Fiche 1 : Générateurs pseudo-aléatoires

L'objectif de cette fiche est d'introduire les méthodes principales de génération de nombres pseudo-aléatoires.

Exercice 1. Générateurs à base de congruences On considère le générateur suivant :

$$x_{n+1} = ax_n \mod 7.$$

Les nombres x_n forment donc une suite de nombre pseudo-aléatoires. On cherche à évaluer le degré d'aléa de cette suite, c'est-à-dire l'imprévisibilité (à l'observation) du prochain nombre. Lorsque cette suite est cyclique, un critère important d'évaluation du générateur est la longueur du cycle observable.

1. Quelle est la longueur maximale du cycle de ce générateur ?
2. Étudier les suites produites par cet algorithme avec $a = 3$ et $a = 4$.

Exercice 2. Période d'un générateur à base de congruences Nous avons le résultat suivant :

Théorème 1 (Hull-Dobell, (1962)) *Soit la suite (x_n) produite par l'algorithme $x_{n+1} = ax_n + b \mod m$. Alors le cycle maximal est de longueur m si et seulement si les trois hypothèses suivantes sont vérifiés :*

1. *$PGCD(a, m) = 1$, $PGCD(b, m) = 1$;*
2. *si un nombre premier p divise m, alors p divise $a - 1$;*
3. *si 4 divise m, alors 4 divise $a - 1$.*

Question 2.1 : Vérifier les conditions du théorème pour les valeurs :

- $a = 4$, $b = 2$, $m = 9$,
- $a = 2$, $b = 2$, $m = 9$,
- $a = 3$, $b = 3$, $m = 9$,
- $a = 1$, $b = 1$, $m = 9$.

Question 2.2 : Des tirages aléatoires successifs doivent être indépendants. Ce n'est évidemment pas le cas pour les générateurs pseudo-aléatoires. Dans certain cas on peut voir apparaître le "déterminisme" de l'algorithme de façon assez flagrante. Un générateur pseudo-aléatoire doit toujours être utilisé avec "méfiance".

Exemple : soit le générateur

$$x_{n+1} = 11x_n + 1 \mod 71.$$

(a) Montrer que la période de ce générateur est 70.

(b) Tracez l'histogramme de $\{x_1, ..., x_{100}\}$.

(c) Tracez l'ensemble des points de coordonnées (x_{n+1}, x_n).

Vous pourrez comparer avec les graphiques obtenus pour le générateur

$$x_{n+1} = 24298x_n + 99991 \mod 199017$$

Conclusion : les propriétés du hasard sont complexes et difficiles à reproduire. "Faire au hasard" ce n'est pas "faire n'importe quoi". C'est bien dommage... pour une fois qu'on aurait pu se le permettre !

Exemples Voici quelques générateurs "connus"

$$x_{n+1} = 7^5 x_n \mod 2^{31} - 1, \text{ (générateur IBM, utilisé dans macos [}\texttt{rand}\text{ de la libc])}$$
$$x_{n+1} = 427419669081x_n \mod 999999999989, \text{ (générateur Maple, 999999999989 est premier)}$$

dont les périodes respectives sont :

$$2^{30} = 1\,073\,741\,824,$$
$$2^{29} = 536\,870\,912,$$

Exercice 3. Décalage de registre Soit $S = \{1, 0, 1, 1\}$ la séquence binaire (le germe). Pour produire le bit suivant (S_5) de la séquence on applique

$$S_1 \,\text{XOR}\, S_3 \quad \text{ce qui donne } 1 \,\text{XOR}\, 1 = 0.$$

Ensuite, on décale et on recommence. On peut décrire cette récurrence par

$$S_{n+1} = S_{n-1} \,\text{XOR}\, S_{n-3}.$$

Question 3.1 : Trouver les 5 prochains bit de la séquence. Quelle est la suite obtenue ? Est-elle bien "aléatoire" ?
Question 3.2 : Étudier les séquences de ce même générateur avec les germes $S = \{1, 0, 1, 0\}$ et $S = \{1, 0, 0, 1\}$.

Question 3.3 : On considère maintenant l'algorithme

$$S_{n+1} = S_{n-2} \,\text{XOR}\, S_{n-3}.$$

Étudier la séquence produite par cet algorithme. Trouver la longueur du cycle de ce générateur. Quel est le comportement de ce générateur sur les autres germes ?

Question 3.4 : Quelle est la longueur maximale du cycle avec un registre à 4 bits ? Quelle est la longueur minimale ? Quelle est la longueur maximale [1] du cycle avec pour un registre de 64 bits ?

Exercice 4. Génération des mots de k bits Proposer des algorithmes de génération des mots de 3 bits. Appliquer aux séquences de l'exercice précèdent.

Exercice 5. Problème : Comment faire une bonne pièce avec une fausse On dispose de pièces de monnaie biaisées, c'est à dire que la fréquence d'apparition de piles ou de faces ne sont pas égales à $\frac{1}{2}$.

On modélise les tirages de ces pièces par des variables aléatoires indépendantes X_i à valeur dans $\{0, 1\}$ et on note $p_i = \mathbb{P}(X_i = 1) = \mathbb{P}(\text{ la pièce } i \text{ tombe sur pile })$.
Question 5.1 : Calculer en fonction de p_1 et p_2 les probabilités :

$$\mathbb{P}((X_1, X_2) = (0, 0)), \qquad \mathbb{P}((X_1, X_2) = (0, 1)), \qquad \mathbb{P}((X_1, X_2) = (1, 0)), \qquad \mathbb{P}((X_1, X_2) = (1, 1)).$$

On note Y_2 la variable aléatoire à valeur dans $\{0, 1\}$ définie par $Y_2 = (X_1 + X_2) \mod 2$
Question 5.2 : Calculer $\pi_2 = \mathbb{P}(Y_2 = 1)$

On suppose maintenant que $p_1 = p_2 = p$.
Question 5.3 : Montrer que $\pi_2 - \frac{1}{2} = \left(p - \frac{1}{2}\right)(1 - 2p)$.
Question 5.4 : Ranger par ordre croissant les 5 nombres $p, 1-p, \pi_2, 1-\pi_2, \frac{1}{2}$. On pourra supposer que $p < \frac{1}{2}$.
Question 5.5 : En déduire de X_1 ou de Y_2 quelle serait la meilleure simulation d'une pièce non biaisée ? Justifier votre réponse.

On pose alors $Y_3 = (X_1 + X_2 + X_3) \mod 2$.
Question 5.6 : Montrer que $Y_3 = (Y_2 + X_3) \mod 2$.
Question 5.7 : Calculer, pour $p_1 = p_2 = p_3 = p$, $\pi_3 = \mathbb{P}(Y_3 = 1)$
Question 5.8 : Exprimer $\left|\pi_3 - \frac{1}{2}\right|$ en fonction de $\left|\pi_2 - \frac{1}{2}\right|$ puis de $\left|p - \frac{1}{2}\right|$. On généralise maintenant le procédé en définissant $Y_n = (X_1 + X_2 + \cdots + X_n) \mod 2$.
Question 5.9 : Montrer que $Y_{n+1} = (Y_n + X_{n+1}) \mod 2$.
Question 5.10 : On suppose que $p_1 = p_2 = \cdots = p_n = p$. Exprimer $\pi_n = \mathbb{P}(Y_n = 1)$ en fonction de π_{n-1} et p.
Question 5.11 : Calculer dans ce cas, $\pi_n - \frac{1}{2}$ en fonction de $p - \frac{1}{2}$.
Question 5.12 : Application : on suppose $p = 0.4$. Pour quelle valeur de n aura-t-on $\left|\pi_n - \frac{1}{2}\right| < 10^{-6}$? Commenter votre résultat.

1. Note historique :

Tausworthe (1965) à étudié les propriétés de l'algorithme suivant :
à partir d'un mot binaire initial (le germe) $x^0 = (x_{-m+1}, \ldots x_{-1}, x_0)$, on produit les éléments de la suite pseudo-aléatoire par récurrence :

$$x_{n+1} = a_1 x_{n-m+1} + a_2 x_{n-m+1} + \ldots + a_m x_n \mod 2$$

Dans l'exemple 1) de l'exercice précédent $m = 4$, $x_{n+1} = x_{n-3} + x_{n-1} \mod 2$ et dans l'exemple 3) $x_{n+1} = x_{n-3} + x_{n-2} \mod 2$.

Il a établi la condition sur les coefficients a_i sous laquelle ce générateur atteint la période maximale $2^m - 1$.