

# GENERALIZATIONS OF ARTIN'S CONJECTURE

KHALED ELGOHARY, MARCUS LAI

**ABSTRACT.** Artin's Conjecture on primitive roots states that a given integer  $a$  that is not a perfect square nor  $-1$  is a primitive root modulo infinitely many primes [1]. Moreover, if  $a$  is not a perfect power and is not congruent to 1 modulo 4, the density of primes  $p$  for which  $a$  is a primitive root modulo  $p$  is described by Artin's Constant [1][2][3]. The Goldmakher–Martin Conjecture attempts to capture a generalized version of this relationship for integers  $a \equiv 3 \pmod{4}$ . In this report, we reinforce the Goldmakher–Martin Conjecture by testing it against empirical data and demonstrating a high statistical correlation.

## 1. INTRODUCTION

Artin's original Conjecture on primitive roots states that any integer that is not a perfect square nor  $-1$  is a primitive root modulo  $p$  for infinitely many primes  $p$  [1]. Further that for any integer  $a$  that is not a perfect power, the density of primes for which  $a$  is a primitive root modulo  $p$  is *independent* of  $a$  and is given by:

$$\prod_q \left(1 - \frac{1}{q(q-1)}\right) \approx 0.3739558136$$

known as Artin's constant where the product is over all primes  $q$ . This conjecture is revised to require that  $a \not\equiv 1 \pmod{4}$  when D.H.Lemer revealed that the formula did not predict the values from empirical data [2][3]. Artin's Conjecture remained widely open until 1967 when Chistopher Hooley published a conditional proof under the assumption of a version of generalized Riemann Hypothesis [2].

The Goldmakher–Martin Conjecture proposes a set of similar infinite products to capture a more generalized version of Artin's Conjecture that considers not only whether  $a$  is a primitive root modulo  $p$ , but also scoring and grouping each prime by *how far* it is such that a fixed  $a$  is a primitive root mod  $p$ . For a similar reason, the Goldmakher–Martin Conjecture is proposed only for integers  $a$  such that  $a$  is not a perfect power and  $a \equiv 3 \pmod{4}$ . Throughout the paper, let:

**Definition 1.**  $\omega(n)$  denote the number of distinct prime factors of  $n$ . Also let  $\Omega(n)$  denote the number of prime factors of  $n$  counted with multiplicity. For example, with  $n = 40 = 2^3 \times 5$ , we have  $\Omega(n) = 4$  but  $\omega(n) = 2$ . By convention,  $\Omega(1) = \omega(1) = 0$ .

For simplicity, let all integers  $a$  be congruent to 3 modulo 4 for the remainder of this report. With these definitions, the Goldmakher–Martin Conjecture can be stated:

**Conjecture 2** (Goldmakher–Martin). *For any complex number  $z$  with  $|z| < 1$ , and for any prime  $p$ ,*

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\Omega(p-1) - \Omega(\text{ord}_p(a))} &= \prod_q \frac{q^3 - q^2 - q + z}{(q-1)(q^2 - z)}; \\ \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega((p-1)/\text{ord}_p(a))} &= \prod_q \left( 1 + \frac{z-1}{q(q-1)} \right); \\ \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega(p-1) - \omega(\text{ord}_p(a))} &= \prod_q \left( 1 + \frac{z-1}{q^2-1} \right).\end{aligned}$$

Here the sums on the left-hand sides are over primes  $p$ , and the products on the right-hand sides are over all primes  $q$ .

In each version of the conjecture, the left-hand side is a generating function whose terms group the primes by the exponent of the sum and count the primes in each group by the coefficient. The division by  $\pi(x)$ , which denotes the number of primes up to  $x$ , normalizes each coefficient into the percentage of primes in each group. The exponents themselves are alternative versions of measuring how far the primes  $p$  are such that a fixed  $a$  is a primitive root mod  $p$ . For instance,  $\Omega(p-1) - \Omega(\text{ord}_p(a)) = 0$  and  $\omega((p-1)/\text{ord}_p(a)) = 0$  are both equivalent to  $\text{ord}_p(a) = p-1$ . Thus, the constant term for those two versions of the conjecture count the number of primes that are exactly primitive root mod a certain  $a$ .

A special case of the conjecture is  $z = 0$ , using the convention that  $0^0 = 1$  while  $0^n = 0$  for  $n \geq 1$ . Then the left-hand sides just count primes with a particular property related to  $\text{ord}_p(a)$ :

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \#\{p \leq x : \Omega(p-1) - \Omega(\text{ord}_p(a)) = 0\} &= \prod_q \frac{q^3 - q^2 - q + 0}{(q-1)(q^2 - 0)} = \prod_q \left( 1 - \frac{1}{q(q-1)} \right); \\ \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \#\{p \leq x : \omega((p-1)/\text{ord}_p(a)) = 0\} &= \prod_q \left( 1 + \frac{0-1}{q(q-1)} \right) = \prod_q \left( 1 - \frac{1}{q(q-1)} \right); \\ \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \#\{p \leq x : \omega(p-1) - \omega(\text{ord}_p(a)) = 0\} &= \prod_q \left( 1 + \frac{0-1}{q^2-1} \right) = \prod_q \left( 1 - \frac{1}{q^2-1} \right).\end{aligned}\tag{1}$$

Indeed, the first two statements is precisely Artin's classical conjecture about primitive roots (which are numbers  $a$  such that  $\text{ord}_p(a) = p-1$ ).

This report seeks to compute the densities described by the conjecture and compare the results to empirical data to test the conjecture.

**Structure of Paper.** In Section 2, we study the Goldmakher–Martin Conjecture and derive a method to compute the proposed densities of primes and error bounds from the infinite products. In Section 3, we describe the process of empirical data collection and provide sample visualizations of empirical results. In Section 4, we compare the empirical and analytical data statically and discuss the result and its implications.

**Acknowledgements.** This research is made possible through the PIMS VXML program and the guidance of Professor Greg Martin and Kyle Yip. Some material from this report is developed collaboratively with the entire PIMS VXML team, such as the coefficients table in section 2.1.

## 2. CONJECTURE ANALYSIS

**2.1. Obtaining Model Coefficients.** The Goldmakher-Martin Conjecture group prime numbers by how far off they are such that a fixed  $a$  is a primitive root mod  $p$  with generating functions and proposes a corresponding set of infinite products that describe the model. However, a method for computing the coefficients of the generating function from the infinite products is necessary to understand the implications of the conjecture and for later comparison against empirical results. This section will derive such a method using the  $\omega((p-1)/\text{ord}_p(a))$  version of the conjecture. Note that this method can be applied similarly to the other two versions of the conjecture to obtain the coefficients from those models. Recall the  $\omega((p-1)/\text{ord}_p(a))$  version of the conjecture:

**Conjecture.** Goldmakher-Martin Conjecture —  $\omega((p-1)/\text{ord}_p(a))$  version

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega((p-1)/\text{ord}_p(a))} = \prod_q \left(1 + \frac{z-1}{q(q-1)}\right)$$

By Fermat's little theorem,  $p-1 \mid \text{ord}_p(a) \implies (p-1)/\text{ord}_p(a) \in \mathbf{Z}$ . Thus, the left-hand sum would only yield terms with integer powers and we can express the left-hand side of the conjecture as a general power series. Let  $f(z)$  denote this general representation and  $c_i$  for  $i \in \mathbf{Z}, i \geq 0$  denote the coefficients of  $f(z)$ .

$$f(z) = \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega((p-1)/\text{ord}_p(a))} = \sum_{i=0}^{\infty} c_i z^i = c_0 + c_1 z^1 + c_2 z^2 \dots$$

Now, we note that if we choose  $z = 0$ , we are left with the constant term of the power series.

$$c_0 = f(0)$$

To get further coefficients, we simply differentiate  $f$ , plug in  $z = 0$ , and divide by the constant factor incurred when performing the differentiation. This gives that:

$$c_i = \frac{f^{(i)}(0)}{i!}$$

Next, we can apply the Goldmakher-Martin Conjecture to express these coefficients in terms of the proposed infinite products which we can compute. The "logarithmic derivative" trick is applied for the first derivative to turn the infinite product into a sum to be linearly differentiated. From there, further derivatives simply employ the product rule. We showcase this iterative process by demonstrating the derivation of the formulas for  $c_0, c_1, c_2$ :

$c_0$ :

$$c_0 = f(0) = \prod_q \left(1 + \frac{0-1}{q(q-1)}\right) = \prod_q \left(1 - \frac{1}{q(q-1)}\right)$$

$c_1$  (obtained with the "logarithmic derivative" trick):

$$c_1 = \frac{f'(0)}{1!} = f(0) \cdot \frac{f'(0)}{f(0)} = f(0) \cdot \left( \frac{d}{dz} \log(f(z)) \right) \Big|_{z=0}$$

$$\begin{aligned}
&= c_0 \cdot \left( \frac{d}{dz} \sum_q \log \left( 1 + \frac{z-1}{q(q-1)} \right) \right) \Big|_{z=0} = c_0 \cdot \left( \sum_q \frac{\frac{1}{q(q-1)}}{1 + \frac{z-1}{q(q-1)}} \right) \Big|_{z=0} \\
&= c_0 \cdot \sum_q \frac{\frac{1}{q(q-1)}}{1 - \frac{1}{q(q-1)}} = c_0 \cdot \sum_q \frac{1}{q^2 - q - 1}
\end{aligned}$$

$c_2$  (obtained with product rule):

$$\begin{aligned}
f'(z) &= f(z) \cdot \frac{f'(z)}{f(z)} = f(z) \cdot \left( \frac{d}{dz} \log(f(z)) \right) \\
\implies f''(z) &= f'(z) \cdot \left( \frac{d}{dz} \log(f(z)) \right) + f(z) \cdot \left( \frac{d^2}{dz^2} \log(f(z)) \right) \\
f''(0) &= c_1 \cdot \left( \frac{d}{dz} \log(f(z)) \right) \Big|_{z=0} + c_0 \cdot \left( \frac{d^2}{dz^2} \log(f(z)) \right) \Big|_{z=0}
\end{aligned}$$

We can substitute in earlier calculations from  $c_1$ , giving us:

$$\begin{aligned}
f''(0) &= c_1 \cdot \sum_q \frac{1}{q^2 - q - 1} + c_0 \cdot \left( \frac{d}{dz} \sum_q \frac{\frac{1}{q(q-1)}}{1 + \frac{z-1}{q(q-1)}} \right) \Big|_{z=0} \\
f''(0) &= c_1 \cdot \sum_q \frac{1}{q^2 - q - 1} + c_0 \cdot \left( \frac{d}{dz} \sum_q \frac{1}{q^2 - q + z - 1} \right) \Big|_{z=0} \\
f''(0) &= c_1 \cdot \sum_q \frac{1}{q^2 - q - 1} + c_0 \cdot \left( \sum_q \frac{1}{(q^2 - q + z - 1)^2} \right) \Big|_{z=0} \\
f''(0) &= c_1 \cdot \sum_q \frac{1}{q^2 - q - 1} + c_0 \cdot \sum_q \frac{1}{(q^2 - q - 1)^2} \\
c_2 &= \frac{f''(0)}{2!} = \frac{1}{2} \left( c_1 \cdot \sum_q \frac{1}{q^2 - q - 1} + c_0 \cdot \sum_q \frac{-1}{(q^2 - q - 1)^2} \right)
\end{aligned}$$

Formulas for coefficients beyond  $c_2$  are obtained similarly to the derivation of the  $c_2$  formula where the function  $f^{(i-1)}(z)$  is obtained and differentiated for  $f^{(i)}(z)$ . Then,  $\frac{f^{(i)}(z)}{i!} = c_i$  would be computed. It's interesting to note that the majority of the terms of a new iteration are composed of previous calculations except for one new derivative.

FIGURE 1. Computed Conjecture Coefficients

$n$	$C_{\Omega}(n)$	$C_{\prime}(n)$	$C_{-}(n)$
0	0.373956	0.373956	0.530712
1	0.387003	0.489829	0.391986
2	0.16705	0.125687	0.0723496
3	0.052465	0.0101645	0.00480193
4	0.0144668	0.000356456	0.000147114
5	0.00377443	$6.48829 \times 10^{-6}$	$2.4291 \times 10^{-6}$
6	0.000961468	$6.98718 \times 10^{-8}$	$2.42589 \times 10^{-8}$
7	0.000242365	$4.86185 \times 10^{-10}$	$1.58808 \times 10^{-10}$
8	0.000060814	$2.32726 \times 10^{-12}$	$7.22469 \times 10^{-13}$
9	0.0000152283	$8.04224 \times 10^{-15}$	$2.38783 \times 10^{-15}$
10	$3.80981 \times 10^{-6}$	$3.039 \times 10^{-17}$	$5.90763 \times 10^{-18}$

**2.2. Obtaining Error Bound.** Of particular interest after we obtain computable formulas for the model coefficients is the accuracy of the formulas. Specifically, what is the error bound on the coefficients at a certain truncation of primes? In this subsection, we showcase a method to extract an error bound for the coefficients.

We recall from the previous subsection that each coefficient is a linear combination of infinite sums and one infinite product. Thus, we require a method to extract error bounds for them to obtain the error bounds. First, let us extract the error bound for one particular infinite sum which appears in the formula of  $c_1$ . The rest of the infinite sums are treated similarly to obtain their respective error bounds. Denote our sample sum as  $A$  and our truncation as  $T$ :

$$A = \sum_q \frac{1}{q^2 - q - 1}, \text{ where } A \text{ sums over all primes } q$$

$$T = \sum_{q \leq x} \frac{1}{q^2 - q - 1}$$

Then, the error,  $E$  is:

$$E = A - T = \sum_q \frac{1}{q^2 - q - 1} - \sum_{q \leq x} \frac{1}{q^2 - q - 1} = \sum_{q > x} \frac{1}{q^2 - q - 1}$$

An observation we can make is that the terms in the infinite sum are strictly positive since  $n^2 > n - 1$  for  $n \in \mathbf{Z}, n \geq 2$ . Thus we can bound:

$$E = \sum_{q > x} \frac{1}{q^2 - q - 1} < \sum_{n > x} \frac{1}{n^2 - n - 1}$$

Next, we approximate the right-hand side sum by an integral. Let  $f(t) = \frac{1}{t^2 - t - 1}$  for  $t \in \mathbf{R}, t \geq 2$ . Then each term of the sum is contained, in order, inside of  $f$ . Now, differentiate:

$$f'(t) = \frac{-(2t - 1)}{(t^2 - t - 1)^2} \text{ for } t \in [2, \infty)$$

Since there is no critical point within the defined range of  $f$ , and  $f'(2) < 0$ ,  $f$  and the sum must be ever decreasing. Note additionally that  $f$  is continuous within the defined range. By these conditions, we can apply the Integral Test to obtain a bound on the error:

$$E = \sum_{q>x} \frac{1}{q^2 - q - 1} < \int_x^\infty f(t)dt = \int_x^\infty \frac{1}{t^2 - t - 1} dt$$

Computing the indefinite integral gives:

$$\begin{aligned} \int_x^\infty \frac{1}{t^2 - t - 1} dt &= \int_x^\infty \frac{1}{(t - \frac{1+\sqrt{5}}{2})(t - \frac{1-\sqrt{5}}{2})} dt \\ &= \int_x^\infty \frac{\frac{1}{\sqrt{5}}}{(t - \frac{1+\sqrt{5}}{2})} - \frac{\frac{1}{\sqrt{5}}}{(t - \frac{1-\sqrt{5}}{2})} dt = \frac{1}{\sqrt{5}} \int_x^\infty \left( t - \frac{1+\sqrt{5}}{2} \right)^{-1} - \left( t - \frac{1-\sqrt{5}}{2} \right)^{-1} dt \\ &= \frac{1}{\sqrt{5}} \left( \log \left( t - \frac{1+\sqrt{5}}{2} \right) \Big|_x^\infty - \log \left( t - \frac{1-\sqrt{5}}{2} \right) \Big|_x^\infty \right) = \frac{1}{\sqrt{5}} \left( \log \left( \frac{t - \frac{1+\sqrt{5}}{2}}{t - \frac{1-\sqrt{5}}{2}} \right) \Big|_x^\infty \right) \\ &= \frac{1}{\sqrt{5}} \left( \lim_{t \rightarrow \infty} \log \left( \frac{t - \frac{1+\sqrt{5}}{2}}{t - \frac{1-\sqrt{5}}{2}} \right) - \log \left( \frac{x - \frac{1+\sqrt{5}}{2}}{x - \frac{1-\sqrt{5}}{2}} \right) \right) \\ &= \frac{-1}{\sqrt{5}} \log \left( \frac{x - \frac{1+\sqrt{5}}{2}}{x - \frac{1-\sqrt{5}}{2}} \right) \end{aligned}$$

This gives us an error bound on the infinite sum in terms of our truncation of choice. Next, let us extract the error bound for the infinite product in  $c_0$ . Denote  $A$  as this product,  $T$  as our truncation, and  $E$  as the error. Then:

$$A = \prod_q \left( 1 - \frac{1}{q(q-1)} \right), \text{ where } A \text{ is a product over all primes } q$$

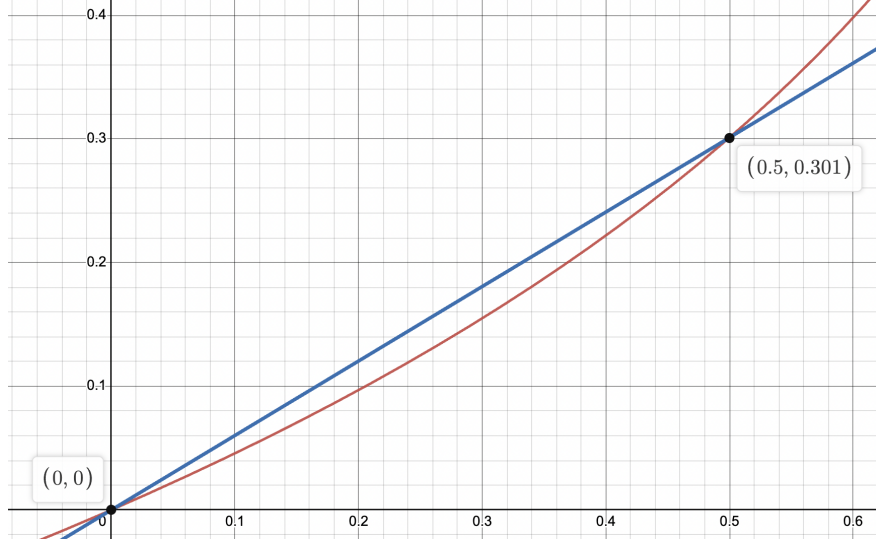
$$\begin{aligned} T &= \prod_{q \leq x} \left( 1 - \frac{1}{q(q-1)} \right) \\ E &= \frac{A}{T} = \prod_{q > x} \left( 1 - \frac{1}{q(q-1)} \right) \\ E^{-1} &= \prod_{q > x} \left( 1 - \frac{1}{q(q-1)} \right)^{-1} \\ -\log E &= \sum_{q > x} -\log \left( 1 - \frac{1}{q(q-1)} \right) \end{aligned}$$

Now, consider a function  $g(t) = -\log(1-t)$ . We observe that  $g(t)$  is a concave up on the interval  $(0, \frac{1}{2}]$ . This can be shown by analyzing the second derivative:

$$g''(t) = \frac{1}{(1-t)^2} > 0, \quad t \neq 1$$

As such, we can bound  $g(t)$  on  $(0, \frac{1}{2}]$  by the secant line between 0 and  $\frac{1}{2}$ :

FIGURE 2. Secant Line (in blue) as an Error Bound for  $-\log(1-t)$  (in red)



$$-2t \log\left(\frac{1}{2}\right) > g(t), 0 < t \leq \frac{1}{2}$$

Similarly, we can verify that  $\frac{1}{q(q-1)} \in (0, \frac{1}{2}]$  and apply our observation, giving:

$$-\log E = -2 \log\left(\frac{1}{2}\right) \sum_{q>x} \frac{1}{q(q-1)}$$

Using the techniques for infinite sums, we obtain:

$$\begin{aligned} -2 \log\left(\frac{1}{2}\right) \sum_{q>x} \frac{1}{q(q-1)} &< -2 \log\left(\frac{1}{2}\right) \int_x^\infty \frac{1}{t(t-1)} dt \\ -2 \log\left(\frac{1}{2}\right) \int_x^\infty \frac{1}{t(t-1)} dt &= -2 \log\left(\frac{1}{2}\right) \int_x^\infty -\frac{1}{t} + \frac{1}{t-1} dt \\ &= -2 \log\left(\frac{1}{2}\right) (-\log(t) + \log(t-1)) \Big|_x^\infty \\ &= -2 \log\left(\frac{1}{2}\right) \left(\log \frac{t-1}{t}\right) \Big|_x^\infty \\ &= 2 \log\left(\frac{1}{2}\right) \log\left(\frac{x-1}{x}\right) \end{aligned}$$

Recall this is a bound on  $-\log E$ :

$$-\log E < 2 \log\left(\frac{1}{2}\right) \log\left(\frac{x-1}{x}\right)$$

$$\log E > -2 \log\left(\frac{1}{2}\right) \log\left(\frac{x-1}{x}\right)$$

$$E > e^{-2 \log(\frac{1}{2}) \log(\frac{x-1}{x})}$$

Finally, we simply follow the rules of error propagation to obtain error bounds on our coefficients.

### 3. DATA COLLECTION / VISUALIZATION

**3.1. Data Collection.** Empirical data of the  $\text{ord}_p(a)$  was collected for many fixed integers  $a$  over many primes  $p$  using Sage[4]. In all, data for  $a \in [-400, 400]$  over primes  $0 < p < 1000000$  (roughly 80000 primes) was collected and analyzed. The code for data collection can be found [here](#).

FIGURE 3. Sample Data -  $a \in [-400, -390], p \in [900001, 900149]$

	-400	-399	-398	-397	-396	-395	-394	-393	-392	-391	-390
900001	1	0	0	0	0	0	0	0	0	0	0
900007	0	1	0	1	2	2	1	2	0	1	0
900019	0	0	2	1	0	2	0	0	2	0	0
900037	1	0	0	0	1	1	1	1	0	1	0
900061	0	1	1	0	0	1	1	2	0	0	0
900089	1	1	0	0	0	0	0	0	0	0	0
900091	1	2	1	1	2	2	2	0	3	1	1
900103	1	1	1	0	0	1	0	1	1	0	0
900121	1	1	1	1	0	2	1	1	0	1	1
900139	0	0	0	0	1	1	0	2	1	1	1
900143	0	0	1	0	0	1	1	0	0	1	1
900149	0	1	0	0	0	1	1	0	0	0	1

**3.2. Data Visualization.** For the sake of visually analyzing the data collected, a simulation was done in R to select a random  $a$  in  $\text{ord}_p(a)$ , in the interval  $[-400, 400]$ , and plot the frequency of the  $\text{ord}_p(a)$  on histograms. A part of the simulation of the code will be shown below along with all the histograms plotted.

```
# Read CSV file
ratio1 <- read.csv(fileName)
# Randomly select a column index that meets a certain condition
randomize <- 2:199
x <- sample(randomize, 1)
while (((ratio1[1, x] - 3) %% 4) != 0) {
  x <- sample(randomize, 1)
}
# Create vector of values and count non-'Inf' elements
```



```

values.a <- ratio1[, x]
counter <- 0
for (i in 2:length(ratio1[, x])) {
  if (values.a[i] != 'Inf') {
    counter <- counter + 1
  }
}
# Create vector of histogram values
hist.values <- c(1:counter)
counter2 <- 1
for (i in 2:length(ratio1[, x])) {
  if (values.a[i] != 'Inf') {
    hist.values[counter2] <- values.a[i]
    counter2 <- counter2 + 1
  }
}
# Plot histogram
hist(hist.values)
}

```

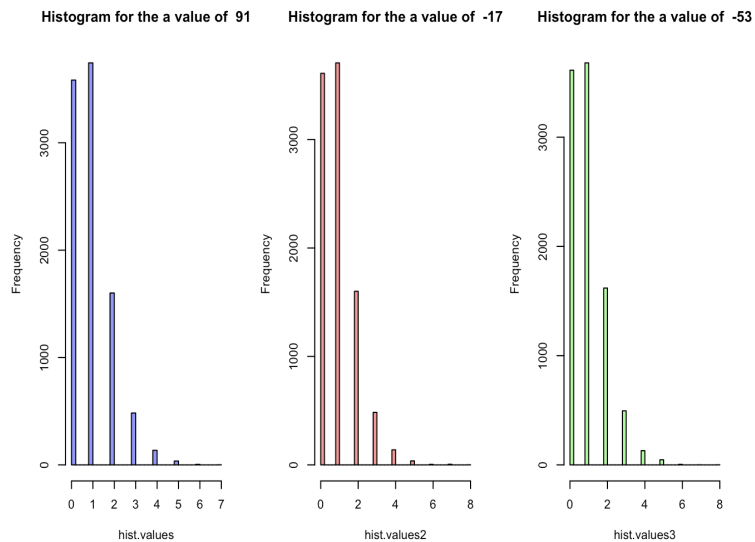
**Description:** The CSV file we are reading contains values of  $a$ , along with their order, what the simulation does is first randomize a number  $a$ , and then make sure that the selected value is  $\equiv 3 \pmod{4}$ , if not another random  $a$  will be generated till it satisfies our congruent module condition. After that, values  $a$  contains all the orders of the value we selected, all we have to do is sort it and remove the 'Inf' values. This process is repeated for 2 other values, and hence in total, we have 3 histograms from each exponent.

**Note:** A total of 3 csv files were used for this simulation, and each file corresponds to the  $a$  values in each of the exponents

$$\begin{aligned}
& \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\Omega(p-1) - \Omega(\text{ord}_p(a))} \\
& \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega((p-1)/\text{ord}_p(a))} \\
& \lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} z^{\omega(p-1) - \omega(\text{ord}_p(a))}
\end{aligned}$$

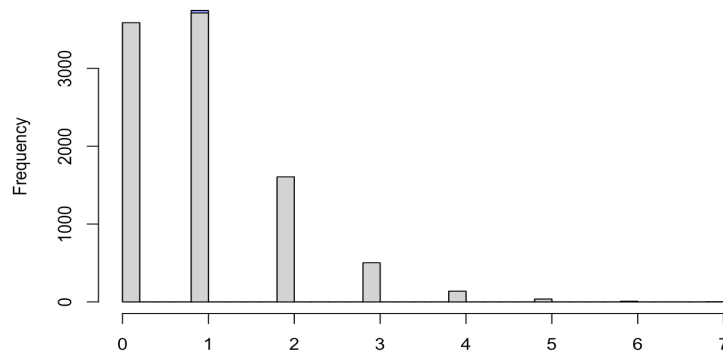
Note: The grey bars are the bars from the histogram generated from the theoretical values, if any other colors appear other than grey, this indicates the height difference between the 2 histograms

## Histograms of first exponent $\Omega(p-1) - \Omega(\text{ord}_p(a))$ generated from the collected data:



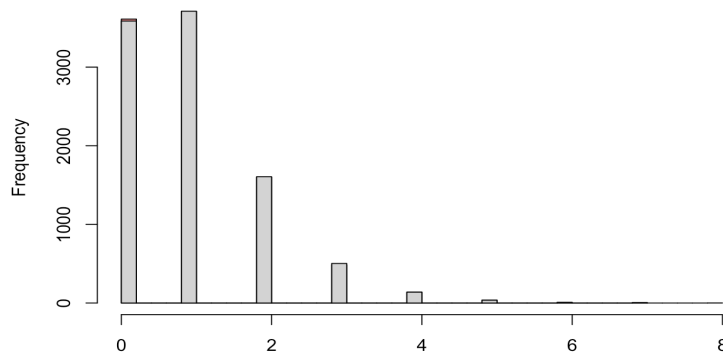
## Overlapping of the histograms generated from the collected data of the first exponent, with the histogram generated from the theoretical values :

- Overlapping of the histogram of a value 91 with the histogram generated from theoretical



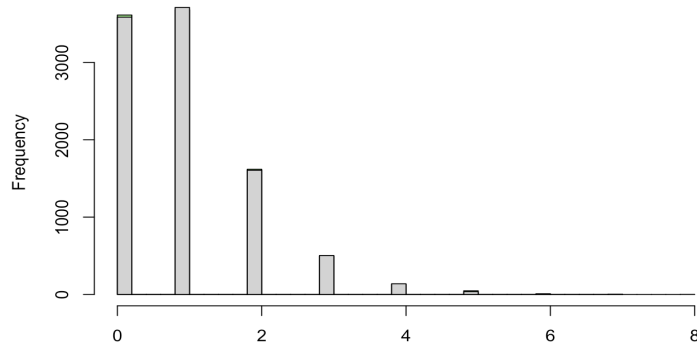
values:

- Overlapping of the histogram of a value  $-17$  with the histogram generated from theoretical



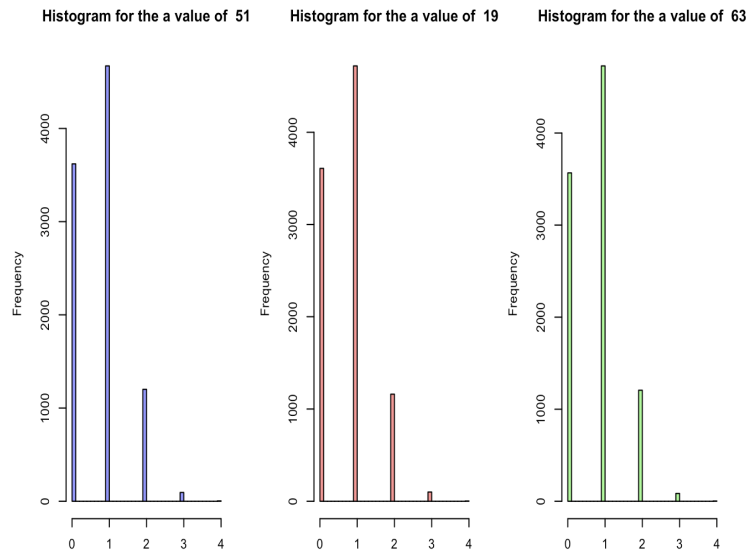
values:

- Overlapping of the histogram of a value 43 with the histogram generated from theoretical



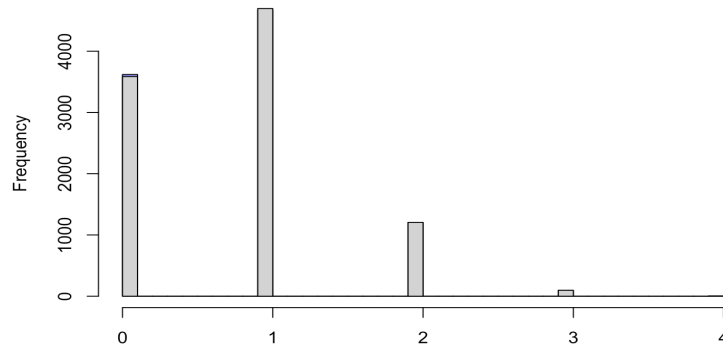
values:

**Histograms of second exponent  $\omega((p-1)/\text{ord}_p(a))$  generated from the collected data:**



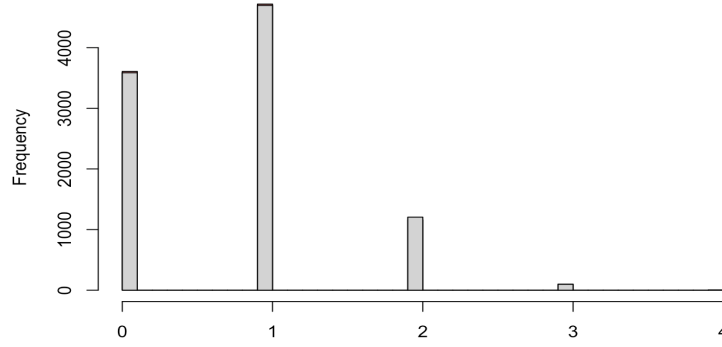
**Overlapping of the histograms generated from the collected data of the second exponent, with the histogram generated from the theoretical values :**

- Overlapping of the histogram of a value 51 with the histogram generated from theoretical



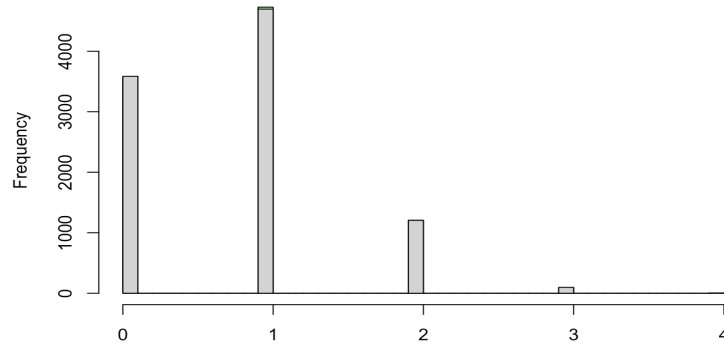
values:

- Overlapping of the histogram of a value 19 with the histogram generated from theoretical



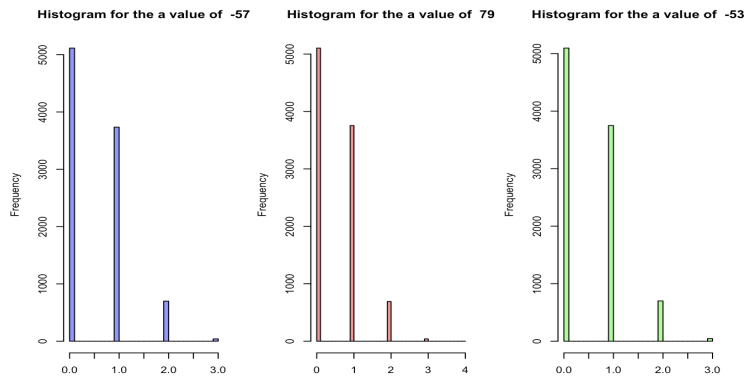
values:

- Overlapping of the histogram of a value 63 with the histogram generated from theoretical



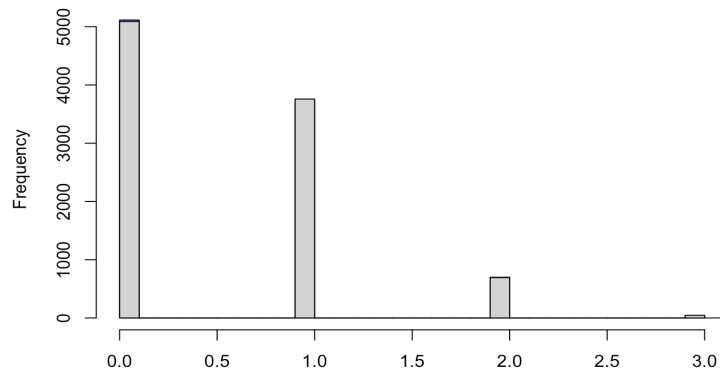
values:

**Histograms of third exponent  $\omega(p-1) - \omega(\text{ord}_p(a))$  generated from the collected data:**



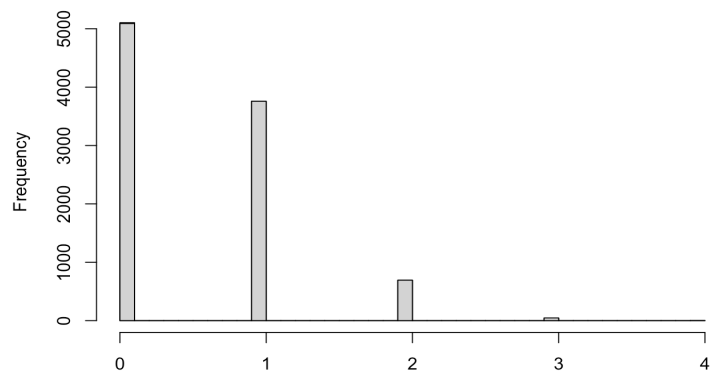
**Overlapping of the histograms generated from the collected data of the third exponent, with the histogram generated from the theoretical values :**

- Overlapping of the histogram of a value  $-57$  with histogram generated from the theoretical



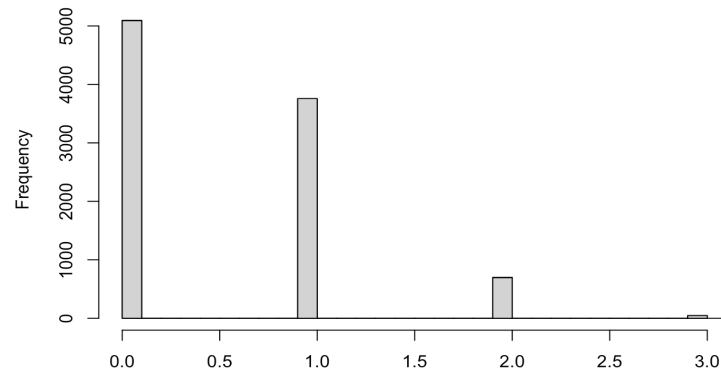
values:

- Overlapping of the histogram of a value  $79$  with histogram generated from the theoretical



values:

- overlapping of the histogram of a value  $-53$  with histogram generated from the theoretical



values:

#### 4. DATA ANALYSIS

For the Data Analysis, we used 2 approaches. The first approach is applying the Kolmogorov-Smirnoff(K-S) test, and the second approach is applying the Mann-Whitney U test.

**Kolmogorov-Smirnoff test:** In the field of statistics, the Kolmogorov-Smirnoff test can be applied to either a 1-dimensional sample or can be applied to 2. In our case we are using the second version since the goal is to compare and prove the correlation of 2 histograms, hence the second version of the test answers the question "How likely is it that we would see two sets of samples like this if they were drawn from same (but unknown) probability distribution (probability of different

possible values of a variable)?" The Kolmogorov-Smirnoff test produces 2 results, the first result is P-value, and the second result is D-value. The P-value indicates the probability of observing the test statistic, or a more extreme value if the null hypothesis is true, and the D-value is the test significance. In the case of the K-S test, the null hypothesis is the hypothesis that the two samples being compared come from the same distribution. Therefore, if the P-value is small (for example if  $P\text{-value} < 0.05$ , where 0.05 is our test significance), we reject the null hypothesis and conclude that two samples come from different distributions, and vice versa if we get a bigger P-value. On the other hand, the D-value is the maximum difference between the cumulative distribution functions of the 2 samples compared. It measures the distance between two distributions and is used to calculate the P-value in the K-S test. The larger the D-value, the greater evidence against the null hypothesis. In summary, if the P-value obtained from the K-S test is  $> D\text{-value}/2$  (test significance), then we reject the null hypothesis that there's a strong correlation between the 2 histograms generated from the datasets.

**Mann-Whitney U test:** Mann-Whitney U test tests whether 2 samples come from the same population or not, if 2 samples come from the same population, then we conclude that if we plot a histogram for each sample, the histograms should share the same characteristics such as (skewness, mean, median, kurtosis), and therefore there would be a very strong correlation between the 2 histograms when it comes to its shape. Just like the Kolmogorov-Smirnov test, the Mann-Whitney U test provides a P-value, and W-value, where the W-value is the sum of the ranks assigned to the observations in one of the 2 groups used for comparison. Note that when generating results for the Mann-Whitney U test, the P-value is calculated from the W-value. We set out test statistic to 0.05, and hence if the  $P\text{-value} > 0.05$ , we don't have sufficient evidence that the 2 datasets come from different populations, or in other words the shape of their histograms is different.

### Kolmogorov-Smirnoff test Results:

- First exponent  $\Omega(p - 1) - \Omega(\text{ord}_p(a))$  results:-
  - KS-test Results for the histogram of a value 91 and histogram generated from theoretical values -> D-value=0.003119, P-value=1, we conclude that  $P\text{-value} > D\text{-value}/2$
  - KS-test Results for the histogram of a value -17 and histogram generated from theoretical values -> D-value=0.0026066, P-value=1, we can conclude that  $P\text{-value} > D\text{-value}/2$ .
  - KS-test Results for the histogram of a value 43 and histogram generated from theoretical values -> D-value=0.0028151, P-value=1, hence we can conclude that  $P\text{-value} > D\text{-value}/2$
- Second exponent  $\omega((p - 1) / \text{ord}_p(a))$  results:-
  - KS-test Results for the histogram of a value 51 and histogram generated from theoretical values -> D-value=0.0034647, P-value=1, we conclude that  $P\text{-value} > D\text{-value}/2$
  - KS-test Results for the histogram of a value 19 and histogram generated from theoretical values -> D-value=0.0043174, P-value=1, we can conclude that  $P\text{-value} > D\text{-value}/2$ .
  - KS-test Results for the histogram of a value 63 and histogram generated from theoretical values -> D-value=0.0020592, P-value=1, hence we can conclude that  $P\text{-value} > D\text{-value}/2$
- Third exponent  $\omega(p - 1) - \omega(\text{ord}_p(a))$  results:-

- KS-test Results for the histogram of a value  $-57$  and histogram generated from theoretical values  $\rightarrow$  D-value=0.0025515, P-value=1, we conclude that  $P\text{-value} > D\text{-value}/2$
- KS-test Results for the histogram of a value  $79$  and histogram generated from theoretical values  $\rightarrow$  D-value=0.001349, P-value=1, we can conclude that  $P\text{-value} > D\text{-value}/2$ .
- KS-test Results for the histogram of a value  $-53$  and histogram generated from theoretical values  $\rightarrow$  D-value=0.0005149, P-value=1, hence we can conclude that  $P\text{-value} > D\text{-value}/2$

### Mann-Whitney U test Results:

- First exponent  $\Omega(p-1) - \Omega(\text{ord}_p(a))$  results:-
  - Mann-Whitney test results for the histogram of a value  $91$  and histogram generated from theoretical values  $\rightarrow$   $W=45879156$ , P-value=0.761, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $-17$  and histogram generated from theoretical values  $\rightarrow$   $W=45830914$ , P-value=0.6518, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $43$  and histogram generated from theoretical values  $\rightarrow$   $W=45896474$ , P-value=0.7877, hence we can conclude that the p-value  $> 0.05$ .
- Second exponent  $\omega((p-1)/\text{ord}_p(a))$  results:-
  - Mann-Whitney test results for the histogram of a value  $51$  and histogram generated from theoretical values  $\rightarrow$   $W=45816972$ , P-value=0.6517, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $19$  and histogram generated from theoretical values  $\rightarrow$   $W=45772524$ , P-value=0.5532, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $63$  and histogram generated from theoretical values  $\rightarrow$   $W=46015156$ , P-value=0.9071, hence we can conclude that the p-value  $> 0.05$ .
- Third exponent  $\omega((p-1)/\text{ord}_p(a))$  results:-
  - Mann-Whitney test results for the histogram of a value  $-57$  and histogram generated from theoretical values  $\rightarrow$   $W=45865982$ , P-value=0.7395, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $79$  and histogram generated from theoretical values  $\rightarrow$   $W=45906432$ , P-value=0.8198, hence we can conclude that the p-value  $> 0.05$ .
  - Mann-Whitney test results for the histogram of a value  $-53$  and histogram generated from theoretical values  $\rightarrow$   $W=45972220$ , P-value=0.9723, hence we can conclude that the p-value  $> 0.05$ .

**Conclusion:** As a conclusion, our statistical analysis shows the conjectural histograms are relatively similar to the actual histograms. Therefore, our computations provide some evidence for Goldmakher and Martin's conjecture.

### REFERENCES

- [1] D. R. Heath-Brown, *Artin's conjecture for primitive roots*, Quart. J. Math. Oxford Ser. (2) **37** (1986), no. 145, 27–38. MR830627

- [2] Christopher Hooley, *On artin's conjecture*. **1967** (1967), no. 225, 209–220.
- [3] Peter Stevenhagen, *The correction factor in Artin's primitive root conjecture*, Journal de théorie des nombres de Bordeaux **15** (2003), no. 1, 383–391 (en). MR2019022
- [4] The Sage Developers, *Sagemath, the Sage Mathematics Software System (Version x.y.z)*, YYYY.  
<https://www.sagemath.org>.