Systèmes dynamiques pour générer des nombres pseudo-aléatoires

Jean-François Couchot

FEMTO-ST Institut, Université de Franche-Comté, France

TIPE 2016

Contexte et problématique

Utilisation des nombres aléatoires

- Vie courante : localisation GPS, cryptage de données (RSA) ;
- Science : statistiques, en probabilité, en simulation numérique.

Génération

- Réellement aléatoire : processus physique $\rightsquigarrow$ non reproductible.
- Pseudo-aléatoire (PRNG) : algorithme $\rightsquigarrow$ reproductible, quid de la sécurité ?

Problématique

Les systèmes dynamiques sont-ils un bon outil mathématique pour générer des nombres pseudo-aléatoires ? Quelles sont les qualités minimums d'un PRNG ?

Plan

1. Résultats principaux et illustrations
2. Points clefs des démonstrations
3. Questions

Plan

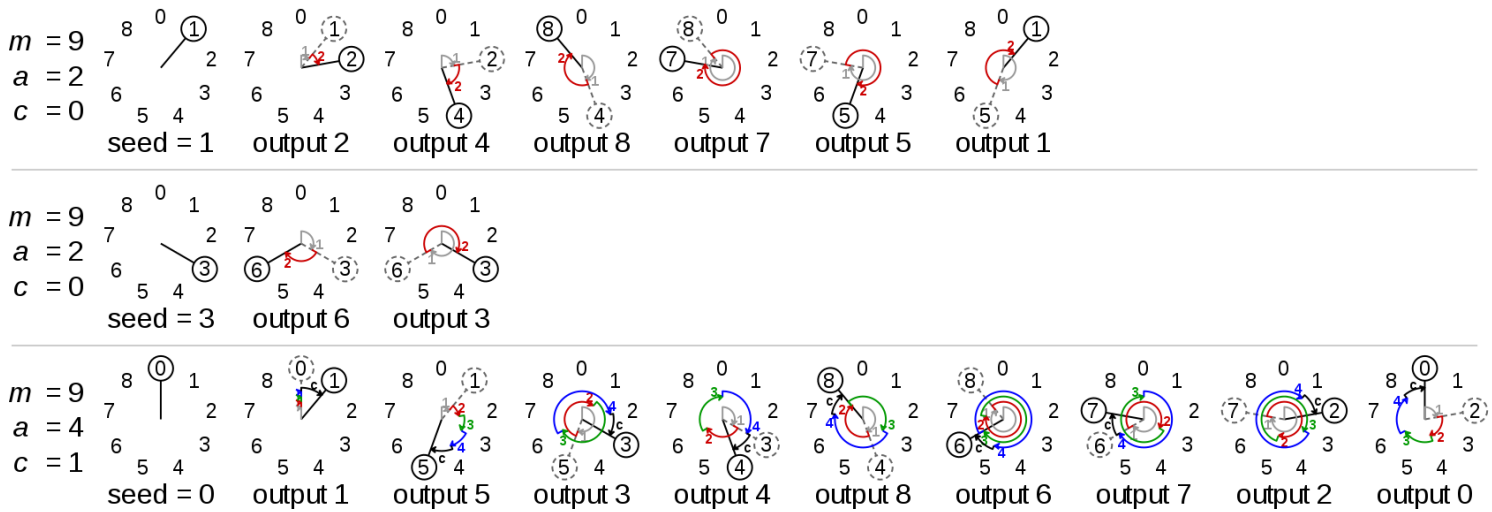
Générateur congruentiel linéaire

Definition

Pour m, a, $c \in \mathbb{N}$ t.q. $m \geq a, c$. La suite d'entiers

$$\begin{cases} x_0 \in \mathbb{N} \text{ t.q. } 0 \leq x_0 \leq m-1 \\ x_i = a \times x_{i-1} + c \mod m. \end{cases} \quad (1)$$

est un *générateur congruentiel linéaire* de *graine* x_0.



- Bien choisir m, a, c et x_0 : primordial !

$c \neq 0$: conditions suff. pour une période max.

(Théorème 1 : Hull & Dobell 1962)

La période de la suite produite par le générateur congruentiel linéaire est m si

1. *c et m : premiers entre eux,*
2. $a \equiv 1 \mod p$ *pour chaque facteur premier p de m et*
3. $a \equiv 1 \mod 4$ *si* 4 *divise m.*

Analyse de l'exemple de la Figure 1

On a $m = 180$, $a = 61$ et $c = 7$. Or $m = 180 = 2^2 \times 3^2 \times 5$.

1. c et m : premiers entre eux,
2. $61 \equiv 1 \mod 2$, $61 \equiv 1 \mod 3$, $61 \equiv 1 \mod 5$ et
3. $61 \equiv 1 \mod 4$.

La période d'un tel générateur est donc maximale.

$c = 0$: conditions suff. pour une période max.

(Théorème 2)

La période de la suite produite par le générateur congruentiel linéaire est $m - 1$ si

1. *x_0 n'est pas nul ;*
2. *m est un nombre premier ;*
3. *a est d'ordre $m - 1$*

(Théorème 3)

Soit m un nombre premier. On peut toujours trouver a d'ordre $m - 1$, $1 \leq a \leq m - 1$.

Analyse de l'exemple de la Figure 2

On a $m = 181$ et $a = 19 \times 80 \times 125 \equiv 131 \mod 181$ et

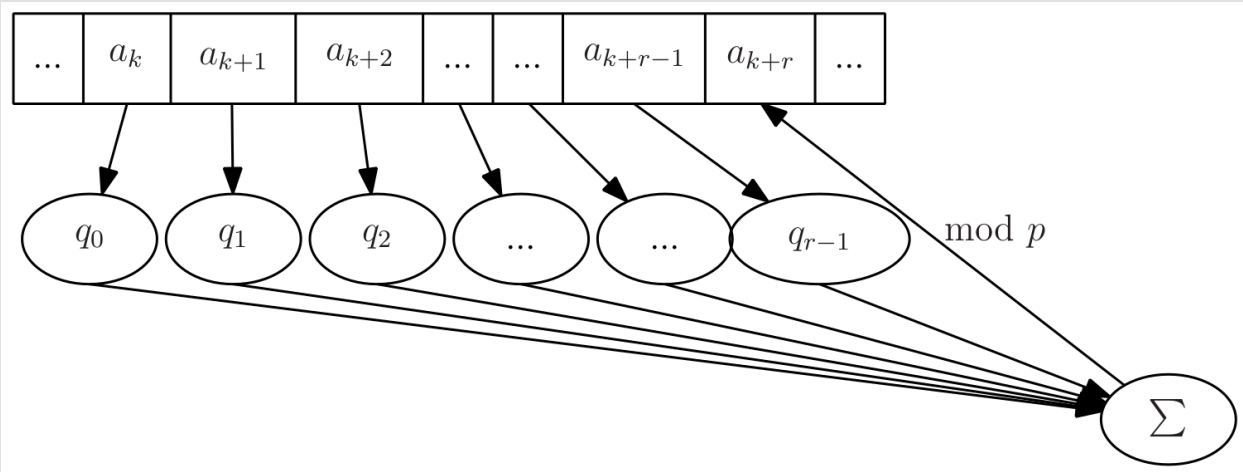
1. $x_0 = 1$ n'est pas nul ;
2. $m = 181$ est un nombre premier ;
3.
   - 19 est d'ordre 4, 80 est d'ordre 9 et 125 est d'ordre 5 ;
   - 4, 5 et 9 premiers entre eux $\rightsquigarrow$ $a = 19 \times 80 \times 125$ est d'ordre 180 ;

La période d'un tel générateur est 180.

Registre à décalage

Definition

Soit p un nombre premier.



Un générateur *à base de registres à décalage linéaires* produit a_{k+r} de $\mathbb{F}_p$ selon

$$a_{k+r} \equiv \sum_{j=0}^{r-1} q_j a_{j+k} \mod p. \tag{9}$$

Conditions suff. pour une période max.

(Théorème 5)

Soit p premier, $q_0, \ldots, q_{r-1} \in \mathbb{F}_p$ tels que $Q(x) = x^r - q_{r-1}x^{r-1} - \cdots - q_1 x - q_0$ est primitif sur $\mathbb{F}_p$ et soit $a_0, \ldots, a_{r-1}$ une graine différente du vecteur nul. Le générateur à base de registres à décalage linéaires engendre une suite de période $p^r - 1$.

Analyse de l'exemple de la Figure 5

On a $p = 3$, $r = 4$, la graine $1, 0, 0, 0$ n'est pas nulle. Il suffi(rai)t de démontrer que

1. $Q(x) = x^4 - 2x^3 - 1 \equiv x^4 + x^3 + 2$ est irréductible (facile)
2. chaque élément de $(\mathbb{F}_p)_{r-1}(x)$ peut s'écrire sous la forme x^i modulo $Q(x)$, $0 \leq i \leq p^r - 2$ (long, mais vrai)

La période d'un tel générateur est donc $3^4 - 1 = 80$.

Répartition des sous-suites

Definition (Fenêtre)

Sous les mêmes conditions que celles du théorème 5, on appelle *fenêtre* toute sous-suite de longueur $p^r - 1$.

(Théorème 6)

Sous les mêmes conditions que celle du théorème 5, soit k un entier, $k \leq r$. Alors dans chaque fenêtre, toute sous-suite de longueur k apparaît p^{r-k} fois, sauf la sous-suite nulle qui, elle, apparaît $p^{r-k} - 1$ fois.

Plan

Démo. du théorème 1 ($a \neq 1$)

La période de la suite produite par le générateur congruentiel linéaire est m si

1. *c et m : premiers entre eux,*
2. $a \equiv 1 \mod p$ *pour chaque facteur premier p de m et*
3. $a \equiv 1 \mod 4$ *si* 4 *divise m.*

1. LEMME 1 $\Rightarrow$ la période est l'entier n minimum qui établit

$$\frac{a^n - 1}{a - 1} \equiv 0 \mod m. \tag{4}$$

2. **Cas** $n = m = p^\alpha$, p **premier et** $\alpha \geq 2$ (LEMME 2) :
   - 2.1 binôme de Newton pour réécrire a^n ;
   - 2.2 analyse de $\frac{p^\alpha}{j}$ dans (6) $\rightsquigarrow$ (4) établie.
3. **Cas** $n < m = p^\alpha$, p **premier et** $\alpha \geq 2$ (LEMME 3) :
   - 3.1 (4) établie $\Rightarrow$ n puissance de p,
   - 3.2 or (4) pas établie pour $n = p^{\alpha-1}$.
4. **Cas** $m = p_1^{\alpha 1} p_2^{\alpha 2} \ldots p_s^{\alpha s}$ **et** $a = 1 + k p_1^{\beta_1} p_2^{\beta_2} \ldots p_s^{\beta_s}$:
   - 4.1 preuve similaire (cas 2. puis 3.).

Démo. du théorème 3

Soit m un nombre premier. On peut toujours trouver a d'ordre $m-1$, $1 \leq a \leq m-1$.

Preuve constructive :

1. Décomposition de $m-1 = p_1^{k_1} \dots p_s^{k_s}$ en produit de facteurs premiers ;
2. $x^{p_i^{k_i}} - 1 \equiv 0 \mod m$ admet une solution g_i sur $\{1, \dots, m-1\}$ d'ordre $p_i^{k_i}$ (LEMME 4) ;
3. $a = g_1 \dots g_s$ est d'ordre $p_1^{k_1} \dots p_s^{k_s} = m-1$ (LEMME 5).

Démo. du théorème 5

(Théorème 5)

Soit p premier, $q_0, \ldots, q_{r-1} \in \mathbb{F}_p$ tels que $Q(x) = x^r - q_{r-1}x^{r-1} - \cdots - q_1 x - q_0$ est primitif sur $\mathbb{F}_p$ et soit $a_0, \ldots, a_{r-1}$ une graine différente du vecteur nul. Le générateur à base de registres à décalage linéaires engendre une suite de période $p^r - 1$.

1. Construction d'un polynôme $b \in (\mathbb{F}_p)_{r-1}(x)$ et d'une fonction $T : (\mathbb{F}_p)_{r-1}(x) \to \mathbb{F}_p$ t.q. $T(b.x^n) = a_n$ (LEMME 6)
2. Comme $x^{p^r-1} \equiv 1 \mod Q(x)$, d'après 1. la période de a_n est au plus $p^r - 1$ (l. 290)
   - $a_{n+p^r+1} \equiv T(b.x^{n+p^r+1}) \equiv T(b.x^n x^{p^r+1}) \equiv a_n$.
3. Si la période était inférieure à $p^r - 1$, alors $Q(x)$ ne serait pas primitif (l. 296)

Plan

1. Résultats principaux et illustrations
   - Générateur congruentiel linéaire
   - Le registre à décalage

2. Points clefs des démonstrations

3. Questions

Question 2 : Culture G.S., mise en confiance

Dans quel(s) contexte(s) avez vous déjà rencontré des générateurs de nombres pseudo-aléatoires.

Réponses : les probabilités, les statistiques, dans des simulations (méthodes de Monte-Carlo), le chiffrage (algorithme RSA par exemple), notamment pour générer des grands nombres aléatoires premiers, la localisation par GPS.

Question 3 : Compréhension immédiate, mise en confiance

A la ligne 30, il est écrit qu'« un générateur congruentiel linéaire produit une suite périodique dont la période ne peut pas excéder m ». Pourquoi ?

Réponse : En construisant $m + 1$ termes successifs, il est nécessaire de rencontrer au moins deux termes x_i et x_j égaux puisqu'il n'y a que m termes différents dans $\{0, \ldots, m-1\}$. Quel que soit $k \in \mathbb{N}$, on a $x_{i+k} = x_{j+k}$.

Question 4 : Compréhension immédiate, mise en confiance

A la ligne 204, on prétend savoir calculer « l'inverse de $x_i - x_{i+1}$ modulo m ». Comment faire ceci ? Ce problème a-t-il toujours une solution ?

Réponses : Il s'agit d'appliquer l'algorithme d'Euclide étendu : on cherche u (et v) tels que $(x_i - x_{i+1}).u + m.v = 1$. Or ceci n'admet une solution que si $(x_i - x_{i+1})$ et m sont premiers entre-eux. (Par exemple, impossible de trouver l'inverse de 2 modulo 6).

Question 6 : Application des théorèmes 1 et 2 de la section 2

Que dire des générateurs suivants :

- $x_i = (129x_{i-1} + 907633385) \mod 2^{32}$ (générateur du Turbo Pascal)
- $x_i = 69069x_{i-1} \mod 2^{32}$ (générateur de Marsaglia)

Réponses :

- Le premier vérifie toutes les hypothèses du théorème 1 (907633385 et 2^{32} sont premiers entre eux, 2 est le seul facteur premier de 2^{32} et comme 129 est impair, $129 \equiv 1 \mod 2$, et, enfin, $129 \equiv 1 \mod 4$). Sa période est donc 2^{32}.
- 2^{32} n'est pas premier. On ne peut pas directement appliquer le théorème 2 et on ne peut donc rien conclure. Remarque : cette condition n'est en effet pas nécessaire car la période de ce générateur est cependant 2^{32}.

Question 8 : Ouverture scientifique

Que démontre la section 2.4 ? Quelle est la portée de ce résultat ?
Réponse. Cette section montre qu'un générateur de nombres pseudo-aléatoires peut certes produire des sorties qui vérifient les critères statistiques de bonne distribution (période longues, sous-suites équitablement réparties), mais une personne peut cependant deviner les paramètres et donc « casser » l'aléa.

Dans une simulation physique, cela n'est pas nécessairement grave. Lorsqu'on génère des grands nombres pseudo aléatoirement pour de la cryptographie ou dans un casino, connaître l'algorithme sous-jacent peut être une faille de sécurité.

Question 9 : Ouverture scientifique

Supposons qu'on dispose d'une fonction qui génère un nombre pseudo-aléatoire dont la sortie j est uniformément distribuée sur l'intervalle $[0, 1[$. Comment peut-on faire pour disposer d'une sortie uniformément distribuée sur l'ensemble des entiers $\{0, \ldots, m\}$.
Réponse : On commence par diviser $[0, 1[$ en $m+1$ intervalles de même longueurs : $I_0 = [0, \frac{1}{m+1}[$, $I_1 = [\frac{1}{m+1}, \frac{2}{m+1}[$, $\ldots$ $I_m = [\frac{m}{m+1}, 1[$. Si la sortie j appartient à I_0, on retourne 0,....
Ceci revient à d'abord multiplier j par $m+1$ et à prendre ensuite sa partie entière, ce qui est fait dans le code Python suivant. La fonction Python `randint(a,b)` génère un entier entre a et b, les deux nombres compris.

```
def genereAleaEntier(m):
    j= random()
    return int(j*(m+1))
```