprimality. This could be extended to find the probability of *n-tuplewise* coprimality. To clarify, *n-tuplewise* coprimality, meaning that there is no common factor between any subset of n integers of the set. This would be done, using the theoretical method, largely in the same way as for pairwise coprimality, however instead of calculating the probability that a prime divides one or no members of the set, the probability of a prime dividing $0, 1, 2, \ldots$ or $n-1$ members of the set.

Of course, a clear further line of inquiry is a proof of my equation. Such a proof would show with certainty that my answer to the research question is correct.

5. Works Cited

Ayoub, Raymond. "Euler and the Zeta Function." The American Mathematical Monthly, vol. 81, no. 10, 1974, pp. 1067-86, doi:10.2307/2319041. Accessed 4 Aug. 2019.

Gambinia, Alessandro, et al. "Probability of digits by dividing random numbers: A ψ and ζ functions approach." Expositiones Mathematicae, vol. 30, no. 3, 2012, pp. 223-38, doi:10.1016/j.exmath.2012.03.001. Accessed 4 Aug. 2019.

Hardy, G. H., et al. An Introduction to the Theory of Numbers. 6th ed., Oxford, Oxford Univ. Press, 2008.

Niven, Ivan Morton, et al. An Introduction to the Theory of Numbers. 5th ed., Singapore, Wiley, 1991.

Parker, Matt, producer. "Generating π from 1,000 random numbers." Stand Up Maths, 13 Mar. 2017, www.youtube.com/watch?v=RZBhSi_PwHU. Accessed 4 Aug. 2019.

Rosen, Kenneth H. Elementary Number Theory. 6th ed., Upper Saddle River, Pearson Education, 2011.