Artin's Conjecture about primes with prescribed primitive roots

Javier López-Contreras

November 2022

In this note, I will give a short summary of Hooley's conditional proof of Artin's Conjecture (AC) [1] and propose Conjecture 2.1, a new self-contained conjecture that, if proven, would reduce the strength of the Riemann Hypothesis (RH) assumed by Hooley. There is strong numerical evidence that the conjecture holds, as shown in Figure 1.

1 Proof of AC conditioned to RH

In his 1967 paper[1], Hooley proves Artin's conjecture about primes with prescribed primitive roots conditioned to the Generalized Riemann Hypothesis for the zeta functions of a family of number fields.

The conjecture is the following.

Conjecture 1.1 (Artin's conjecture of primes with a prescribed primitive root)**.** *Let $a \in \mathbb{Z}_{>1}$ not a perfect square. Then, there are infitely many primes $p \in \mathbb{Z}$ such that $\overline{a} \in \mathbb{Z}/p\mathbb{Z}$ is a primitive root. Moreover, the set of such primes has positive Dirichlet density, denoted $A(a)$.*

An the paper proves the following statement

Theorem 1.1 (Hooley)**.** *Given $a > 1$ not a square, let h be the maximum integer such that a is an h-th power. For a given k square free, let $k_1 = k/(h,k)$. If the Generalized Riemann Hypothesis is true for the Zeta-functions of the number fields $L_k = \mathbb{Q}(\sqrt[k_1]{a}, \zeta_k)$ for all k square free, then Artin's Conjecture is true for a.*

This section will give a sketch of the strategy used in this paper, needed to understand the improvement on such techniques that we propose in the second part of the document.

1.1 Preparation

For the whole of this document, we fix the following notations. Let $a \in \mathbb{Z}$, $a > 1$ not a square, p, q distinct primes, k a square free integer, h the maximum integer such that a is an h-th power and $k_1 = k/(h, k)$. Also $w(N)$ is the number of distinct primes dividing N.

Proposition 1.1. *a is a primitive root modulo p if and only if there are no primes q with both*

1. $p = 1 \mod q$
2. $a^{\frac{p-1}{q}} = 1 \mod p$

In such case, we will say that q is a witness of a not being a p.r. modulo p.

Definition 1.1. *We will use the following notations.*

1. $R_a(q, p) = \begin{cases} 1 & q \textit{ is a witness} \\ 0 & \textit{otherwise} \end{cases}$
2. $N_a(x) = \#\{p < x \mid a \textit{ is a p.r.} \mod p\}$
3. $N_a(x, \xi) = \#\{p < x \mid \nexists\, q \textit{ witness in the range } q < \xi\}$
4. $M_a(x, \xi_1, \xi_2) = \#\{p < x \mid \exists\, q \textit{ witness in the range } \xi_1 < q \leq \xi\}$
5. $P_a(x, k) = \#\{p < x \mid \forall q | k, q \textit{ is a witness}\}$

Proposition 1.2 (Basic observations of the newly defined functions)**.**

1. $N_a(x) = N_a(x, x - 1)$
2. $N_a(x) \leq N_a(x, \xi)$
3. $N_a(x) \geq N_a(x, \xi) - M_a(x, \xi, x - 1)$
4. $M_a(x, \xi_1, \xi_2) \leq \sum_{\xi_1 < q \leq \xi_2} P_a(x, q)$

Proposition 1.3. $N_a(x, \xi) = \sum_{l'} \mu(l') P_a(x, l')$, *where the second sum is over all l' square free with factors $\leq \xi$. Note that*

$$l' \leq \prod_{q \leq \xi} q = e^{\sum_{q \leq \xi} \log q} \leq e^{2\xi}$$

where in the last inequality we have used the prime number theorem.

Proposition 1.4. *Let $\xi_1 = \frac{1}{6}\log x, \xi_2 = x^{1/2}\log^{-2} x, \xi_3 = x^{1/2}\log x$. From the previous observations, we get*

$$\begin{aligned} N_a(x) = N_a(x,\xi_1) + O(M_a(x,\xi_1,\xi_2)) +& \\ + O(M_a(x,\xi_2,\xi_3)) + O(M_a(x,\xi_3,x-1))& \end{aligned} \tag{1}$$

Hooley proves that the first is the leading term, being $\sim A(a)\frac{x}{\log x}$ for an explicit constant $A(a)$. Moreover he proves that, the other 3 terms will be asymptotically smaller, upper bounded by $O\left(\frac{x}{\log^2 x}\right)$. This concludes that $N_a(x) \sim A(a)\frac{x}{\log x}$, which is precisely Artin's conjecture. The choice of ξ_i is taken carefully to fulfill the estimates.

Comment. *The bounds of terms 3 and 4 use elementary techniques. For terms 1 and 2, the R.H. is needed. As we will detail in the following section, the estimation of term 1 only needs the* 2/3*-zero free region but the upper bounding of term 2 will need the full* 1/2 *R.H.*

The conjecture that we propose gives a equally good bound for term 2 using less strength of the R.H. We do so by improving the bound on term 4, which makes it possible to choose a lower ξ_3, which at its turn makes it possible to choose lower ξ_2 without disrupting the bound of term 3. Having a lower ξ_2 gives the possibility of conserving the bound of the second term but using less strength of the R.H.

The estimation of the first term still needs the 2/3 *R.H., so the best this possible improvement can hope to do is lower the conditions, but not give a condition-less proof.*

1.2 Bounds on the 3rd and 4th term

Proposition 1.5 (Bound of the 4th term)**.** *Let $\xi_3 = x^{1/2}\log x$, then*

$$M_a(x,\xi_3,x-1) = O\left(\frac{x}{\log^2 x}\right)$$

Proof. If q is a witness, in particular $a^{\frac{p-1}{q}} = 1 \mod p$. Hence, if there is a witness $q > \xi_3$, there will be an $m < \frac{x}{\xi_3}$ such that $p | a^m - 1$. All the primes counted on $M_a(x,\xi_3,x-1)$ need to be divisors of

$$S_a(x/\xi_3) := \prod_{m<x/\xi_3} (a^m - 1)$$

Hence, $2^{M_a(x,\xi_3,x-1)} < S_a(x/\xi_3)$ which implies $M_a(x,\xi_3,x-1) < \log S_a(x/\xi_3) < \log a \sum_{m<x/\xi_3} m = O\left((x/\xi_3)^2\right) = O\left(\frac{x}{\log^2 x}\right)$.

Comment. *One is forced to choose $\xi_3 = x^{1/2}\log x$ for the last equality to be true. Yet, in this document we conjecture a refined upper bound for the number of primes diving $S_a(n) = \prod_{m<n}(a^m - 1)$. Using our conjecture, one will be able to choose a lower ξ_3.*

Proposition 1.6 (Bound of the 3rd term)**.** *Let $\xi_2 = x^{1/2}\log^{-2} x$ and $\xi_3 = x^{1/2}\log x$. Then $M_a(x, \xi_2, \xi_3) = O\left(\frac{x}{\log^2 x}\right)$.*

Proof. By Proposition 1.2, we may express $M_a(x, \xi_2, \xi_3) \leq \sum_{\xi_2<q\leq\xi_3} P_a(x, q)$.

Now, if q is a witness, then in particular $p \equiv 1 \mod q$. By Brun's method, which is a inequality related to Dirichlet's Theorem, we have

$$P_a(x, q) \leq \sum_{\substack{p\leq x \\ p\equiv 1 \mod q}} 1 \leq \frac{A_1 x}{(q-1)\log(x/q)}$$

From this we obtain the bound

$$\begin{aligned} M_a(x, \xi_2, \xi_3) &= O\left(\frac{x}{\log x}\sum_{\xi_2<q\leq\xi_3}\frac{1}{q}\right) = \\ &= O\left(\frac{x}{\log^2 x}\left(\log\frac{\xi_3}{\xi_2} + O(1)\right)\right) = O\left(\frac{x\log\log x}{\log^2 x}\right) \end{aligned} \tag{2}$$

Comment. *This proposition forces to choose the polynomial degree of ξ_2 to be the same as ξ_3, a priori $1/2$. Yet a key takeaway from this proposition is that the bound only depends on the ratio ξ_3/ξ_2. If we manage to lower ξ_3, we can automatically lower ξ_2 without disturbing this bound.*

1.3 Artin's observation

When Artin proposed the conjecture, he had an intuition for what the Dirichlet density of the primes with a prescribed primitive root at a had to be. He developed this intuition by translating the problem to an Algebraic Number Theory setting. This change in point of view is explained by the following proposition.

Proposition 1.7. *Let $a > 1$ not a square, k as square-free integer and p a prime. All $q|k$ are witnesses if and only if p is completely split in the extension $L_k/\mathbb{Q}$, with $L_k = \mathbb{Q}(a^{1/k_1}, \zeta_k)$, with ζ_k a primitive k-root of unity.*

Definition 1.2. *We will use the notation $n(k) := [L_k : \mathbb{Q}]$.*

Comment. *By Chebotarev's theorem, we get $P_a(x,k) = n(k)\frac{x}{\log x}$. The explicit value of $[L_k : \mathbb{Q}]$ is computed by Hooley [1] but it is not essential for the exposition on this note. Artin deduced what the constant in $N_a(x)$ should be by imagining an type of "infinite" inclusion-exclusion lemma but formalizing this lemma is the main difficulty in the conjecture.*

1.4 Reduction to counting primes

The point of view found by Artin gives a clearer line of attack to the conjecture. This is exemplified by the following propositions, linking the prime counting function to the sums we are interested in computing.

Definition 1.3 (Prime counting function)**.**

$$\pi(x,k) := \#\{\mathfrak{p} \textit{ prime ideal of } L_k | N\mathfrak{p} \leq x\}$$

Proposition 1.8.

$$n(k)P_a(x,k) = \pi(x,k) + O(n(k)w(k)) + O(n(k)x^{1/2}) \tag{3}$$

Proof. This is an implication of elementary ramification theory applied to L_k, check the article[1] for the details.

1.5 Prime counting theorem

By Proposition 1.8, an estimate of $\pi(x,k)$ will give an estimate of $P_a(x,k)$ and hence, by Propositions 1.2 and 1.3, which in turn will give an estimate of the first and second term in Equation 1. The final part of Hooley's article deduces a good enough prime counting theorem.

Theorem 1.2. *Assuming the GRH for ζ_{L_k}, we have the estimate*

$$\pi(x,k) = \frac{x}{\log x} + O(n(k)x^{1/2}\log kx) \tag{4}$$

Proof. Hooley starts from the classical idea that π can be expressed in terms of the zeroes of ζ_{L_k}. He deduces a theorem about the vertical distribution of zeroes and, together with the assumption that the zeroes are in the 1/2 line, we is able to deduce the desired bound.

Comment. *If you follow Hooley's proof only assuming the zero-free region Re*$(s) > f$*, you get the estimate*

$$\pi(x,k) = \frac{x}{\log x} + O(n(k)x^f \log kx) \tag{5}$$

From the rest of the document, f *will note the value up to which the RH is assumed.*

1.6 Bounds for the 1st and 2nd term

By Proposition 1.8, one gets an estimate of P_a and unrolling Propositions 1.2 and 1.3 one gets estimates of the first and second term in Equation 1. They are explained in the following propositions.

Proposition 1.9 (Estimation of the 1st term)**.**

$$\begin{aligned} N_a(x,\xi_1) &= \sum_{l'} \mu(l') \left(\frac{x}{\log x \cdot n(l')} + O(x^f \log x) \right) = \\ &\underset{l'<e^{2\xi_1} \textit{by Prop. 1.3}}{=} \frac{x}{\log x} \sum_{l'} \frac{\mu(l')}{n(l')} + O\left(\sum_{l<e^{2\xi_1}} x^f \log x \right) = \\ &= A(a)\frac{x}{\log x} + O(e^{2\xi_1} x^f \log x) = \\ &= A(a)\frac{x}{\log x} + O(x^{f+1/3} \log x) \end{aligned} \tag{6}$$

Comment. *Very significantly, note that for the extra term to be irrelevant, we only need* f *to be* $f < 2/3$*. For this, it is sufficient to assume a* $R(s) \geq 2/3$ *zero-free region.*

Proposition 1.10 (Bound of the 2nd term)**.**

$$\begin{aligned} M_a(x,\xi_2,\xi_3) &\leq \sum_{\xi_1<q\leq\xi_2} \left(\frac{x}{\log x \cdot q(q-1)} + O(x^f \log x) \right) = \\ &= O\left(\frac{x}{\log x} \sum_{q>\xi_2} \frac{1}{q^2} \right) + O\left(x^f \log x \sum_{q\leq\xi_2} 1 \right) = \\ &= O\left(\frac{x}{\xi_1 \log x} \right) + O\left(\frac{x^f \xi_2 \log x}{\log \xi_2} \right) = O\left(\frac{x}{\log^2 x} \right) \end{aligned} \tag{7}$$

Comment. *Note that in the last equality we did need* $f = 1/2$ *because* $\xi_2 = x^{1/2}\log^{-2} x$*. If we manage to lower the polynomial degree of* ξ_2*, we would be able to conserve the bound using a higher* f*, hence reducing the conditions in Hooley's proof.*

2 Proposed improvement

We propose the following self-contained conjecture.

Conjecture 2.1. *Let $S_a(n) := \prod_{m<n} (a^m - 1)$. Let $w(N) = \#\{\text{distinct primes } p|N\}$. Is is true that $w(S_a(n)) = O(n \cdot \text{poly-log})$?*

We state that this would reduce the conditions on Hooley's conditional proof from the full R.H to a a $R(s) \geq 2/3$ zero free region. The weaker conjecture $w(S_a(n)) = O(n^{2-\epsilon} \cdot \text{poly-log})$ for $\epsilon > 0$ would already improve the conditions to a $R(s) \geq 1/2 + \epsilon/3$ zero-free region.

Comment. *For the application on AC, the value of a can be asked to be a non-square. Yet, numerical evidence in Figures 3 seems to imply that the conjecture is true regardless. This doesn't contradict the necessary condition in AC as a being a non-square is still used in Artin's observation.*

Comment. *The polylogarithmic part will take no paper in the application to AC, can be taken as large as one wants.*

Comment. *Note that, following the factorization $a^m - 1 = \prod_{d|m} \Phi_d(a)$, the conjecture is very related to the values of $w(\Phi_d(a))$, where Φ_d is the d-th cyclotomic polynomial. There seems to be a conjecture by Erdős [2] on $P(\Phi(a))$, the largest prime divisor which has a very similar flavor.*

2.1 Upper bound $w(S_a(n)) = O(n^2)$

It is not hard to prove $w(S_a(n)) = O(n^2)$. For example, $2^{w(S_a(n))} < S_a(n)$, from which the desired bound follows. This bound can be improved by logarithmic factors in a number of ways. For instance using the well-known bound $w(N) = O\left(\frac{\log N}{\log \log N}\right)$, which can be proven by looking at $N = \prod_{p<n} p$ the primordials.

2.2 Lower bound $w(S_a(n)) = \Omega(n)$

A trivial application of Zsigmondy's theorem[3] shows $w(S_a(n)) = \Omega(n)$.

2.3 Numerical evidence

We believe that the strong conjecture is true. Numerical evidence is shown in Figure 1, for $a = 2$.

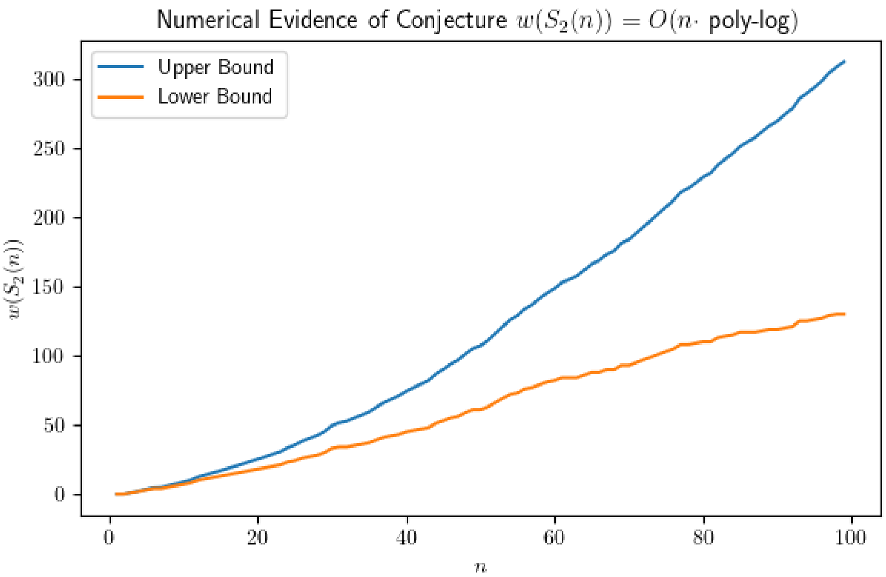

Figure 1: Numerical Evidence of Conjecture 2.1. The lower bound w' is the number of distinct primes in $S_2(n)$ in the range $< 10^8$. The upper bound has an extra correction term of $\frac{n(n-1)}{2}\log_{10^8}(2)$ which over counts the number of primes that $S_2(n)$ can have on the range $\geq 10^8$.

The limitation of these numerical computations is the number of primes can be saved in a computer in practice. The current program, foudn in the Appendix, checks for primes up to $L = 10^8$ through an Eratosthenes Sieve. Yet $S_2(n)$ grows very quickly so, a priori, it could start having prime factors larger than our range. We can only give an exact value of $w(S_a(n))$ for n relatively small (~ 10). For higher values, we compute a lower and higher bound for $w(S_2(n))$.

The lower bound $w'(S_2(n))$ is just the number of distinct primes dividing $S_2(n)$ that are in the range $p < L$ which we compute by counting. The upper bound is $w' + \frac{n(n-1)}{2}\log_L(2)$. This is an upper bound because any extra prime of $S_2(n)$ not in the our range is at least $\geq L$, hence there can only be, at most, $\log_L(S_2(n)) \leq \log_L(2^{\sum_{m<n} m}) = \frac{n(n-1)}{2}\log_L(2)$.

2.4 Improvement on Artin's conjecture

Conjecture 2.1 gives a finer upper bound for the 4th term in Equation 1. This will let us choose a smaller ξ_3'. For this section, we assume Conjecture 2.1 and, to simplify the computations, we let the polylogarithmic part be trivial $L(n) = 1$. Hence, suppose

$w(S_a(n)) \leq C_a \cdot n$

Proposition 2.1 (New Bound of the 4th Term)**.** *Let* $\xi_3' = \log^2 x$*, then*

$$M_a(x, \xi_3', x-1) = O\left(\frac{x}{\log^2 x}\right)$$

Proof. As seen in the original proof $M_a(x, \xi_3, x-1) \leq w(S_a(x/\xi_3))$. Now Hooley uses the trivial bound $w(S_a(n)) = O(n^2)$ and concludes that $M_a(x, \xi_3, x-1) = O\left((x^2/\xi_3^2)\right) = O\left(\frac{x}{\log^2 x}\right)$. In the new case, $M_a(x, \xi_3', x-1) = O(w(S_a(x/\xi_3'))) = O(x/\xi_3') = O\left(\frac{x}{\log^2 x}\right)$.

Now let $\xi_2' = \log^{-3} x$, which makes the ratio $\xi_3'/\xi_2' = \log^5 x$. Proposition 1.6 still holds with these new brackets. But now, having $\xi_2' = \log^{-3} x$ makes the bound of the 2 term condition-free. This can be seen in the last equality of Proposotion 1.10.

Hence, the only condition that remains is the $R(s) \geq 2/3$ zero-free region used for estimation the first term.

References

[1] Christopher Hooley. “On Artin’s conjecture.” In: *Journal für die reine und angewandte Mathematik* 225 (1967), pp. 209–220. URL: http://eudml.org/doc/150785.

[2] M. Ram Murty and François Séguin. “Prime divisors of sparse values of cyclotomic polynomials and Wieferich primes”. In: *Journal of Number Theory* 201 (2019), pp. 1–22. ISSN: 0022-314X. DOI: https://doi.org/10.1016/j.jnt.2019.02.016. URL: https://www.sciencedirect.com/science/article/pii/S0022314X19300927.

[3] K. Zsigmondy. “Zur Theorie der Potenzreste”. In: *Monatshefte für Mathematik und Physik* 3 (1892), pp. 265–284. DOI: https://doi.org/10.1007/BF01692444. URL: https://link.springer.com/article/10.1007/BF01692444#citeas.

A Estimates for $a \in \{3, 4, 5, 6\}$

Estimates for $a = 2$ are given in Figure 1. For $a \in \{3, 4, 5, 6\}$, they are given below. These values include a square and a composite number. A detailed explanation of the lower and upper bounds can be found in Figure 1.

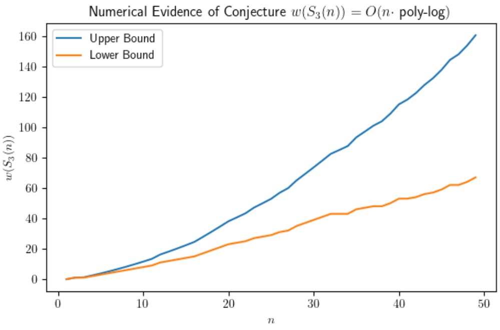

Figure 2: Computations $w(S_3(n))$

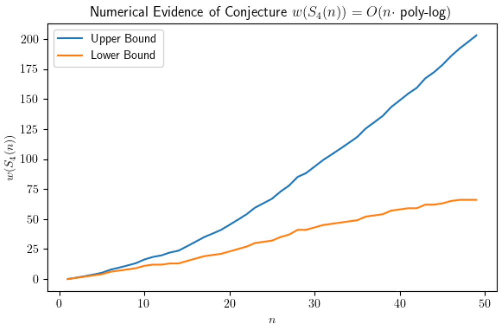

Figure 3: Computations $w(S_4(n))$

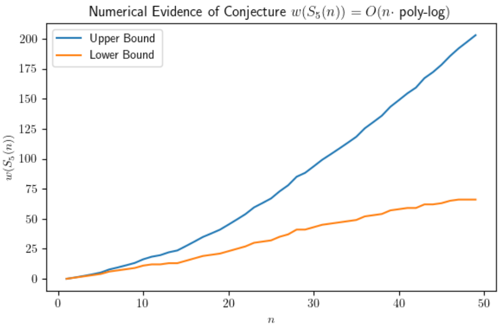

Figure 4: Computations $w(S_5(n))$

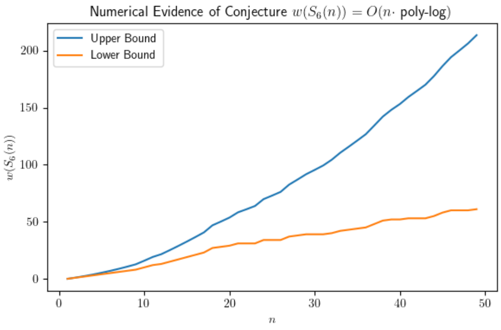

Figure 5: Computations $w(S_6(n))$

B Program computing estimations of $w(S_a(n))$

This is the exact version of the program used to compute the data in Figure 1.

```cpp
#include <bits/stdc++.h>
using namespace std;

typedef long long ll;

ll L = 1e8;

// Eratosthenes Sieve
vector<ll> sieve(ll n) {
  vector<ll> primes;
  vector<bool> prime(n, true);
    for (ll i=2; i<n; i++) {
    if (prime[i]) {
      primes.push_back(i);
      for (ll m=2*i; m < n; m += i) prime[m] = false;
    }
  }
  return primes;
}

// Fast exponentiation
ll poww(ll a, ll n, ll p) {
  if (n == 0) return 1LL;
  ll mid = poww(a, n/2, p);
  ll twomid = (mid*mid)%p;
  if (n%2 == 0) return twomid;
  else return (a * twomid)%p;
}

int main() {
  vector<ll> primes = sieve(L);
  ll a = 2;
  ll N = 100;

  for (ll n=1; n<N; n++) {
    cerr << n << endl;
    ll count = 0;
    for (ll p : primes) {
      ll num = 1;
      for (ll m=1; m<n; m++) {
        num *= (poww(a, m, p) + p - 1) % p;
        num %= p;
      }
      if (num == 0) count += 1;
    }
    long double logVal = 0;
    for (ll m=1; m<n; m++) logVal += m*log(a);
    cout << n << "," << count << "," << count + logVal / log(L) << endl;
  }
}
```