

Résumé de la thèse

Propriétés aléatoires des suites d'entiers

1 Introduction

La génération des nombres pseudoaléatoires joue un rôle important dans beaucoup de domaines des mathématiques et de la physique, en particulier dans les problèmes de cryptographie ou d'analyse numérique.

En cryptographie, l'un des algorithmes de chiffrement les plus sûrs est le « one-time pad » : nous convertissons d'abord le message en une suite d'entiers binaires (des « bits ») $A_N = \{a_1, \dots, a_N\} \in \{0, 1\}^N$ (par exemple, à chaque lettre de l'alphabet, nous assignons un nombre représenté en forme dyadique avec le même nombre de chiffres), donc A_N s'appelle le *texte clair*. Après, nous construisons (par exemple par l'observation d'un phénomène physique) une suite binaire aléatoire $E_N = \{e_1, \dots, e_N\} \in \{0, 1\}^N$ (appelée la *clé*) qui a la même longueur que le texte clair. Nous chiffons le texte clair A_N bit par bit, en ajoutant modulo 2 ses éléments aux éléments de E_N .

Un inconvénient évident du *one-time pad* est que la clé doit être aussi longue que le texte clair, ce qui augmente la difficulté de la construction, de

la gestion et de la distribution de la clé. Ces difficultés motivent la conception d'un système de chiffrement pour lequel la clé est construite de manière pseudoaléatoire à partir d'une petite clé secrète, avec l'intention que la clé semble aléatoire à un adversaire disposant de moyens de calculs limités. De tels systèmes de chiffrement n'offrent pas une sécurité sans condition, mais on peut conserver l'espoir est qu'ils sont suffisamment sûrs.

Jusqu'à récemment, il n'y avait pas de tentative réussie de construction d'une suite *finie* pseudoaléatoire. Par exemple, les définitions du terme « pseudoaléatoire » qui sont basées sur la non-existence d'un algorithme en temps polynômial pourraient être contestées immédiatement dans le cas des suites finies, puisque dans les algorithmes en temps polynômial, il n'y a aucune restriction pour le degré ou les coefficients du polynôme. Comment dépendent-ils de la longueur de la suite ?

Dans le cas fini, la *complexité linéaire* (voir par exemple [13]) est une définition plus intéressante du pseudoaléatoire, qui repose sur la longueur du « linear feedback shift register » le plus court (récurrence linéaire sur \mathbb{F}_2) qui construit une suite ayant E_N pour les N premiers éléments. Mais encore le fait que la complexité linéaire soit grande n'est pas suffisant pour considérer que la suite E_N est aléatoire. La complexité linéaire constitue une bonne mesure pseudoaléatoire, mais elle est loin d'être satisfaisante.

Nouvelles mesures du caractère pseudoaléatoire

En 1997, C. Mauduit et A. Sárközy [11] ont introduit de nouvelles mesures du caractère pseudoaléatoire des suites binaires. Considérons une suite finie

$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N$ (pour des raisons techniques, nous préférons, pour des raisons de symétrie, travailler avec ± 1 plutôt que 0 ou 1).

Définition 1 *La mesure de bonne-distribution de E_N est définie par*

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

où le maximum est pris sur tous les a, b, t tels que $a, b, t \in \mathbb{N}$ et $1 \leq a \leq a + (t-1)b \leq N$.

Définition 2 *La mesure de corrélation d'ordre k de E_N est définie par*

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=1}^M e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} \right|$$

où le maximum est pris sur tous les $D = (d_1, d_2, \dots, d_k)$ et M tels que $1 \leq d_1 < d_2 < \dots < d_k < M + d_k \leq N$.

Cassaigne, Mauduit et Sárközy ont montré que, pour la majorité des suites $E_N \subseteq \{-1, +1\}^N$, les mesures $W(E_N)$ et $C_k(E_N)$ sont $\ll \sqrt{N}(\log N)^c$. Ainsi une suite E_N peut être considérée comme une bonne suite pseudoaléatoire si $W(E_N)$ et $C_k(E_N)$ (au moins pour petit k) sont petites en termes de N .

Il y a plusieurs manières de définir des mesures pseudoaléatoires comme W et C_k , mais il n'y a aucune mesure universelle parfaite du caractère pseudoaléatoire. On peut imposer d'autres critères (et dans certaines applications, on est forcé de le faire), et également, on peut introduire de nouvelles mesures pseudoaléatoires. Cependant, il serait de plus en plus difficile de manipuler ces mesures. Ainsi nous devons fixer une limite et nous concentrer sur certains critères pseudoaléatoires qui sont les plus importants dans les

applications. Nous choisirons donc à l'intérieur de cette limite la mesure de bonne-distribution et de corrélation, et dans quelques chapitres de cette thèse nous étudierons également une nouvelle mesure pseudoaléatoire : la mesure de symétrie.

Les cinq chapitres suivants de la thèse sont les articles [8], [6], [7], [9] et [5]. Excepté [7] tous les articles sont reproduits dans leur totalité, mais dans le quatrième chapitre j'ometts une partie de [7], puisqu'une version améliorée de certains résultats peut être trouvée dans [5], qui est le dernier chapitre de la thèse.

2 Sur une propriété pseudoaléatoire des suites binaires

Dans cette section nous définissons une nouvelle mesure basée sur l'observation suivante de Mauduit et Sárközy : si une suite finie contient une suite symétrique relativement grande, alors cette structure symétrique peut causer des difficultés dans certaines applications.

Définition 3 *La mesure de symétrie de E_N est définie par*

$$S(E_N) = \max_{a < b} \left| \sum_{j=0}^{\lfloor (b-a)/2 \rfloor - 1} e_{a+j} e_{b-j} \right| = \max_{1 \leq a < b \leq N} |H(E_N, a, b)|.$$

D'abord nous montrons que la mesure de symétrie de E_N est autour de \sqrt{N} pour presque tous $E_N \in \{-1, +1\}^N$.

Théorème 1 *Il y a un nombre entier N_0 , tel que pour tous $N > N_0$ on a*

$$S(E_N) > \frac{7}{20} \sqrt{N}.$$

Tandis que pour N assez grand, $S(E_N)$ est toujours plus grand que \sqrt{N} multiplié par une constante, une majoration comparable (à un facteur log près) n'a lieu seulement que pour la majorité des suites $E_N \in \{-1, +1\}^N$:

Théorème 2 *Pour tout $\varepsilon > 0$ il existe un nombre entier $N_0 = N_0(\varepsilon)$ tel que pour $N > N_0$ on a*

$$P(S(E_N) < 4.25 (N \log N)^{1/2}) > 1 - \varepsilon.$$

Puisque pour tout $1 \leq k \leq \frac{p-1}{2}$ nous avons $\left(\frac{p-k}{p}\right) = \left(\frac{-1}{p}\right) \left(\frac{k}{p}\right)$, la mesure de symétrie du symbole de Legendre E_{p-1} est $(p-1)/2$. Nous montrons que la mesure de symétrie de la première partie de la suite E_{p-1} est petite.

Théorème 3 *Si p est un nombre premier impair, et nous posons*

$$E_{(p-1)/2} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \dots, \left(\frac{(p-1)/2}{p}\right) \right)$$

on a

$$S(E_{(p-1)/2}) \leq 18p^{1/2} \log p.$$

Dans les deux théorèmes suivants, nous donnons une forme quantitative des relations entre la mesure de bonne-distribution et la mesure de symétrie.

Théorème 4 *Pour tout $N \in \mathbb{N}$ et $E_N \in \{-1, +1\}^N$ on a*

$$W(E_N) \leq 3(NS(E_N))^{1/2}.$$

Enfin, nous voyons que ce résultat est précis : il existe une suite dont la mesure de bonne-distribution est grande, mais les mesures de corrélation et de symétrie sont petites.

Théorème 5 *Si $k, N \in \mathbb{N}$, $N > N_0$ et*

$$N^{3/4} \leq k \leq N$$

il existe une suite $E_N \in \{-1, +1\}^N$ avec

$$W(E_N) \geq k$$

et

$$\max\{C_2(E_N), S(E_N)\} < 120 \max\left\{\frac{k^2}{N}, (N \log N)^{1/2}\right\}.$$

3 Sur une famille de suites pseudoaléatoires

Le caractère pseudoaléatoire de nombreuses suites binaires a été étudié par J. Cassaigne, S. Ferenczi, C. Mauduit, J. Rivat et A. Sárközy. Cependant, ces constructions produisent seulement peu de suites pseudoaléatoires distinctes, alors que parfois on a besoin des grandes familles des suites pseudoaléatoires.

L. Goubin, C. Mauduit et A. Sárközy ont réussi à construire de grandes familles de suites pseudoaléatoires généralisant une construction basée sur le symbole de Legendre.

Ici, en généralisant une construction de Sárközy [14], nous construisons un autre type de grande famille de suites pseudoaléatoires basées sur la notion de l'indice (logarithme discret). Soit p un nombre premier fixé, g une racine primitive modulo p , $(a, p) = 1$, et désignons par $\text{ind } a$ l'indice de a :

$$g^{\text{ind } a} \equiv a \pmod{p}$$

avec $1 \leq \text{ind } a \leq p - 1$. Donc

Construction 1 Définissons la suite $E_{p-1} = \{e_1, \dots, e_{p-1}\}$ par

$$e_n = \begin{cases} +1 & \text{si } 1 \leq \text{ind } f(n) \leq (p-1)/2 \\ -1 & \text{si } (p+1)/2 \leq \text{ind } f(n) \leq p-1 \text{ ou } p \mid f(n). \end{cases} \quad (1)$$

Nous donnons d'abord des estimations pour les mesures de bonne-distribution, de corrélation et de symétrie de la suite E_{p-1} .

Théorème 6 Pour tout $f(x) \in F_p[x]$ de degré $k \geq 1$ on a

$$W(E_{p-1}) < 38kp^{1/2}(\log p)^2.$$

Le cas de la mesure de corrélation est plus difficile :

Théorème 7 Supposons qu'au moins une des quatre conditions suivantes est vérifiée :

- a) f est irréductible.
- b) Si f a la factorisation $f = \varphi_1^{\alpha_1} \varphi_2^{\alpha_2} \dots \varphi_u^{\alpha_u}$ où $\alpha_i \in \mathbb{N}$ et φ_i est irréductible sur F_p , il existe un β tels que exactement un ou deux φ_i ont le degré β ;
- c) $\ell = 2$;
- d) $(4\ell)^k < p$ ou $(4k)^\ell < p$.

Alors on a

$$C_\ell(E_{p-1}) < 10k\ell 4^\ell p^{1/2}(\log p)^{\ell+1}.$$

Dans cette section, on étudie également la mesure de symétrie et la complexité des familles (une notion importante introduite par Ahlswede, Mauduit, Khachatryan et Sárközy [1]).

4 Une version rapide d'un générateur pseudoaléatoire

Dans le troisième chapitre, on a construit une grande famille des suites pseudoaléatoires en utilisant le logarithme discret. Les suites dans cette construction ont des propriétés pseudoaléatoires fortes, mais elles peuvent être construites seulement très lentement, puisqu'on ne connaît pas d'algorithme rapide pour calculer $\text{ind } n$. Dans ce chapitre nous améliorons la construction (1) en modifiant la suite de manière à pouvoir la construire plus vite.

Soit p un nombre premier impair, $f \in \mathbb{F}_p[x]$ un polynôme de degré $k \geq 1$, soit m tel que

$$m \mid p - 1,$$

et soit x un nombre premier avec $m : (x, m) = 1$. L'idée cruciale de la construction est de réduire $\text{ind } n$ modulo m :

Construction 2 On définit $\text{ind}^* n$ de la manière suivante : pour tout $1 \leq n \leq p - 1$

$$\text{ind } n \equiv x \cdot \text{ind}^* n \pmod{m}$$

($\text{ind}^* n$ existe car $(x, m) = 1$.) Définissons la suite $E_{p-1} = \{e_1, \dots, e_{p-1}\}$ par

$$e_n = \begin{cases} +1 & \text{si } 1 \leq \text{ind}^* f(n) \leq \frac{m}{2}, \\ -1 & \text{si } \frac{m}{2} < \text{ind}^* f(n) \leq m \text{ ou } p \mid f(n). \end{cases} \quad (2)$$

Nous montrons que cette construction a presque autant que bonnes propriétés pseudoaléatoires que celle décrite dans le troisième chapitre, cependant elle peut être construite beaucoup plus rapidement, car nous pouvons

calculer $\text{ind}^* f(n)$ par $c(\log p)^4 \max_{p \text{ nombre premier, } p|m} p$ opérations. Si les facteurs premiers de m sont plus petits que $\log p$ alors $\text{ind}^* f(n)$ peut être calculé par $c(\log p)^6$ opérations.

5 Sur la corrélation des ordres binaires

Alon, Kohayakawa, Mauduit, Moreira et V. Rödl [2] ont montré que si $1 \leq k \leq N$ sont entiers, on a $C_{2k}(E_N) > \sqrt{\frac{1}{2} \left\lfloor \frac{N}{2k+1} \right\rfloor}$ pour $E_N \in \{-1, +1\}^N$.

La mesure de corrélation d'ordre impair peut être petite : pour la suite $E_N = \{-1, +1, -1, +1, \dots\}$ on a $C_{2k+1}(E_N) = 1$. Pour cette suite E_N on a $C_2(E_N) = \sum_{n=1}^{N-1} e_n e_{n+1} = N-1$. Cassaigne, Mauduit et Sárközy [3] ont posé le problème suivant :

Problème 1 Pour $N \rightarrow \infty$, existe-t-il des suites E_N tels que $C_2(E_N) = O(\sqrt{N})$ et $C_3(E_N) = O(1)$ simultanément ?

Récemment, Mauduit [10] a posé une autre question apparentée :

Problème 2 Est-il vrai que pour tout $E_N \in \{-1, +1\}^N$ on a

$$C_2(E_N)C_3(E_N) \gg N$$

ou au moins

$$C_2(E_N)C_3(E_N) \gg N^c \tag{3}$$

avec une constante $\frac{1}{2} < c \leq 1$?

Dans le cinquième chapitre nous réglerons le problème 1 et partiellement le problème 2 en montrant (3) avec la constante $c = 2/3$.

Théorème 8 Si $k, \ell \in \mathbb{N}$, $\log N \geq 2k + 1 > 2\ell$, $N \in \mathbb{N}$ et $N > 67k^4 + 400$,

$E_n \in \{-1, +1\}^N$ et si

$$C_{2\ell}(E_N) < \frac{1}{40\sqrt{k(2\ell+1)}} N^{1-\ell/(2k+1)},$$

alors on a

$$C_{2k+1} > \frac{1}{7} \left(\frac{2\ell}{17(2k+1)} \right)^{\ell/2} N^{1/2}.$$

En particulier pour $\ell = 1$ si

$$C_2(E_N) < \frac{N^{2/3}}{50\sqrt{\log N}},$$

alors on a

$$C_3(E_N), C_5(E_N), \dots \gg \sqrt{N}.$$

6 Sur la généralisation d'une inégalité entre les mesures pseudoaléatoires

Mauduit et Sárközy [12] ont prouvé l'inégalité suivante impliquant les mesures pseudoaléatoires W et C_2 : Pour toute suite $E_N \in \{-1, +1\}^N$ on a $W(E_N) \leq 3\sqrt{NC_2(E_N)}$. Dans le dernier chapitre (voir également [5]), nous généralisons cette inégalité à la mesure de corrélation d'ordre supérieur :

Théorème 9 Pour tout $E_N \in \{-1, +1\}^N$ on a

$$W(E_N) \ll N^{1-1/(2\ell)} (C_{2\ell}(E_N))^{1/(2\ell)}.$$

Mauduit et Sárközy [12] ont montré que leur inégalité est précise en employant des arguments probabilistes. Ici, nous présentons une construction explicite (une suite de la famille décrite dans le quatrième chapitre) pour laquelle même l'inégalité généralisée est précise à un facteur constant près.

Références

- [1] R. Ahlswede and L. Khachatrian and C. Mauduit and A. Sárközy, *A complexity measure for families of binary sequences*, Periodica Mathematica Hungarica 46 (2003), 107-118.
- [2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira et V. Rödl, *Measures of pseudorandomness for finite sequences : minimum values*, Combinatorics, Probability and Computation, à paraître.
- [3] J. Cassaigne, C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences VII : The measures of pseudorandomness*, Acta Arith. 103 (2002), 97-118.
- [4] L. Goubin, C. Mauduit, A. Sárközy, *Construction of large families of pseudorandom binary sequences*.
- [5] K. Gyarmati, *An inequality between the measures of pseudorandomness*, Annales Univ. Sci. Budapest. Eötvös, à paraître.
- [6] K. Gyarmati, *On a family of pseudorandom binary sequences*, Periodica Math. Hungar., 49 (2), 2004 pp. 1-19.
- [7] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer et Combinatorics, Conference Proceedings, à paraître.
- [8] K. Gyarmati, *On a pseudorandom property of binary sequences*, The Ramanujan Journal, Vol. 8, No. 3 (2004), 289-302.
- [9] K. Gyarmati, *On the correlation of binary sequences*, Studia Sci. Math. Hungar., à paraître.

- [10] C. Mauduit, *Construction of pseudorandom finite sequences*, notes de conférence “Information Theory and Some Friendly Neighbours – ein Wunschkonzert”, non publié, Bielefeld, 2003.
- [11] C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences I : Measures of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365-377.
- [12] C. Mauduit, A. Sárközy, *On the measures of pseudorandomness of binary sequences*, Discrete Math. 271 (2003), 195-207.
- [13] A. J. Menezes, P. C. Oorschot, S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996.
- [14] A. Sárközy, *A finite pseudorandom binary sequence*, Studia Sci. Math. Hungar. 38 (2001), 377-384.