Sidon sets and statistics of the ElGamal function

Lucas Boppré Niehues, Joachim von zur Gathen,
Lucas Pandolfo Perin, Ana Zumalacárregui

September 26, 2018

Abstract

In the ElGamal signature and encryption schemes, an element x of the underlying group $G = \mathbb{Z}_p^\times = \{1, \ldots, p-1\}$ for a prime p is also considered as an exponent, for example in g^x, where g is a generator of G. This *ElGamal map* $x \mapsto g^x$ is poorly understood, and one may wonder whether it has some randomness properties. The underlying map from G to $\mathbb{Z}_{p-1}$ with $x \mapsto x$ is trivial from a computer science point of view, but does not seem to have any mathematical structure.

This work presents two pieces of evidence for randomness. Firstly, experiments with small primes suggest that the map behaves like a uniformly random permutation with respect to two properties that we consider. Secondly, the theory of Sidon sets shows that the graph of this map is equidistributed in a suitable sense.

It remains an open question to prove more randomness properties, for example, that the ElGamal map is pseudorandom.

1 Introduction

In the ElGamal signature scheme [5] with parameter n, we take an n-bit number d and a cyclic group $G = \langle g \rangle$ of order d. In ElGamal's original proposal, p is an n-bit prime number, $G = \mathbb{Z}_p^\times = \{1, \ldots, p-1\}$, $d = p - 1$, and $\mathbb{Z}_d = \{1, \ldots, d\}$ is the *exponent group*. More commonly, one takes $\mathbb{Z}_d = \{0, \ldots, d-1\}$, but both are valid set of representatives. We let g be a generator of G, so that $G = \{g^b \colon b \in \mathbb{Z}_d\}$. The object of this paper is to investigate randomness properties of the *ElGamal map* from G to G with $x \mapsto g^x$, where $x \in \mathbb{Z}_d$ on the right hand side. Since g^x determines x uniquely, this is a permutation of G. If we consider $x \in \mathbb{Z}_d$ on the left hand side, it is the discrete exponentiation map in base g.

A secret global key $a \in \mathbb{Z}_d$ and session key $k \in \mathbb{Z}_d^\times$ are chosen uniformly at random, and their public versions $A = g^a$ and $K = g^k$ in G are published. The signature of a message $m \in \mathbb{Z}_d$ is (K, b) with $b = k^{-1}(m - aK) \in \mathbb{Z}_d$.

The private key is easily broken if discrete logarithms in G can be calculated efficiently; see Figure 1. For more details, see von zur Gathen [7], Sections 8.2 and 9.8.
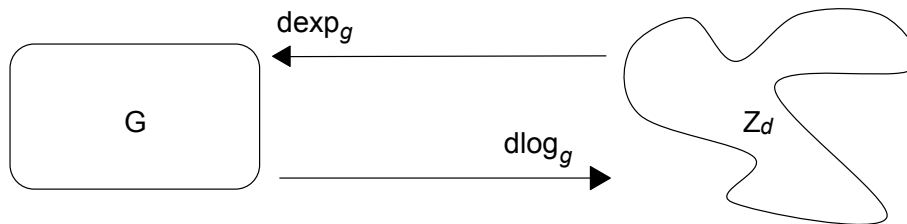
Figure 1: Public group and exponent group in ElGamal Signature scheme

The Decisional Diffie-Hellman (DDH) problem is to decide whether, given a triple $(x, y, z) \in G^3$, there exist $a, b, c \in \mathbb{Z}_d$ so that $x = g^a$, $y = g^b$, and $z = g^{ab}$; then (x, y, z) is a *Diffie-Hellman triple*.

If such triples are indistinguishable from uniformly random triples, for uniformly random a and b, then the ElGamal encryption scheme is indistinguishable by public key only attacks. The results of Canetti, Friedlander, Konyagin, Larsen, Lieman, and Shparlinski [1], indicate that the most significant and least significant bits of each element in DDH triples are indeed distributed uniformly. Do pairs (x, g^x), for uniformly random x, exhibit a similar behavior?

This paper first gives some experimental evidence in favor of this. We take some small primes, just above 1000, and consider two parameters of permutations: the number of cycles and the number of k-cycles for given k. Their averages for random permutations are well-known, and we find that the average values for the ElGamal function are reasonably close to those numbers. Secondly, we use the theory of Sidon sets to prove an equidistributional property with appropriate parameters; see also Cobeli, Vâjâitu & Zaharescu [4] for a different approach to show equidistribution.

Martins & Panario [11] study similar questions, but for general polynomials that need not be permutations, and for different parameters. Konyagin, Luca, Mans, Mathieson, Sha & Shparlinski [8] consider enumerative and algorithmic questions about (non-)isomorphic functional graphs, and Mans, Sha, Shparlinski & Sutantyo [10] provide statistics, conjectures, and results about cycle lengths of quadratic polynomials over finite prime fields. Kurlberg, Luca & Shparlinski [9] and Felix & Kurlberg [6] deal with fixed points of the map $x \mapsto x^x$ modulo primes.

2 Experiments in $\mathbb{F}_p$

The pictorial representation in Figure 2 shows the cycle structure of the permutation $x \mapsto g^x$ in $\mathbb{F}_p$ with $p = 1009$ and $g = 11$, the smallest generator. Each circle corresponds to a cycle, whose length is proportional to the circle's circumference. Next, Figure 3 shows together 12 permutations in $\mathbb{F}_p$ using the 12 smallest generators of $\mathbb{F}_p$.

In the following subsections, we take the cycle structures for all $\phi(1008) = 288$ generators of $\mathbb{F}_{1009}$, and then of all generators for the first fifty primes larger
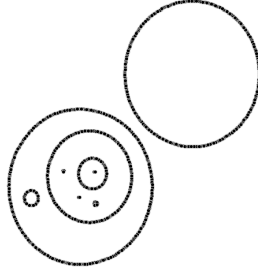
Figure 2: Representation of the cycles generated with $g = 11$ in $\mathbb{F}_{1009}$

than 1000. We calculate the averages for the number of cycles and the number of k-cycles and compare them to the known values for random permutations.
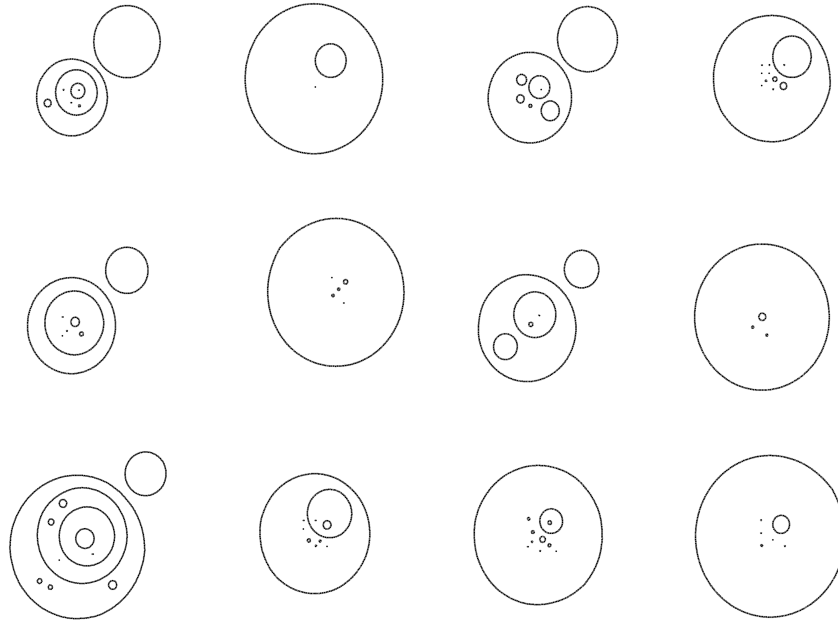

Figure 3: Graphical presentation of permutations of g^x in $\mathbb{F}_{1009}$

2.1 Number of cycles in permutations

We study in detail the number of cycles in the permutations. The number of permutations in S_n with c cycles equals the Stirling number $s(n,c)$ of the first kind, and thus is the coefficient of x^c in the falling factorial $x^{\underline{n}} = x \cdot (x - 1) \cdots (x - n + 1)$. Figure 4 shows the distribution of the number of cycles for uniformly random permutations of n elements, that is, the fraction $s(n,c)/n!$ (in percent) for $n = 1009$ and $1 \le c \le 20$, as a continuous line. In the same figure, the experimental statistics for 288 permutations chosen uniformly at random are presented as dots. This was done in order to calibrate our expectations. Theory and experiments match quite well.

Figure 5 shows the same continuous line, but now the dots represent the counts for the 288 generators of $\mathbb{F}_{1009}$. The result looks quite similar to Figure 4.
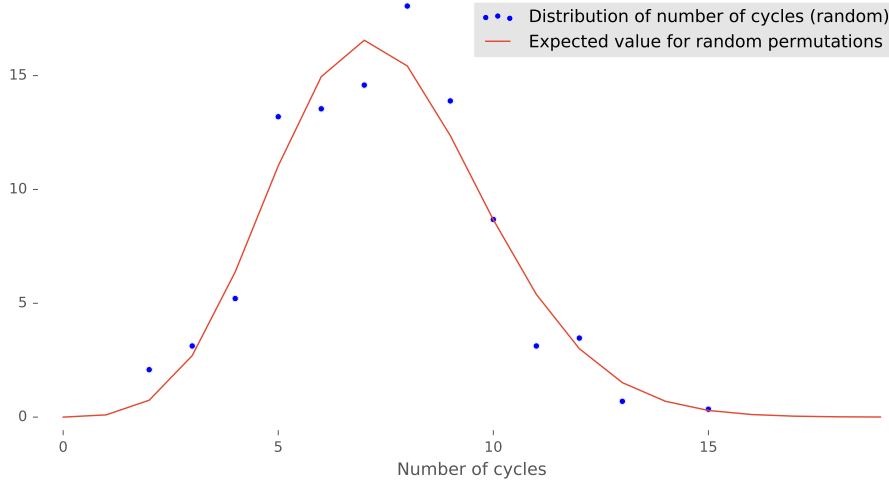

Figure 4: Distribution in percent of number of cycles for 288 random permutations in S_{1009}.

2.2 Number of k-cycles in permutations

Given a random permutation of elements, the number of cycles of length k is on average $1/k$ [12]. In Figure 6, we give the average number of cycles of length k for all 288 generators of the multiplicative group in dots. The experimental results are reasonably close to the theoretical values.

For the specific case $k = 1$, the average number of fixed points in random permutations is 1. The results in Figure 6 are very close, by a small error margin. Therefore, to better illustrate this property, Figure 7 shows the average number of fixed points for all generators in the multiplicative group for all prime numbers from 2 to 2111. As expected, the average of fixed points is closely distributed
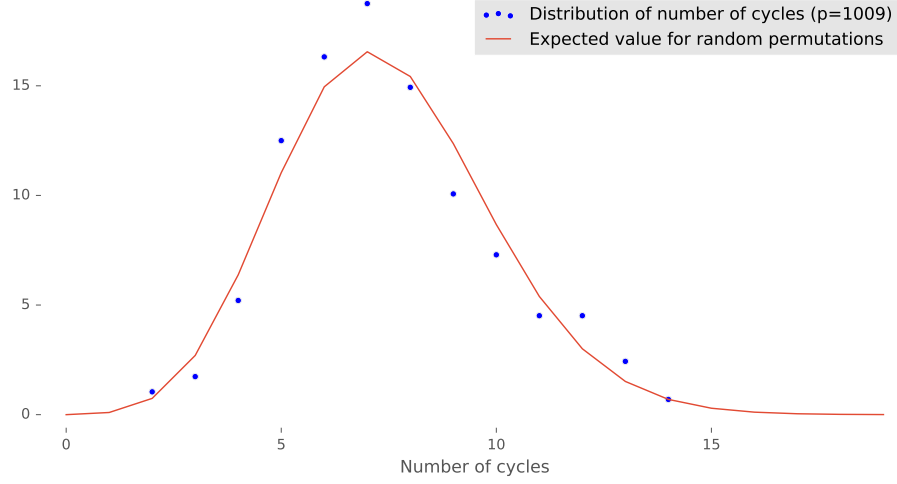
4

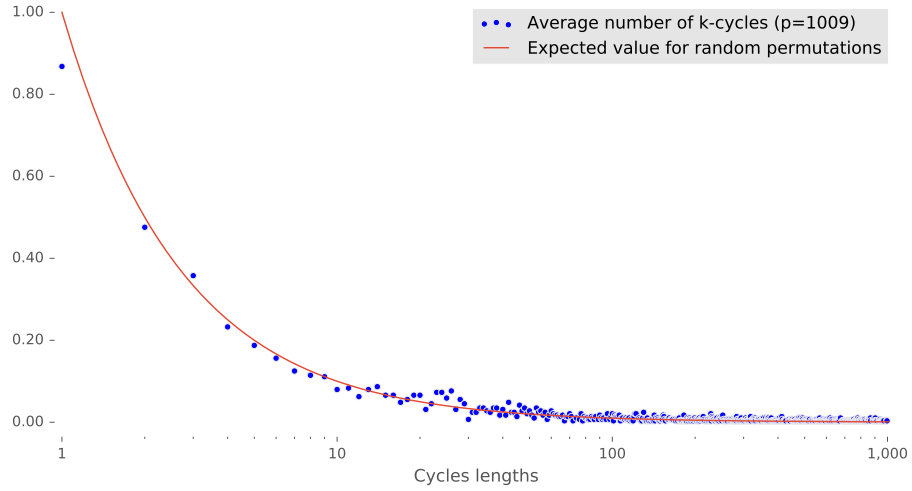Figure 5: Distribution of number of cycles in ElGamal functions on $\mathbb{F}_{1009}$

Figure 6: Average number of k-cycles in in ElGamal functions on $\mathbb{F}_{1009}$

to the theoretical value. We also note that by increasing p, the average of fixed points in the experiments gets closer to the expected theoretical value.

3 Sidon sets

A subset A of an abelian group G (written additively) is a *Sidon set* if for every $y \in G \setminus \{0\}$ there exists at most one pair $(a, b) \in A^2$ such that $y = a - b$. Clearly,
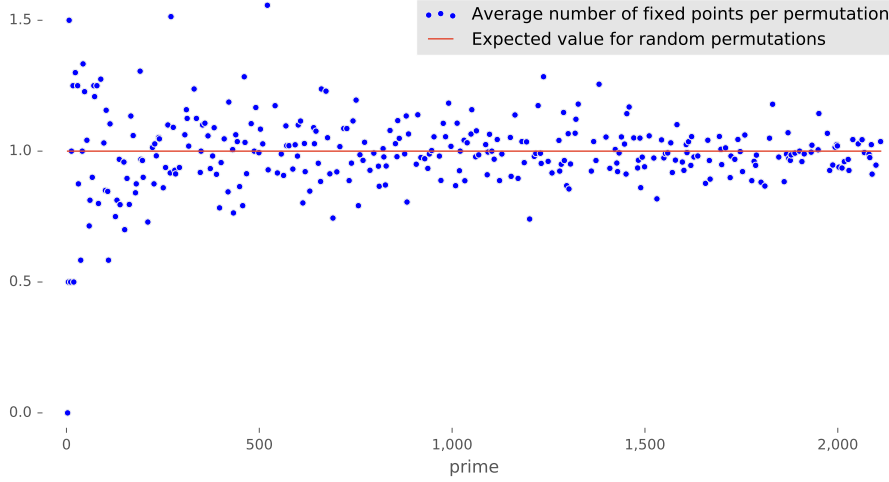
5

Figure 7: Average number of fixed points for all generators of $\mathbb{F}_p$ with $2 \leq p \leq 2111$

for any set A there are exactly $\#A$ pairs $(a, b) \in A^2$ for which $0 = a - b$, where $\#A$ is the cardinality of A.

Let p be a prime, $g \in \mathbb{Z}_p^\times$ a generator of the multiplicative group $\mathbb{Z}_p^\times$, and identify $\mathbb{Z}_{p-1} = \{0, 1, \ldots, p-2\}$ and $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$. We consider the additive group $G = \mathbb{Z}_{p-1} \times \mathbb{Z}_p$ (using the additive structure of both factors), and the subset

$$S = \{(g^x, x) : x \in \mathbb{Z}_{p-1}\}. \tag{3.1}$$

Thus S is the graph of the discrete logarithm function modulo p, since $S = \{(y, \log_g y) : y \in \mathbb{Z}_p \setminus \{0\}\}$, and, after swapping the coordinates, the graph of the ElGamal function.

The following result is well known; see Cilleruelo [2], Example 2. We include a proof for the sake of completeness.

Lemma 3.1. *The set S in (3.1) is a Sidon set in $\mathbb{Z}_p \times \mathbb{Z}_{p-1}$.*

Proof. For some $(u, v) \neq (0, 0)$ in $\mathbb{Z}_p \times \mathbb{Z}_{p-1}$ and $c_1, c_2 \in \mathbb{Z}_{p-1}$, suppose that $(g^{c_1}, c_1) - (g^{c_2}, c_2) = (u, v)$. Then

$$\begin{aligned} c_1 - c_2 &\equiv v \bmod p - 1, \\ g^{c_1} - g^{c_2} &\equiv u \bmod p. \end{aligned} \tag{3.2}$$

In particular $v \not\equiv 0 \bmod p - 1$, since otherwise $u \equiv g^{c_1} - g^{c_1} \equiv 0 \bmod p$ which contradicts the assumption.

From the first equation in (3.2) we know that

$$g^{c_1 - v} \equiv g^{c_2} \bmod p,$$

6

hence in the second equation we have

$$g^{c_1}\left(1 - g^{-v}\right) \equiv u \bmod p.$$

Since $(1 - g^{-v}) \not\equiv 0 \bmod p$ by the above, we conclude that

$$g^{c_1} \equiv \left(1 - g^{-v}\right)^{-1} u \bmod p,$$

and thus the pair (c_1, c_2) is uniquely determined by (u, v). $\qquad\square$

Lemma 3.2. *Let φ be a nontrivial character of $G = \mathbb{Z}_p \times \mathbb{Z}_{p-1}$ and S be the set in (3.1). Then*

$$\left|\sum_{a \in S} \varphi(a)\right| < (3(p-1))^{1/2}.$$

Proof. Any nontrivial character φ of G satisfies $\sum_{x \in G} \varphi(x) = 0$. Thus, for the set $S - S = \{x \in G : x = a - b \text{ for some } a, b \in S\}$ we have

$$\sum_{x \in S-S} \varphi(x) = -\sum_{x \notin S-S} \varphi(x). \tag{3.3}$$

Since $|z| = (z \cdot \bar{z})^{1/2}$ for a complex number z and $\overline{\varphi(x)} = \varphi(-x)$ for every $x \in G$, where $\bar{z}$ denotes the complex conjugate of z, it follows that

$$\left|\sum_{a \in S} \varphi(a)\right|^2 = \left(\sum_{a \in S} \varphi(a)\right)\left(\sum_{b \in S} \varphi(-b)\right) = \sum_{a,b \in S} \varphi(a - b)$$

$$= \sum_{y \in G} \varphi(y) \cdot \#\{(a, b) \in S^2 : y = a - b\}. \tag{3.4}$$

Since S is a Sidon set by Lemma 3.1, we know that

$$\#\{(a, b) \in S^2 : y = a - b\} = \begin{cases} \#S & \text{if } y = 0, \\ 1 & \text{if } y \in S - S \setminus \{0\}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$\left|\sum_{a \in S} \varphi(a)\right|^2 = \#S - 1 + \sum_{y \in S-S} \varphi(y)$$

$$= \#S - 1 - \sum_{y \notin S-S} \varphi(y)$$

$$\leq \#S - 1 + \left|\sum_{y \notin S-S} \varphi(y)\right|. \tag{3.5}$$

Luckily, we have a complete description of the set $S - S$, since every pair $(a, b) \in S^2$ is uniquely determined by the difference $a - b$ unless $a - b = 0$, for which we have exactly $\#S = p - 1$ options; hence

$$\#(S - S) = (\#S)^2 - \#S + 1 = (p-1)^2 - (p-1) + 1 = \#G - 2\#S + 1 \tag{3.6}$$

7

since $\#G = p(p-1)$. Clearly we have from (3.6) that

$$\left| \sum_{y \notin S-S} \varphi(y) \right| \leq \#G - \#(S-S) = 2\#S - 1. \qquad (3.7)$$

Combining equations (3.4), (3.5) and (3.7) we have

$$\left| \sum_{a \in S} \varphi(a) \right|^2 \leq 3\#S - 2,$$

which concludes the proof. $\qquad \square$

The following classical result is only included here for the sake of completeness.

Lemma 3.3. *Let n and N be positive integers with $1 \leq N < n$. Then, for any integer h*

$$\sum_{0 \leq a < n} \left| \sum_{h \leq x < N+h} \exp(2\pi i a x/n) \right| < 5n \log n.$$

Proof. Without loss of generality we will assume that $h = 0$, since

$$\left| \sum_{h \leq x < N+h} \exp(2\pi i a x/n) \right| = \left| \sum_{0 \leq x < N} \exp(2\pi i a (x+h)/n) \right|$$

$$= \left| \exp(2\pi i a h/n) \sum_{0 \leq x < N} \exp(2\pi i a x/n) \right|$$

$$= \left| \sum_{0 \leq x < N} \exp(2\pi i a x/n) \right|.$$

The contribution of $a = 0$ to the sum is precisely $N < n$.
Observe that for a given $1 \leq a \leq n-1$ the sum

$$\sum_{0 \leq x \leq N} \exp(2\pi i a x/n) = 1 + \exp(2\pi i a x/n) + \cdots + \exp(2\pi i a x/n)^{N-1}$$

is in fact a geometric progression with ratio $q = \exp(2\pi i a/n) \neq 1$ thus

$$\sum_{0 \leq x \leq N} \exp(2\pi i a x/n) = \left| \frac{q^N - 1}{q - 1} \right| \leq \frac{2}{|q-1|}.$$

We have

$$|q - 1| = |\exp(2\pi i a/n) - 1| = |\exp(\pi i a/n) - \exp(-\pi i a/n)| = 2|\sin(\pi a/n)|.$$

Then

$$|\sin(\pi a/n)| = |\sin(\pi(a-n)/n)| \geq \frac{2\min\{a, n-a\}}{n}$$

8

because $\sin(\alpha) \geq 2\alpha/\pi$ for $0 \leq \alpha \leq \pi/2$. Therefore

$$\sum_{0 \leq a < n} \Big| \sum_{0 \leq x < N} \exp(2\pi i a x/n) \Big| \leq N + \sum_{0 < a < n} \frac{n}{\min\{a, n-a\}}$$

$$\leq N + 2n \sum_{1 \leq a \leq n/2} \frac{1}{a}. \qquad (3.8)$$

The proof follows from (3.8) and the inequality

$$\sum_{1 \leq a \leq n/2} \frac{1}{a} < 1 + \log(n),$$

which holds for any integer $n \geq 2$. □

Theorem 3.1. *Let* $S = \{(g^x, x) : x \in \mathbb{Z}_{p-1}\}$. *For any box* $B = [h+1 .. h + N] \times [k+1 .. k+M] \subseteq \mathbb{Z}_p \times \mathbb{Z}_{p-1}$ *we have*

$$\left| \#(S \cap B) - \frac{\#B}{p} \right| \leq 50 p^{1/2} \log^2 p.$$

Furthermore, if $\#B \in \omega(p^{3/2} \log^2 p)$, *then* $\#(S \cap B) \sim \#B/p$.

Here,
$$\omega(f) = \{g \colon \mathbb{R} \to \mathbb{R}^+ : g(x)/|f(x)| \to 0 \text{ if } x \to \infty\}$$

for some $f \colon \mathbb{R} \to \mathbb{R}^+$.

Proof. By the orthogonality of characters and separating the contribution of the trivial character $\varphi_0 = 1$, we have

$$\#(S \cap B) = \frac{1}{p(p-1)} \sum_{\varphi} \sum_{a \in S} \sum_{b \in B} \varphi(a - b)$$

$$= \frac{\#B}{p} + \frac{1}{p(p-1)} \sum_{\varphi \neq \varphi_0} \sum_{a \in S} \sum_{b \in B} \varphi(a - b).$$

Thus

$$\left| \#(S \cap B) - \frac{\#B}{p} \right| = \frac{1}{p(p-1)} \Big| \sum_{\varphi \neq \varphi_0} \sum_{a \in S} \sum_{b \in B} \varphi(a - b) \Big|$$

$$\leq \frac{1}{p(p-1)} \sum_{\varphi \neq \varphi_0} \Big| \sum_{a \in S} \varphi(a) \Big| \Big| \sum_{b \in B} \varphi(b) \Big|$$

$$\leq \frac{1}{p(p-1)} \Big(\max_{\varphi \neq \varphi_0} \Big| \sum_{a \in S} \varphi(a) \Big| \Big) \sum_{\varphi \neq \varphi_0} \Big| \sum_{b \in B} \varphi(b) \Big|. \qquad (3.9)$$

The characters of G act as follows:

$$\varphi((x, y)) = \exp\left(2\pi i \left(\frac{sx}{p} + \frac{ty}{p-1} \right) \right), \qquad \text{for some } (s, t) \in G.$$

9

Hence we have

$$\sum_{\varphi \neq \varphi_0} \left| \sum_{b \in B} \varphi(b) \right| \leq \Big(\sum_{0 \leq s < p} \big| \sum_{h < x \leq h+N} \exp(2\pi i s x / p) \big| \Big)$$
$$\times \Big(\sum_{0 \leq t < p-1} \big| \sum_{k < y \leq k+M} \exp(2\pi i t y / (p-1)) \big| \Big),$$

which implies, by Lemma 3.3, that

$$\sum_{\varphi \neq \varphi_0} \left| \sum_{b \in B} \varphi(b) \right| < 25 p (p-1) \log^2 p. \tag{3.10}$$

By Lemma 3.2,

$$\max_{\varphi \neq \varphi_0} \left| \sum_{a \in S} \varphi(a) \right| < \sqrt{3(p-1)},$$

which combined with (3.10) in (3.9) concludes the proof. $\qquad\qquad\square$

One can show, with a bit more of work, see Cilleruelo & Zumalacárregui [3], that in fact

$$\left| \#(S \cap B) - \frac{\#B}{p} \right| \in O\big(p^{1/2} \log_+^2(|B| p^{-3/2}) \big),$$

which extends slightly the asymptotic range for $\#B$, where our "big-Oh" notation, for a real function $f \colon \mathbb{R} \to \mathbb{R}^+$, $O(f)$ denotes the following set of functions:

$$O(f) = \{ g \colon \mathbb{R} \to \mathbb{R} \mid \exists\, C > 0 \ \text{ with } |g(x)| \leq C f(x) \text{ for sufficiently large } x \},$$

and $\log_+(x) = \max\{\ln(x), 1\}$ for $x \in \mathbb{R}^+$. The implied asymptotics are for growing p. In fact, in [3] this result was obtained for a much larger family of dense Sidon sets.

Igor Shparlinski has pointed out to us that one can obtain similar asymptotic results with the exponential sum machinery. However, that method is unlikely to yield explicit estimates, without "O"-term.

4 Conclusion

We have shown, both experimentally and theoretically, some randomness properties of the ElGamal function over $\mathbb{Z}_p$ for a prime p. Many questions along these lines remain open:

- stronger results, perhaps even pseudorandomness,

- other groups for G, for example, elliptic curves,

- similar questions about the Schnorr function, where G is a "small" subgroup of a "large" group $\mathbb{Z}_p$.

5 Acknowledgements

References

[1] Ran Canetti, John Friedlander, Sergei Konyagin, Michael Larsen, Daniel Lieman, and Igor Shparlinski. On the statistical properties of Diffie-Hellman distributions. *Israel Journal of Mathematics*, 120:23–46, 2000.

[2] J. Cilleruelo. Combinatorial problems in finite fields and Sidon sets. *Combinatorica*, 32(5):497–511, 2012.

[3] Javier Cilleruelo and Ana Zumalacárregui. Saving the logarithmic factor in the error term estimates of some congruence problems. *Math. Z.*, 286(1-2):545–558, 2017.

[4] Cristian Cobeli, Marian Vâjâitu, and Alexandru Zaharescu. On the set $ax + bg^x \pmod{p}$. *Port. Math. (N.S.)*, 59(2):195–203, 2002.

[5] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, IT-31(4):469–472, 1985.

[6] Adam Tyler Felix and Pär Kurlberg. On the fixed points of the map $x \mapsto x^x$ modulo a prime, II. 2016. `arXiv:1607.04948`.

[7] Joachim von zur Gathen. *CryptoSchool*. Springer, 2015.

[8] Sergei V. Konyagin, Florian Luca, Bernard Mans, Luke Mathieson, Min Sha, and Igor E. Shparlinski. Functional graphs of polynomials over finite fields. *J. Combin. Theory Ser. B*, 116:87–122, 2016.

[9] Pär Kurlberg, Florian Luca, and Igor E. Shparlinski. On the fixed points of the map $x \mapsto x^x$ modulo a prime. *Mathematical Research Letters*, 22(01):141–168, 2015.

[10] Bernard Mans, Min Sha, Igor E. Shparlinski, and Daniel Sutantyo. On functional graphs of quadratic polynomials, 2017. `arXiv:1706.04734`.

[11] Rodrigo S. V. Martins and Daniel Panario. On the heuristic of approximating polynomials over finite fields by random mappings. *International Journal of Number Theory*, 12:1987–2016, 2016. Erratum pages 2041-2042.

[12] H. Wilf. *Generatingfunctionology*. Academic Press, New York, 1990.

Author addresses:

Lucas Boppré Niehues and Lucas Pandolfo Perin, LabSEC, Universidade Federal de Santa Catarina, Brazil.
`lucasboppre@gmail.com` and `lucas.perin@posgrad.ufsc.br`

Joachim von zur Gathen, B-IT, Universität Bonn, Germany.
`gathen@bit.uni-bonn.de`

Ana Zumalacárregui, University of New South Wales, Sydney, Australia.
`a.zumalacarregui@unsw.edu.au`