

Cours MO102 - Utilisation de fonctions aléatoires

En informatique, il est courant de vouloir modéliser et simuler des phénomènes *aléatoires*. Or, l'obtention de données aléatoires est réalisée sur un ordinateur à partir d'algorithmes de nature *déterministe*, ce qui peut sembler contradictoire. . .

Sans trop rentrer dans les détails mathématiques, on va essayer dans les exercices qui vont suivre de se familiariser avec les outils de génération de variables pseudo-aléatoires disponibles en Matlab et vérifier leurs propriétés. Pour plus de précision sur la théorie des probabilités et sur les statistiques, on pourra se référer au cours de probabilité de première année de l'ENSTA, disponible sur le site de l'auteur (cermics.enpc.fr/~delmas/Enseig/ensta_cours.pdf).

1 Introduction

Les notions de probabilité et de variable aléatoire sont en mathématiques des concepts très précis. On en présente ici une version intuitive. Pour cela, on se donne un ensemble Ω (ensemble des *états*) et une mesure \mathbb{P} (qui associe à certains sous-ensembles de Ω leur *probabilité*). Une *variable aléatoire réelle* X est alors une fonction définie sur Ω à valeurs dans \mathbb{R} , ayant des propriétés supplémentaires de mesurabilité (que nous ne détaillerons pas). Une *réalisation* x de la variable aléatoire X est la valeur $X(\omega)$ prise par la variable aléatoire X en un état de la nature ω .

Dire que la variable aléatoire réelle X a une densité de probabilité signifie qu'il existe une fonction f_X qui fournit la probabilité que X prenne ses valeurs dans un intervalle quelconque $[a, b]$:

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx .$$

À titre d'exemple, on dit que la variable aléatoire X suit une loi *uniforme* sur le segment $[0, 1]$ (loi notée $\mathcal{U}([0, 1])$) si X est à valeurs dans $[0, 1]$ et si sa densité est identiquement égale à 1. On dit que X est une variable aléatoire *gaussienne centrée réduite* (loi notée $\mathcal{N}(0, 1)$: on parle aussi de loi normale) si X est à valeurs dans $]-\infty, +\infty[$, de densité $f_X(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$.

On peut alors facilement définir les moments de la variable aléatoire X . Le premier moment est appelé *espérance* (on dit aussi moyenne), et représente la "valeur moyenne" de la variable aléatoire :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx ;$$

le second moment, appelé *variance*, représente l'écart quadratique moyen de X à son espérance :

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f_X(x) dx .$$

L'intérêt des moments d'une variable aléatoire est qu'ils permettent de représenter cette variable. Ainsi, la loi d'une variable aléatoire gaussienne est *caractérisée* par ses deux premiers moments.

Enfin, une notion importante en probabilité est celle d'*indépendance* entre variables aléatoires. Intuitivement, deux variables aléatoires X et Y (éventuellement de même loi) sont indépendantes si la loi de l'une n'est pas influencée par la loi de l'autre.

2 Génération de nombres pseudo-aléatoires

En fait, on ne sait générer sur un ordinateur que des variables *pseudo-aléatoires*, c'est-à-dire des variables dont les propriétés statistiques s'approchent des propriétés statistiques des variables aléatoires que l'on cherche à représenter. L'outil de base consiste à savoir générer une suite de nombres pseudo-aléatoires $\{x_n\}_{n \in \mathbb{N}}$ simulant le tirage de nombres suivant la loi uniforme sur $[0,1]$. Une méthode pour engendrer une telle suite $\{x_n\}_{n \in \mathbb{N}}$ est de définir une suite d'entiers $\{y_n\}_{n \in \mathbb{N}}$ par la relation de récurrence :

$$y_{n+1} = ay_n + b \pmod{c},$$

où a , b et c sont des entiers bien choisis, et de poser $x_n = y_n/c$. Le caractère aléatoire vient de ce que les x_n ainsi engendrés vérifient les propriétés suivantes :

- (i) x_n est une réalisation d'une variable aléatoire de loi uniforme sur $[0, 1]$,
- (ii) les x_n successifs sont des réalisations de variables aléatoires indépendantes.

On sait ensuite générer d'autres lois de probabilité à partir de la loi uniforme (voir §3.2).

3 Représentation de l'aléatoire en Matlab

À partir des principes exposés au paragraphe précédent, Matlab dispose en standard de deux générateurs de nombres aléatoires :

- `rand` pour la loi uniforme $\mathcal{U}([0, 1])$,
- `randn` pour la loi gaussienne centrée réduite $\mathcal{N}(0, 1)$.

Il faut savoir que ces générateurs sont par défaut toujours initialisés de la même façon, si bien que l'exécution de `rand(1, 10)` au démarrage de Matlab renverra toujours le même vecteur de dix valeurs. C'est pourquoi il est recommandé d'initialiser le générateur à l'aide de la commande

```
rand('state', sum(100*clock))
```

afin que l'initialisation de l'algorithme dépende de l'heure à laquelle elle est réalisée.

3.1 Loi des grands nombres

Écrire un script permettant de générer une séquence aléatoire (x_1, \dots, x_N) de taille N suivant la loi normale $\mathcal{N}(0, 1)$ et calculer ses moyenne et variance empiriques m_N et σ_N^2 :

$$m_N = \frac{1}{N} \sum_{n=1}^N x_n,$$

$$\sigma_N^2 = \frac{1}{N} \sum_{n=1}^N (x_n - m_N)^2.$$

On vérifiera que lorsque N tend vers $+\infty$, les valeurs empiriques de la moyenne et de la variance tendent vers les celles du phénomène que l'on simule (*loi des grands nombres*). On tracera les courbes m_N et σ_N^2 en fonction de N .

3.2 Génération de lois de probabilité

On cherche maintenant à générer une séquence aléatoire suivant une loi de probabilité donnée en n'utilisant que le générateur uniforme `rand`.

3.2.1 Méthode de l'inverse

Définition 1 Soit X une variable aléatoire réelle et soit F_X sa fonction de répartition :

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(x) dx .$$

On appelle inverse généralisée de F_X la fonction F_X^{-1} définie par :

$$F_X^{-1}(p) = \inf_{x \in \mathbb{R}} \{x \mid F_X(x) \geq p\} .$$

Dans les “bons cas”, la fonction F_X est continue strictement croissante et son inverse généralisée correspond alors à son inverse au sens usuel. La méthode de l'inverse permettant de simuler une variable aléatoire de fonction de répartition donnée est basée sur le théorème suivant.

Théorème 1 Soit X une variable aléatoire réelle de fonction de répartition F_X , et soit U une variable aléatoire de densité $\mathcal{U}([0, 1])$. Alors, la variable aléatoire $F_X^{-1}(U)$ suit la même loi que X .

On va appliquer ce théorème pour simuler une variable aléatoire réelle X suivant une loi de probabilité *exponentielle* de paramètre $\lambda > 0$ (notée \mathcal{E}_λ), c'est-à-dire une variable aléatoire à valeurs dans $[0, +\infty[$ dont la densité est : $f_X(x) = \lambda \exp(-\lambda x)$.¹

Écrire un script permettant de générer une séquence aléatoire (x_1, \dots, x_N) de taille N suivant la loi exponentielle \mathcal{E}_λ . Calculer et tracer les moyennes empiriques successives en fonction de N et vérifiez expérimentalement la loi des grands nombres.

3.2.2 Partie optionnelle : simulation d'une loi gaussienne

La fonction inverse F_X^{-1} n'étant pas connue analytiquement dans le cas de la distribution gaussienne, on propose une méthode différente pour la simuler à partir de lois uniformes.

Théorème 2 Une variable aléatoire réelle X suit la loi gaussienne centrée réduite $\mathcal{N}(0, 1)$ si et seulement si on a $X = R \cos(\Theta)$, avec R et Θ deux variables aléatoires indépendantes telles que R^2 suive une loi exponentielle $\mathcal{E}_{\frac{1}{2}}$ et Θ une loi uniforme $\mathcal{U}([0, 2\pi])$

Il découle de ce théorème et du précédent que si U et V sont deux variables aléatoires indépendantes suivant chacune la loi uniforme $\mathcal{U}([0, 1])$, alors $X = \sqrt{-2 \log(U)} \cos(2\pi V)$ est une variable aléatoire suivant la loi normale $\mathcal{N}(0, 1)$.

Écrire un script permettant de générer une séquence aléatoire (x_1, \dots, x_N) de taille N suivant une loi gaussienne $\mathcal{N}(m, \sigma^2)$ de moyenne m et de variance $\sigma^2 > 0$ données, en utilisant uniquement le générateur uniforme `rand`.

On tracera l'histogramme associé et on comparera les caractéristiques de ce générateur à celles du générateur `randn` de Matlab.

1. La fonction de répartition associée est : $F_X(x) = 1 - \exp(-\lambda x)$, et son inverse est : $F_X^{-1}(p) = -\frac{1}{\lambda} \log(1 - p)$.