Cours n° 2 : Simulation de variables et vecteurs aléatoires

1 Variables discrètes

Dans le cas d'une variable aléatoire discrète, la méthode est canonique. Soit X une variable aléatoire à valeurs dans $\{x_1, \ldots, x_r\}$ de loi $(p_1, \ldots, p_r)$, avec $\mathbb{P}\{X = x_j\} = p_j$. Montrer que si U est une variable uniforme sur $[0,1]$ alors la variable aléatoire

$$\sum_{j=1}^{r} x_j \, \mathbf{1}_{]p_0+\ldots+p_{j-1},p_1+\ldots+p_j]}(U)$$

(où l'on note $p_0 = 0$) suit la même loi que X. Donc, pour simuler un échantillon $(y_1, \ldots, y_k)$ d'une telle variable X, on simule un échantillon $(u_1, \ldots, u_k)$ de variables de loi uniforme sur $[0,1]$ et on pose $y_i = x_j$ si $u_i \in]p_0 + \ldots + p_{j-1}, p_1 + \ldots + p_j]$.

2 Simulation par inversion

La méthode de simulation par inversion repose sur le résultat suivant.

Proposition 1. *Soit X une variable aléatoire réelle de fonction de répartition $F_X(t) = \mathbb{P}\{X \leq t\}$. On définit l'inverse généralisée F_X^{-1} de F_X sur $]0,1[$ par*

$$F_X^{-1}(x) = \inf\{t \in \mathbb{R} \,|\, F_X(t) \geq x\}.$$

Alors, si U est une variable aléatoire de loi uniforme sur $]0,1[$, $F_X^{-1}(U)$ a même loi que X.

Ainsi, si on tire n nombres au hasard uniformément répartis entre 0 et 1, $(u_1, u_2, \ldots, u_n)$, l'échantillon recherché, $(x_1, x_2, \ldots, x_n)$, de loi celle de X, sera déterminé par $x_i = F_X^{-1}(u_i)$.

3 Algorithme de Box-Muller

Soient U et V deux variables aléatoires indépen-dantes de loi uniforme sur $[0,1]$. Les variables X et Y définies par

$$\begin{aligned} X &= \sqrt{-2\ln(U)}\cos(2\pi V)\ , \\ Y &= \sqrt{-2\ln(U)}\sin(2\pi V) \end{aligned}$$

sont indépendantes et de loi $\mathcal{N}(0,1)$.

4 Simulation par méthode du rejet

La méthode d'inversion nécessitait la connaissance explicite de F_X^{-1}, ce qui n'est pas toujours le cas (notamment pour les lois normales). Une méthode alternative est la méthode de rejet. Celle-ci sert d'abord à simuler des lois conditionelles ou des loi uniformes sur des domaines de $\mathbb{R}^d$ en utilisant le résultat suivant.

Proposition 2. *Soit (Y_n) une suite de variables aléatoires indépendantes de loi μ et B un borélien tel que $\mu(B) > 0$. Soit T le plus petit entier $n \geq 1$ tel que $Y_n \in B$. Alors*

(i) T est une variable aléatoire de loi géométrique de paramètre $\mu(B)$,

(ii) Y_T est une variable aléatoire ayant pour loi la loi conditionnelle $\mu(.|B)$. En particulier lorsque μ est la loi uniforme sur un borélien C contenant B et tel que $0 < \lambda(B) < \lambda(C)$, cette loi conditionnelle est la loi uniforme sur B.

On a noté λ la mesure de Lebesgue. En pratique on peut donc simuler une variable X de loi uniforme sur B de la façon suivante : on tire Y_1 uniforme sur C, si $Y_1 \in B$, on pose $X = Y_1$, sinon on retire une variable Y_2 uniforme sur C et on recommence.

Cette méthode permet également de simuler des lois à densités en considérant le domaine sous le graphe de la densité. La clé est le "théorème" suivant :

Théorème fondamental de la simulation. Soit f une densité sur $\mathbb{R}^d$. Alors simuler X de densité f est équivalent à simuler (X, U) de loi uniforme (i.e., Lebesgue) sur

$$\{(x, u) : (x, u) \in \mathbb{R}^{d+1},\ 0 < u < f(x)\}.$$

En effet, la loi marginale du couple (X, U) admet bien la densité f, car par le théorème de Fubini–Tonelli, pour tout g mesurable positive,

$$\mathbb{E}[g(X')] = \int_{\mathbb{R}^{d+1}} g(x)\mathbf{1}_{0<u<f(x)}\, dxdu = \int_{\mathbb{R}^d} \left(\int \mathbf{1}_{0<u<f(x)}\, du \right) g(x)\, dx = \int_{\mathbb{R}^d} f(x)g(x)\, dx.$$

Inversement, conditionnellement à $X = x$, la variable aléatoire U suit la loi uniforme dans l'intervalle $[0, f(x)]$ qui peut facilement être simulée.

Méthode du rejet comparatif. On souhaite simuler X de densité f sur $\mathbb{R}^d$. On suppose qu'il existe une densité de probabilité g telle que

— on sait simuler une variable de densité g,

— il existe une constante K ($K \geq 1$!) telle que $f \leq Kg$.

On utilise alors la procédure suivante, qui termine presque sûrement et simule une variable X de densité f :

1. On simule une variable U de densité g et une variable W indépendante de X de loi uniforme sur $[0, 1]$. On pose $V = K\,W\,g(U)$.
2. Si $V \leq f(U)$ on pose $X = U$, sinon on retourne en (1).