THE AVERAGE ORDER OF ELEMENTS IN THE MULTIPLICATIVE GROUP OF A FINITE FIELD

YILAN HU AND CARL POMERANCE

ABSTRACT. We consider the average multiplicative order of a nonzero element in a finite field and compute the mean of this statistic for all finite fields of a given degree over their prime fields.

1. INTRODUCTION

For a cyclic group of order n, let $\alpha(n)$ denote the average order of an element. For each $d \mid n$, there are exactly $\varphi(d)$ elements of order d in the group (where φ is Euler's function), so

$$\alpha(n) = \frac{1}{n} \sum_{d \mid n} d\varphi(d).$$

It is known (von zur Gathen, et al. [2]) that

$$\frac{1}{x} \sum_{n \le x} \alpha(n) = \frac{3\zeta(3)}{\pi^2} x + O\left((\log x)^{2/3} (\log\log x)^{4/3}\right).$$

We are interested here in obtaining an analogous result where n runs over the orders of the multiplicative groups of finite fields. Let p denote a prime number. We know that up to isomorphism, for each positive integer k, there is a unique finite field of p^k elements. The multiplicative group for this field is cyclic of size $p^k - 1$. We are concerned with the average order of an element in this cyclic group as p varies. We show the following results.

Theorem 1. *For each positive integer k there is a positive constant K_k such that the following holds. For each number $A > 0$, each number $x \ge 2$, and each positive integer k with $k \le (\log x)/(2 \log\log x)$, we have*

$$\frac{1}{\pi(x)} \sum_{p \le x} \frac{\alpha(p^k - 1)}{p^k - 1} = K_k + O_A\left(\frac{1}{\log^A x}\right).$$

This theorem in the case $k = 1$ appears in Luca [3]. Using Theorem 1 and a partial summation argument we are able to show the following consequence.

Corollary 2. *For all numbers $A > 0$, $x \ge 2$, and for any positive integer $k \le (\log x)/(2 \log\log x)$, we have*

$$\frac{1}{\pi(x)} \sum_{p \le x} \alpha(p^k - 1) = K_k \frac{\mathrm{li}\,(x^{k+1})}{\mathrm{li}\,(x)} + O_A\left(\frac{x^k}{\log^A x}\right),$$

where K_k is the constant from Theorem 1 and $\mathrm{li}\,(x) := \int_2^x \mathrm{d}t/\log t$.

2000 *Mathematics Subject Classification.* 11B05, 11B75.

This paper is based on the first author's 2010 Dartmouth honors thesis, written under the direction of the second author. They gratefully acknowledge Florian Luca, who suggested the problem. The second author was supported in part by NSF grant DMS-1001180.

Since $\mathrm{li}(x^{k+1})/\mathrm{li}(x) \sim x^k/(k+1)$ as $x \to \infty$, Corollary 2 implies that

$$\frac{1}{\pi(x)} \sum_{p \le x} \alpha(p^k - 1) \sim \frac{K_k}{k+1} x^k, \text{ as } x \to \infty.$$

We identify the constants K_k as follows. Let $N_k(n)$ denote the number of solutions to the congruence $s^k \equiv 1 \pmod{n}$.

Proposition 3. *For each prime p and positive integer k let*

$$S_k(p) = \sum_{j=1}^{\infty} \frac{N_k(p^j)}{p^{3j-1}}.$$

Then $S_k(p) < 1$ and

$$K_k := \prod_p (1 - S_k(p))$$

is a real number with $0 < K_k < 1$.

2. Preliminary results

In this section we prove Proposition 3 and we also prove a lemma concerning the function $N_k(n)$.

Proof of Proposition 3. We clearly have $N_k(n) \le \varphi(n)$ for every n, since $N_k(n)$ counts the number of elements in the group $(\mathbb{Z}/n\mathbb{Z})^*$ with order dividing k and there are $\varphi(n)$ elements in all in this group. Thus, we have

$$S_k(p) \le \sum_{j=1}^{\infty} \frac{\varphi(p^j)}{p^{3j-1}} = \left(1 - \frac{1}{p}\right) \sum_{j=1}^{\infty} \frac{p}{p^{2j}} = \left(1 - \frac{1}{p}\right) \frac{p}{p^2 - 1} = \frac{1}{p+1}.$$

This proves the first assertion, but it is not sufficient for the second assertion. For p an odd prime, the group $(\mathbb{Z}/p^j Z)^*$ is cyclic so that the number of elements in this group of order dividing k is

$$N_k(p^j) = \gcd(k, \varphi(p^j)). \tag{1}$$

The same holds for $p^j = 2$ or 4, or if $p = 2$ and k is odd. Suppose now that $p = 2, j \ge 3$, and k is even. Since $(\mathbb{Z}/2^j\mathbb{Z})^*$ is the direct product of a cyclic group of order 2 and a cyclic group of order 2^{j-2}, we have

$$N_k(2^j) = 2 \cdot \gcd(k, 2^{j-2}) = \gcd(2k, \varphi(2^j)). \tag{2}$$

Thus, we always have $N_k(p^j) \le 2k$, and so

$$S_k(p) \le \sum_{j=1}^{\infty} \frac{2k}{p^{3j-1}} = \frac{2kp}{p^3 - 1}.$$

In particular, we have $S_k(p) = O_k(1/p^2)$, which with our first assertion implies that the product for K_k converges to a positive real number that is less than 1. This completes the proof. $\square$

Lemma 4. *For every positive integer k and each real number $x \ge 1$ we have*

$$\sum_{n \le x} \frac{N_k(n)}{n} \le 2(1 + \log x)^k.$$

Proof. Let $\omega(n)$ denote the number of distinct primes that divide n and let $\tau_k(n)$ denote the number of ordered factorizations of n into k positive integral factors. Since $k^{\omega(n)}$ is the number of ordered factorizations of n into k pairwise coprime factors, we have $k^{\omega(n)} \le \tau_k(n)$ for all n. Further, from (1), (2) and the fact that $N_k(n)$ is multiplicative in the variable n, we have $N_k(n) \le 2k^{\omega(n)}$, so that $N_k(n) \le 2\tau_k(n)$. Thus, it suffices to show that

$$\sum_{n\le x} \frac{\tau_k(n)}{n} \le (1+\log x)^k. \tag{3}$$

We prove (3) by induction on k. It holds for $k=1$ since $\tau_1(n)=1$ for all n, so that

$$\sum_{n\le x} \frac{N_1(n)}{n} = \sum_{n\le x} \frac{1}{n} \le 1 + \int_1^x \frac{\mathrm{d}t}{t} = 1+\log x.$$

Assume now that $k\ge 1$ and that (3) holds for k. Since $\tau_{k+1}(n) = \sum_{d|n} \tau_k(n)$,

$$\begin{aligned}\sum_{n\le x} \frac{\tau_{k+1}(n)}{n} &= \sum_{n\le x} \frac{1}{n} \sum_{d|n} \tau_k(d) = \sum_{d\le x} \frac{\tau_k(d)}{d} \sum_{m\le x/d} \frac{1}{m} \\ &\le \sum_{d\le x} \frac{\tau_k(d)}{d}(1+\log x) \le (1+\log x)^{k+1},\end{aligned}$$

by the induction hypothesis. This completes the proof. □

Corollary 5. *For k a positive integer and y a positive real with $k \le 1+\log y$, we have*

$$\sum_{n>y} \frac{N_k(n)}{n^2} \le 2(k+1)\frac{(1+\log y)^k}{y}.$$

Proof. By partial summation, Lemma 4, and integration by parts, we have

$$\begin{aligned}\sum_{n>y} \frac{N_k(n)}{n^2} &= \int_y^\infty \frac{1}{t^2} \sum_{y<n\le t} \frac{N_k(n)}{n}\,\mathrm{d}t \le 2\int_y^\infty \frac{(1+\log t)^k}{t^2}\,\mathrm{d}t \\ &= \frac{2}{y}\left((1+\log y)^k + k(1+\log y)^{k-1} + k(k-1)(1+\log y)^{k-2} + \cdots + k!\right) \\ &\le 2(k+1)\frac{(1+\log y)^k}{y},\end{aligned}$$

using $k \le 1+\log y$. This completes the proof. □

3. The main theorem

Proof of Theorem 1. The function

$$\frac{\alpha(m)}{m} = \frac{1}{m^2}\sum_{n|m} n\varphi(n)$$

is multiplicative and so by Möbius inversion, we may write

$$\frac{\alpha(m)}{m} = \sum_{n|m} \gamma(n),$$

where γ is a multiplicative function. It is easy to compute that

$$\gamma(p^j) = -\frac{p-1}{p^{2j}} \tag{4}$$

for every prime p and positive integer j. If $\mathrm{rad}(n)$ denotes the largest squarefree divisor of n, we thus have

$$\gamma(n) = (-1)^{\omega(n)}\frac{\varphi(\mathrm{rad}(n))}{n^2} \tag{5}$$

for each positive integer n. Note that (4), (5) are also in [3].

For n a positive ineger, label the $N_k(n)$ roots to the congruence $s^k \equiv 1 \pmod n$ as $s_{k,1}, s_{k,2}, \dots, s_{k,N_k(n)}$. We have

$$\sum_{p\le x}\frac{\alpha(p^k-1)}{p^k-1} = \sum_{p\le x}\sum_{n\mid p^k-1}\gamma(n) = \sum_{n\le x^k-1}\gamma(n)\sum_{\substack{p\le x\\ n\mid p^k-1}}1$$
$$= \sum_{n\le x^k-1}\gamma(n)\sum_{i=1}^{N_k(n)}\pi(x;n,s_{k,i}),$$

where $\pi(x;q,a)$ denotes the number of primes $p\le x$ with $p\equiv a \pmod q$.

If q is not too large in comparison to x and if a is coprime to q, we expect $\pi(x;q,a)$ to be approximately $\frac{1}{\varphi(q)}\pi(x)$. With this thought in mind, let $E_{q,a}(x)$ be defined by the equation

$$\pi(x;q,a) = \frac{1}{\varphi(q)}\pi(x) + E_{q,a}(x).$$

Further, let $y = x^{1/2}/\log^{A+4} x$, where A is as in the statement of Theorem 1. From the above, we thus have

$$\begin{aligned}\sum_{p\le x}\frac{\alpha(p^k-1)}{p^k-1} &= \sum_{n\le x^k-1}\gamma(n)\sum_{i=1}^{N_k(n)}\pi(x;n,s_{k,i})\\ &= \sum_{n\le y}\frac{\gamma(n)N_k(n)}{\varphi(n)}\pi(x) + \sum_{n\le y}\gamma(n)\sum_{i=1}^{N_k(n)}E_{n,s_{k_i}}(x) + \sum_{y<n\le x^k-1}\gamma(n)\sum_{i=1}^{N_k(n)}\pi(x;n,s_{k,i})\\ &= T_1+T_2+T_3, \quad \text{say.}\end{aligned}$$

We further refine the main term T_1 as

$$T_1 = \pi(x)\sum_{n=1}^{\infty}\frac{\gamma(n)N_k(n)}{\varphi(n)} - \pi(x)\sum_{n>y}\frac{\gamma(n)N_k(n)}{\varphi(n)}.$$

The first sum here has an Euler product as

$$\sum_{n=1}^{\infty}\frac{\gamma(n)N_k(n)}{\varphi(n)} = \prod_p\left(1+\sum_{j=1}^{\infty}\frac{\gamma(p^j)N_k(p^j)}{\varphi(p^j)}\right) = \prod_p\left(1-\sum_{j=1}^{\infty}\frac{N_k(p^j)}{p^{3j-1}}\right) = K_k,$$

where we used (4). For the second sum in the expression for T_1, we have by (5) and Corollary 5,

$$\left|\sum_{n>y}\frac{\gamma(n)N_k(n)}{\varphi(n)}\right| \le \sum_{n>y}\frac{N_k(n)}{n^2} \le 2(k+1)\frac{(1+\log y)^k}{y}.$$

Here we have used $k \le (\log x)/(2\log\log x)$ and $y = x^{1/2}/\log^{A+4} x$, so that $k \le 1 + \log y$ for all sufficiently large x depending on the choice of A. Further, with these choices for k, y we have $(1+\log y)^k < x^{1/2}$ for x sufficiently large, so that

$$\pi(x)\left|\sum_{n>y}\frac{\gamma(n)N_k(n)}{\varphi(n)}\right| \le \pi(x)\frac{2(k+1)(1+\log y)^k}{y} \le \frac{\pi(x)}{\log^A x}$$

for all sufficiently large values of x depending on A. Thus, $T_1 = K_k\pi(x) + O_A(\pi(x)/\log^A x)$.

Thus, it remains to show that both T_2 and T_3 are $O_A(\pi(x)/\log^A x)$. Using the elementary estimate $\pi(x;q,a) \le 1 + x/q$, we have

$$|T_3| \le \sum_{y<n\le x^k-1}|\gamma(n)|N_k(n)\left(1+\frac{x}{n}\right) \le \sum_{y<n\le x^k-1}\frac{N_k(n)}{n} + x\sum_{y<n\le x^k-1}\frac{N_k(n)}{n^2},$$

by (5). We have seen that the second sum here is negligible, and the first sum is bounded by $2(1+k\log x)^k$ using Lemma 4. This last expression is smaller than

$$\left(\frac{\log^2 x}{\log\log x}\right)^k = \frac{x}{\exp(\log x\log\log\log x/(2\log\log x))} = O_A\left(\frac{\pi(x)}{\log^A x}\right)$$

for any fixed choice of A.

To estimate T_2, note that

$$|T_2| \le \sum_{n\le y}|\gamma(n)|N_k(n)\max_{(a,n)=1}\left|\pi(x;n,a) - \frac{1}{\varphi(n)}\pi(x)\right| \le \sum_{n\le y}\max_{(a,n)=1}\left|\pi(x;n,a) - \frac{1}{\varphi(n)}\pi(x)\right|,$$

since $|\gamma(n)| \le \varphi(n)/n^2 \le 1/n$ and $N_k(n) \le \varphi(n) \le n$. Thus, by the Bombieri–Vinogradov theorem, see [1, Ch. 28], we have $|T_2| = O_A(\pi(x)/\log^A x)$, by our choice of y. These estimates conclude our proof of Theorem 1. □

4. Proof of Corollary 2 and more on the constants K_k

In this section we prove Corollary 2 and we numerically compute a few of the constants K_k.

Proof of Corollary 2. By partial summation, we have

$$\begin{aligned}\sum_{p\le x}\alpha(p^k-1) &= \sum_{p\le x}\frac{\alpha(p^k-1)}{p^k-1}(p^k-1)\\ &= (x^k-1)\sum_{p\le x}\frac{\alpha(p^k-1)}{p^k-1} - \int_2^x kt^{k-1}\sum_{p\le t}\frac{\alpha(p^k-1)}{p^k-1}\,\mathrm{d}t.\end{aligned}$$

Thus, by Theorem 1, the prime number theorem, and integration by parts, we have

$$\begin{aligned}\sum_{p\le x}\alpha(p^k-1) &= (x^k-1)K_k\pi(x) - \int_2^x kt^{k-1}K_k\pi(t)\,\mathrm{d}t + O\left(\frac{\pi(x)x^k}{\log^A x}\right)\\ &= (x^k-1)K_k\mathrm{li}(x) - \int_2^x kt^{k-1}K_k\mathrm{li}(t)\,\mathrm{d}t + O\left(\frac{\pi(x)x^k}{\log^A x}\right)\\ &= \int_2^x K_k\frac{t^k}{\log t}\,\mathrm{d}t + O\left(\frac{\pi(x)x^k}{\log^A x}\right).\end{aligned}$$

This last integral is $K_k \mathrm{li}(x^{k+1}) - K_k \mathrm{li}(2^{k+1})$, so the corollary now follows via one additional call to the prime number theorem. $\square$

We now examine the constants K_k for $k \le 4$. Since $N_1(p^j) = 1$ for all p^j, we have

$$K_1 = \prod_p \left(1 - \sum_{j\ge 1} \frac{p}{p^{3j}}\right) = \prod_p \left(1 - \frac{p}{p^3 - 1}\right) = 0.5759599689\ldots.$$

(This constant is also worked out in [3].) For K_2 we note that $N_2(p^j) = 2$ for all prime powers p^j except that $N_2(2) = 1$ and $N_2(2^j) = 4$ for $j \ge 3$. Thus,

$$\sum_{j\ge 1} \frac{N_2(2^j)}{2^{3j-1}} = \frac{1}{4} + \frac{2}{32} + \frac{1}{56} = \frac{37}{112},$$

and so

$$K_2 = \frac{75}{112} \prod_{p>2} \left(1 - \frac{2p}{p^3 - 1}\right) = 0.4269891575\ldots.$$

For K_3, we have $N_3(p^j) = 3$ for $p \equiv 1 \pmod 3$ and for $p = 3$ and $j \ge 2$. Otherwise, $N_3(p^j) = 1$. Thus,

$$K_3 = \frac{205}{234} \prod_{p \,\equiv\, 1 \ (\mathrm{mod}\ 3)} \left(1 - \frac{3p}{p^3 - 1}\right) \prod_{p \,\equiv\, 2 \ (\mathrm{mod}\ 3)} \left(1 - \frac{p}{p^3 - 1}\right) = 0.6393087751\ldots.$$

For K_4, we have $N_4(p^j) = 4$ for $p \equiv 1 \pmod 4$, $N_4(p^j) = 2$ for $p \equiv 3 \pmod 4$, $N_4(2) = 1$, $N_4(2^2) = 2$, $N_4(2^3) = 4$, and $N_4(2^j) = 8$ for $j \ge 4$. Thus,

$$K_4 = \frac{299}{448} \prod_{p \,\equiv\, 1 \ (\mathrm{mod}\ 4)} \left(1 - \frac{4p}{p^3 - 1}\right) \prod_{p \,\equiv\, 3 \ (\mathrm{mod}\ 4)} \left(1 - \frac{2p}{p^3 - 1}\right) = 0.3775394971\ldots.$$

These calculations were done with the aid of Mathematica. With a little effort other constants K_k may be computed, but if k has many divisors, then the calculation gets a bit more tedious.

We close with the observation that there is an infinite sequence of numbers k on which $K_k \to 0$. In particular, if $k = k_m$ is the least common multiple of all numbers up to m, then $N_k(p) = p - 1$ for every prime $p \le m + 1$, so that

$$K_k < \prod_p \left(1 - \frac{N_k(p)}{p^2}\right) < \prod_{p \le m+1} \left(1 - \frac{p-1}{p^2}\right).$$

Since $\sum (p-1)/p^2 = +\infty$, it follows that as $m \to \infty$, $K_{k_m} \to 0$. Using the theorem of Mertens, we in fact have $\liminf K_k \log\log k < +\infty$.

REFERENCES

[1] H. Davenport, Multiplicative number theory, third edition, Springer, New York, 2000.

[2] J. von zur Gathen, A. Knopmmacher, F. Luca, L. G. Lucht, and I. E. Shparlinski, Average order in cyclic groups, J. Théor. Nombres Bordeaux **16** (2004), 107–123.

[3] F. Luca, Some mean values related to average multiplicative orders of elements in finite fields, Ramanujan J. **9** (2005), 33–44.

Yilan Hu
Yilan.Hu.10@Alum.Dartmouth.org

Carl Pomerance
Department of Mathematics
Dartmouth College
Hanover, NH 03755, USA
carl.pomerance@dartmouth.edu