Construction of large families of pseudorandom subsets using elliptic curves

Zhixiong Chen and Chenhuang Wu
Key Laboratory of Applied Mathematics, Putian University,
Putian, Fujian 351100, P.R.China
ptczx@126.com, wuchenhuang2008@126.com

February 16, 2009

Abstract

Recently, Dartyge and Sárközy investigated the measures, i.e., the well distribution measure and the correlation measure of order k, of pseudorandomness of subsets of the set $\{1, 2, \ldots, N\}$, and they presented several constructive examples for subsets with strong pseudorandom properties when N is a prime number. In this article, we present a construction of pseudorandom subsets using elliptic curves over finite fields and estimate the pseudorandom measures. Character sums play an important role in the proofs.

Keywords Pseudo-random - Subsets - Elliptic curves - Character sums
Mathematics Subject Classification (2000) 05A05 - 11Z05

1 Introduction

In many applications (cryptography, simulation, etc.) we need a random subset $\mathcal{R}$ of the positive integers not exceeding a certain fixed integer N. Recently, Dartyge, Mosaki and Sárközy introduced and studied the pseudo-random measures of subsets of the set of the integers not exceeding N [4, 5, 6]. These measures are closely related to the measures of pseudorandomness of binary sequences introduced by Mauduit and Sárközy [13] and of the p-pseudorandom binary sequences defined by Hubert and Sárközy [10].

For a subset $\mathcal{R}$ of $\{1, 2, \ldots, N\}$, define the associated sequence E_N by

$$E_N = E_N(\mathcal{R}) = \{e_1, \ldots, e_N\} \in \left\{1 - \frac{|\mathcal{R}|}{N}, -\frac{|\mathcal{R}|}{N}\right\}^N$$

with

$$e_m = \begin{cases} 1 - \frac{|\mathcal{R}|}{N} & \text{for } m \in \mathcal{R} \\ -\frac{|\mathcal{R}|}{N} & \text{otherwise} \end{cases} \quad (m = 1, \ldots, N). \tag{1}$$

Then the *well-distribution measure* of the subset $\mathcal{R}$ of $\{1, 2, \ldots, N\}$ is defined by

$$W(\mathcal{R}, E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb} \right|$$

where the maximum is taken over all a, b, t such that $a, b, t \in \mathbb{N}$ and $1 \leq a \leq a + (t-1)b \leq N$, while the *correlation measure of order* k of $\mathcal{R}$ is defined as

$$C_k(\mathcal{R}, E_N) = \max_{M,D} \left| \sum_{m=1}^{M} e_{m+d_1} e_{m+d_2} \cdots e_{m+d_k} \right|$$

where the maximum is taken over all $D = (d_1, \ldots, d_k)$ with non-negative integers $0 \leq d_1 < \cdots < d_k$ and M such that $M + d_k \leq N$. One would expect that these measures are "small". Thus we may consider a subset $\mathcal{R}$ of $\{1, 2, \ldots, N\}$ as a "good" pseudo-random subset if $W(\mathcal{R}, E_N)$ and $C_k(\mathcal{R}, E_N)$ (at least for small k) are small; they must be $o(N)$ and ideally, they are $O(N^{1/2+\varepsilon})$ [7].

Dartyge, Mosaki and Sárközy present some good constructions of pseudo-random subsets when N is a prime number in [4, 5, 6, 7]. However in applications one usually needs large families of pseudo-random subsets. It is an interesting to design "good" pseudo-random subsets for different N (for example, non-prime numbers) and using different algebraic systems. It is a natural way to choose elliptic curves over finite fields, partially for the elliptic curve cryptography for extensive use. We will apply elliptic curves to construct some families of pseudo-random subsets and analyze their pseudorandomness in the present paper.

We first introduce some notions and basic facts of elliptic curves over finite fields. Let $p > 3$ be a (large) prime, $\mathbb{F}_p$ the finite field of p elements which we identify with the set $\{0, 1, \cdots, p-1\}$, $\mathbb{F}_p^*$ the set of non-zero elements of $\mathbb{F}_p$. Let $\mathcal{E}$ be an elliptic curve over $\mathbb{F}_p$, given by an affine Weierstrass equation of the standard form

$$y^2 = x^3 + Ax + B$$

with coefficients $A, B \in \mathbb{F}_p$ and nonzero discriminant, see [8]. It is known that the set $\mathcal{E}(\mathbb{F}_p)$ of $\mathbb{F}_p$-rational points of $\mathcal{E}$ forms an Abelian group under an appropriate composition rule denoted by $\oplus$ and with the point at infinity O as the neutral element. We recall that

$$|\#\mathcal{E}(\mathbb{F}_p) - p - 1| \leq 2p^{1/2},$$

where $\#\mathcal{E}(\mathbb{F}_p)$ is the number of $\mathbb{F}_p$-rational points, including the point at infinity O. The translation map by $W \in \mathcal{E}(\mathbb{F}_p)$ on $\mathcal{E}(\mathbb{F}_p)$ is defined as

$$\tau_W : P \mapsto P \oplus W.$$

It is obvious that $(f \circ \tau_W)(P) = f(\tau_W(P)) = f(P \oplus W)$.

In this article, for convenience, we always suppose that $\mathcal{E}(\mathbb{F}_p)$ is a cyclic group of order N and $G \in \mathcal{E}(\mathbb{F}_p)$ is a generator, i.e., $\mathcal{E}(\mathbb{F}_p) = \langle G \rangle$. In particular, $N \sim p$ in this case. A multiple of G is taken by $nG = \oplus_{i=1}^{n} G$. We write $nG = (x_n, y_n) \in \mathbb{F}_p \times \mathbb{F}_p$ on $\mathcal{E}$ for all $1 \leq n \leq N-1$ and set $X(nG) = x_n$ and $Y(nG) = y_n$.

We would like to study the pseudorandom properties of the subset $\mathcal{R}$ of $\{1, 2, \ldots, N\}$ defined by

$$\mathcal{R} := \{n \mid 1 \leq n \leq N, \ X(nG) \equiv h \pmod{p} \text{ for any } h \in H\} \tag{2}$$

where $r \in \mathbb{Z}$, $s \in \mathbb{N}$, $s < p/2$ and $H = \{r, r+1, \ldots, r+s-1\}$.

We remark that $\mathcal{R}$ can be defined in several different ways using elliptic curves, we refer to a preprint version of [9], which is available at

http://iml.univ-mrs.fr/editions/preprint2002/preprint2002.html,

and [1, 2, 3] for related issues.

2 The cardinality of $\mathcal{R}$

Exponential sums play an important role in the proofs to estimate the cardinality of $\mathcal{R}$ and its pseudo-random measures.

For any positive integer m, we identify $\mathbb{Z}_m$ with the residue ring modulo m. Put

$$e_m(z) = \exp(2\pi i z/m).$$

The exponential sums enter into our problem by means of the following well known basic identity.

Lemma 1 ([12]) *For any element $c \in \mathbb{Z}_m$, we have*

$$\sum_{z \in \mathbb{Z}_m} e_m(cz) = \begin{cases} m, & \text{if } c = 0 \\ 0, & \text{otherwise.} \end{cases}$$

We also need the following statement.

Lemma 2 ([12]) *The bound*

$$\sum_{c=0}^{m-1} \left| \sum_{z=u+1}^{u+v} e_m(cz) \right| \le m(1 + \log m)$$

holds for any integers u and $1 \le v \le m$.

Let $\psi(z) = \exp(2\pi iz/p)$ be a classical additive character of $\mathbb{F}_p$. We also need the following upper bound which is a special case of [11, Corollary 1].

Lemma 3 *Let f be a nonconstant rational function and $G \in \mathcal{E}(\mathbb{F}_p)$ be a rational point of order N. Then the bound*

$$\left| \sum_{\substack{z=0 \\ f(zG) \neq \infty}}^{N-1} \psi(\lambda f(zG)) e_N(\eta z) \right| \le 2\deg(f) p^{1/2}$$

holds for all $\lambda \in \mathbb{F}_p^*$ and $\eta \in \mathbb{Z}_N$. Hence the bound on incomplete sums

$$\left| \sum_{\substack{z=u \\ f(zG) \neq \infty}}^{v} \psi(\lambda f(zG)) \right| \le 2\deg(f) p^{1/2}(1 + \log N)$$

holds for all $\lambda \in \mathbb{F}_p^*$ and integers $0 \le u < v \le N - 1$.

We now present a bound on the cardinality of $\mathcal{R}$.

Theorem 1 *Let $\mathcal{R}$ be defined as in (2). Then the cardinality of $\mathcal{R}$ satisfies*

$$\left| |\mathcal{R}| - \frac{sN}{p} \right| \le 4p^{1/2}(1 + \log p).$$

Proof. From the definition of $\mathcal{R}$ in (2) and Lemma 1, we have

$$\sum_{h=r}^{r+s-1}\sum_{\lambda\in\mathbb{F}_p}\psi(\lambda(X(nG)-h)) = \begin{cases} p, & \text{if } n\in\mathcal{R} \\ 0, & \text{otherwise.}\end{cases}$$

Hence by Lemmas 2 and 3 we obtain

$$\begin{aligned}
|\mathcal{R}| &= \sum_{n=1}^{N}\frac{1}{p}\sum_{h=r}^{r+s-1}\sum_{\lambda\in\mathbb{F}_p}\psi(\lambda(X(nG)-h)) \\
&= \frac{s(N-1)}{p}+\frac{1}{p}\sum_{\lambda\in\mathbb{F}_p^*}\sum_{h=r}^{r+s-1}\psi(-\lambda h)\sum_{n=1}^{N}\psi(\lambda X(nG)) \\
&\le \frac{s(N-1)}{p}+\frac{1}{p}\sum_{\lambda\in\mathbb{F}_p^*}\left|\sum_{h=r}^{r+s-1}\psi(-\lambda h)\right|\cdot\left|\sum_{n=1}^{N}\psi(\lambda X(nG))\right| \\
&\le \frac{sN}{p}+4p^{1/2}(1+\log p).
\end{aligned}$$

We complete the proof of Theorem 1.

3 Pseudo-random measures of $\mathcal{R}$

Now we present upper bounds on the well-distribution measure and the correlation measure of order k of $\mathcal{R}$ defined in (2). The associated sequence E_N defined by (1) is

$$e_m = \begin{cases} 1-\alpha & \text{for } m\in\mathcal{R} \\ -\alpha & \text{otherwise,}\end{cases}$$

where

$$\alpha = \frac{|\,\mathcal{R}\,|}{N} = \frac{s}{p}+8\theta p^{-1/2}(1+\log p)$$

with some θ satisfying $|\theta|<1$, since $N\sim p$. Let $\beta=\frac{s}{p}-\alpha$.

Throughout this paper, the implied constant in the symbol " $\ll$ " is absolute.

Theorem 2 *Let $\mathcal{R}$ be a subset of $\{1,\dots,N\}$ defined as in (2), we have*

$$W(\mathcal{R},E_N)\ll p^{1/2}(1+\log p)(1+\log N).$$

Theorem 3 *Let $\mathcal{R}$ be a subset of $\{1,\dots,N\}$ defined as in (2), for $k<p$ we have*

$$C_k(\mathcal{R},E_N)\ll kp^{1/2}(2+\log p)^k(1+\log N).$$

3.1 Proofs

For $1 \le n \le N-1$, it is easy to see that

$$\begin{aligned} e_n &= (1-\alpha)\frac{1}{p}\sum_{h=r}^{r+s-1}\sum_{\lambda\in\mathbb{F}_p}\psi(\lambda(X(nG)-h)) \\ &\quad -\alpha\left(1-\frac{1}{p}\sum_{h=r}^{r+s-1}\sum_{\lambda\in\mathbb{F}_p}\psi(\lambda(X(nG)-h))\right) \\ &= \frac{1}{p}\sum_{h=r}^{r+s-1}\sum_{\lambda\in\mathbb{F}_p}\psi(\lambda(X(nG)-h))-\alpha \\ &= \frac{s}{p}-\alpha+\frac{1}{p}\sum_{\lambda\in\mathbb{F}_p^*}\sum_{h=r}^{r+s-1}\psi(-\lambda h)\psi(\lambda X(nG)) \\ &= \beta+\frac{1}{p}\sum_{\lambda\in\mathbb{F}_p^*}\sum_{h=r}^{r+s-1}\psi(-\lambda h)\psi(\lambda X(nG)). \end{aligned} \tag{3}$$

While $e_N = -\alpha$, since $NG = O$.

Proof of Theorem 2. Assume that $a, b, t \in \mathbb{N}$ and $1 \le a \le a + b(t-1) \le N$. According to (3), we obtain

$$\begin{aligned} \left|\sum_{i=0}^{t-1} e_{a+ib}\right| &\le \left|\frac{1}{p}\sum_{i=0}^{t-1}\sum_{\lambda\in\mathbb{F}_p^*}\sum_{h=r}^{r+s-1}\psi(-\lambda h)\psi(\lambda X((a+ib)G))\right| + \left|\sum_{i=0}^{t-1}\beta\right| + 1 \\ &\le \frac{1}{p}\sum_{\lambda\in\mathbb{F}_p^*}\left|\sum_{h=r}^{r+s-1}\psi(-\lambda h)\right|\cdot\left|\sum_{i=0}^{t-1}\psi(\lambda X((a+ib)G))\right| + |t\beta| + 1 \\ &\le 4p^{1/2}(1+\log p)(1+\log N) + |t\beta| + 1 \end{aligned}$$

by Lemma 2 and Lemma 3 or [1, Lemma 5], which is derived from Lemma 3. While

$$|t\beta| = 8t\theta p^{-1/2}(1+\log p) \le 16p^{1/2}(1+\log p)$$

Since $t \le N \sim p$. So we have

$$\left|\sum_{i=0}^{t-1} e_{a+ib}\right| \ll p^{1/2}(1+\log p)(1+\log N).$$

We complete the proof of Theorem 2.

Proof of Theorem 3. Assume that integers $d_1, \ldots, d_k$ and $M \in \mathbb{N}$ with

$$0 \le d_1 < \cdots < d_k, M + d_k \le N.$$

Now using (3), we obtain

$$\begin{aligned}
& \left| \sum_{m=1}^{M} e_{m+d_1} e_{m+d_2} \cdots e_{m+d_k} \right| \\
\le & \left| \sum_{m=1}^{M} \prod_{i=1}^{k} \left(\beta + \frac{1}{p} \sum_{\lambda \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-\lambda h) \psi(\lambda X((m+d_i)G)) \right) \right| + 1 \\
= & \frac{1}{p^k} \left| \sum_{m=1}^{M} \prod_{i=1}^{k} \left(p\beta + \sum_{\lambda \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-\lambda h) \psi(\lambda X((m+d_i)G)) \right) \right| + 1 \\
= & \frac{1}{p^k} \left| \sum_{m=1}^{M} \sum_{u=0}^{k} \sum_{1 \le j_1 < \ldots < j_u \le k} (p\beta)^{k-u} \prod_{i=1}^{u} \left(\sum_{\lambda \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-\lambda h) \psi(\lambda X((m+d_{j_i})G)) \right) \right| + 1 \\
= & \frac{1}{p^k} \left| \sum_{u=0}^{k} (p\beta)^{k-u} \sum_{1 \le j_1 < \ldots < j_u \le k} \sum_{\lambda_1 \in \mathbb{F}_p^*} \cdots \sum_{\lambda_u \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-h(\lambda_1 + \ldots + \lambda_u)) \right. \\
& \left. \sum_{m=1}^{M} \psi(\lambda_1 X((m+d_{j_1})G) + \ldots + \lambda_u X((m+d_{j_u})G)) \right| + 1 \\
= & \frac{1}{p^k} \left| \sum_{u=0}^{k} (p\beta)^{k-u} \sum_{1 \le j_1 < \ldots < j_u \le k} \sum_{\lambda_1 \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-h\lambda_1) \cdots \sum_{\lambda_u \in \mathbb{F}_p^*} \sum_{h=r}^{r+s-1} \psi(-h\lambda_u) \right. \\
& \left. \sum_{m=1}^{M} \psi((\lambda_1 X \circ \tau_{d_{j_1} G} + \ldots + \lambda_u X \circ \tau_{d_{j_u} G})(mG)) \right| + 1 \\
\le & \frac{1}{p^k} \sum_{u=0}^{k} \left(\begin{array}{c} k \\ u \end{array} \right) (p\beta)^{k-u} p^u (1 + \log p)^u Z + 1 \\
& \text{(where } \left| \sum_{m=1}^{M} \psi((\lambda_1 X \circ \tau_{d_{j_1} G} + \ldots + \lambda_u X \circ \tau_{d_{j_u} G})(mG)) \right| \le Z)
\end{aligned}$$

$$= \frac{1}{p^k}(p\beta + p(1+\log p))^k Z + 1$$

$$= (\beta + 1 + \log p)^k Z + 1 \le (2 + \log p)^k Z + 1.$$

It suffices to estimate the value of Z, i.e., the upper bound of

$$\sum_{m=1}^{M} \psi((\lambda_1 X \circ \tau_{d_{j_1}G} + \ldots + \lambda_u X \circ \tau_{d_{j_u}G})(mG))$$

for any $\lambda_1, \ldots, \lambda_u \in \mathbb{F}_p^*$ and $1 \le u \le k$. By [1, Lemma 1],

$$\lambda_1 X \circ \tau_{d_{j_1}G} + \ldots + \lambda_u X \circ \tau_{d_{j_u}G}$$

is a nonconstant rational function of degree at most $2u$. So by Lemma 3 again, we obtain

$$\begin{aligned} & \left| \sum_{m=1}^{M} \psi((\lambda_1 X \circ \tau_{d_{j_1}G} + \ldots + \lambda_u X \circ \tau_{d_{j_u}G})(mG)) \right| \\ \le \; & 4up^{1/2}(1+\log N) \le 4kp^{1/2}(1+\log N). \end{aligned}$$

We complete the proof of Theorem 3 by setting

$$Z = 4kp^{1/2}(1+\log N).$$

Acknowledgements

The research was partially supported by the Natural Science Foundation of Fujian Province of China under grant 2007F3086 and 2008F5049 and the Funds of the Education Department of Fujian Province under grant JA07164.

References

[1] Z. Chen, G. Xiao. 'Good' pseudo-random binary sequences from elliptic curves. Cryptology ePrint Archive, Report 2007/275, 2007, http://eprint.iacr.org/.

[2] Z. Chen. Elliptic curve analogue of Legendre sequences. Monatsh. Math. 154 (1) (2008) 1–10.

[3] Z. Chen, S. Li, G. Xiao. Construction of pseudo-random binary sequences from elliptic curves by using discrete logarithm. In: G. Gong et al. (Eds.) SETA'06. LNCS, vol. 4086, Springer-Verlag, Berlin, 2006, pp. 285–294.

[4] C. Dartyge, A. Sárközy. On pseudo-random subsets of the set of the integers not exceeding N. Periodica Mathematica Hungarica 54 (2007) 183–200.

[5] C. Dartyge, A. Sárközy. Large families of pseudorandom subsets formed by power residues. Unif. Distrib. Theory 2(2) (2007) 73–88.

[6] C. Dartyge, E. Mosaki, A. Sárközy. On large families of subsets of the set of the integers not exceeding N. Ramanujan J. (2008) DOI 10.1007/s11139-008-9135-z.

[7] C. Dartyge, A. Sárközy. On pseudo-random subsets of $\mathbb{Z}_n$. Monatsh. Math. (2008) DOI 10.1007/s00605-008-0072-0.

[8] A. Enge. Elliptic Curves and Their Applications to Cryptography : An Introduction. Kluwer Academic Publishers, Dordrecht, 1999.

[9] L.Goubin, C. Mauduit, A. Sárközy. Construction of large families of pseudorandom binary sequences. J. Number Theory, 106(1) (2004) 56-69.

[10] P. Hubert, A. Sárközy. On p-pseudorandom binary sequences. Periodica Mathematica Hungarica 49 (2004) 73–91.

[11] D. Kohel, I. E. Shparlinski. Exponential sums and group generators for elliptic curves over finite fields. Proc. Algorithmic Number Theory Symposium, LNCS, vol.1838, Springer-Verlag, Berlin, 2000, pp.395-404.

[12] R. Lidl, H. Niederreiter. Finite Fields. Reading, MA: Addison-Wesley, 1983.

[13] C. Mauduit, A. Sárközy. On finite pseudorandom binary sequences I: measures of pseudorandomness, the Legendre symbol. Acta Arithmetica 82 (1997) 365–377.