Générateurs de nombres aléatoires
Cours 1 - Générateurs algorithmiques

A. Ridard

A propos de ce document

- Pour naviguer dans le document, vous pouvez utiliser :
  - le menu (en haut à gauche)
  - les différents liens
- Pour signaler une erreur, vous pouvez envoyer un message à l'adresse suivante : anthony.ridard@univ-ubs.fr

Plan du cours

On a besoin d'un procédé permettant de produire des nombres « au hasard » :

- une suite de 0 et de 1 pour simuler des lancers de pièces
- une suite d'entiers de 1 à 6 pour simuler des lancers de dés
- une suite de nombres réels entre 0 et 1 pour simuler une suite de variables aléatoires i.i.d. [1] uniformes sur [0,1]
- ...

1. indépendantes et identiquement distribuées

On a besoin d'un tel procédé :

- en **simulation stochastique** [2] pour modéliser un système (biologique, financier...) afin de comprendre, prévoir, gérer son comportement
- dans les **jeux d'argent liés au hasard** (loterie, casino)
- en **cryptographie** [3]

Ces besoins n'ont pas les mêmes exigences en terme de qualité de l'aléa

2. A partir d'une suite $(U_n)_{n\in\mathbb{N}}$ de variables aléatoires i.i.d. uniformes sur $[0,1]$, on peut en fait simuler une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires i.i.d. de fonction de répartition F en transformant les U_n à l'aide de « l'inverse » F^{-1} : $X_n = F^{-1}(U_n)$

3. Ceci fera l'objet du cours 2

Ces sources physiques peuvent être par exemple :

- des lancers de pièces, de dés, de roulettes
- du bruit thermique dans les résistances de circuits électroniques
- des capteurs de radiations
- des microsecondes de l'horloge de l'ordinateur
- ...

Ces générateurs possèdent une « bonne » entropie (incertitude) MAIS ils sont :

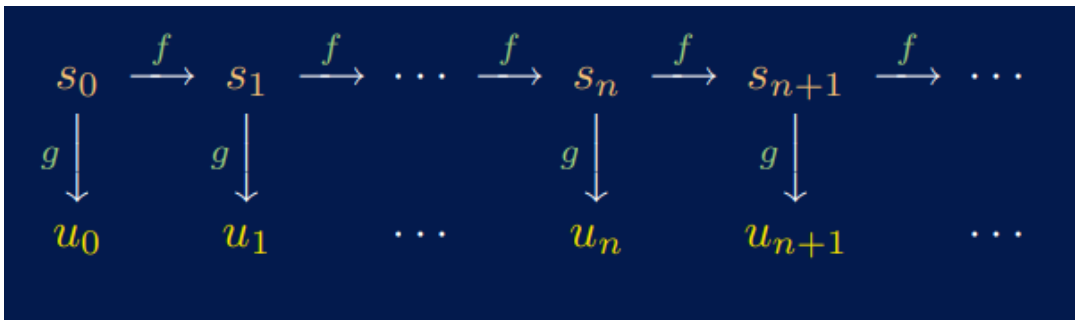
- « encombrants »
- pas toujours fiables
- peu analysables (mathématiquement)
- **non reproductibles** [a]

a. Caractéristique indispensable :

- en simulation stochastique pour pouvoir contrôler, comparer, optimiser les modèles
- en cryptographie pour pouvoir déchiffrer !

Un GPA est caractérisé par un quadruplet (S, f, U, g) :

- S est **l'ensemble** (très grand mais fini) **des états** contenant la graine s_0
- $f : S \to S$ est **la fonction de transition** permettant de calculer $s_n = f(s_{n-1})$
- U est **l'ensemble des valeurs de sortie** (on se limitera à $U = [0,1]$)
- $g : S \to U$ est **la fonction de sortie** permettant d'obtenir $u_n = g(s_n)$



Une fois la graine choisie (au hasard), toutes les sorties supposées mimer des variables aléatoires i.i.d. uniformes sur [0,1] sont entièrement déterminées !
Un GPA est en fait un « amplificateur de hasard »...

Un GCL est défini par :

- $S = \{0, 1, \ldots, m-1\}$
- $f(s) = as + b \mod m$
- $U = [0,1]$ ou plus exactement $U = [0,1[$
- $g(s) = \dfrac{s}{m}$

Deux exemples avec $b \neq 0$:

- la fonction `rand()` du langage C ANSI est définie par :

$$m = 2^{31},\ a = 1103515245 \text{ et } b = 12345$$

- la fonction `drand48()` du langage C ANSI est définie par :

$$m = 2^{48},\ a = 25214903917 \text{ et } b = 11$$

Deux exemples avec $b = 0$:

- le générateur `RANDU` implanté sur les IBM des années 60 est défini par :

$$m = 2^{31},\ a = 65539 \text{ et } b = 0$$

- le générateur du logiciel MAPLE est défini par :

$$m = 10^{12} - 11,\ a = 427419669081 \text{ et } b = 0$$

C'est le cas $b = 0$ que nous allons maintenant généraliser...

Les vecteurs seront notés **en gras**

Un GMR d'ordre k est défini par :

- $S = \{0,1, \ldots, m-1\}^k$
- $f(\mathbf{s}) = f\left(s^{(1)}, \ldots, s^{(k)}\right) = \left(s^{(2)}, \ldots, s^{(k)}, a_1 s^{(1)} + \ldots + a_k s^{(k)} \mod m\right)$
- $U = [0,1]$ ou plus exactement $U = [0,1[$
- $g(\mathbf{s}) = \dfrac{s^{(k)}}{m}$

- Calcul de u_n pour $n \geq k$:

$$u_n = g(\mathbf{s}_n) = \frac{s_n^{(k)}}{m} \quad \text{avec } s_n^{(k)} = a_1 s_{n-1}^{(1)} + \ldots + a_k s_{n-1}^{(k)} \mod m$$

- Un GMR avec $k = 1$ est un GCL où $b = 0$
- Un GMR avec $m = 2$ est un LFSR [a]

a. Ceci fera l'objet du cours 2

Le générateur **MRG32k3** s'obtient en combinant deux GMRs d'ordre 3 :

- $S = S_1 \times S_2$ avec $S_i = \{0, 1, \ldots, m_i - 1\}^3$
- $f(\mathbf{s}) = f\left(\mathbf{s}_1, \mathbf{s}_2\right) = \left(f_1(\mathbf{s}_1), f_2(\mathbf{s}_2)\right)$ avec :

$$f_i(\mathbf{s}_i) = \left(s_i^{(2)}, s_i^{(3)}, a_{i1} s_i^{(1)} + a_{i2} s_i^{(2)} + a_{i3} s_i^{(3)} \mod m_i\right)$$

- $g(\mathbf{s}) = \dfrac{s_1^{(3)} - s_2^{(3)} \mod m_1}{m_1}$
- $a_{11} = 0,\ a_{12} = 1403580,\ a_{13} = -810728$ et $m_1 = 2^{32} - 209$
- $a_{21} = 527612,\ a_{22} = 0,\ a_{23} = -1370589$ et $m_2 = 2^{32} - 22853$

Un générateur digital d'ordre k est défini par :

- $S = \{0,1\}^k = \mathbb{F}_2^k$
- $f(\mathbf{s}) = A\mathbf{s} \mod 2$ où A est une matrice $k \times k$
- $U = [0,1]$ ou plus exactement $U = [0,1[$
- $g(\mathbf{s}) = \sum_{i=1}^{w} \left((B\mathbf{s})_i \mod 2 \right) 2^{-i}$ où B est une matrice $w \times k$

Un LFSR (d'ordre k) est un tel générateur avec :

$$A = \begin{pmatrix} & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ a_1 & a_2 & a_3 & \dots & a_k \end{pmatrix} \text{ et } B = \begin{pmatrix} 1 & & & 0 & \dots & 0 \\ & 1 & & & & \\ & & \ddots & & & \\ & & & 10 & \dots & 0 \end{pmatrix}$$

Un générateur inversif est défini par :

- $S = \{0,1, \ldots, p-1\}$ avec p premier
- $f(s) = as^{p-1} + b \mod p$
- $U = [0,1]$ ou plus exactement $U = [0,1[$
- $g(s) = \frac{s}{p}$

Ce type de générateur ne sera pas étudié dans ce cours

Une suite $(s_n)_{n\in\mathbb{N}}$ de nombres produits par un GPA :

- mime des variables aléatoires i.i.d. uniformes sur $\left\{\frac{0}{K}, \frac{1}{K}, \ldots, \frac{K}{K}\right\}$, et non sur [0,1] à cause de la précision nécessairement finie des nombres réels dans un ordinateur
- est **ultimement périodique** c'est à dire périodique à partir d'un certain rang :

$$\exists T \in \mathbb{N}^*,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ s_{n+T} = s_n$$

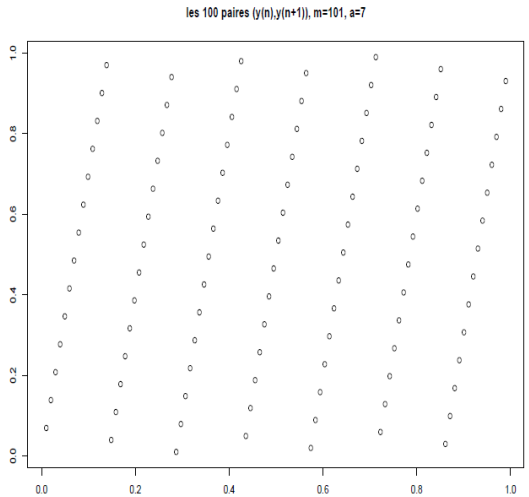
Le plus petit entier T vérifiant cette condition est alors la **période** de la suite.

- Un générateur se doit [a] de posséder une période d'au moins 2^{100}
- La période est sensible à la forme et aux paramètres du GPA
- Choisir un « bon » GPA nécessite une analyse théorique (mathématique)

a. compte tenu de la puissance de calcul des ordinateurs aujourd'hui

- La discrépance évalue les corrélations entre les valeurs successives de la suite
- Elle mesure le caractère serré et homogène du recouvrement
- Elle prend des formes différentes selon le type du générateur
- Pour un GCL, on s'intéresse à la distance entre les hyperplans

Une représentation des hyperplans (droites) en dimension 2 [4]



Comparaison (rapport) avec la distance minimale théorique (en dimension 2, ..., 8)

rand	0,84	0,52	0,63	0,49	0,68	0,43	0,54
drand48	0,51	0,80	0,45	0,58	0,66	0,80	0,60
RANDU	0,93	0,012	0,059	0,16	0,29	0,45	0,62
MAPLE	0,75	0,74	0,65	0,73	0,63	0,56	0,56

4. On évalue alors la corrélation entre deux termes consécutifs

- Contrairement à la discrépance, cette validité est empirique
- Les tests statistiques permettent d'invalider une hypothèse [5] en détectant un biais
- On peut au moins vérifier que les principales caractéristiques ne font pas défaut

5. Après avoir fixé l'erreur de première espèce, disons $\alpha = 5\%$, si l'hypothèse nulle H_0 (celle que l'on cherche à invalider) est vraie, alors la variable de décision du test dont on connait la distribution (exacte ou approchée) n'a que 5% de chance de prendre une valeur dans la zone dite critique ou de rejet (valeurs extrêmes). Si c'est le cas avec l'échantillon d'étude, alors on rejette H_0, mais dans le cas contraire, rien ne garantit sa validité. C'est la puissance du test qui mesure la confiance que l'on peut avoir en notre décision lorsque l'on n'a aucune raison de rejeter H_0...

- Pour tester l'espérance (resp. la variance) qui doit être égale à 1/2 (resp. 1/12), on peut faire un test de conformité de moyenne (resp. de variance)
- Pour tester l'adéquation entre la loi empirique et la loi théorique, on peut faire un test d'ajustement (test du chi 2 ou test de Kolmogorov-Smirnov)
- On peut aussi penser au test de collision et au test d'espacement des anniversaires

Ce cours s'appuie principalement sur cet article