Texas A&M University

Summer Research Paper 1

Exploring the Euclidean Algorithm

Authors:
Stephen Capps
Sarah Sahibzada
Taylor Wilson

Supervisor:
Dr. Sarah Pollock

June 10, 2015

Contents

1 Introduction

The Euclidean Algorithm is familiar to most, and is widely used in many fields, such as cryptography and number theory. Aging over two millennia, it has raised many questions over its lifetime. This paper will specifically investigate the number of iterations it takes to complete the Euclidean Algorithm.

For example, the number of iterations it take to complete $\gcd(42, 36)$ is 2, as shown below.

$$\gcd(42, 36) = 6 :$$

$$42 = 1 * 36 + 6 \tag{1}$$

$$36 = 6 * 6 + 0 \tag{2}$$

Some other examples are as follows:

$$\gcd(689, 456) = 1 \quad | \quad \text{Iterations} : 6$$

$$\gcd(78, 45) = 3 \quad | \quad \text{Iterations} : 5$$

$$\gcd(8394, 238) = 2 \quad | \quad \text{Iteration} : 7$$

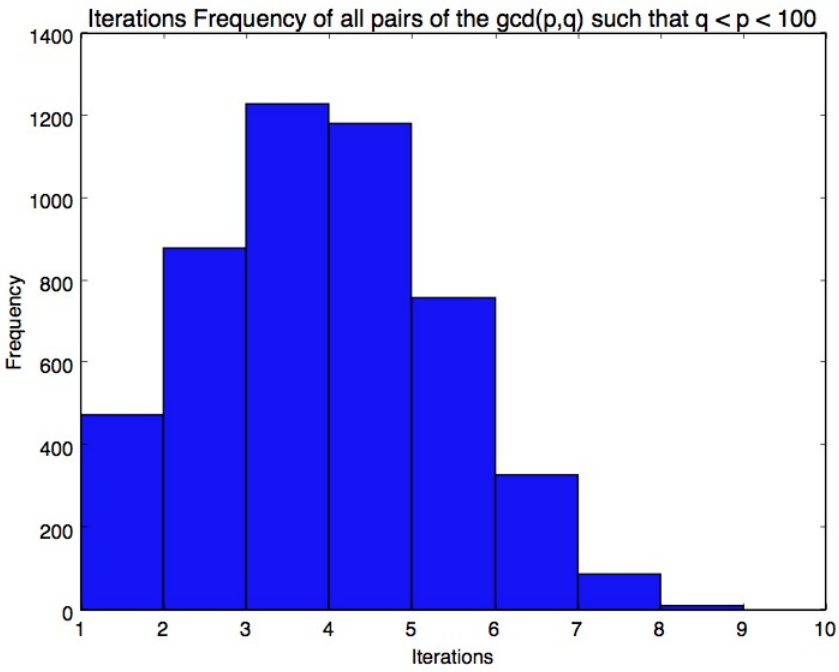
This leads us to an investigation into what numbers yield the longest iterations, and the distribution of these iterations.

2 Theoretical Analysis

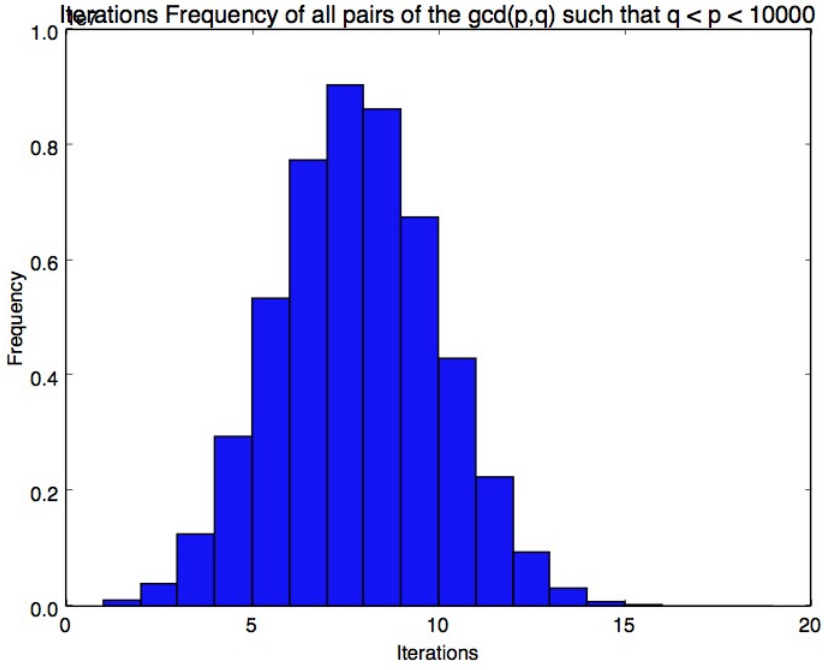
3 Computational Approaches

4 Results

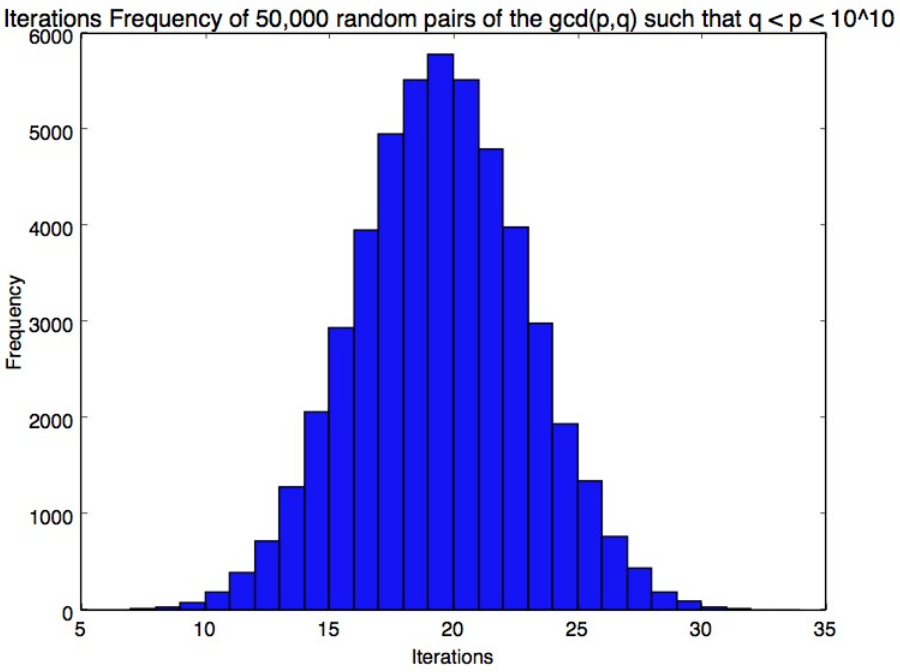
The results were quiet surprising, as most distributions were almost perfectly Gaussian. The following figures are different distributions of different iterations of various gcd combinations.



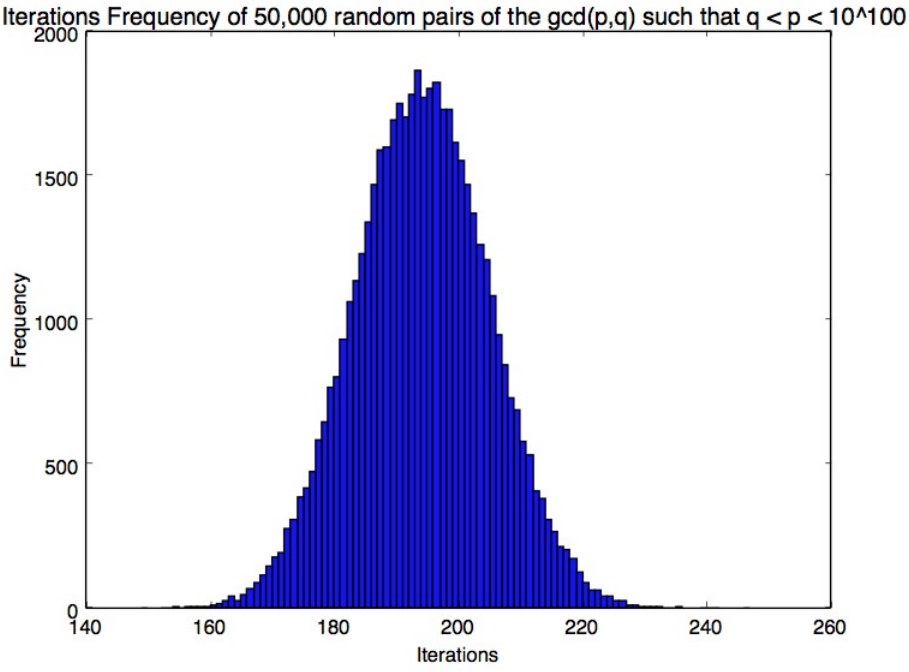
(Figure 1)



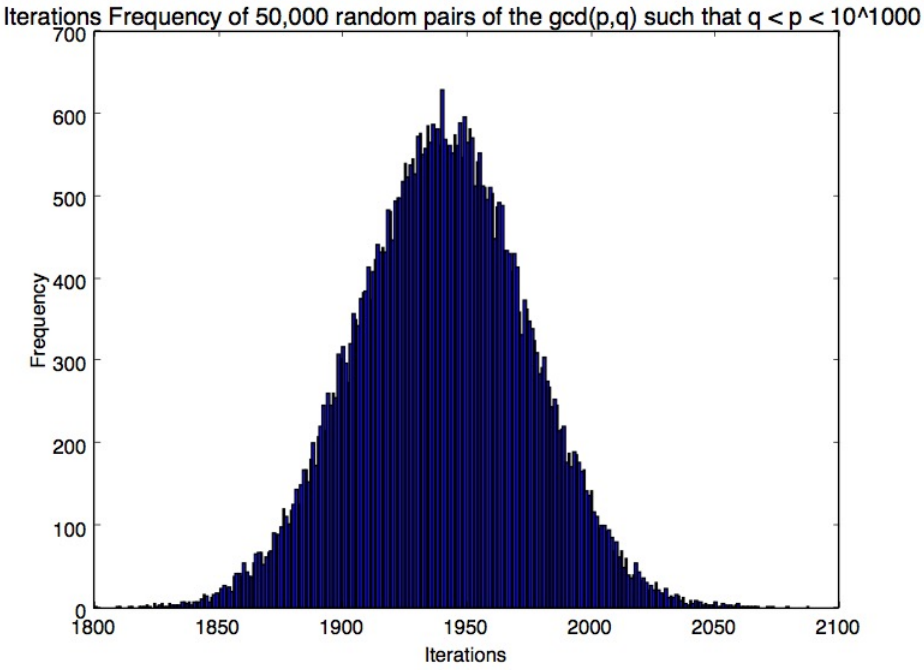
(Figure 2)



(Figure 3)



(Figure 4)


Iterations Frequency of 50,000 random pairs of the gcd(p,q) such that q < p < 10^1000
Frequency
Iterations
0
100
200
300
400
500
600
700
1800
1850
1900
1950
2000
2050
2100

(Figure 5)

As the number of digits we consider increases, the distribution becomes continuously more Gaussian. However, after a point our computational resources restrict us from checking all pairs p,q with some large upper bound. Thus, we restrict ourselves by picking a set amount of pairs and calculating their distribution. You can see this progression as the figures continue.

There are a few minor observations to be made:

First, in (Figure 1), it must be noted that the lack of normality here is due to the small sample size. The size of this data set was no more than 4950, and spanned across 9 bins. As the number of bins needed increases, the more normal the graph becomes.

Second, in (Figure 5), the inconsistencies in the normal distribution can be attributed to a too small a sample size. If given the computing power and time, one could compute all pairs less than 10^{1000}. Note the increasing mean iterations as we climb the upper bound.

5 Discussion

As it is obvious, the distribution of these gcd lengths is almost perfectly Gaussian. The approach we used seemed to work awfully well, as computers handle modular operations and subtraction very well. The main constraints we had were as we went past 5 digit numbers. Calculating all pairs becomes exponentially difficult as the digits you allow increases.

6 Individual Contributions