THE CORRECTION FACTOR IN ARTIN'S PRIMITIVE ROOT CONJECTURE

P. Stevenhagen

Abstract. In 1927, E. Artin proposed a conjectural density for the set of primes p for which a given integer g is a primitive root modulo p. After computer calculations in 1957 by D. H. and E. Lehmer showed unexpected deviations, Artin introduced a correction factor to explain these discrepancies. The modified conjecture was proved by Hooley in 1967 under assumption of the generalized Riemann hypothesis.

This paper discusses two recent developments with respect to the correction factor. The first is of historical nature, and is based on letters between Artin and the Lehmers from 1957–58 that were discovered in the Lehmer archives in Berkeley in December 2000. The second concerns a new interpretation of the correction factor in terms of local contributions by H. W. Lenstra, P. Moree and the author that is well-suited to deal with many generalizations of Artin's original primitive root problem.

1. Introduction

In a conversation with Hasse that took place in September 1927, Emil Artin conjectured that the set Σ_g of primes p for which a given integer $g \in \mathbf{Z}$ that is not an exact power is a primitive root modulo p is infinite, and that it has a natural density equal to *Artin's constant* $A = \prod_{q \text{ prime}}(1 - \frac{1}{q(q-1)}) \approx .3739558$. In particular, the density should be independent of g.

The heuristical argument underlying Artin's conjecture is simple and well-known. A prime number $p \nmid 2g$ is in Σ_g if and only if there is no prime number q dividing $p-1$ such that g is a q-th power modulo p. Writing ζ_q for a primitive q-th root of unity and $K_q = \mathbf{Q}(\zeta_q, \sqrt[q]{g})$ for the splitting field of $X^q - g$ over $\mathbf{Q}$, this condition means that for no prime $q < p$, the prime p splits completely in the number field K_q. It follows from the Chebotarev density theorem (or one of its 19th century predecessors, see [9]) that for fixed q, the set of primes p that do not split completely in K_q has density $1 - \frac{1}{[K_q:\mathbf{Q}]}$. If we assume the splitting conditions in the various fields K_q to be 'independent', it seems reasonable to expect Σ_g to have density $\prod_{q \text{ prime}}(1 - \frac{1}{[K_q:\mathbf{Q}]})$. If g is not an exact power in $\mathbf{Z}$, one has $[K_q : \mathbf{Q}] = q(q-1)$ for all primes q and we obtain Artin's constant. For arbitrary $g \neq \pm 1$, there is a largest integer $h \in \mathbf{Z}$ for which g is an h-th power in $\mathbf{Z}$, and the conjectured density becomes the rational multiple

$$(1.1) \qquad A_h = \prod_{q \text{ prime}}(1 - \frac{1}{[K_q : \mathbf{Q}]}) = \prod_{q|h}(1 - \frac{1}{q-1}) \prod_{q\nmid h}(1 - \frac{1}{q(q-1)})$$

1991 *Mathematics Subject Classification.* Primary 11R45; Secondary 11L03, 11N13.
Key words and phrases. Artin's conjecture, primitive roots.

of Artin's constant $A = A_1$. Note that A_h vanishes if and only if h is even, i.e., if g is a square; in this case it is clear that g does not generate $\mathbf{F}_p^*$ for a single odd prime p.

2. History of the conjecture

In the 1930s, there were various efforts to prove Artin's conjecture. Dealing with splitting conditions in an infinite number of fields K_q proved to be too hard for Hasse and his student Bilharz. In 1935, Hasse learned via Davenport that Erdős believed to have a proof, and he turned to the analogous problem for the field $k(X)$ of rational functions over a finite field k. Erdős's 'proof' turned out to depend on the generalized Riemann hypothesis, and on a statement in the distribution of prime numbers he was unable to justify. Hasse and Bilharz were successful in proving Artin's conjecture for $k(X)$, and even for arbitrary function fields in one variable over k. In these easier cases, the Riemann hypothesis was known to hold or was being established at the same time by A. Weil.

The original conjecture remained open, and was discussed by Hasse in his 1950 textbook [3] and in a 1952 paper [4]. In 1957, Derrick H. Lehmer and his wife Emma numerically tested Artin's conjecture for various values of g on a computer and sent a table with their findings to Artin in Princeton. This is the beginning of a short and amusing exchange of letters [2] that were discovered in December 2000 in the Lehmer archives in the Bancroft library in Berkeley. In a letter dated December 19, 1957, Artin writes to Emma Lehmer that he had tested his conjecture at the time he made it using tables of Kraitchik, and goes on to explain in detail the heuristic argument given in our introduction. As for the references to the literature Emma Lehmer had asked for, Artin has 'no idea' but 'Hasse who is very much interested should know', and Hasse's mail address follows. D. H. Lehmer replied on New Year's Eve of 1957 that he and his wife had thought that Artin had obtained his constant from the expression $A = \sum_{p \text{ prime}} \frac{\varphi(p-1)}{p-1}$, which counts for each p the fraction of elements that are generators of the finite field $\mathbf{F}_p^*$, and he asks for a reference for the density statement for the primes splitting in K_q. Artin's January 6 reaction consisted of three densely handwritten pages starting as follows:

> Dear Professor Lehmer:
>
> Since you are interested in the density of primes connected with the factorisation of polynomials I would like to stress the fact that the root of these questions belongs to algebraic number theory and should be viewed from this point of view. Any interpretation in terms of elementary number theory hides very essential thoughts into the nature of the questions. If you have the patience to study the following explanations I should think that you will agree. As a matter of fact one knows the most general density law of this nature.

A crash course in algebraic number theory leading up to the Chebotarev density theorem follows. By the end of the month, on January 28, 1958, a fourth letter follows. It is interesting enough to be quoted in full, as it is the birth certificate of the correction factor in Artin's conjecture.

> Dear Mrs Lehmer!
>
> Your run away 5 started me thinking again and I can tell you the reason. Please return to my first letter. Let us first consider the primitive root 2 (that's all I did when I made my conjecture). Consider the field $K_q = R(\zeta_q, \sqrt[q]{2})$ which has degree $q(q-1)$ and accounts for the factor $1 - \frac{1}{q(q-1)}$ in the formula. There is no difficulty showing that a finite number of the probabilities are independent. We need however an infinite number and that is extremely

hard. That a finite number is independent means merely that the fields K_q have no common subfield $\neq R$, that the Galois group of a compositum is a direct product.

But replace now 2 by a prime $p \equiv 1 \bmod 4$. Then I was extremely hasty. Namely $K_2 = R(-1, \sqrt{p}) = R(\sqrt{p})$ is a subfield of $R(\zeta_p)$ hence of K_p. (This is the only dependence that occurs, but it destroys my conjecture). Now what does this mean? We did want to exclude the primes modulo which $X^p - p$ splits completely, the primes which split completely in K_p and have density $\frac{1}{p(p-1)}$. But such a prime splits then automatically in the subfield $R(\sqrt{p})$. Since another condition was not to split in $K_2 = R(\sqrt{p})$, a prime which satisfies this will already *not* split in K_p. In elementary terms: if p is quadratic non-residue of a prime then $X^p - p$ cannot split into linear factors modulo that prime.

This means merely that the factor $1 - \frac{1}{p(p-1)}$ in the product has to be dropped, for $p = 5$ that means that the expected value is by the factor 20/19 larger. I turn to your results. In the column $16 \cdot 10^3$ you give as expected value 696 which is very good for the primitive root 2. For 5 it has to be replaced by 733 which is close enough to 744. In the column $23 \cdot 10^3$ the expected value is 959 and is good for 2, for 5 it should be 1010 which is again very good. For $p \equiv 3 \bmod 4$ nothing has to be changed and for composite numbers I did not work out the change. In the column $20 \cdot 10^3$ you get for 5 the value 890, right on the dot of the expected value 890.

So I was careless but the machine caught up with me.

Cordially, E. Artin.

The Lehmers mention Artin's correction for primes $g \equiv 1 \bmod 4$ without attribution in their 1963 paper [6], but do not appear to have digested Artin's crash course. Even Hasse provides a correction factor in his 1964 edition of [3] that is incorrect if $g \equiv 1 \bmod 4$ is *not* prime. Lang and Tate, who were Artin's students in Princeton in 1958, list the correction factor for arbitrary non-square g with $d = \mathrm{disc}(\mathbf{Q}(\sqrt{g})) \equiv 1 \bmod 4$ in their 1965 preface to Artin's collected works [1] as

$$E = E(d) = 1 - \mu(|d|) \prod_{\substack{q|d \\ q|h}} \frac{1}{q-2} \prod_{\substack{q|d \\ q\nmid h}} \frac{1}{q^2-q-1}. \tag{2.1}$$

Thus, Σ_g should have density $E(d) \cdot A_h$ for g satisfying $d = \mathrm{disc}(\mathbf{Q}(\sqrt{g})) \equiv 1 \bmod 4$. If $K_2 = \mathbf{Q}(\sqrt{g})$ is of even discriminant, no correction factor is necessary and the conjectured density simply equals the value A_h from (1.1).

In 1967, Hooley [5] proved Artin's corrected conjecture under the assumption of the generalized Riemann hypothesis. The hypothesis is used to obtain sufficient control of the error terms in the density statements for the sets of primes that split completely in the fields K_q. So far, there is not a single value of g for which one can show unconditionally that Σ is infinite.

3. The correction factor

Hooley attributes the correction factor (2.1) to Heilbronn, and derives it from the inclusion-exclusion expression $\sum_{n=1}^{\infty} \frac{\mu(n)}{[K_n:\mathbf{Q}]}$ for the density in terms of the degrees of the splitting fields K_n of $X^n - g$ over $\mathbf{Q}$. If the field $K_2 = \mathbf{Q}(\sqrt{g})$ is *not* quadratic of odd discriminant, then $n \longmapsto [K_n : \mathbf{Q}]$ is a multiplicative function and the expression reduces to the product A_h given in (1.1). If K_2 is quadratic of odd discriminant d, then $[K_n : \mathbf{Q}]$ is no longer multiplicative as it equals $\frac{1}{2}\prod_{q|n}[K_q : \mathbf{Q}]$ for all squarefree n that are divisible by d. In this case, a 'rather harder' calculation [5, p. 219] leads to the correction factor $E(d)$ from (2.1).

We note that, just like A_h itself, the factor $E(d)$ has a 'multiplicative structure' of the form $E(d) = 1 + \prod_{q \text{ prime}} E_q$, with 'local factors' E_q that can be described in terms of K_q as

$$E_q = \begin{cases} \frac{-1}{[K_q:\mathbf{Q}]-1} & \text{if } q \text{ divides } 2d; \\ 1 & \text{otherwise.} \end{cases} \tag{3.1}$$

The local factors $1 - 1/[K_q : \mathbf{Q}]$ going into A_h have a clear interpretation: they represent the fraction of the primes having the 'right' local behavior in K_q. The local factors E_q from (3.1) also admit an interpretation of a local nature, as we will show now.

The requirement we impose on our primes p in order to guarantee that g is a primitive root modulo p is that p does not split completely in the number field K_q for all primes $q < p$. In other words, the Frobenius symbol of p in $G_q = \mathrm{Gal}(K_q/\mathbf{Q})$ has to lie in the set $S_q = G_q \setminus \{1\}$ of non-trivial elements for these q; this happens for a fraction $\#S_q/\#G_q = 1 - 1/[K_q : \mathbf{Q}]$ of primes p when we look at a single prime q. If the fields K_q are linearly independent over $\mathbf{Q}$, then for any squarefree number n, the Galois group $G_n = \mathrm{Gal}(K_n/\mathbf{Q})$ is simply the direct product $\prod_{q|n} G_q$ of the Galois groups at q, and we want the primes p that have Frobenius element in $S_n = \prod_{q|n} S_q$. Such p clearly form a set of density $\#S_n/\#G_n = \prod_{q|n} \#S_q/\#G_q$. Letting n tend to infinity as in [7] and assuming GRH, we are left with a set of primes p having density A_h as in (1.1).

As Artin observed, the only case where the fields K_q are *not* linearly independent over $\mathbf{Q}$ occurs when $K_2 = \mathbf{Q}(\sqrt{g})$ is quadratic of odd discriminant d. Indeed, K_2 is then a subfield of $\mathbf{Q}(\zeta_d)$, hence a subfield of K_n for all squarefree values of n that are divisible by d. This leads to a difference in the computation of the fraction of 'good' Frobenius elements inside $G_n = \mathrm{Gal}(K_n/\mathbf{Q})$ for all squarefree n that are divisible by $2d$. Indeed, suppose d is odd, and fix such an n. Then have an exact sequence

$$1 \longrightarrow G_n = \mathrm{Gal}(K_n/\mathbf{Q}) \longrightarrow \textstyle\prod_{q|n} G_q \xrightarrow{\chi} \{\pm 1\} \longrightarrow 1,$$

Here $\chi = (\chi_q)_{q|n}$ is the quadratic character on $G = \prod_{q|n} G_q$ having as its 2-component the isomorphism $\chi_2 : G_2 \xrightarrow{\sim} \{\pm 1\}$, and having components χ_q at odd primes q defined by $\chi_{K_2} = \prod_{q>2} \chi_q$. In other words, χ_q is the non-trivial quadratic character modulo q at the primes $q|d$, and $\chi_q = 1$ at all other odd primes q.

We now need to count the fraction of elements in G_n that lie in the set $S = \prod_{q|n} S_q \subset G = \prod_{q|n} G_q$ of 'good' Frobenius elements, i.e., in the subset $S_n = G_n \cap S$ of G_n. The correction factor $E(d)$ arises as the factor by which $\#S_n/\#G_n$ differs from the 'uncorrected value' $\#S/\#G = \prod_{q|n} \#S_q/\#G_q$.

Clearly, we have $\#G = 2 \cdot \#G_n$. We can relate $\#S$ to $\#S_n$ using a character sum. Observe first that the function $\frac{1}{2}(1+\chi) : G \to \{0,1\}$ is the characteristic function of G_n on G. Let $\psi : G \to \{0,1\}$ be the characteristic function of $S = \prod_{q|n} S_q$ on G. Then $\frac{1}{2}(1+\chi)\psi$ is the characteristic function of S_n on G, so we have

$$\frac{\#S_n}{\#G_n} = \frac{1}{\#G_n} \sum_{x \in G} \frac{1}{2} \cdot \psi(x)[1 + \chi(x)] = \frac{\#S}{\#G}[1 + \sum_{x \in G} \frac{\psi(x)\chi(x)}{\#S}].$$

As the functions ψ, χ and $\#S$ on $G = \prod_{q|n} G_q$ are simply products of functions ψ_q, χ_q and $\#S_q$ defined on the components G_q, we can rewrite this as

$$\frac{\#S_n}{\#G_n} = \frac{\#S}{\#G}[1 + \prod_{q|n} \sum_{x \in G_q} \frac{\psi_q(x)\chi_q(x)}{\#S_q}]. \tag{3.2}$$

From (3.2) we see that the quotient of $\#S_n/\#G_n$ and $\#S/\#G$ is indeed of the form $1+\prod_{q|n} E_q$. Note that this quotient does not change if we replace n by any other squarefree multiple of $2d$: it is *constant* if n tends to infinity.

As $\psi_q : G_q \to \{0,1\}$ is the characteristic function of S_q on G_q, we have

$$E_q = \sum_{x\in G_q} \frac{\psi_q(x)\chi_q(x)}{\#S_q} = \frac{1}{\#S_q}\sum_{x\in S_q}\chi_q(x),$$

i.e., E_q is the *average value* of χ_q on S_q. At all 'non-critical' primes $q \nmid 2d$, we find $E_q = 1$ as χ_q is then the trivial character on G_q. At the primes $q|2d$, we are summing a non-trivial character χ_q over the set $S_q = G_q \setminus \{1\}$, which yields

$$\sum_{x\in S_q}\chi_q(x) = (\sum_{x\in G_q}\chi_q(x)) - \chi_q(1) = -\chi_q(1) = -1.$$

We find $E_q = -1/\#S_q = -1/([K_q:\mathbf{Q}]-1)$ in accordance with (3.1). This 'explains' the multiplicative structure of the correction factor $E(d)$.

4. Generalizations

As indicated in [7], Artin's original conjecture on primitive roots admits many variants for which Hooley's proof (under GRH) essentially goes through unchanged. The main problem becomes the explicit computation of the correction factor that transforms the product $\delta = \prod_q \#S_q/\#G_q$ of the local densities into the actual density. The character sum approach just given works in great generality, and in this final section we will formulate a rather general result along these lines.

In the original conjecture, we determined the density of the set Σ_g of primes p with the property that for all primes $q < p$, the Frobenius element of p in $G_q = \mathrm{Gal}(K_q/\mathbf{Q})$ is in the subset $S_q \subset G_q$ of non-trivial elements. Here we have $K_q = \mathbf{Q}(\zeta_q, \sqrt[q]{g})$ for all q.

For our generalization, we allow K_q and S_q to be different from the choice above for finitely many primes q. Thus, for every prime q, we assume that K_q is an extension of $\mathbf{Q}(\zeta_q, \sqrt[q]{g})$ that is finite and normal over $\mathbf{Q}$ and contained in the field $\mathbf{Q}(\zeta_{q^\infty}, \sqrt[q^\infty]{g})$ obtained by adjoining all q-power roots of g to $\mathbf{Q}$. We write $G_q = \mathrm{Gal}(K_q/\mathbf{Q})$, and take for S_q any subset of G_q that is stable under conjugation. For all but finitely many q, we suppose that we have $K_q = \mathbf{Q}(\zeta_q, \sqrt[q]{g})$ and $S_q = G_q\setminus\{1\}$. The modified question now becomes: what is the density of the set Σ of primes $p \nmid g$ with the property that for every prime $q < p$, the Frobenius element of p in $G_q = \mathrm{Gal}(K_q/\mathbf{Q})$ is in S_q?

The modification we have given may appear to be only slight, but it already covers a number of important generalizations. For instance, it enables us to deal with primes p that lie in a prescribed residue class or for which g generates a subgroup of given index in $(\mathbf{Z}/p\mathbf{Z})^*$.

The situation we are in is still very similar to that of the original conjecture. As the fields $\mathbf{Q}(\zeta_{q^\infty}, \sqrt[q^\infty]{g})$ for *odd* primes q are linearly disjoint over $\mathbf{Q}$, the only possible dependency between the fields K_q occurs when K_2 contains a quadratic field F of odd discriminant d, which is then contained in $\mathbf{Q}(\zeta_{|d|})$. As in the case of Artin's conjecture, this leads to two rather different situations. If K_2 does not contain a quadratic field F of odd discriminant, then all fields K_q are linearly independent, and we find as in [7] that under GRH, the density of Σ equals the product

$$\delta = \prod_q \frac{\#S_q}{\#G_q}$$

of all local densities $\#S_q/\#G_q$. By our assumptions, this is a rational multiple of Artin's constant. It vanishes if and only if S_q is empty for some q.

If K_2 does contain a quadratic field F of odd discriminant d, we take $\chi_2 = \chi_F$ the quadratic character of conductor d, and define χ_q for the primes $q > 2$ by $\chi_F = \prod_{q>2} \chi_q$. In particular, χ_q is trivial for all $q \nmid 2d$. We let

$$E_q = \frac{1}{\#S_q} \sum_{x \in S_q} \chi_q(x)$$

be the average value of χ_q on S_q. Clearly, we have $E_q = 1$ for $q \nmid 2d$. Now the argument of the previous section shows that, again under GRH, the density of Σ equals

$$\delta \cdot [1 + \prod_q E_q],$$

with δ the product of all local densities defined above. Thus, the correction factor $1 + \prod_q E_q$ has exactly the same shape for these generalizations as it has for the original Artin problem.

We finish this section by a concrete example that shows how the character sum method is applied in practice. Suppose that $g \in \mathbf{Z}$ not an exact power, and that $F = \mathbf{Q}(\sqrt{g})$ has odd discriminant d. We want to determine the density (under GRH) of the set Σ of primes $p \equiv a \bmod f$ for which g is a primitive root.

Let $f = \prod_{q|f} q^{e_q}$ be the factorization of our modulus f. Then we can make the standard choice $K_q = \mathbf{Q}(\zeta_q, \sqrt[q]{g})$ and $S_q = \mathrm{Gal}(K_q/\mathbf{Q}) \setminus \{1\}$ at all primes $q \nmid f$. At these q, the local density has the standard value $1 - 1/(q^2 - q)$ from (1.1).

At $q|f$, we take $K_q = \mathbf{Q}(\zeta_{q^{e_q}}, \sqrt[q]{g})$ and let the set S_q of 'good' Frobenius elements be the intersection of the 'congruence set' of elements raising the root of unity $\zeta_{q^{e_q}}$ to its ath power and the 'Artin set' of elements that are non-trivial on the subfield $\mathbf{Q}(\zeta_q, \sqrt[q]{g}) \subset K_q$. If we have $a \not\equiv 1 \bmod q$, then the first set is contained in the second and the local density simply equals $1/\varphi(q^{e_q})$. If we have $a \equiv 1 \bmod q$ then it becomes $(1 - 1/q) \cdot 1/\varphi(q^{e_q})$ as a fraction $1/q$ of the elements in the first set is outside the second. This yields a product

$$\delta = \frac{1}{\varphi(f)} \prod_{q \nmid f} (1 - \frac{1}{q^2 - q}) \prod_{q|\gcd(a-1,f)} (1 - \frac{1}{q}) \tag{4.1}$$

of all local densities. As K_2 contains the quadratic field $F = \mathbf{Q}(\sqrt{g})$ of odd discriminant d, there is a correction factor to be taken into account. At the primes $q|2d$ that do not divide f, the local correction factor equals $E_q = -1/(q^2 - q - 1)$ as in (3.1). We have $E_2 = -1$ as χ_2 equals -1 on S_2. For $q|\gcd(d, f)$ the character χ_q has the constant value $\chi_q(a) = \left(\frac{a}{q}\right)$ on S_q, so we find $E_q = \left(\frac{a}{q}\right)$. This leads to the global correction factor

$$E = 1 - \prod_{q|\gcd(d,f)} \left(\frac{a}{q}\right) \prod_{\substack{q|d \\ q\nmid f}} \frac{-1}{q^2 - q - 1}, \tag{4.2}$$

and we find (under GRH) that Σ has density $E \cdot \delta$ with E as in (4.2) and δ as in (4.1).

As the preceding example illustrates, the character sum method reduces the calculation of the correction factor to a number of fairly mechanical local computations. It can be further extended to situations that are not directly covered by the generalization considered in this section, such as the case in which one replaces the base field $\mathbf{Q}$ by an arbitrary number field. We refer the reader to [8] for a fuller treatment.

Acknowledgements. I thank Pieter Moree for his encouragement in getting hold of the letters in [2] and for providing some of the historical details in section 2, and the Bancroft Library in Berkeley for its courteous help in accessing the Lehmer archives.

References

1. E. Artin, *Collected papers* (S. Lang, J. T. Tate, eds.), Addison-Wesley, 1965.
2. E. Artin, D.H. Lehmer, E. Lehmer, *Correspondence 1957–58*, Archives of D.H. Lehmer, Bancroft Library, Berkeley.
3. H. Hasse, *Vorlesungen über Zahlentheorie*, Akademie-Verlag, 1950.
4. H. Hasse, *Über die Artinsche Vermuting und verwandte Dichtefragen*, Ann. Acad. Sci. Fennicae Ser. A. I. Math. Phys. (1952), no. 116.
5. C. Hooley, *On Artin's conjecture*, J. reine u. angew. Math. **225** (1967), 209–220.
6. D. H. Lehmer, E. Lehmer, *Heuristics, anyone?*, Studies in Mathematical Analysis and Related Topics, Stanford University Press, 1962.
7. H. W. Lenstra, Jr, *On Artin's conjecture and Euclid's algorithm in global fields*, Inv. Math. **42** (1977), 201–224.
8. H. W. Lenstra, Jr, P. Moree, P. Stevenhagen, *Character sums for primitive root densities*, in preparation.
9. P. Stevenhagen, H. W. Lenstra, Jr,, *Chebotarëv and his density theorem*, Math. Intelligencer **18** (1996), no. 2, 26–37.

Mathematisch Instituut, Universiteit Leiden, Postbus 9512, 2300 RA Leiden, The Netherlands

E-mail address: psh@math.leidenuniv.nl