Chapitre 2 : Simulation de lois non uniformes, par des méthodes générales

Université Hassiba Benbouali de Chlef

Plan du Cours

Ce chapitre traite de la simulation de variables aléatoires par des méthodes générales ou particulières.

- Simulation des lois discrètes
- Simulation des lois continues
- Méthodes particulières

Supposons que nous voulons simuler une variable aléatoire discrète prenant les valeurs $\{x_1, x_2, \ldots, x_k, \ldots\}$ avec les probabilités $\{p_1, p_2, \ldots, p_k, \ldots\}$ respectivement. Notons que

$$\sum_{i=1}^{\infty} p_i = 1$$

Si U est de loi uniforme $\mathcal{U}[0,1]$ alors

$$X = \sum_{k=1}^{\infty} x_k \mathbb{1}_{\left\{\sum_{i=1}^{k-1} p_i < U \leq \sum_{i=1}^{k} p_i\right\}}$$

est la variable discrète que nous désirons simuler.

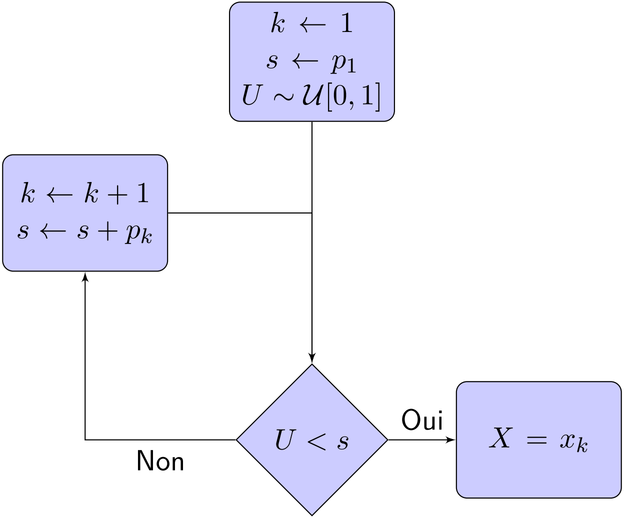
Remarque

La somme $\sum_{i=1}^{0} p_i$ est posée égale à zéro.

Preuve

$$\begin{aligned}
\mathbb{P}\left[X = x_k\right] &= \mathbb{P}\left[\sum_{i=1}^{k-1} p_i < U \leq \sum_{i=1}^{k} p_i\right] \\
&= \mathbb{P}\left[U \leq \sum_{i=1}^{k} p_i\right] - \mathbb{P}\left[U \leq \sum_{i=1}^{k-1} p_i\right] \\
&= \sum_{i=1}^{k} p_i - \sum_{i=1}^{k-1} p_i \\
&= p_k.
\end{aligned}$$

Algorithme


$k \leftarrow 1$
$s \leftarrow p_1$
$U \sim \mathcal{U}[0,1]$
$k \leftarrow k+1$
$s \leftarrow s + p_k$
$U < s$
Oui
Non
$X = x_k$

Loi de Bernoulli

On veut simuler $X \sim \mathcal{B}(p)$:

- On sait que si $U \sim \mathcal{U}[0,1]$, alors $\mathbb{1}_{\{U \leq P\}}$ suit la loi $\mathcal{B}(p)$.
- Donc $X = \mathbb{1}_{\{U \leq p\}}$. On simule u de la loi $\mathcal{U}[0,1]$:
  - Si $u \leq p$, alors $x = 1$
  - Si $u > p$, alors $x = 0$

Loi Binomiale

On veut simuler $X \sim \mathcal{B}(n, p)$.

- On sait qu'une variable binomiale représente la somme de n variables indépendantes de Bernoulli de paramètre p
- Il suffit donc de simuler n variables aléatoires indépendantes de loi $\mathcal{B}(p)$ et d'en faire la somme :

$$X = \sum_{i=1}^{n} \mathbb{1}_{\{U_i \leq p\}}$$

  - On simule $u_1, \ldots, u_n$ de la loi $\mathcal{U}[0, 1]$:
  - On détermine le nombre k de $u_i \leq p$
  - On pose $x = k$

Loi de Poisson

On veut simuler $X \sim \mathcal{P}(\lambda)$. On sait que

$$\begin{aligned} p_k &= \mathbb{P}(X = k) = \frac{\lambda^k}{k} \mathrm{e}^{-\lambda},\ k \in \mathbb{N} \\ p_{k+1} &= \mathbb{P}(X = k+1) = \frac{\lambda^{k+1}}{(k+1)} \mathrm{e}^{-\lambda} = \frac{\lambda}{(k+1)} \frac{\lambda^k}{k} \mathrm{e}^{-\lambda} = \frac{\lambda}{(k+1)} p_k \end{aligned}$$

On note $P_k = \sum_{j=0}^{k} p_j$, on remarque $P_{k+1} = P_k + \frac{\lambda}{k+1} p_k$
On simule donc une variable aléatoire de Poisson de paramètre λ en prenant

$$X = \sum_{k \geq 0} k \mathbb{1}_{\{P_{k-1} \leq u < P_k\}} \text{ avec la convention } P_{-1} = 0.$$

Supposons que la fonction de répartition $\mathbb{F} : \mathbb{R} \to [0,1]$ de la distribution que l'on veut simuler est continue et strictement croissante. Alors $\mathbb{F}$ admet un inverse que l'on note $\mathbb{F}^{-1}$.

Transformation de v.a. uniforme

Théorème 2.1

Soient X une variable aléatoire et $\mathbb{F}_X$ sa fonction de répartition continue et strictement croissante.

1. *La v.a. $U = \mathbb{F}_X(X)$ suit la loi $\mathcal{U}[0,1]$.*
2. *Si U est une variable aléatoire de loi $\mathcal{U}[0,1]$ alors $Y = \mathbb{F}_X^{-1}(U)$ a la même loi que X (ou admet pour fonction de répartition $\mathbb{F}_X$).*

Preuve

1. Comme $\mathbb{F}_X \;:\; \mathbb{R} \to [0,1]$ alors
$$\mathbb{P}\left[\mathbb{F}_X \leq u\right] = 0 \text{ pour tout } u < 0$$
et
$$\mathbb{P}\left[\mathbb{F}_X \geq u\right] = 1 \text{ pour tout } u > 1$$

   Pour $0 < u < 1$,
$$\mathbb{P}\left[\mathbb{F}_X(X) \leq u\right] = \mathbb{P}\left[X \leq \mathbb{F}_X^{-1}(u)\right] = \mathbb{F}_X\left(\mathbb{F}_X^{-1}(u)\right) = u.$$

2. La fonction de répartition $\mathbb{F}_Y$ de Y satisfait, pour tout $y \in \mathbb{R}$,
$$\begin{aligned}\mathbb{F}_Y(y) = \mathbb{P}\left[Y \leq y\right] &= \mathbb{P}\left[\mathbb{F}_X^{-1}(U) \leq y\right] \\ &= \mathbb{P}\left[U \leq \mathbb{F}_X(y)\right] \\ &= \mathbb{F}_X(y).\end{aligned}$$

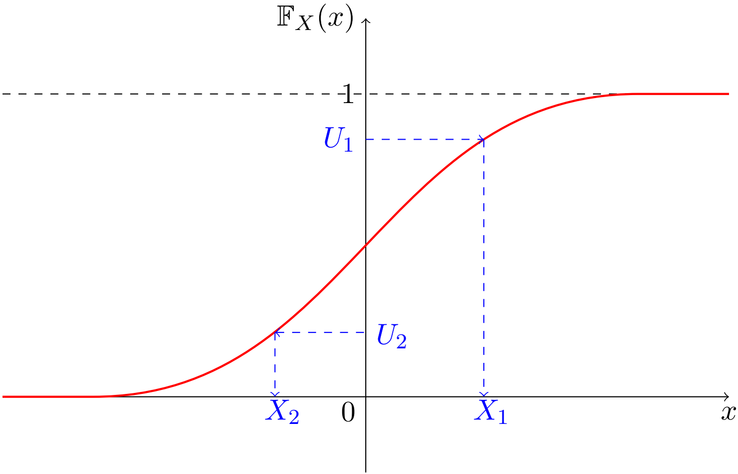


FIGURE 1 – Illustration graphique.

Loi Exponentielle

On veut simuler $X \sim \mathcal{E}(\lambda)$. On sait que :

- $\mathbb{F}_X(x) = 1 - \mathrm{e}^{-\lambda x}$.
- On résout l'équation $U = \mathbb{F}_X(X)$:

$$\begin{aligned} U = \mathbb{F}(X) \quad &\Leftrightarrow \quad U = 1 - \mathrm{e}^{-\lambda X} \\ &\Leftrightarrow \quad 1 - U = \mathrm{e}^{-\lambda X} \\ &\Leftrightarrow \quad X = -\frac{1}{\lambda} \ln(1 - U) \end{aligned}$$

Comme U et $(1-U)$ suivent la même loi (loi uniforme sur $[0,1]$). Donc :

$$X = -\frac{1}{\lambda} \ln U.$$

Remarque

L'hypothèse de la connaissance de $\mathbb{F}_X^{-1}$ n'a de sens que si $\mathbb{F}_X$ est strictement croissante. Cependant, même dans ce cas, il se peut que $\mathbb{F}_X^{-1}$ n'ait pas d'expression analytique simple, c'est le cas par exemple pour la loi normale.

Méthode de Box et Müller

1. Simuler U et V indépendants et de loi $\mathcal{U}[0,1]$.
2. On pose
$$\left\{\begin{array}{l} X = \sqrt{-2\ln U}\cos(2\pi V) \\ Y = \sqrt{-2\ln U}\sin(2\pi V) \end{array}\right.$$
alors X et Y sont des variables aléatoires indépendantes et de loi $\mathcal{N}(0,1)$.

Méthode de Marsaglia-Bray

1. Simuler U_1 et U_2 indépendants et de loi $\mathcal{U}[0,1]$.
2. On pose $V_1 = 2U_1 - 1$ et $V_2 = 2U_2 - 1$.
3. On accepte V_1 et V_2 si $W = V_1^2 + V_2^2 < 1$ et on rejette les points si $W > 1$.
4. On pose
$$\begin{cases} Z_1 = V_1\sqrt{-2\dfrac{\ln W}{W}} \\ Z_2 = V_2\sqrt{-2\dfrac{\ln W}{W}} \end{cases}$$
alors Z_1 et Z_2 sont des variables aléatoires indépendantes et de loi $\mathcal{N}(0,1)$.

Remarque

De la simulation de la loi $\mathcal{N}(0,1)$, on déduit celle de la loi normale d'espérance m et de variance σ^2 quelconques par une transformation affine. Si X suit la loi $\mathcal{N}(0,1)$, alors $Y = m + \sigma X$ suit la loi $\mathcal{N}(m, \sigma^2)$.

La méthode du rejet

Dans le cas où la densité f a support contenu dans l' intervalle $[a, b]$ et dont la valeur maximale est égale à M ie :

$$\forall x \in [a, b] \;\; f(x) \leq M$$

Il est possible de représenter la surface entre le graphe de la densité et l'axe des X à l'intérieur d'un rectangle R : $[a, b] \times [0, M]$ de surface $M(b-a)$.
On Choisit au hasard un point dans ce rectangle et on détermine s'il est situé au-dessus ou au-dessous de la courbe $y = f(x)$. Cette idée est la clé de la méthode de rejet.

Algorithme

1. Considérons le U_1 et U_2 deux variables aléatoires indépendantes et uniformes sur $[0, 1]$.
   On choisit un point $A\,(x, y)$ uniformément distribué sur le rectangle R telle que : $\begin{cases} X & = (b-a)\,U_1 + a \\ Y & = MU_2 \end{cases}$
2. Si $Y \leq f\,(X)$ on garde X, sinon on choisit un autre point.
3. On réitère jusqu'à ce que la condition $Y \leq f\,(X)$ soit vérifiée.

Simulation de la loi Beta

La densité de la loi Beta est :

$$f_X(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1} \mathbb{1}_{[0,1]}(x)$$

La fonction Beta est de la forme :

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

On applique la méthode de rejet dans ce cas sur la loi de $Beta(3,5)$:

$$f_X(x) = 105x^2(1-x)^4 \mathbb{1}_{[0,1]}(x)$$

Simulation de la loi Beta (suite)

Trouvons l'abscisse du point où f est maximale :

$$\begin{array}{rcl} f_X(x) & = & 105x^2(1-x)^4 \mathbb{1}_{[0,1]}(x) \\ f'_X(x) & = & 105x(1-x)^3(2-6x) \\ f'_X(x)=0 & \iff & x=0 \vee \ x=1 \vee \ x=\dfrac{1}{3} \end{array}$$

D'où l'abscisse du point ou f maximal est $x=\dfrac{1}{3}$. La valeur à l'optimum M :

$$M = f_X(x) = f\left(\frac{1}{3}\right) = 2.305$$

Cas d'une densité quelconque

Le principe peut se généraliser, dans ce cas en remplaçant le rectangle par une surface délimitée par le graphe d'une fonction non négative opportunément choisie. Il est préférable d'utiliser la densité $g(x)$ facile à simuler de forme semblable à $f(x)$ définie sur le même intervalle, on définira la courbe enveloppante par $Mg\left(x\right)$ telle que :

$$\forall x \ f\left(x\right) \leq Mg\left(x\right)$$

Cette valeur de M sera aussi la plus petite possible telle que :

$$M = \sup \frac{f\left(x\right)}{g\left(x\right)},\ M > 1 \left(\text{car } 1 = \int f(x) \ \mathrm{d}x \leq \int Mg\left(x\right) \ \mathrm{d}x = M \right)$$

Algorithme

1. On simule Y selon g, soit Y_1 la valeur obtenue .
2. On simule U selon $\mathcal{U}[0,1]$ d'une manière indépendante de Y. Soit U_1 la valeur trouvée.
3. Si $U_1 \leq \dfrac{f(Y_1)}{Mg(Y_1)}$ on accepte Y_1 comme une valeur simulée de Y selon f, sinon on le rejette et on recommence à l'étape 1.

Simulation de la loi gamma

La loi gamma de paramètre $a > 0$ est définie par sa densité

$$f(t) = \frac{1}{\Gamma(a)} t^{a-1} \exp(-t) \mathbb{1}_{]0,+\infty[}(t)$$

Le cas $a = 1$ est celui de la loi exponentielle ie :
$f(t) = \exp(-t) \mathbb{1}_{]0,+\infty[}(t)$.
On vérifie par un simple calcule de maximum que pour tout $t > 0$:

$$\frac{f(t)}{g(t)} = \frac{1}{\Gamma(a)} t^{a-1} \exp t(\lambda - 1)$$

On obtient :

$$t = \frac{1-a}{\lambda - 1}$$

Simulation de la loi gamma (suite)

Donc :

$$\begin{aligned}\frac{f(t)}{g(t)} &= \frac{1}{\Gamma(a)}\left(\frac{1-a}{\lambda-1}\right)^{a-1}\exp\left(\frac{1-a}{\lambda-1}\right)(\lambda-1)\\ &= \frac{1}{\Gamma(a)}\left(\frac{1-a}{\lambda-1}\right)^{a-1}\exp(1-a)\end{aligned}$$

Alors :

$$M = \frac{1}{\Gamma(a)}\left(\frac{1-a}{\lambda-1}\right)^{a-1}\exp(1-a)$$