Astérisque

HARALD NIEDERREITER

Nombres pseudo-aléatoires et équirépartition

Astérisque, tome 61 (1979), p. 155-164

<http://www.numdam.org/item?id=AST_1979__61__155_0>

Société Mathématique de France
Astérisque 61 (1979) p. 115-164

NOMBRES PSEUDO-ALÉATOIRES ET ÉQUIRÉPARTITION

par

Harald NIEDERREITER

Nous voulons, tant ce feu nous brûle le cerveau,
Plonger au fond du gouffre, Enfer ou Ciel, qu'
importe? Au fond de l'Inconnu pour trouver du
nouveau!

Baudelaire, *Le Voyage*

La nécessité de produire des "nombres aléatoires" résulte dans le cadre de la simulation de processus complexes et, en particulier, dans la méthode de Monte-Carlo que l'on peut décrire brèvement comme une technique numérique basée sur des procédés d'échantillonnage. Un problème se pose naturellement, à savoir comment mettre à exécution l'échantillonnage concret dans une application spécifique d'une telle méthode. Une définition adéquate d'une suite de nombres aléatoires serait utile pour résoudre cette question pratique.

Il y a plusieurs tentatives bien connues d'arriver à une définition satisfaisante d'une suite de nombres aléatoires (voir [3], [7], [8]). Pour la plupart, ces tentatives sont fondées sur le principe de Venn [21] qui postule qu'une telle suite satisfasse certaines propriétés de distribution. Normalisons les termes d'une suite et supposons désormais qu'ils appartiennent à l'intervalle unité $I = [0, 1]$. Une condition *sine qua non* pour le caractère aléatoire d'une suite

$x_0, x_1, \ldots$ est l'équirépartition dans I. Mais cela ne suffit pas, car il faut aussi avoir égard à l'exigence que les termes successifs de la suite soient indépendants. Ceci nous amène à considérer la "suite-chenille" des s-uplets $\underline{x}_n = (x_n, x_{n+1}, \ldots, x_{n+s-1})$, $n = 0, 1, \ldots$, et la condition que cette suite soit équirépartie dans l'hypercube unité $I^s = [0, 1]^s$. Si l'on impose cette condition pour tout entier $s \geq 1$, il revient au même de demander que la suite scalaire $x_0, x_1, \ldots$ soit *complètement équirépartie*.

Par conséquent, on peut regarder une suite complètement équirépartie comme modèle de suite de nombres aléatoires, au moins pour les applications de la méthode de Monte-Carlo dans lesquelles interviennent seulement les propriétés de distribution, par exemple pour l'intégration numérique. On connaît de nombreuses constructions de suites complètement équiréparties, la plus utile du point de vue numérique étant celle de Knuth [2] qui emploie des fractions dyadiques. Rauzy [20] a montré le résultat suivant; si f est une fonction entière non polynomiale qui prend des valeurs réelles sur l'axe réel et verifie la condition de croissance

$$\overline{\lim_{r \to \infty}} \frac{\log \log M(f;r)}{\log \log r} < \frac{5}{4} \quad \text{où} \quad M(f;r) = \sup_{|z| \leq r} |f(z)| ,$$

alors la suite des parties fractionnaires $\{f(1)\}, \{f(2)\}, \ldots, \{f(n)\}, \ldots$ est complètement équirépartie. Levine [6] a construit, pour tout nombre transcendant $\beta > 1$, un nombre $\alpha \in \mathbb{R}$ tel que la suite des parties fractionnaires $\{\alpha\beta\}, \{\alpha\beta^2\}, \ldots, \{\alpha\beta^n\}, \ldots$ est complètement équirépartie. Voir [19] pour les constructions classiques de suites complètement équiréparties.

Quoique les suites complètement équiréparties soient intéressantes en ce qui concerne l'aspect théorique, il est préférable, pour les calculs très étendus de Monte-Carlo, d'utiliser des suites qui sont fabriquées dans l'ordinateur par des algorithmes rapides et simples. Bien entendu, une suite déterministe produite de cette manière ne se qualifie pas de suite de nombres aléatoires. Néanmoins, elle peut respecter plusieurs critères de caractère aléatoire adaptés à des besoins

particuliers. Dans ce cas, les termes d'une telle suite déterministe s'appellent *nombres pseudo-aléatoires*.

L'algorithme le plus efficace et commode pour l'obtention de nombres pseudo-aléatoires est le *générateur multiplicatif* (*homogène* ou *mixte*) proposé par Lehmer [5]. Soient m, λ, r et y_0 des entiers avec $m \geq 3$, $2 \leq \lambda < m$, $0 \leq y_0 < m$, λ et m premiers entre eux et $(\lambda - 1)y_0 + r \not\equiv 0 \pmod m$. Alors on engendre une suite d'entiers $y_0, y_1, \ldots$ par la récurrence

$$y_{n+1} \equiv \lambda y_n + r \pmod m, \quad n = 0,1, \ldots,$$

en observant la limitation $0 \leq y_n < m$ pout tout n. Une suite de nombres $x_0, x_1, \ldots$ appartenant à l'intervalle unité I est dérivée en posant $x_n = y_n/m$ pour tout n. Les nombres x_n sont déjà les nombres pseudo-aléatoires engendrés par le générateur multiplicatif. Habituellement, on prend pour m un grand nombre premier ou une grande puissance de 2. On appelle m le *module* et λ le *multiplicateur*.

On peut étudier les propriétés de distribution d'une suite $x_0, x_1, \ldots$ engendrée par le générateur multiplicatif en utilisant un modèle probabiliste. Fixons $\Theta \in \mathbb{R}$ et considérons la récurrence

$$x_{n+1} = \{\lambda x_n + \Theta\}, \quad n = 0,1, \ldots,$$

où $x_0 \in I$ est arbitraire. Or, la transformation

$$T : x \in I \mapsto \{\lambda x + \Theta\}$$

est ergodique par rapport à la mesure de Lebesgue. On en déduit que, pour presque toute valeur de départ $x_0 \in I$, la suite $(x_n) = (T^n x_0)$, $n = 0,1, \ldots$, est équirépartie dans I.

Quant à l'épreuve d'indépendance des termes successifs de la suite $x_0, x_1, \ldots$ ci-dessus, on voit aisément que la suite $\underline{x}_n = (x_n, x_{n+1}, \ldots, x_{n+s-1})$, $n = 0,1, \ldots,$

n'est jamais équirépartie dans I^s pour $s \geq 2$. Mais cette suite est en un certain sens presque sûrement "asymptotiquement" équirépartie dans I^s. Écrivons $\underline{x}_n(\lambda)$ au lieu de $\underline{x}_n$ pour souligner que $\underline{x}_n$ dépend de λ. Alors Franklin [1] a montré le résultat suivant: si f est une fonction continue de I^s dans $\mathbb{R}$, on a pour presque tout $x_0 \in I$

$$\lim_{\lambda \to \infty} \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\underline{x}_n(\lambda)) = \int_{I^s} f(\underline{t}) d\underline{t}.$$

Pour $\Theta = 0$ le théorème central-limite "asymptotique" de Yermakov [22] précise ce résultat. Soit f une fonction continue de I^s dans $\mathbb{R}$ et dénotons par

$$R_N(f, x_0, \lambda) = \frac{1}{N} \sum_{n=0}^{N-1} f(\underline{x}_n(\lambda)) - \int_{I^s} f(\underline{t}) d\underline{t}$$

l'erreur d'intégration. Alors,

$$\lim_{N \to \infty} \lim_{\lambda \to \infty} \text{mes} \left\{ x_0 \in I : R_N(f, x_0, \lambda) < \frac{\sigma u}{\sqrt{N}} \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt$$

pour tout $u \in \mathbb{R}$, où la constante σ désigne un certain écart-type dépendant de f. Ces résultats suggèrent que les nombres pseudo-aléatoires engendrés par un générateur multiplicatif devraient montrer un bon comportement concernant les épreuves de distribution, au moins pour certains grands multiplicateurs λ.

Naturellement, il est plus difficile d'établir des résultats effectifs pour des suites spécifiques de nombres pseudo-aléatoires. Le problème de la répartition dans I des termes d'une telle suite étant déjà étudié en détail (voir [9], [10], [12]), considérons maintenant le problème de répartition pour la dimension $s \geq 2$. L'écart entre la fonction de distribution des points $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_{N-1}$ de la suite-chenille et l'équirépartition sur I^s est mesuré par la *discrépance*

$$D_N^{(s)} = \sup_J |F_N(J) - V(J)|,$$

où le sup est étendu à tous les sous-intervalles J de I^s, $F_N(J)$ est N^{-1} fois le nombre d'entiers n tels que $0 \leq n < N$ et $\underline{x}_n \in J$, et V(J) désigne le volume de J. Il faut remarquer que pour la valeur de départ rationnelle $x_0 = y_0/m$ les suites $x_0, x_1, \ldots$ et $\underline{x}_0, \underline{x}_1, \ldots$ sont périodiques avec la même période τ. À cause de cela, il est évident que l'on n'utilise les terms x_n et $\underline{x}_n$ de ces suites que pour $0 \leq n < \tau$. Par conséquent, on ne considère $D_N^{(s)}$ que pour $1 \leq N \leq \tau$.

Pour estimer $D_N^{(s)}$, il faut établir au préalable une inégalité (voir [14]) qui met en évidence le lien entre la discrépance de certaines suites et les sommes exponentielles associées. Ce résultat améliorera dans ce cas spécial une inégalité de type général (voir [4], [18]). Soit M l'ensemble des treillis $\underline{h} = (h_1, \ldots, h_s) \in \mathbb{Z}^s$ avec $-m/2 < h_j \leq m/2$ pour $1 \leq j \leq s$ et $\underline{h} \neq \underline{0}$. Posons $r(h,m) = m \sin(\pi|h|/m)$ pour $h \neq 0$, $r(0,m) = 1$, $r(\underline{h},m) = r(h_1,m) \ldots r(h_s,m)$, et $e(t) = e^{2\pi i t}$ pour $t \in \mathbb{R}$.

LEMME. - *Pour tout générateur multiplicatif avec module* m *et toute dimension* $s \geq 2$ *on a*

$$D_N^{(s)} \leq \frac{s}{m} + \sum_{\underline{h} \in M} \frac{1}{r(\underline{h},m)} \left| \frac{1}{N} \sum_{n=0}^{N-1} e(\underline{h} \cdot \underline{x}_n) \right|$$

quel que soit l'entier positif N.

Ainsi le problème de l'estimation de $D_N^{(s)}$ est réduit au traitement des sommes exponentielles qui surgissent dans le résultat précédent. Les sommes exponentielles de ce type étaient déjà étudiées par l'auteur (voir [11], [12]). On peut alors donner certaines améliorations de ces résultats antérieurs. Dans le cas d'un générateur multiplicatif homogène (c'est-à-dire $r \equiv 0 \pmod m$) on arrive à l'inégalité suivante.

THÉORÈME 1. - *Soient* μ *l'ordre de* $\lambda + m\mathbb{Z}$ *dans le groupe multiplicatif* $(\mathbb{Z}/m\mathbb{Z})^*$ *et* b *un entier avec* b *et* m *premiers entre eux. Alors on a*

$$\left| \sum_{n=0}^{\mu-1} e(b\lambda^n/m) \right| \leq \sqrt{m} - \frac{\mu}{\phi(m)}(\sqrt{m} - 1)$$

et

$$\left| \sum_{n=0}^{N-1} e(b\lambda^n/m) \right| < \sqrt{m}\left(\frac{2}{\pi}\log \mu + \frac{2}{5}\right) + N\left(\frac{\sqrt{m}}{\mu} - \frac{\sqrt{m}-1}{\phi(m)}\right)$$

pour $1 \leq N < \mu$.

Dans le cas d'un générateur multiplicatif mixte (c'est-à-dire $r \not\equiv 0 \pmod m$) on applique le résultat suivant.

THÉORÈME 2. - *Soient $y_0, y_1, \ldots$ la suite d'entiers engendrés par un générateur multiplicatif et b un entier avec b et m premiers entre eux. Alors on a*

$$\left| \sum_{n=0}^{\tau-1} e(by_n/m) \right| \leq \left(\frac{m\tau - \tau^2}{\mu}\right)^{1/2}$$

et

$$\left| \sum_{n=0}^{N-1} e(by_n/m) \right| < \left(\frac{m\tau}{\mu}\right)^{1/2}\left(\frac{2}{\pi}\log \tau + \frac{2}{5}\right) + \frac{N}{\tau}\left(\frac{m\tau - \tau^2}{\mu}\right)^{1/2}$$

pour $1 \leq N < \tau$.

Ces inégalités conduisent à des bornes supérieures de la discrépance $D_N^{(s)}$. On introduit une légère modification d'une définition donnée antérieurement (voir [13], [14]). En effet, posons

$$\rho^{(s)}(\lambda, m) = \min r(\underline{h}),$$

où le minimum est étendu à tous les treillis $\underline{h} = (h_1, \ldots, h_s) \in M$ avec $h_1 + h_2\lambda + \ldots + h_s\lambda^{s-1} \equiv 0 \pmod m$ et $r(\underline{h})$ désigne l'entier positif $r(\underline{h}) = \max(1, 2|h_1|) \ldots$

$\max(1, 2|h_s|)$. Utilisons C_s pour dénoter une constante positive qui dépend seulement de la dimension s, mais qui peut atteindre des valeurs différentes selon le cas.

THÉORÈME 3. - Si le module m est un nombre premier, on a

$$D_\tau^{(s)} \leq \frac{(m-\tau)}{\tau}^{1/2}\left(\frac{2}{\pi}\log m + \frac{7}{5}\right)^s + \frac{C_s \log^s m}{\rho^{(s)}(\lambda,m)}$$

et

$$D_N^{(s)} \leq C_s\left(\frac{m^{1/2}(\log m)^{s+1}}{N} + \frac{\log^s m}{\rho^{(s)}(\lambda,m)}\right) \quad \text{pour} \quad 1 \leq N < \tau.$$

La borne supérieure de $D_\tau^{(s)}$ suggère que l'on choisisse un multiplicateur qui donne la valeur maximale de τ, c'est-à-dire un multiplicateur qui est une racine primitive mod m. Dans ce cas, on obtient

$$D_\tau^{(s)} \leq \frac{C_s \log^s m}{\rho^{(s)}(\lambda,m)}.$$

On en déduit qu'une suite de nombres pseudo-aléatoires engendrés par un générateur multiplicatif pour laquelle s termes successifs sont stochastiquement presque indépendants est obtenue en choisissant un grand module premier m et un multiplicateur λ qui est une racine primitive mod m et donne une grande valeur de $\rho^{(s)}(\lambda,m)$. D'ailleurs, on peut montrer qu'il existe, pour tout $s \geq 2$ et tout module premier m, une racine primitive λ_0 mod m telle que

$$D_\tau^{(s)} \leq C_s\, m^{-1} \log^s m \log\log m,$$

où $\tau = m - 1$. Il est remarquable que cet ordre de grandeur est très proche de la valeur la plus faible connue (et conjecturée) de la discrépance de $m-1$ points de I^s (voir [4], [17]).

On peut établir des résultats analogues pour un module m qui est une puissance d'un nombre premier (voir [14], [15], [17]) ou une puissance de 10 (voir [16]). De plus, il y a une borne inférieure de $D_N^{(s)}$ qui indique que le nombre $\rho^{(s)}(\lambda,m)$ fournit une mesure appropriée de la qualité des paramètres λ et m à l'égard de la dimension s.

THÉORÈME 4. - *Pour tout module* m *et tout multiplicateur* λ *on a*

$$D_N^{(s)} \geq \frac{C_s}{\rho^{(s)}(\lambda,m)}$$

quel que soit l'entier N, $1 \leq N \leq \tau$.

RÉFÉRENCES

[1] J.N. FRANKLIN. - Deterministic simulation of random processes. Math. Comp. 17, 28-59 (1963).

[2] D.E. KNUTH. - Construction of a random sequence. Nordisk Tidskr. Informations-Behandling (BIT) 5, 246-250 (1965).

[3] D.E. KNUTH. - The Art of Computer Programming, Vol.2: Seminumerical Algorithms. Addison-Wesley, Reading, Mass., 1969.

[4] L. KUIPERS et H. NIEDERREITER. - Uniform Distribution of Sequences. Wiley-Interscience, New York, 1974.

[5] D.H. LEHMER. - Mathematical methods in large-scale computing units. Proc. 2nd Sympos. on Large-Scale Digital Calculating Machinery (Cambridge, Mass., 1949), pp. 141-146. Harvard University Press, Cambridge, Mass., 1951.

[6] M.B. LEVINE. - Sur l'équirépartition de la suite $\{\alpha\lambda^x\}$ (en russe). Mat. Sb. 98, 207-222 (1975).

[7] P. MARTIN-LÖF. - The definition of random sequences. Information and Control 9, 602-619 (1966).

[8] M. MENDÈS FRANCE. - Suites de nombres au hasard (d'après Knuth). Sém. Théorie des Nombres 1974-1975, Univ. Bordeaux, Exp. 6.

[9] H. NIEDERREITER. - On the distribution of pseudo-random numbers generated by the linear congruential method. Math. Comp. 26, 793-795 (1972).

[10] H. NIEDERREITER. - On the distribution of pseudo-random numbers generated by the linear congruential method. II. Math. Comp. 28, 1117-1132 (1974).

[11] H. NIEDERREITER. - Some new exponential sums with applications to pseudo-random numbers. Topics in Number Theory (Debrecen, 1974), Colloq. Math. Soc. János Bolyai, Vol. 13, pp. 209-232. North-Holland, Amsterdam, 1976.

[12] H. NIEDERREITER. - On the distribution of pseudo-random numbers generated by the linear congruential method. III. Math. Comp. 30, 571-597 (1976).

[13] H. NIEDERREITER. - Statistical independence of linear congruential pseudo-random numbers. Bull. Amer. Math. Soc. 82, 927-929 (1976).

[14] H. NIEDERREITER. - Pseudo-random numbers and optimal coefficients. Advances in Math. 26, 99-181 (1977).

[15] H. NIEDERREITER. - The serial test for linear congruential pseudo-random numbers. Bull. Amer. Math. Soc. 84, 273-274 (1978).

[16] H. NIEDERREITER. - Statistical tests for linear congruential pseudo-random numbers. COMPSTAT 1978: Proceedings in Computational Statistics (Leiden, 1978), pp. 398-404. Physica-Verlag, Vienne, 1978.

[17] H. NIEDERREITER. - Quasi-Monte Carlo methods and pseudo-random numbers. Bull. Amer. Math. Soc. 84 (à paraître).

[18] H. NIEDERREITER et W. PHILIPP. - Berry-Esseen bounds and a theorem of Erdös and Turán on uniform distribution mod 1. Duke Math. J. 40, 633-649 (1973).

[19] A.G. POSTNIKOV. - Modelage arithmétique de processus aléatoires (en russe). Trudy Mat. Inst. Steklov. 57 (1960).

[20] G. RAUZY. - Fonctions entières et répartition modulo un, II. Bull. Soc. Math. France 101, 185-192 (1973).

[21] J. VENN. - The Logic of Chance. Macmillan, London, 1876.

[22] S.M. YERMAKOV. - Note sur les suites pseudo-aléatoires (en russe). J.Vyčisl. Mat. i Mat. Fiz. 12, 1077-1082 (1972).

Harald NIEDERREITER
Chair in Pure Mathematics
University of the West Indies
KINGSTON 7
Jamaïque

Ce travail était subventionné par U.S. National Science Foundation Grant MCS-7701699A01.