ON THE FIXED POINTS OF THE MAP $x \mapsto x^x$ MODULO A PRIME, II

ADAM TYLER FELIX AND PÄR KURLBERG

Abstract. We study number theoretic properties of the map $x \mapsto x^x \pmod p$, where $x \in \{1, 2, \ldots, p-1\}$, and improve on some recent upper bounds, due to Kurlberg, Luca, and Shparlinski, on the number of primes $p < N$ for which the map only has the trivial fixed point $x = 1$. A key technical result, possibly of independent interest, is the existence of subsets $\mathscr{N}_q \subset \{2, 3, \ldots, q-1\}$ such that almost all k-tuples of distinct integers $n_1, n_2, \ldots, n_k \in \mathscr{N}_q$ are multiplicatively independent (if k is not too large), and $|\mathscr{N}_q| = q \cdot (1 + o(1))$ as $q \to \infty$. For q a large prime, this is used to show that the number of solutions to a certain large and sparse system of $\mathbb{F}_q$-linear forms $\{\mathscr{L}_n\}_{n=2}^{q-1}$ "behaves randomly" in the sense that $|\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) = 1, n = 2, 3, \ldots, q-1\}| \sim q^d(1-1/q)^q \sim q^d/e$. (Here $d = \pi(q-1)$ and the coefficents of $\mathscr{L}_n$ are given by the exponents in the prime power factorization of n.)

1. Introduction

For a prime p, let $\psi_p : \{1, 2, \ldots, p-1\} \to \{1, 2, \ldots, p-1\}$ be the remainder of x^x divided by p. The function ψ_p has cryptographic applications related to variations of the ElGamal signature scheme (see [9, Notes 11.70 and 11.71]); our main focus is studying the number of non-trivial fixed points of ψ_p as p varies. Let

$$F(p) := \#\{x \in \{1, 2, \ldots, p-1\} : \psi_p(x) = x\}$$

denote the number of fixed points of ψ_p. For convenience, we will slightly abuse notation and simply write $\psi_p(x) = x^x \pmod p$ (note that x^x is not well defined modulo p.) As 1 is always a fixed point of ψ_p we will say it is trivial; all other fixed points are said to be *nontrivial*.

Date: June 26, 2017.

2010 *Mathematics Subject Classification.* 11N37, 11N36, 11N25.

A.F. supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine. P.K. was partially supported by grants from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, and the Swedish Research Council (621-2011-5498).

Kurlberg, Luca and Shparlinski [7] gave bounds on the number of primes p for which ψ_p only has trivial fixed points. More specifically, they show most primes p have at least one fixed point besides 1: with $\mathcal{A}(N) = \{p \le N : F(p) = 1\}$ they proved that (cf. [7, Theorem 1])

$$\#\mathcal{A}(N) \le \frac{\pi(N)}{(\log_3 N)^{\vartheta+o(1)}} \tag{1.1}$$

as $N \to \infty$, where $\pi(x) := \#\{p \le x : p \text{ is prime}\}$ is the prime counting function and

$$\vartheta = \frac{1}{\zeta(2)} - \frac{1}{2\zeta(2)^2} = \frac{6\pi^2 - 18}{\pi^4} \approx 0.4231394212\cdots, \tag{1.2}$$

$\log x := \max\{\ln x, 2\}$, and $\log_k := \log(\log_{k-1} x)$ for $k \in \mathbb{N}$ and $k \ge 2$.

In (1.1), the exponent ϑ is related to the number of solutions to a certain system of linear forms modulo q, where q is a prime. For the convenience of the reader, we briefly describe how solutions to linear forms modulo q are related to fixed points of ψ_p (cf. [7, Section 2] for more details): For primes $p \equiv 1 \mod q$, it turns out that ψ_p has a nontrivial fixed point if n/q is a q-th power modulo p, for some integer $n \in [1, q-1]$. This in turn can be characterised in terms of the image of Frobenius, acting on $\mathrm{Gal}\left(\mathbb{Q}(\sqrt[q]{1}, \sqrt[q]{2}, \ldots, \sqrt[q]{q-1}, e^{2\pi i/q})/\mathbb{Q}(e^{2\pi i/q})\right)$, lying in a certain union of conjugacy classes. The cardinality of said union is related to the number of solutions, modulo q, to the following system of linear equations. Let $d = \pi(q-1)$, and for $1 \le n \le q-1$, let

$$n = \prod_{i=1}^{d} p_i^{\mu_i(n)}$$

be the prime power factorization of n, where we have ordered the primes $p \le q-1$ so that $p_1 < p_2 < \cdots < p_d < q$. For $n \in \mathbb{Z} \cap [1, q-1]$, define linear forms $\mathscr{L}_n : \mathbb{F}_q^d \to \mathbb{F}_q$ by

$$\mathscr{L}_n(\mathbf{v}) := \sum_{i=1}^{d} \mu_i(n) v_i, \tag{1.3}$$

where $\mathbf{v} := (v_1, v_2, \ldots, v_d) \in \mathbb{F}_q^d$. For $x_0 \in \mathbb{F}_q^\times$ fixed, let

$$N_q := N_q(x_0) = \#\Big\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) \ne x_0 \text{ for all } n \in \{1, 2, 3, \ldots, q-1\}\Big\}$$

and put $c(q) := N_q/q^d$. Kurlberg, Luca and Shparlinski showed that $\#\mathcal{A}(N) \le \frac{\pi(N)}{(\log_3 N)^{1-c(q)+o(1)}}$, gave the bound (cf. [7, Lemma 3])

$$c(q) \le 1 - \vartheta + o(1) \tag{1.4}$$

and conjectured[1] that $c(q) = e^{-1} + o(1)$. The basis for the conjecture is the following probabilistic heuristic: if q is large, $\mathbf{v} \neq 0$, and the linear forms $\{\mathscr{L}_n\}_{n=2}^{q-1}$ are *random*, then the probability that $\mathscr{L}_n(\mathbf{v}) \neq x_0$ for all n equals $(1-1/q)^{q-2} = 1/e + o(1)$. Summing over all nonzero $\mathbf{v}$ and using the linearity of expectations, we find that the expected value of $N_q(x)$ is $q^d \cdot (1/e + o(1))$.

Of course the collection of linear forms is far from random, e.g., the number of nonzero coefficients of $\mathscr{L}_n$ equals $\omega(n)$ (the number of distinct prime divisors of n); for $n < q$ we find that $\omega(n) \leq \log q = d^{o(1)}$ and hence $\{\mathscr{L}_n\}_{n=2}^{q-1}$ is a collection of quite sparse linear forms (in the sense that most coefficients are zero). Moreover, as $\mu_i(n) \leq 1$ if $p_i > \sqrt{q}$, most coefficients of the linear forms are very small. Nonetheless, the above heuristic turns out to give the correct answer.

Theorem 1.1. *As $q \to \infty$,*

$$c(q) = \frac{1}{e} + O\left(\frac{1}{\log_2 q}\right). \tag{1.5}$$

Remark 1.1. The method of proof would give a similar result in (roughly) the following setting. Assume that L_q is a finite collection of non-zero distinct linear forms modulo q having the properties that (1): there exists a subset $L'_q \subset L_q$ such that $|L'_q| = (1 + o(1/q))|L_q| = (1 + o(1))q$. (2): for almost all k-tuples $L_{k,q}$ of distinct forms in L'_q, the forms in $L_{k,q}$ are linearly independent, for $2 \leq k \leq K_q$, where K_q (slowly) tends to infinity with q. (3): The number of k tuples of distinct forms $L_{k,q}$ whose rank $r \leq k-1$ is $|L_q|^{r-o(1)}$.

We have the following corollary of Theorem 1.1 and [7, pp. 154-155]:

Corollary 1.2. *As $N \to \infty$,*

$$\#\mathcal{A}(N) \leq \frac{\pi(N)}{(\log_3 N)^{1-\frac{1}{e}+o(1)}}.$$

For comparison with (1.2), note that $1-\frac{1}{e} \approx 0.63212\cdots$. Also, if one wishes to be explicit, then $o(1)$ in the exponent becomes $O\left(\frac{\log_5 N}{\log_4 N}\right)$. For more details, see [7, §2].

[1] The conjecture was mistakenly stated for any $x_0 \in \mathbb{F}_q$, but it is essential to assume that $x_0 \neq 0$ since the form $\mathscr{L}_1$ is the zero form, and hence $\mathscr{L}_1(\mathbf{v}) = 0$ for all $\mathbf{v} \in \mathbb{F}_q^d$. The upper bound (1.4) is valid without any assumption on x_0, as it is based on examining square-free values of $n \geq 2$.

1.1. **Outline of the proof.** Since $\mathscr{L}_1$ is the zero form and $x_0 \neq 0$, it is enough to consider $\mathbf{v} \in \mathbb{F}_q^d$ such that $\mathscr{L}_n(\mathbf{v}) \neq x_0$ for all $n \in \{2, \dots, q-1\}$. In §3, we then reduce the problem of determining $N_q(x_0)$ for $x_0 \in \mathbb{F}_q^\times$ to that of finding $N_q := N_q(1)$. We further note that for any subset $\mathscr{N} \subset \{2, 3, 4, \dots, q-1\}$,

$$\begin{aligned} N_q &= \#\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) \neq 1 \text{ for all } n \in \{2, \dots, q-1\}\} \\ &\leq \#\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) \neq 1 \text{ for all } n \in \mathscr{N}\} \\ &= M_{q,\mathscr{N}} = \sum_{k=0}^{N} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S|=k}} M_{q,S}, \end{aligned}$$

where

$$M_{q,S} := \#\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) = 1 \text{ for all } n \in S\}.$$

In particular, truncating the inclusion/exclusion at an odd, or even, number of terms gives the following bounds on $M_{q,\mathscr{N}}$, for any $K \in \mathbb{N}$:

$$\sum_{k=0}^{2K-1} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S|=k}} M_{q,S} \leq M_{q,\mathscr{N}} \leq \sum_{k=0}^{2K} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S|=k}} M_{q,S}.$$

(These combinatorial bounds appears in many places in number theory, e.g. in Brun's pure sieve.) Let

$$\Sigma := \Sigma_K := \sum_{k=0}^{K} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S|=k}} M_{q,S}.$$

Observe that, if S is a set of $\mathbb{F}_q$-independent linear forms, then $M_{q,S} = q^{d-|S|}$ and this quickly yields the main term. Estimating the error term is more difficult; it amounts to determining the contribution from $M_{q,S}$ as S ranges over sets of $\mathbb{F}_q$-dependent forms. Our strategy is to first reduce the problem of $\mathbb{F}_q$-independence of subsets of forms $\{\mathscr{L}_n\}_{n=2}^{q-1}$ to multiplicative independence of subsets of $\{2, 3, \dots, q-1\}$ (see Lemma 3.2). A key technical result, perhaps of independent interest, is then that there exists large subsets $\mathscr{N}_q \subset \{2, 3, \dots, q-1\}$ such that essentially all k-tuples of distinct elements of $\mathscr{N}_q$ are multiplicatively independent, provided k is not too large. Before stating the result we introduce the following convenient notation: given a set $\mathscr{A}$ and $k \in \mathbb{N}$, let $\mathscr{A}^{[k]} := \{\mathscr{B} \subset \mathscr{A} : |\mathscr{B}| = k\}$.

Theorem 1.3. *For each integer q there exists $\mathscr{N}_q \subset \{2, 3, \dots, q-1\}$ such that, as $q \to \infty$,*

$$\#\mathscr{N}_q = q + O(q/\log_2 q)$$

(where the implied constant is less than 1) and

$$\begin{aligned}(1.6)\quad &\#\left\{\mathscr{S} \in \mathscr{N}_q^{[k]} : \mathscr{S} \text{ is multiplicatively independent}\right\} \\ &= \binom{\#\mathscr{N}_q}{k} + O((\#\mathscr{N}_q)^{k-3/2+o(1)}) = \frac{q^k}{k!} + O\left(\frac{q^k}{(k-1)!\log_2 q}\right),\end{aligned}$$

provided that $k = o(\sqrt{\log_2 q})$.

Using Theorem 1.3 we easily obtain a sufficiently good upper bound on N_q. To obtain a lower bound we remove all $\mathbf{v} \in \mathbb{F}_q^d$ such that $\mathscr{L}_n(\mathbf{v}) = 1$ for some n in the complementary set $\mathscr{N}_q^c = \{2, 3, \ldots, q-1\} \setminus \mathscr{N}_q$. As $\#\mathscr{N}_q^c = O(q/\log_2 q)$, a sufficient upper bound on the number of removed $\mathbf{v}$ follows easily (see §5.2.)

Remark 1.2. For recent results on asymptotics for the number of multiplicatively dependent k-tuples (not necessarily distinct) whose coordinates are algebraic numbers of bounded height, see [10]. In particular, [10, Theorem 1.1] gives an asymptotic for the number of multiplicatively dependent k-tuples, though not uniform in k. On the other hand, using [8, Corollary 3.2] (due to K. Yu) to find "short" exponent vectors in multiplicative relations leads to a good upper bound with a significant improvement in the level of uniformity in k. We thank Igor Shparlinski for pointing this out.

1.2. **Related results.** Little is known about the dynamics and distribution of ψ_p. The proof technique for [1, Theorem 4] implies $F(p) \le p^{\frac{1}{3}+o(1)}$. In [6], Friedrichsen and Holden introduced a probabilistic model for $F(p)$: the distribution of $F(p)$ should be closely related to $\sum_{d|p-1} X_d$, where X_d ranges over independent random variables having binomial distributions with parameters $(\phi(d), 1/d)$; they also gave numerical evidence for the validity of this model. See §6 for further numerical investigations. Further, in [7, Section 3], a heuristic argument that $\sum_{p \le N} F(p) = (1 + o(1))N$ was given.

As for lower and upper bounds on the size of the image, by Crocker [5] and Somer [11], we know that

$$\left\lceil \sqrt{\frac{p-1}{2}} \right\rceil \le \#\Big\{\psi_p(x) : x \in \{1, 2, \ldots, p-1\}\Big\} \le \frac{3}{4}p + O\left(p^{\frac{1}{2}+o(1)}\right).$$

There are also upper bounds on the cardinality of preimages: with

$$N(p, a) := \#\Big\{x \in \{1, 2, \ldots, p-1\} : \psi_p(x) \equiv a (\bmod\ p)\Big\},$$

and

$$M(p) := \#\Big\{(x, y) \in \{1, 2, \ldots, p-1\}^2 : \psi_p(x) = \psi_p(y)\Big\},$$

Balog, Broughan and Shparlinski [1, Corollary 5, Theorem 7 and Theorem 8] showed the following uniform bounds for a with $\gcd(a,p)=1$ and multiplicative order t:

$$N(p,a) \le \min\left\{p^{\frac{1}{3}+o(1)}t^{\frac{2}{3}}, p^{1+o(1)}t^{-\frac{1}{12}}\right\} \tag{1.7}$$

and

$$M(p) \le p^{\frac{48}{25}+o(1)}.$$

Let $a=1$. Then, as noted in [1], (1.7) implies $N(p,1) \le p^{\frac{1}{3}+o(1)}$. Cilleruelo and Garaev [4, 3] improve these bounds to $N(p,1) \le p^{\frac{27}{82}+o(1)}$ and $M(p) \le p^{\frac{23}{12}+o(1)}$.

Acknowledgements. We would like to thank Florian Luca and Igor Shparlinski for their comments on an early version of the paper. We would also like to thank the two anonymous referees for their careful reading of the paper and for comments that greatly improved the exposition, as well as leading to a sharper formulation of Theorem 1.3.

2. Notation

The letters p, q and ℓ denote prime numbers. The letters d, k, m, n, r, s and t denote natural numbers. Letters of the form $\mathbf{v}$ and $\mathbf{w}$ denote vectors in $\mathbb{F}_q^d$. For $n \in \mathbb{N}$, $\mathrm{rad}(n)$ and $P(n)$ respectively denote the largest squarefree divisor and the largest prime divisor of n. We write $p^\alpha \| n$ if $p^\alpha \mid n$ and $p^{\alpha+1} \nmid n$, and the function $\nu_\ell(n)$ denotes the maximum power of ℓ that divides n. That is, $\nu_\ell(n)=k$ means $\ell^k \| n$. We say that $n_1, n_2, \ldots, n_r$ are **multiplicatively independent** if $\alpha_1=\alpha_2=\cdots=\alpha_r=0$ is the only integer solution to $n_1^{\alpha_1}n_2^{\alpha_2}\cdots n_r^{\alpha_r}=1$. Otherwise, $n_1, n_2, \ldots, n_r$ are **multiplicatively dependent**. The linear form $\mathscr{L}_n$, where $n \in \mathbb{N}$, is defined in (1.3). We say $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \ldots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$**-independent** if $\alpha_1=\alpha_2=\cdots=\alpha_k=0$ with $\alpha_i \in \mathbb{F}_q$ for all $i \in \{1,2,\ldots,k\}$ is the only solution to

$$\alpha_1\mathscr{L}_{n_1}(\mathbf{v}) + \alpha_2\mathscr{L}_{n_2}(\mathbf{v}) + \cdots + \alpha_k\mathscr{L}_{n_k}(\mathbf{v}) = 0(\mathbf{v}) = 0$$

for all $\mathbf{v} \in \mathbb{F}_q^k$. Otherwise, $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \ldots, \mathscr{L}_{n_k}$ are called $\mathbb{F}_q$**-dependent**.

Recall that $\pi(x) := \#\{p \le x\}$, and that we define $\log_k(x)$ for $x \in \mathbb{R}_{>0}$ and $k \in \mathbb{N}$ iteratively: $\log x = \log_1 x = \max\{\ln x, 2\}$ and $\log_k x = \log(\log_{k-1} x)$ for $k \in \mathbb{N}$ and $k \ge 2$. Let $f : X \to \mathbb{C}$ and $g : X \to \mathbb{R}_{\ge 0}$ be functions. By the equivalent notations $f(x) = O(g(x))$ or $f \ll g$, we mean there exists a constant C such $|f(x)| \le Cg(x)$ for all $x \in X$. The constant C is called the **implied constant** when writing $f(x) = O(g(x))$. If the implied constant is dependent on some parameter P, then we write $f(x) = O_P(g(x))$ or $f(x) \ll_P g(x)$. We write $f(x) \asymp$

$g(x)$, $f(x) \sim g(x)$ and $f(x) = o(g(x))$ to signify $f(x) \ll g(x) \ll f(x)$, $f(x)/g(x) \to 1$ and $f(x)/g(x) \to 0$ as $x \to \infty$ with $x \in X$, respectively.

3. Lemmata

We first reduce the problem using the following lemmas.

Lemma 3.1. *If $x_0 \in \mathbb{F}_q^\times$ then $N_q(x_0) = N_q(1)$.*

Proof. The statements follow since $T_{x_0} : \mathbb{F}_q^d \to \mathbb{F}_q^d$ defined by $T_{x_0}(\mathbf{v}) = x_0 \cdot \mathbf{v}$ is an isomorphism if $x_0 \in \mathbb{F}_q^\times$. □

As such, denote $N_q = N_q(1)$.

Lemma 3.2. *Let $k \in \mathbb{N}$.*

(a) If $n_1, n_2, \dots, n_k$ are multiplicatively dependent, then the forms $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \dots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$-dependent.

(b) Suppose $k < \frac{\log q}{10 \log_2 q}$. Then, $n_1, n_2, \dots, n_k \in \{2, 3, \dots, q-1\}$ are multiplicatively independent if and only if $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \dots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$-independent.

Proof. Let $n_1, n_2, \dots, n_k \in \{2, 3, \dots, q-1\}$ be distinct. Suppose n_i has prime power factorization $n_i = p_1^{e_{i,1}} p_2^{e_{i,2}} \cdots p_d^{e_{i,d}}$, where $e_{i,j} = 0$ is permissible.

(a) Suppose $n_1, n_2, \dots, n_k$ are multiplicatively dependent. Then, there exist integers $\alpha_1, \alpha_2, \dots, \alpha_k$ such that $n_1^{\alpha_1} n_2^{\alpha_2} \cdots n_k^{\alpha_k} = 1$. In particular,

$$\begin{aligned} 1 &= \left(p_1^{e_{1,1}} p_2^{e_{1,2}} \cdots p_d^{e_{1,d}}\right)^{\alpha_1} \left(p_1^{e_{2,1}} p_2^{e_{2,2}} \cdots p_d^{e_{2,d}}\right)^{\alpha_2} \cdots \left(p_1^{e_{k,1}} p_2^{e_{k,2}} \cdots p_d^{e_{k,d}}\right)^{\alpha_k} \\ &= p_1^{\alpha_1 e_{1,1} + \alpha_2 e_{2,1} + \dots + \alpha_k e_{k,1}} p_2^{\alpha_1 e_{1,2} + \alpha_2 e_{2,2} + \dots + \alpha_k e_{k,2}} \cdots p_d^{\alpha_1 e_{1,d} + \alpha_2 e_{2,d} + \dots + \alpha_k e_{k,d}}. \end{aligned}$$

So, $\alpha_1 e_{1,m} + \alpha_2 e_{2,m} + \dots + \alpha_k e_{k,m} = 0$ for each $m \in \{1, 2, \dots, d\}$. As such,

$$0 = \sum_{j=1}^{d} \left(\sum_{i=1}^{k} \alpha_i e_{i,j} \right) v_j = \sum_{i=1}^{k} \alpha_i \sum_{j=1}^{d} e_{i,j} v_j = \sum_{i=1}^{k} \alpha_i \mathscr{L}_{n_i}(\mathbf{v})$$

for all $\mathbf{v} = (v_1, v_2, \dots, v_d) \in \mathbb{F}_q^d$. That is, $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \dots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$-dependent.

(b) Suppose $k < \frac{\log q}{10 \log_2 q}$. By (a), it suffices to show that multiplicative independence implies $\mathbb{F}_q$-independence. Suppose that $n_1, n_2, \dots, n_k$ are multiplicatively independent. If we let $E := (e_{i,j})_{i=1,2,\dots,k}^{j=1,2,\dots,d}$, then $\operatorname{rank}_{\mathbb{Z}}(E) = \operatorname{rank}_{\mathbb{Q}}(E) = k$. In particular, there exists an invertible $k \times k$ matrix E' which consists of k independent columns of E. Without loss of generality, the first k

columns of E are independent. Suppose $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \ldots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$-dependent. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{F}_q^k \setminus \{0\}$ be such that $\alpha_1 \mathscr{L}_{n_1} + \alpha_2 \mathscr{L}_{n_2} + \cdots + \alpha_k \mathscr{L}_{n_k} = 0$. Then, $E'\alpha = 0$. In particular, $q \mid \det(E')$. Recall that Hadamard's inequality states

$$|\det(E')| \leq \prod_{j=1}^{k} \|\mathbf{e}_j\|,$$

where $\mathbf{e}_j$ is the j^{th} row of E' and $\|\cdot\|$ is the Euclidean norm (e.g., see [2, §2.11].) Note that $\|\mathbf{e}_j\| \leq k^{1/2} \frac{\log q}{\log p_j}$. Thus,

$$q \text{ divides } |\det(E')| \leq k^{k/2} \prod_{j=1}^{k} \frac{\log q}{\log p_j} < \frac{k^{k/2} (\log q)^k}{\log 2} < q$$

since $k < \frac{\log q}{10 \log_2 q}$. Thus, $\det(E') = 0$, which implies $\operatorname{rank}(E') < k$, which is a contradiction. So, no such α exists and the forms $\mathscr{L}_{n_1}, \mathscr{L}_{n_2}, \ldots, \mathscr{L}_{n_k}$ are $\mathbb{F}_q$-independent.

$\square$

4. Proof of Theorem 1.3

To simplify the notation we will denote $\mathscr{N} := \mathscr{N}_q$, and let

$$N := \#\mathscr{N}. \tag{4.1}$$

4.1. **The subset $\mathscr{N}$.** Recall the following notation: for $m \in \mathbb{N}$ and ℓ a fixed prime,

$$P(m) := \max\{p : p|m\} \qquad \text{(the largest prime divisor of } m)$$
$$\nu_\ell(m) := \max\{\alpha \in \mathbb{N} \cup \{0\} : \ell^\alpha | m\} \qquad \text{(the } \ell\text{-adic valuation of } m)$$

The following parameters will be determined later: B, respectively $f(q)$, are parameters giving bounds on the exponents of large, respectively small, primes dividing elements of $\mathscr{N}$.

Let

$$\mathscr{N} := \Big\{ n \in \{2, 3, \ldots, q-1\} : n = sr, \text{ where } s \in \mathscr{S} \text{ and } r \in \mathscr{R} \Big\},$$

where

$$\mathscr{S} := \Big\{ s \in \{1, 2, \ldots, q-1\} : P(s) \leq B \text{ and } \nu_p(s) \leq f(q) \text{ for all primes } p \Big\}$$

and

$$\mathscr{R} := \Big\{ r \in \{1, 2, \ldots, q-1\} : p|r \text{ and } p \geq B \text{ implies } p \| r \Big\}.$$

We then find (recall that $N = \#\mathscr{N}$, cf. (4.1))

$$q - 2 - N \leq \#\Big\{n < q : p \mid n \text{ for some } p \leq B \text{ and } \nu_p(n) \geq f(q)\Big\} \\ + \#\Big\{n < q : p^2 \mid n \text{ for some } p > B\Big\}.$$

These quantities can be bounded as follows:

$$\begin{aligned}\#\Big\{n < q : p \mid n &\text{ for some } p \leq B \text{ and } \nu_p(n) \geq f(q)\Big\} \\ &\leq \sum_{p \leq B} \#\Big\{n < q : p \mid n \text{ implies } \nu_p(n) \geq f(q)\Big\} \\ &\leq q \sum_{p \leq B} \frac{1}{p^{f(q)}} \leq \frac{q}{2^{f(q)}} \pi(B).\end{aligned}$$

and

$$\#\Big\{n < q : p^2 \mid n \text{ for some } p > B\Big\} \leq \sum_{p > B} \frac{q}{p^2} \leq \frac{3q}{B \log B}$$

for all $B \geq 1$. In particular,

$$q - 2 - N \leq \frac{q\pi(B)}{2^{f(q)}} + \frac{3q}{B \log B} \leq \frac{qB}{2^{f(q)-1} \log B} + \frac{3q}{B \log B}.$$

Define

$$B := c_1 \log_2 q, \quad f(q) := c_2 \log_3 q, \tag{4.2}$$

where $c_1, c_2 > 0$ are constants to be chosen later. Then,

$$\begin{aligned} q - 2 - N &\leq \frac{c_1 q \log_2 q}{2^{c_2 (\log\log\log q) - 1}} + \frac{3q}{c_1 (\log_2 q)(\log c_2 + \log_3 q)} \\ &\leq 2c_1 q \exp\left(\log_3 q - (\log_3 q)^{c_2 \log 2}\right) + \frac{3q}{c_1 (\log_2 q)(\log c_2 + \log_3 q)} \\ &\leq c_3 \frac{q}{\log_2 q}, \end{aligned}$$

where c_3 is a constant and $c_3 \in (0, 1)$ if $c_2 > 2/\log 2$. In particular, for $c_2 > 2/\log 2$,

$$N = q + O\left(\frac{q}{\log_2 q}\right), \tag{4.3}$$

where the implied constant in (4.3) is less than 1.

4.2. **Multiplicatively dependent k-tuples of $\mathscr{N}$.** Assume that we are given distinct multiplicatively dependent integers $n_1, n_2, \ldots, n_k \in \mathscr{N}$, and suppose that $r < k$ is the (multiplicative) rank of these integers. That is, there exists $m_1, m_2, \ldots, m_r \in \{n_1, n_2, \ldots, n_k\}$ such that

(a) $m_1, m_2, \ldots, m_r$ are multiplicatively independent and
(b) for any $n \in \{n_1, n_2, \ldots, n_k\} \setminus \{m_1, m_2, \ldots, m_r\}$, the enlarged set $\{m_1, m_2, \ldots, m_r, n\}$ is multiplicatively dependent.

Without loss of generality, $n_i = m_i$ for all $i \in \{1, 2, \ldots, r\}$. Then, for every $j \in \{r+1, r+2, \ldots, k\}$, there exists $\alpha_j \in \mathbb{N}$ and $\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{rj} \in \mathbb{Z}$ such that

$$n_j^{\alpha_j} = n_1^{\alpha_{1j}} n_2^{\alpha_{2j}} \cdots n_r^{\alpha_{rj}}.$$

For convenience, let $j = r+1$, $\alpha = \alpha_{r+1}$, $\alpha_{ij} = \alpha_i$ and $n = n_{r+1}$. Then,

$$n^{\alpha} = n_1^{\alpha_1} n_2^{\alpha_2} \cdots n_r^{\alpha_r}. \tag{4.4}$$

Let $J_+ = \{j \in \{1, 2, \ldots, r\} : \alpha_j > 0\}$, $J_- = \{j \in \{1, 2, \ldots, r\} : \alpha_j < 0\}$ and $J_0 = \{j \in \{1, 2, \ldots, r\} : \alpha_j = 0\}$. Note that $|J_+| + |J_-| + |J_0| = r$ and

$$\operatorname{rad}(n) \,\Big|\, \operatorname{rad}\left(\prod_{j \in J_+} n_j\right)$$

as $n_1, n_2, \ldots, n_r \in \mathbb{N}$.

Case 1: $|J_-| = 0$. In this case, $\operatorname{rad}(n_j) \in \{d \in \mathbb{N} : d \mid \operatorname{rad}(n)\}$. Thus, there are $\tau(\operatorname{rad}(n))^{|J_+|}$ choices for the radicals of elements corresponding to J_+. There are also $N^{|J_0|}$ choices for elements corresponding to J_0.

For any squarefree number $n_0 \in \mathscr{N}$, the number of elements in $m \in \mathscr{N}$ with radical n_0 is bounded as follows: recall $m \in \mathscr{N}$ satisfies the condition that $\nu_p(m) \le 1$ for all $p > B$. So, the only place where $\operatorname{rad}(m) = n_0$ and m differ is in the prime factors $p \le B$. Thus, by the definition of $\mathscr{N}$, the number of choices for the difference of m and n_0 is bounded by (recall (4.2))

$$\pi(B)^{f(q)} \ll \left(\frac{\log_2 q}{\log_3 q}\right)^{c_2 \log_3 q} \ll \exp\left(c_2 (\log_3 q)^2\right).$$

So, the number of choices for n_j with $j \in J_+$ corresponding to n is

$$\ll \tau(\operatorname{rad}(n))^{|J_+|} \exp\left(c_2 |J_+| (\log_3 q)^2\right).$$

The classical bound

$$\tau(m) \ll \exp(C_\tau \log m / \log_2 m), \tag{4.5}$$

where $C_\tau > 0$ is a constant, yields

$$\tau(\mathrm{rad}(n)) \ll \exp\left(C_\tau \frac{\log q}{\log_2 q}\right).$$

As such, the number of choices of n_j with $j \in J_+$ given n is

$$\ll \exp\left(C_\tau |J_+| \frac{\log q}{\log_2 q} + c_2 |J_+| (\log_3 q)^2\right).$$

So, the number of overall choices is

$$\ll N^{1+|J_0|} \exp\left(C_\tau |J_+| \frac{\log q}{\log_2 q} + c_2 |J_+| (\log_3 q)^2\right).$$

Case 2: $|J_-| > 0$. Then,

$$n \prod_{j \in J_-} n_j^{\alpha_j} = \prod_{j \in J_+} n_j^{\alpha_j} =: n_{J_+}$$

Let $m = \mathrm{rad}(n_{J_+})$. Hence, $\mathrm{rad}(n) \mid \mathrm{rad}(m)$ and $\mathrm{rad}(n_j) \mid \mathrm{rad}(m)$ for all $j \in J_-$. As before, the number of choices for the radical of n, n_j with $j \in J_-$ is bounded by $\tau(m)^{1+|J_-|}$. Also, from the computation for the number of elements in $\mathscr{N}$ with radical m, we have that the number of overall choices in this case is bounded by

$$\begin{aligned} &\ll N^{|J_0|+|J_+|} \exp\left(|J_-| C_\tau (r-1) \frac{\log q}{\log_2 q} + c|J_-|(\log_3 q)^2\right) \\ &\ll N^{r-|J_-|} \exp\left(C_\tau r^2 \frac{\log q}{\log_2 q} + cr(\log_3 q)^2\right). \end{aligned}$$

Note that, for the remaining elements m_1 in $\{n_{r+2}, n_{r+3}, \ldots, n_k\}$ in Cases 1 and 2, we have

$$\mathrm{rad}(m_1) \mid \mathrm{rad}\left(\prod_{i=1}^{r} n_i\right).$$

In particular, there are $\tau(n_1 n_2 \cdots n_r)^{k-r-1}$ choices for the radical of m_1. Using the previous bound on the number of ways an element in $\mathscr{N}$ can have a fixed radical, we find that the total number of ways to chose the remaining $n_{r+2}, n_{r+3}, \ldots, n_k$ is

$$\ll \exp\left(k^2 C_\tau \frac{\log q}{\log_2 q}\right).$$

From the bounds in the two different cases it follows that the number of distinct k-tuples of elements in $\mathscr{N}$, having multiplicative rank $r < k$, and J_-, J_0, J_+ fixed, is

$$\ll \begin{cases} N^{1+|J_0|} \exp\left(C_\tau (k^2 + |J_+|)\frac{\log q}{\log_2 q} + c_2 |J_+| (\log_3 q)^2\right) & \text{if } |J_-| = 0, \\ N^{r-|J_-|} \exp\left(C_\tau (k^2 + r^2)\frac{\log q}{\log_2 q} + cr (\log_3 q)^2\right) & \text{if } |J_-| > 0. \end{cases} \tag{4.6}$$

We claim that (4.6) is $O(N^{r-1/2+o(1)})$ for suitably small r; this clearly holds if $|J_-| > 0$. If $|J_-| = 0$ the supposition would hold if $|J_0| < r-1$. Suppose, on the contrary, that $|J_0| = r-1$. Then, (4.4) yields $n_{r+1}^{\alpha_{r+1}} = n_i^{\alpha_i}$ for some $i \le r$, and $\alpha_{r+1}, \alpha_i > 0$. Without loss of generality we can assume that $(\alpha_{r+1}, \alpha_i) = 1$, and $\alpha_{r+1} > \alpha_i$ (the case $\alpha_{r+1} < \alpha_i$ is similar.) Since $(\alpha_{r+1}, \alpha_i) = 1$ we must have $n_{r+1} = M^{\alpha_i}$ and $n_i = M^{\alpha_{r+1}}$ for some integer $M > 1$; as $\alpha_{r+1} \ge 2$ and $n_i < q$, there are at most $q^{1/2}$ choices for M, and consequently there are a total of $O(q^{1/2+o(1)})$ choices for n_{r+1} and n_i, and at most N^{r-1} choices for the remaining $n_1, n_2, \ldots, n_{i-1}, n_{i+1}, n_r$.

Thus (4.6) is $O(N^{r-1/2+o(1)})$ if r is sufficiently small, and since $r \le k$, a choice of $k = o(\sqrt{\log_2 q})$ will suffice. For more explicit error terms we will argue as follows. Recall that an initial choice of a basis of size r was chosen. Now, for r fixed, the number of possible choices of triples, J_+, J_- and J_0 are bounded by the combinatorial factor 3^r. We thus find that

$$\#\{\mathscr{S} \in \mathscr{N}^{[k]} : \operatorname{rank}(\mathscr{S}) = r\} \ll \binom{k}{r} 3^r N^{r-1/2+o(1)} \exp\left(2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 r (\log_3 q)^2\right), \tag{4.7}$$

and hence

$$\begin{aligned} &\#\{\mathscr{S} \in \mathscr{N}^{[k]} : \mathscr{S} \text{ is multiplicatively dependent}\} \\ &\qquad \ll \sum_{r=1}^{k-1} \binom{k}{r} 3^r N^{r-1/2+o(1)} \exp\left(2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 r (\log_3 q)^2\right) \\ &\qquad \ll N^{k-3/2+o(1)} \exp\left(2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 k (\log_3 q)^2\right) \sum_{r=0}^{\infty} \frac{(3k)^r}{r!} \\ &\qquad \ll N^{k-3/2+o(1)} \exp\left(3k + 2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 k (\log_3 q)^2\right). \end{aligned}$$

In particular, for $k = o(\sqrt{\log_2 q})$,

$$\begin{aligned}
&\#\left\{\mathscr{S} \in \mathscr{N}^{[k]} : \mathscr{S} \text{ is multiplicatively independent}\right\} \\
&\qquad = \binom{N}{k} + O\left(N^{k-3/2+o(1)} \exp\left(3k + 2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 k (\log_3 q)^2\right)\right) \\
&\qquad = \binom{N}{k} + O\left(N^{k-3/2+o(1)}\right),
\end{aligned}$$

thus proving the first equality in (1.6). Then, (4.3) and the comment following it imply, for $k = o(\log_2 q)$, that

(4.8)

$$\begin{aligned}
\binom{N}{k} &= \frac{1}{k!} N(N-1)(N-2)\cdots(N-k+1) = \frac{1}{k!}\left(q + O\left(\frac{q}{\log_2 q}\right)\right)^k \\
&= \frac{q^k}{k!}(1+O(1/\log_2 q))^k = \frac{q^k}{k!}(1+O(k/\log_2 q)) = \frac{q^k}{k!} + O\left(\frac{q^k}{(k-1)!\log_2 q}\right).
\end{aligned}$$

Moreover, if $k = o(\sqrt{\log_2 q})$, then

$$\begin{aligned}
&\#\left\{\mathscr{S} \in \mathscr{N}^{[k]} : \mathscr{S} \text{ is multiplicatively independent}\right\} \\
&\qquad = \frac{q^k}{k!} + O\left(\frac{q^k}{(k-1)!\log_2 q}\right) + O\left(N^{k-2} \exp\left(3k + 2C_\tau k^2 \frac{\log q}{\log_2 q} + c_2 k (\log_3 q)^2\right)\right) \\
&\qquad = \frac{q^k}{k!} + O\left(\frac{q^k}{(k-1)!\log_2 q}\right),
\end{aligned}$$

where the implied constant is absolute. The proof of Theorem 1.3 is thus concluded.

5. Proof of Theorem 1.1

Denote $\mathscr{N} = \mathscr{N}_q$ and $N = \#\mathscr{N}$, with $\mathscr{N}_q$ as in Theorem 1.3. Recall from §1.1 that $N_q \le M_{q,\mathscr{N}}$, and that

$$\sum_{k=0}^{2K-1} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S| = k}} M_{q,S} \le M_{q,\mathscr{N}} \le \sum_{k=0}^{2K} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S| = k}} M_{q,S},$$

where $M_{q,S} := \#\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) = 1 \text{ for all } n \in S\}$. Let

$$\Sigma := \Sigma_K := \sum_{k=0}^{K} (-1)^k \sum_{\substack{S \subset \mathscr{N} \\ |S| = k}} M_{q,S}.$$

Then,

$$\begin{aligned}\Sigma &= \sum_{k=0}^{K}(-1)^k \sum_{\substack{S=\{n_1,n_2,\ldots,n_k\}\in\mathscr{N}^{[k]}\\ \{\mathscr{L}_{n_1},\mathscr{L}_{n_2},\ldots,\mathscr{L}_{n_k}\}\text{ is }\mathbb{F}_q\text{-independent}}} M_{q,S} \\ &\quad+\sum_{k=0}^{K}(-1)^k \sum_{\substack{S=\{n_1,n_2,\ldots,n_k\}\in\mathscr{N}^{[k]}\\ \{\mathscr{L}_{n_1},\mathscr{L}_{n_2},\ldots,\mathscr{L}_{n_k}\}\text{ is }\mathbb{F}_q\text{-dependent}}} M_{q,S} \\ &=: \Sigma_1+\Sigma_2,\end{aligned} \tag{5.1}$$

say. Now, for $K < \log q/(10\log_2 q)$, Lemma 3.2 together with the rank-nullity theorem of linear algebra implies that

$$\Sigma_1 = \sum_{k=0}^{K}(-1)^k q^{d-k} \sum_{\substack{\{n_1,n_2,\ldots,n_k\}\in\mathscr{N}^{[k]}\\ n_1,n_2,\ldots,n_k\text{ are}\\ \text{multiplicatively independent}}} 1 \tag{5.2}$$

and

$$\Sigma_2 = \sum_{k=0}^{K}(-1)^k \sum_{\substack{S=\{n_1,n_2,\ldots,n_k\}\in\mathscr{N}^{[k]}\\ n_1,n_2,\ldots,n_k\text{ are}\\ \text{multiplicatively dependent}}} M_{q,S}. \tag{5.3}$$

5.1. **The Upper Bound.** For K even (recall (5.1)-(5.3)), we have

$$N_q \leq \#\left\{\mathbf{v}\in\mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v})\neq 1 \text{ for all } n\in\mathscr{N}\right\} \leq \Sigma_1+\Sigma_2.$$

Theorem 1.3, together with (5.2), gives

$$\begin{aligned}\Sigma_1 &= \sum_{k=0}^{K}(-1)^k q^{d-k}\left(\frac{q^k}{k!}+O\left(\frac{q^k}{(k-1)!\log_2 q}\right)\right) \\ &= \frac{q^d}{e}+O\left(q^d\sum_{k>K}\frac{1}{k!}\right)+O\left(\frac{q^d}{\log_2 q}\right) \\ &= \frac{q^d}{e}+O\left(\left(\frac{2}{K}\right)^{\frac{K}{2}}q^d\right)+O\left(\frac{q^d}{\log_2 q}\right)\end{aligned} \tag{5.4}$$

for K growing with q so that $K = o\left(\sqrt{\log_2 q}\right)$.

For Σ_2, there are no multiplicatively dependent sets of size 1 unless $n = 1$. By (4.7), together with the rank-nullity theorem, we have

(5.5)
$$\begin{aligned}
\Sigma_2 &\ll \sum_{k=2}^{K} \sum_{\substack{S=\{n_1,n_2,\dots,n_k\}\in\mathscr{N}^{[k]} \\ n_1,n_2,\dots,n_k \text{ are} \\ \text{multiplicatively dependent}}} M_{q,S} \le \sum_{k=2}^{K}\sum_{r=1}^{k-1} \sum_{\substack{S=\{n_1,n_2,\dots,n_k\}\in\mathscr{N}^{[k]} \\ \mathrm{rank}_{\mathbb{Z}}(S)=r}} M_{q,S} \\
&\ll \sum_{k=2}^{K}\sum_{r=1}^{k-1} \binom{k}{r} 3^r q^{d-r} N^{r-1} \exp\left(2C_\tau r^2 \frac{\log q}{\log_2 q} + c_2 r (\log_3 q)^2\right) \\
&\ll q^{d-1} \exp\left(3K + 2C_\tau K^2 \frac{\log q}{\log_2 q} + c_2 K (\log_3 q)^2\right).
\end{aligned}$$

Thus, by (5.4) and (5.5),

$$\begin{aligned}
N_q \le M_{q,\mathscr{N}} &\le \frac{q^d}{e} + O\left(\left(\frac{2}{K}\right)^{\frac{K}{2}} q^d\right) + O\left(\frac{q^d}{\log_2 q}\right) \\
&\qquad + O\left(q^{d-1}\exp\left(3K + 2C_\tau K^2 \frac{\log q}{\log_2 q} + c_2 K (\log_3 q)^2\right)\right) \\
&= \frac{q^d}{e} + O\left(\left(\frac{2}{K}\right)^{\frac{K}{2}} q^d\right) + O\left(\frac{q^d}{\log_2 q}\right)
\end{aligned}$$

for $K = o\left(\sqrt{\log_2 q}\right)$, and taking $K \asymp (\log_2 q)^{1/3}$ yields

$$M_{q,\mathscr{N}} \le \frac{q^d}{e} + O\left(\frac{q^d}{\log_2 q}\right).$$

A similar argument with K odd gives a lower bound of the same form, and thus

$$N_q \le M_{q,\mathscr{N}} = \frac{q^d}{e} + O\left(\frac{q^d}{\log_2 q}\right). \tag{5.6}$$

5.2. **The Lower Bound.** In §5.1, we proved

$$\begin{aligned}
N_q &:= \#\left\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) \ne 1 \text{ for all } n \in \{2,3,\dots,q-1\}\right\} \\
&\le M_{q,\mathscr{N}} = \frac{q^d}{e} + O\left(\frac{q^d}{\log_2 q}\right)
\end{aligned}$$

by restricting the set $\{2,3,\dots,q-1\}$ to $\mathscr{N}$ and showing (cf. (5.6))

$$\#\left\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_n(\mathbf{v}) \ne 1 \text{ for all } n \in \mathscr{N}\right\} = \frac{q^d}{e} + O\left(\frac{q^d}{\log_2 q}\right).$$

Let $\mathscr{N}_{\text{bad}} = \{2, 3, \ldots, q-1\} \setminus \mathscr{N}$. Then,

$$N_q = \frac{q^d}{e} + O\left(\frac{q^d}{\log_2 q}\right) + O\Big(\{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_m(\mathbf{v}) = 1 \text{ for some } m \in \mathscr{N}_{\text{bad}}\}\Big).$$

Note that $N_{\text{bad}} := \#\mathscr{N}_{\text{bad}} \ll \frac{q}{\log_2 q}$ by (4.3). The number of $\mathbf{w} \in \{\mathbf{v} \in \mathbb{F}_q^d : \mathscr{L}_m(\mathbf{v}) = 1$ for some $m \in \mathscr{N}_{\text{bad}}\}$ is bounded by

$$\ll \sum_{m \in \mathscr{N}_{\text{bad}}} \#\left\{v \in \mathbb{F}_q^d : \mathscr{L}_m(\mathbf{v}) = 1\right\} \ll q^{d-1} N_{\text{bad}} \ll \frac{q^d}{\log_2 q}.$$

In particular, $N_q = q^d/e + O(q^d/\log_2 q)$; after dividing by q^d the proof of Theorem 1.1 is concluded.

6. Statistics

We have compared the model introduced by Friedrichsen and Holden with the data from the problem. Below (cf. Figures 1 and 2) are the histograms and the quantile-quantile plots for some seven and ten-digits primes. The quantile-quantile plots compare the theoretical quantiles (red line, Gaussian with mean 0 and standard deviation 1) with the observed ones (coloured dots) from our experiment. The data is broken up based on how large $\omega(p-1)$ is. The datasets sizes are 7216 (seven) and 241 148 (ten). The red curve in the histograms is the Gaussian with mean and standard deviation μ and σ, respectively, as reported.

The model for the problem is as follows: we wish to count

$$F(p) := \#\Big\{x \in \{1, 2, \ldots, p-1\} : x^x \equiv x \pmod p\Big\}.$$

Consider the following lemma:

Lemma 6.1. *Let p be a prime, $y \in \{1, 2, \ldots, p(p-1)\}$ an integer such that $p \nmid y$, and let d be a divisor of $p-1$ such that $\operatorname{ord}_p y \mid d$. Then,*

$$\#\Big\{x \in \{1, 2, \ldots, p(p-1)\} : p \nmid x, x^x \equiv y \pmod p, \operatorname{ord}_p x = d\Big\} = \frac{p-1}{d}.$$

Here $\operatorname{ord}_p(y)$ denotes the multiplicative order of y modulo p, i.e., the smallest integer $k > 0$ such that $y^k \equiv 1 \mod p$.

This lemma implies that the number of solutions to $x^x \equiv x \pmod p$ with $1 \leq x \leq p(p-1)$ is

$$(p-1) \sum_{d \mid n} \frac{\varphi(d)}{d}$$

(see Friedrichesen and Holden [6]). The above result suggests that $F(p)$ should be distributed as a binomial random variable with mean μ and variance σ^2, where

$$\mu = \sum_{d|p-1} \frac{\varphi(d)}{d} \qquad \text{and} \qquad \sigma^2 = \sum_{d|p-1} \frac{\varphi(d)(d-1)}{d^2}.$$

The histograms in Figure 1 represent the normalized statistic for $F(p)$ according to this model.

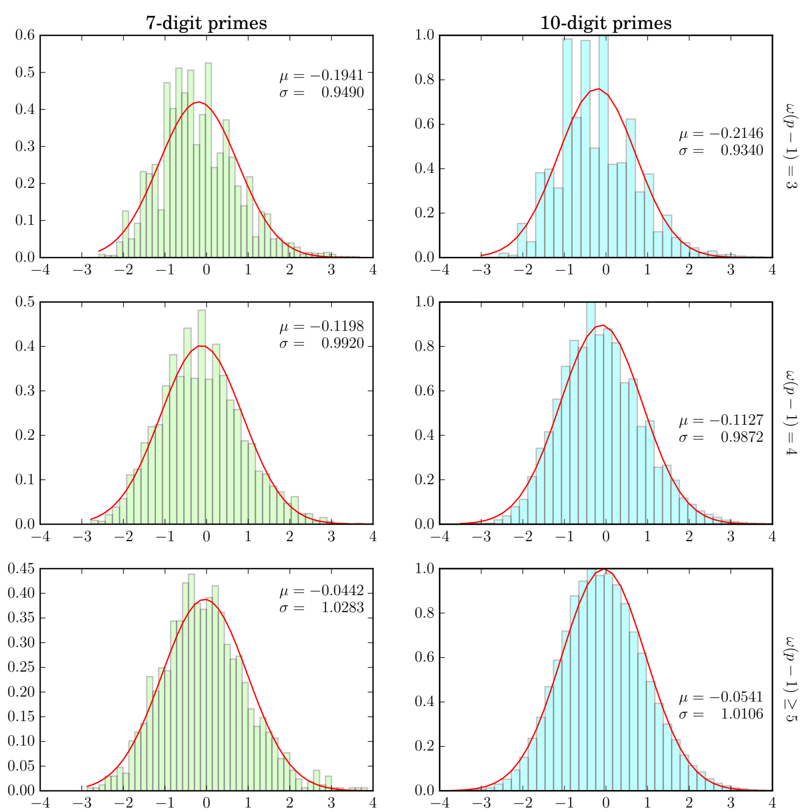

FIGURE 1. Histograms for seven- and ten-digit primes broken up into subgroups ($\omega(p-1) = 3$, $\omega(p-1) = 4$ and $\omega(p-1) \geq 5$).

That is, for a prime p, we compute $F(p)$ using primitive roots and index calculus. Then, we normalize $F(p)$ to $z = (F(p) - \mu)/\sigma$, where μ and σ are as above. The resulting histograms are presented in Figure 1. As can be seen from the histograms, the data seems to be tending to a normal distribution $N(0,1)$, especially in the mean μ.

The probability plots below compare our observed data with the theoretical model $N(0,1)$ as follows. The i^{th} order (descending) statistic for the theoretical values is defined according to Filliben's estimate:

$$i^{\text{th}} \text{ order statistic} = \begin{cases} (0.5)^{1/n} & \text{if } i = n, \\ \frac{i-0.3175}{n+0.365} & \text{if } i \in \{2, 3, \dots, n-1\}, \\ 1-(0.5)^{1/n} & \text{if } i = 1, \end{cases}$$

where n is the size of the dataset. As the quantile function is the inverse of the cumulative distribution function, we obtain the red line in Figure 2. For the observed data, we sort the corresponding values for $z = (F(p) - \mu)/\sigma$ and plot these values according to their values on the y-axis (observed values).

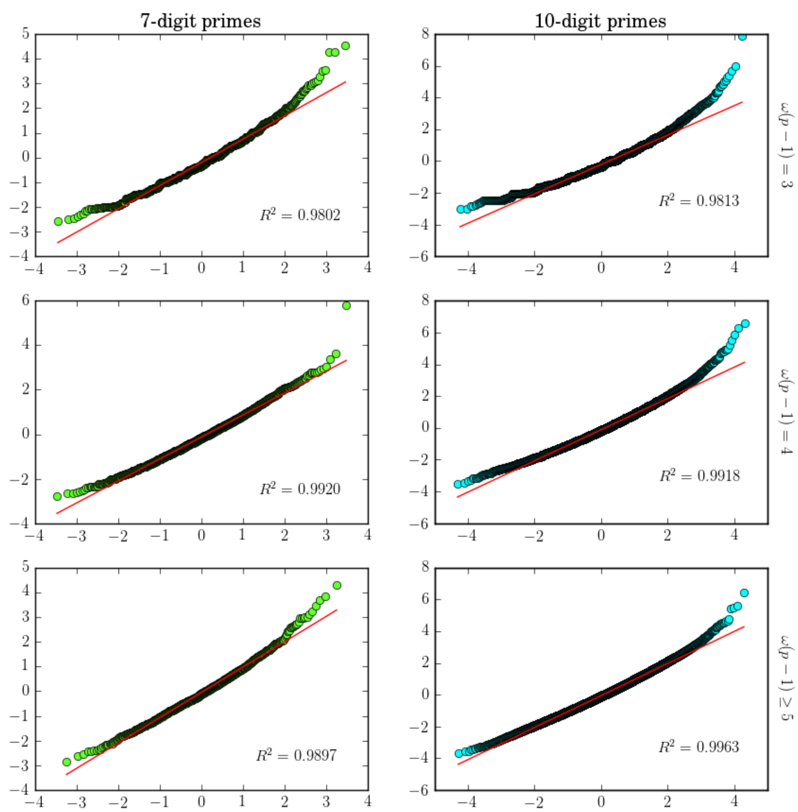

FIGURE 2. Probability plots for seven- and ten-digit primes broken up into subgroups ($\omega(p-1) = 3$, $\omega(p-1) = 4$ and $\omega(p-1) \geq 5$).

The high values of R^2 in Figure 2 indicate the model explains the observed variation very well. We note that, as can be seen in all the probability plots, there is a tendency for the data to have a higher standard deviation on the tails. We have not been able to determine a satisfactory explanation for this behaviour. We finally remark that the group $\omega(p-1) = 2$ was computed but the data gives rise to many outliers for reasons that are readily ascertainable.

References

[1] Antal Balog, Kevin A. Broughan, and Igor E. Shparlinski. On the number of solutions of exponential congruences. *Acta Arith.*, 148(1):93–103, 2011.

[2] Edwin F. Beckenbach and Richard Bellman. *Inequalities*. Second revised printing. Ergebnisse der Mathematik und ihrer Grenzgebiete. Neue Folge, Band 30. Springer-Verlag, New York, Inc., 1965.

[3] J. Cilleruelo and M. Z. Garaev. Congruences involving product of intervals and sets with small multiplicative doubling modulo a prime and applications. *Math. Proc. Cambridge Philos. Soc.*, 160(3):477–494, 2016.

[4] Javier Cilleruelo and Moubariz Z. Garaev. On the congruence $x^x \equiv \lambda \pmod{p}$, 2015.

[5] Roger Crocker. On residues of n^n. *Amer. Math. Monthly*, 76:1028–1029, 1969.

[6] Matthew Friedrichsen and Joshua Holden. Statistics for fixed points of the self-power map, arxiv:1403.5548, 2014.

[7] Pär Kurlberg, Florian Luca, and Igor E. Shparlinski. On the fixed points of the map $x \mapsto x^x$ modulo a prime. *Math. Res. Lett.*, 22(1):141–168, 2015.

[8] Thomas Loher and David Masser. Uniformly counting points of bounded height. *Acta Arith.*, 111(3):277–297, 2004.

[9] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography*. CRC Press Series on Discrete Mathematics and its Applications. CRC Press, Boca Raton, FL, 1997. With a foreword by Ronald L. Rivest.

[10] Francesco Pappalardi, Min Sha, Igor E. Shparlinski, and Cameron L. Stewart. On multiplicatively dependent vectors of algebraic numbers, arxiv:1606.02874, 2016.

[11] Lawrence Somer. The residues of n^n modulo p. *Fibonacci Quart.*, 19(2):110–117, 1981.

Department of Mathematics, KTH, Royal Institute of Technology, 100 44 Stockholm, Sweden

E-mail address: atfelix@kth.se

Department of Mathematics, KTH, Royal Institute of Technology, 100 44 Stockholm, Sweden

E-mail address: kurlberg@kth.se