Probabilités V

STEP, MINES ParisTech*

6 janvier 2025 (#47ea1e5)

Table des matières

*Ce document est un des produits du projet paulinebernard/CDIS issu de la collaboration de (P)auline Bernard (CAS) et (T)homas Romary (GEOSCIENCES). Il dérive du projet boisgera/CDIS, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS).

Objectifs d'apprentissage

Cette section s'efforce d'expliciter et de hiérarchiser les acquis d'apprentissages associés au chapitre. Ces objectifs sont organisés en paliers :

(○) Prérequis (•) Fondamental (••) Standard (•••) Avancé (••••) Expert

Sauf mention particulière, la connaissance des démonstrations du document n'est pas exigible [1]

Intégrale de Monte Carlo

- • connaître le principe de l'intégration par la méthode Monte Carlo
- • savoir que cette approche se justifie par la loi des grands nombres
- • savoir que l'approximation fournie par le TCL offre un contrôle de l'erreur

Génération de nombres pseudo-aléatoires

- • connaître le principe de la génération de nombres pseudo-aléatoires par la méthode des congruences

Méthodes de simulation de v.a.

- • connaître et savoir implémenter en python la méthode d'inversion
- • connaître et savoir implémenter en python la méthode de rejet
- • connaître la méthode de Box-Muller

1. l'étude des démonstrations du cours peut toutefois contribuer à votre apprentissage, au même titre que la résolution d'exercices.

— • connaître et savoir implémenter la simulation de vecteurs gaussien par la méthode de Cholesky

Echantillonnage d'importance

— • connaître la définition de la méthode d'échantillonnage d'importance

— • savoir qu'un bon choix de densité instrumentale permet de minimiser la variance d'estimation

Intégrale de Monte-Carlo

Les méthodes de Monte-Carlo ont été développées initialement par des physiciens dans les années 1950 (notamment par les travaux de Metropolis, Ulam, Hastings, Rosenbluth) pour calculer des intégrales (déterministes) à partir de méthodes probabilistes numériquement assez économiques. Le nom a été donné en référence au célèbre casino du fait du caractère aléatoire de ces méthodes.

Les méthodes de simulation sont basées sur la production de nombres aléatoires, distribués selon une certaine loi de probabilité. Dans de nombreuses applications, pour une certaine fonction h, on souhaite calculer, pour une variable aléatoire X de loi $\mathbb{P}_X$

$$\mathcal{I} = \mathbb{E}\left(h(X)\right) = \int h(x)\, \mathbb{P}_X(dx).$$

En général, même si on sait évaluer h en tout point, on ne peut pas calculer formellement l'intégrale $\mathcal{I}$. Le calcul d'intégrale par la méthode Monte-Carlo consiste dans sa version la plus simple à générer un *échantillon* $(X_1, \ldots, X_n) \sim_{i.i.d.} \mathbb{P}_X$, et à estimer $\mathcal{I}$ par la moyenne empirique

$$M_n(h) = \frac{1}{n}\sum_{i=1}^{n} h(X_i),$$

où i.i.d signifie indépendant et identiquement distribué. En effet, d'après la loi forte des grands nombres, si $h(x)$ est $\mathbb{P}_X$-intégrable, on a l'assurance que quand $n \to +\infty$,

$$M_n(h) \to \int h(x)\mathbb{P}_X(dx) \text{ p.s.}$$

Si de plus $h(X)^2$ est intégrable, la vitesse de convergence de $M_n(h)$ peut être évaluée, puisque la variance

$$\mathbb{V}(M_n(h)) = \frac{1}{n}\int \left(h(x) - \mathcal{I}\right)^2 \mathbb{P}_X(dx)$$

peut également être estimée à partir de l'échantillon $(X_1, \ldots, X_n)$ par la quantité

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} \left(h(X_i) - M_n(h)\right)^2 .$$

Le théorème central limite assure alors que pour n grand,

$$\frac{M_n(h) - \mathcal{I}}{\sigma_n}$$

suit approximativement une loi $\mathcal{N}(0,1)$ [2]. Cette propriété conduit à la construc-

2. ce résultat sera démontré dans le cours de science des données au second semestre.

tion de tests de convergence et de bornes de confiance asymptotiques pour $M_n(h)$. Par exemple, on aura

$$\mathbb{P}\left(\mathcal{I} \in [M_n(h) - 1.96\sigma_n, M_n(h) + 1.96\sigma_n]\right) \approx 0.95,$$

où $1.96 \approx \Phi^{-1}(0,975)$ avec Φ la fonction de répartition de la loi normale centrée réduite.

En outre, cette propriété indique que la vitesse de convergence de $M_n(h)$ est de l'ordre de $\sqrt{n}$, et ce indépendamment de la dimension du problème. Cela explique la supériorité de cette méthode par rapport aux méthodes d'intégration numérique déterministes dont les vitesses de convergence décroissent rapidement (exponentiellement) avec la dimension du problème.

Exercice – Calcul numérique d'une probabilité (••) Calculer une estimation de la probabilité de l'événement $\sin(X) > 1/2$ pour X de loi $\mathcal{N}(0,1)$.

Indication : si x est un vecteur, `y=(np.sin(x)>1/2)` désigne le vecteur de booléens dont les coordonnées valent `True` ou `False` selon que les coordonnées correspondantes de `np.sin(x)` soient ou non $> 1/2$. Par ailleurs, les vecteurs de booléens se convertissent naturellement en vecteurs de 0 et de 1 si on leur applique une fonction dont les arguments sont des nombres (ex : `np.mean`). (Solution p. 18.)

Génération de nombres pseudo-aléatoires

Les ordinateurs sont des machines déterministes. Il peut sembler paradoxal de leur demander de générer des nombres aléatoires [3]. En réalité, les algorithmes de génération de nombres aléatoires vont produire des séquences de nombres déterministes qui vont avoir l'aspect de l'aléatoire. On s'intéresse ici à la génération de nombres uniformes sur $]0,1[$ (on exclut les bornes par commodité), puisque, comme on le verra dans la suite, il s'agit de l'ingrédient de base de toutes les méthodes de simulation stochastique.

Définition – Générateur de nombres uniformes pseudo-aléatoires

Un *générateur de nombres uniformes pseudo-aléatoires* est un algorithme qui étant donné une valeur initiale u_0 et une transformation T produit une séquence $u_i = T^i(u_0)$, $i \in \mathbb{N}^*$, de valeurs dans $]0,1[$.

Pour tout $n \in \mathbb{N}^*$, les valeurs $(u_1, \ldots, u_n)$ reproduisent le comportement d'une suite de variables aléatoires $(V_1, \ldots, V_n)$ i.i.d de loi uniforme sur]0,1[, lorsqu'on les compare au travers d'un ensemble de tests statistiques [4], vérifiant par exemple que la corrélation entre deux nombres successifs est suffisamment faible.

---

3. Von Neumann (1951) résume ce problème très clairement : "Any one who considers arithmetical methods of reproducing random digits is, of course, in a state of sin. As has been pointed out several times, there is no such thing as a random number — there are only methods of producing random numbers, and a strict arithmetic procedure of course is not such a method."

4. Par exemple, la suite de tests Die Hard, due à Marsaglia.

Exemple – la méthode des congruences Cet algorithme, dû à Lehmer (1951), est l'un des premiers à avoir été proposé et implémenté. Il repose sur 2 paramètres :

— le multiplicateur $a \in \mathbb{N}^*$,
— le modulo $m \in \mathbb{N}^*$.

Etant donné un entier k_0 tel que $1 \leq k_0 \leq m-1$, la séquence de nombres est générée par la transformation suivante : pour tout $n \in \mathbb{N}^*$ on définit l'entier $k_{n+1} = a\, k_n \bmod m$ puis on pose

$$u_{n+1} = \frac{k_{n+1}}{m}, \quad u_0 = \frac{k_0}{m}.$$

On peut remarquer que l'algorithme va produire une séquence de valeurs qui sera périodique, c'est-à-dire qu'après un certain nombre d'itérations, la suite se répétera, et qu'il pourra fournir au plus $m-1$ valeurs différentes. Ce trait est commun à tous les générateurs de nombre aléatoires et est lié aux limitations matérielles des ordinateurs (on ne peut représenter qu'un nombre fini de nombres). Le choix des valeurs de a et m est par conséquent crucial. Il existe des critères qui permettent de s'assurer du bon comportement de cette suite :

— m est un nombre premier (le plus grand possible),
— $p = (m-1)/2$ est un nombre premier,
— $a^p = -1 \bmod m$.

Ce critère assure une période pleine et donc que tous les nombres $(1/m, \ldots, (m-1)/m)$ seront générés. Cependant, les nombres générés ne sont pas indépendants, pas même non corrélés. On peut montrer que la corrélation entre 2 nombres successifs vaut approximativement $1/a$. Il convient donc de choisir a suffisamment grand pour que celle-ci devienne négligeable. Par exemple, en prenant $a = 1000$ et $m = 2001179$, on obtient une période de 2001178 et une corrélation de l'ordre de 10^{-3}.

Ce générateur de nombres uniformes pseudo-aléatoires, bien que rudimentaire, a ouvert la voie à des générateurs plus sophistiqués, toujours basés sur des opérations arithmétiques. Le plus couramment utilisé à ce jour est l'algorithme de Mersenne-Twister. Il s'agit du générateur par défaut de la plupart des logiciels tels que Python, R, MATLAB, Julia, MS Excel,... Il passe notamment avec succès toute la batterie de tests Die Hard. En particulier, sa période vaut $2^{19937} - 1$.

Pour certains usages, cet algorithme n'est cependant pas recommandé du fait de sa prédictibilité. C'est notamment un défaut rédhibitoire pour les applications en cryptographie. Il existe des variantes mieux adaptées à ce cas de figure. On notera enfin qu'il existe des générateurs de nombres aléatoires basés sur des phénomènes physiques comme un bruit électrique ou des phénomènes quantiques et donc parfaitement imprévisibles.

Méthodes de simulation de variables aléatoires réelles

On a vu au chapitre II du cours de Probabilités que l'on pouvait transformer des variables aléatoires réelles suivant certaines lois pour en obtenir de nouvelles.

Par exemple, si $X_1, \ldots, X_n$ sont $n \in \mathbb{N}^*$ variables gaussiennes centrées réduites indépendantes, alors $X_1^2 + \ldots, X_n^2$ suit une loi du χ^2 à n degrés de liberté. Dans le même esprit, on va voir ici comment simuler des v.a.r. de lois diverses à partir de la simulation de variables uniformes sur $]0,1[$. On introduit une notation qui sera utile dans la suite : pour spécifier que deux v.a.r. X et Y ont même loi, on écrira $X \stackrel{\mathcal{L}}{=} Y$.

Méthode d'inversion

L'objectif de ce paragraphe est de définir quand et comment il est possible de simuler une variable aléatoire réelle X de fonction de répartition (f.d.r.) F_X en transformant la simulation d'une variable aléatoire U de loi uniforme sur $]0,1[$. En d'autres termes, on cherche à déterminer les conditions sous lesquelles il est possible d'identifier une fonction borélienne $\psi :]0,1[\to \mathbb{R}$ telle que $X \stackrel{\mathcal{L}}{=} \psi(U)$.

Commençons par un cadre simple, où F_X est **bijective** d'un intervalle non vide de $\mathbb{R}$ sur $]0,1[$.

Proposition – Cas où F_X est bijective

Soient X une variable aléatoire réelle de fonction de répartition F_X et U une variable uniforme sur $]0,1[$. S'il existe un intervalle non vide $]a,b[\subset \mathbb{R}$ tel que $F_X :]a,b[\to]0,1[$ est bijective, de bijection réciproque $F_X^{-1} :]0,1[\to]a,b[$, alors $F_X^{-1}(U) \stackrel{\mathcal{L}}{=} X$ et $F_X(X) \stackrel{\mathcal{L}}{=} U$.

Démonstration Le premier résultat est immédiat : pour tout $x \in \mathbb{R}$, par croissance de F_X et donc de F_X^{-1}, on a

$$\mathbb{P}\left(F_X^{-1}(U) \le x\right) = \mathbb{P}\left(U \le F_X(x)\right) = F_X(x).$$

Concernant le second, notons G la fonction de répartition de la variable aléatoire $F_X(X)$. La croissance de F_X garantit que $\mathbb{P}\left(F_X(X) \in]0,1[\right) = 1$. Ainsi, pour tout $x \in \mathbb{R}$ on a bien

$$G(x) = \left| \begin{array}{ll} 1 & \text{si } x \ge 1, \\ \mathbb{P}\left(F_X(X) \le x\right) = \mathbb{P}\left(X \le F_X^{-1}(x)\right) = x & \text{si } 0 < x < 1, \\ 0 & \text{sinon.} \end{array} \right.$$

■

Exercice – Exemples d'application 1 (•) Donner un algorithme de simulation d'une v.a.r. X suivant une loi

— uniforme sur un intervalle $I \subset \mathbb{R}$,
— Exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$,
— de Cauchy, de densité $x \in \mathbb{R} \mapsto \left(\pi\left(1+x^2\right)\right)^{-1}$,
— de Laplace de paramètres $\mu \in \mathbb{R}$ et $s \in \mathbb{R}_+^*$, de densité $x \in \mathbb{R} \mapsto \frac{1}{2s}\exp\left\{-\frac{|x-\mu|}{s}\right\}$,
— Logistique de paramètres $\mu \in \mathbb{R}$ et $s \in \mathbb{R}_+^*$, de fonction de répartition $x \in \mathbb{R} \mapsto \left(1+\exp\left\{-\frac{x-\mu}{s}\right\}\right)^{-1}$?

(Solution p. 18.)

Dans cette situation idéale, $\psi = F_X^{-1}$ est une solution à notre problème. Que se passe-t-il en revanche si F_X n'est pas bijective ?

Exercice – Pile ou face (•) Proposer une méthode pour simuler un tir à pile ou face à partir de la simulation d'une variable uniforme sur $]0, 1[$. (Solution p. 18.)

On a déjà vu au chapitre II que les fonctions de répartition de v.a.r. possèdent un nombre au plus dénombrable de points de discontinuité. Sur chaque intervalle où elles sont continues, on peut alors considérer qu'elles sont bijectives, quitte à réduire les zones de palier à un point. Cela permet de généraliser la notion de bijection réciproque pour ces fonctions.

Définition – Réciproque généralisée

Soit F une fonction de répartition. On définit sa *réciproque généralisée* (aussi appelée *inverse généralisée* ou *pseudo-inverse*) comme la fonction

$$F^- : u \in]0, 1[\mapsto \inf \{x \in \mathbb{R} : F(x) \geq u\} \in \mathbb{R}.$$

Remarque – Conséquences

- Cette fonction est bien définie sur tout $]0, 1[$, car quel que soit u dans cet intervalle, l'ensemble $\{x \in \mathbb{R} : F(x) \geq u\}$ est non vide et minoré. S'il était vide ou non minoré pour un certain $u_0 \in]0, 1[$, pour tout $x \in \mathbb{R}$ on aurait dans le premier cas $F(x) < u_0 < 0$ et dans le second $F(x) \geq u_0 > 1$. L'une comme l'autre de ces inégalités est impossible pour une fonction de répartition.
- La réciproque généralisée de la f.d.r. F_X d'une v.a.r. X est aussi appelée *fonction quantile*. On pourra notamment remarquer que $F_X^-\left(\frac{1}{2}\right)$ est la médiane de X.
- Lorsque F réalise une bijection d'un intervalle non vide $I \subset \mathbb{R}$ sur $]0, 1[$, sa réciproque généralisée coïncide avec sa bijection réciproque.

On a alors le résultat suivant, qui stipule que $\psi = F_X^-$ est une solution universelle à notre problème. La preuve détaillée est donnée en Annexe.

Théorème – Méthode d'inversion

Soient U une variable uniforme sur $]0, 1[$ ainsi que X une variable aléatoire réelle de fonction de répartition F_X et de réciproque généralisée F_X^-. Alors $F_X^-(U) \overset{\mathcal{L}}{=} X$.

Exercice – Exemples d'application 2 (•) Donner un algorithme de simulation d'une v.a.r. X suivant une loi

- Binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in]0, 1[$,
- Uniforme sur l'union de deux segments non vides et disjoints $[a, b], [c, d] \subset \mathbb{R}$, de densité $x \in \mathbb{R} \mapsto (b - a + d - c)^{-1} \, 1_{[a,b]\cup[c,d]}(x)$?

(Solution p. 18.)

Limitations

La méthode d'inversion peut sembler universelle pour simuler une v.a.r. X à partir de $U \sim \mathcal{U}_{]0,1[}$. Cependant, elle nécessite en pratique de disposer d'une expression analytique de F_X pour en déduire sa réciproque généralisée. Ce n'est typiquement pas le cas de nombreuses lois usuelles comme la loi Normale. On va donc déterminer d'autres procédures pour simuler des variables suivant de telles lois.

Méthode de rejet

La méthode de rejet est une alternative populaire à la méthode d'inversion, lorsque cette dernière ne peut être utilisée directement et que **la loi cible possède une densité**. On la doit à Von Neumann (1951). Pour en comprendre le fondement, il nous faut d'abord introduire une généralisation naturelle de la loi uniforme dans $\mathbb{R}$ à tout $\mathbb{R}^d$ ($d \in \mathbb{N}^*$). On notera λ la mesure de Lebesgue sur $\mathbb{R}^d$.

Définition – Loi uniforme sur un borélien

La loi uniforme sur un borélien $A \subset \mathbb{R}^d$ de volume $\lambda(A) > 0$ est une loi de probabilité admettant pour densité

$$f : x \in \mathbb{R}^d \mapsto \frac{1_A(x)}{\lambda(A)}.$$

Exercice – Loi uniforme sur un pavé (•) Comment simuler un vecteur aléatoire $(U_1, \ldots, U_d)$ de loi uniforme sur un pavé non vide $[a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$? (Solution p. 19.)

Une propriété fondamentale de cette loi est qu'elle reste stable par conditionnement, dans le sens suivant.

Proposition – Stabilité par conditionnement

Soit U un vecteur aléatoire de loi uniforme sur un borélien $A \subset \mathbb{R}^d$ de volume $\lambda(A) > 0$. Alors pour tout borélien $B \subset A$ de volume $\lambda(B) > 0$, la loi conditionnelle $\mathbb{P}_{U|U\in B}$ de U sachant que $U \in B$ est uniforme sur B.

Exercice – Démonstration (•) (Solution p. 19.)

Exercice – Simulation (•) Déduire de la proposition précédente (p. 8) une méthode pour simuler un vecteur aléatoire de loi uniforme sur un borélien $B \subset \mathbb{R}^d$ **borné** et de volume $\lambda(B) > 0$. (Solution p. 20.)

Il est aussi possible de simuler un vecteur aléatoire de loi uniforme sur certains boréliens non vides et non bornés, comme l'établit la proposition ci-dessous.

Proposition – Loi uniforme sur le sous-graphe d'une densité

Soient une densité $f : \mathbb{R} \to \mathbb{R}$, une v.a.r. X et une variable U uniforme sur $]0, 1[$, indépendante de X. On note $A_f := \{(x, y) \in \mathbb{R} \times \mathbb{R}_+ : f(x) \geq y\}$ le domaine limité par le graphe de f et l'axe des abscisses (souvent appelé sous-graphe de f). Si X est de densité f, alors le couple $(X, Uf(X))$ suit une loi uniforme sur A_f.

Démonstration Commençons par remarquer que $\lambda(A_f) = 1$. Quel que soit $(z, v) \in \mathbb{R}^2$, par indépendance de X et U et par Fubini on a

$$\begin{aligned}
\mathbb{P}(X \leq z, Uf(X) \leq v) &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{]-\infty,z]}(x)\, 1_{]-\infty,v]}(uf(x))\, 1_{]0,1[}(u)\, f(x)\, du\, dx \\
&= \int_{-\infty}^{z} \left(\int_{\mathbb{R}} 1_{]-\infty,v]\cap]0,f(x)[}(uf(x))\, f(x)\, du \right) dx \\
&= \int_{-\infty}^{z} \int_{-\infty}^{v} 1_{]0,f(x)[}(u)\, du\, dx \\
&= \int_{-\infty}^{z} \int_{-\infty}^{v} 1_{A_f}(x, u)\, du\, dx.
\end{aligned}$$

Ainsi, $(X, Uf(X))$ admet pour densité 1_{A_f}, qui correspond bien à celle d'une loi uniforme sur A_f. ■

Pour simuler un vecteur uniforme (X, Y) sur un ensemble A_f tel que défini à la proposition précédente, il suffit donc de simuler une v.a.r. X de densité f, puis une variable U uniforme sur $]0, 1[$, et de poser $Y = f(X)U$.

En combinant les résultats précédents, on obtient la méthode de rejet, illustrée sur la figure ci-dessous, où $A_Y := \{(x, y) \in \mathbb{R} \times \mathbb{R}_+ : y \leq f_Y(x)\}$.

Méthode de rejet

On souhaite simuler une variable aléatoire réelle X de densité f_X. Supposons que l'on sait simuler une variable U uniforme sur $]0, 1[$, ainsi qu'une v.a.r. Y (par exemple avec la méthode d'inversion) de densité f_Y telle qu'il existe un réel $a > 0$ pour lequel on a $\forall x \in \mathbb{R} : f_X(x) \leq a\, f_Y(x)$. Il suffit alors de suivre l'algorithme suivant :

1. simuler Y et U,
2. si $aUf_Y(Y) > f_X(Y)$, recommencer à l'étape 1.
3. poser $X = Y$.

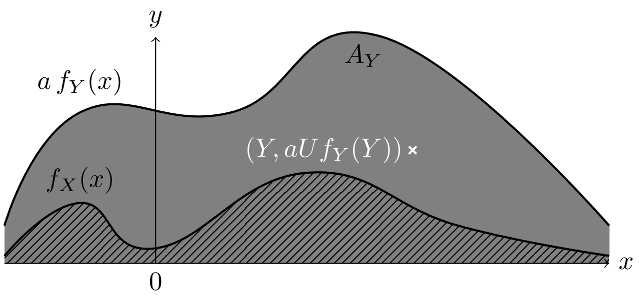

FIGURE 1 – Méthode de rejet

Limitations

La méthode de rejet a l'avantage de permettre de simuler des variables aléatoires à densité dont la fonction de répartition n'a pas de forme analytique, rendant la méthode d'inversion inapplicable. Néanmoins, pour pouvoir l'appliquer il faut connaître une densité auxiliaire qui, multipliée par un réel positif, majore la densité cible, et que l'on sait simuler. Le taux de rejet, c'est-à-dire la probabilité de l'événement $\{aU f_Y(Y) > f_X(Y)\}$, peut parfois être élevé, notamment lorsque la dimension de X est grande, ce qui limite l'efficacité de la méthode.

Exercice – Taux de rejet (••) Calculer le taux de rejet de la méthode proposée ci-dessus. (Solution p. 20.)

Simulation de variables aléatoires gaussiennes : Box-Muller

On a vu que la méthode d'inversion est inappropriée pour simuler une variable gaussienne, puisqu'elle requiert l'expression analytique de la fonction de répartition cible. Il existe des méthodes basées sur une intégration numérique de la densité gaussienne puis une inversion de cette approximation de la f.d.r. mais elle ne sont pas optimales en temps de calcul. La méthode de rejet est quant à elle sous-optimale, dans le sens où toutes les variables uniformes générées ne sont pas directement utilisées (une partie, potentiellement grande, est rejetée). La loi normale étant fondamentale en probabilité, il est plus que souhaitable de pouvoir en trouver une méthode de simulation exacte et efficace.

George E. P. Box et Mervin E. Muller ont proposé en 1958 une telle méthode (G.E.P. Box and M.E. Muller (1958)). Elle exploite la propriété d'invariance par rotation de la densité d'un couple de variables gaussiennes indépendantes centrées réduites.

Proposition – Proposition

Soient U et V deux variables indépendantes, de loi uniforme sur $]0,1[$. Alors les variables aléatoires $X = \sqrt{-2\ln(U)}\cos(2\pi V)$ et $Y = \sqrt{-2\ln(U)}\sin(2\pi V)$ sont indépendantes et suivent toutes deux une loi normale centrée réduite.

Démonstration On considère $(\widetilde{X}, \widetilde{Y})$ un vecteur aléatoire dont les deux composantes sont indépendantes, de loi normale centrée réduite. On note ses coordonnées polaires aléatoires $\widetilde{R}$ et $\widetilde{\Theta}$. On a vu dans le cours de Probabilités II que dans ce cas, $\widetilde{R}$ et $\widetilde{\Theta}$ sont indépendantes, la première de densité $f_{\widetilde{R}} : r \in \mathbb{R} \mapsto r\, e^{-\frac{r^2}{2}} 1_{\mathbb{R}_+^*}(r)$ et la seconde de loi uniforme sur $]0, 2\pi]$.

X et Y sont les coordonnées cartésiennes obtenues à partir d'un rayon et d'un angle : en posant $R = \sqrt{-2\ln(U)}$ et $\Theta = 2\pi V$ on obtient $X = R\cos(\Theta)$ et $Y = R\sin(\Theta)$. Par indépendance de U et V, R et Θ sont indépendantes. Pour que le vecteur (X, Y) ait la même distribution que $(\widetilde{X}, \widetilde{Y}) = \left(\widetilde{R}\cos(\widetilde{\Theta}), \widetilde{R}\sin(\widetilde{\Theta})\right)$, il suffit de montrer que $R \overset{\mathcal{L}}{=} \widetilde{R}$ et $\Theta \overset{\mathcal{L}}{=} \widetilde{\Theta}$.

— Commençons par étudier la loi de R, de fonction de répartition F_R. On remarque que la fonction $u \in]0,1[\mapsto \sqrt{-2\ln(u)} \in \mathbb{R}_+^*$ est bijective, strictement décroissante. Ainsi, pour tout $r \in \mathbb{R}_-$ on a $\mathbb{P}(R \leq r) = 0$ et pour tout $r \in \mathbb{R}_+^*$ on a

$$\mathbb{P}(R \leq r) = \mathbb{P}\left(\sqrt{-2\ln(U)} \leq r\right) = \mathbb{P}\left(U \geq e^{-\frac{r^2}{2}}\right) = 1 - e^{-\frac{r^2}{2}}.$$

qui correspond à la fonction de répartition associée à la densité $f_{\widetilde{R}}$.

$$\int_{-\infty}^{r} f_{\widetilde{R}}(x)\,dx = \left| \begin{array}{ll} \displaystyle\int_0^r x\,e^{-\frac{x^2}{2}}\,dx = \left[-e^{-\frac{x^2}{2}}\right]_0^r = 1 - e^{-\frac{r^2}{2}} & \text{si } r > 0, \\ 0 & \text{sinon.} \end{array} \right.$$

— En ce qui concerne la loi de Θ, de fonction de répartition F_Θ, puisque la fonction $v \in]0,1[\mapsto 2\pi v \in]0, 2\pi[$ est bijective strictement croissante, on a que pour tout $\theta \in \mathbb{R}$

$$F_\Theta(\theta) = \left| \begin{array}{ll} 1 & \text{si } \theta \geq 2\pi, \\ \mathbb{P}\left(V \leq \frac{\theta}{2\pi}\right) = \dfrac{\theta}{2\pi} & \text{si } \theta \in]0, 2\pi[, \\ 0 & \text{si } \theta \leq 0, \end{array} \right.$$

qui est bien la fonction de répartition d'une loi uniforme sur $]0, 2\pi[$.

■

Cette méthode permet de simuler directement deux variables gaussiennes centrées réduites indépendantes à partir de deux variables uniformes indépendantes. Pour simuler une variable gaussienne d'espérance $m \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+^*$ quelconques, on se rappelera que si X suit une loi normale centrée réduite, alors $\sigma X + m$ suit une loi normale d'espérance m et de variance σ^2.

Simulation d'un vecteur gaussien à densité

On souhaite simuler un vecteur gaussien $X = (X_1, \dots, X_d)$ à valeurs dans $\mathbb{R}^d$ d'espérance m et de matrice de covariance C définie positive (et donc inversible) données.

Puisque la matrice C est définie positive, elle admet une racine carrée, c'est-à-dire qu'il existe une matrice N telle que $C = N\,N^t$. En effet, on peut par exemple décomposer C de la manière suivante :

$$C = V\,D\,V^t,$$

où V est une matrice orthogonale et D est la matrice diagonale dont les termes diagonaux sont les valeurs propres (toutes strictement positives) de C. Il suffit alors de prendre $N = V\,D^{1/2}$, où $D^{1/2}$ est la matrice diagonale dont les termes diagonaux sont les racines carrées des valeurs propres.

En pratique, il est coûteux numériquement d'effectuer le calcul des valeurs propres et des vecteurs propres de C. On va plutôt calculer sa *décomposition* ou *factorisation de Cholesky* qui permet d'écrire

$$C = L\,L^t$$

avec L une matrice triangulaire inférieure.[5]

Soit maintenant un autre vecteur gaussien $Y = (Y_1, \dots, Y_d)$ à valeurs dans $\mathbb{R}^d$ et de matrice de covariance identité, notée I_d. Autrement dit, les Y_i sont des variables aléatoires gaussiennes centrées, réduites et indépendantes.

Alors, le vecteur $Z = m + LY$ est gaussien, d'espérance m et de matrice de covariance C. En effet, Z est gaussien comme combinaison linéaire de variables aléatoires gaussiennes, $\mathbb{E}(Z) = \mathbb{E}(m+LY) = m$ et $\mathbb{V}(Z) = \mathbb{E}((LY)^2) = LI_dL^t = C$.

Echantillonnage d'importance

On introduit dans cette section la méthode d'échantillonnage d'importance (importance sampling en anglais), que l'on appelle aussi, de manière plus intuitive, échantillonnage préférentiel, pour les lois à densité. Pour ce faire, nous allons commencer par un exemple qui montre qu'il peut être plus efficace de simuler des valeurs selon une loi différente de celle d'intérêt, autrement dit de modifier la représentation de l'intégrale $\mathcal{I}$ sous la forme d'une espérance calculée selon une autre densité.

Exemple – Loi de Cauchy conditionnelle Supposons que l'on s'intéresse à calculer la probabilité p qu'une variable X de loi de Cauchy standard soit plus grande que 2 (on peut le calculer directement et $p = 0.15$)

$$p = \int_2^{+\infty} \frac{1}{\pi(1+x^2)} dx.$$

Si on estime p directement à partir d'un échantillon $(X_1, \dots, X_n)$ simulé selon la loi de Cauchy standard, soit

$$\widehat{p}_1 = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i > 2},$$

la *variance de l'estimateur* $\mathbb{V}(\widehat{p}_1) = p(1-p)/n = 0.127/n$, puisque $\widehat{p}_1$ suit une loi binomiale de paramètre (n, p). On peut réduire cette variance (et donc améliorer la qualité de l'estimateur) en tirant parti de la symétrie de la densité de la loi de Cauchy, en formant un second estimateur

$$\widehat{p}_2 = \frac{1}{n} \sum_{i=1}^{n} 1_{|X_i| > 2},$$

dont la variance vaut $\mathbb{V}(\widehat{p}_2) = p(1-p/2)/2n = 0.052/n$. La relative inefficacité de ces méthodes est due au fait que la majeure partie des valeurs simulées seront

5. Cette décomposition est très utile dans la résolution de systèmes linéaires de la forme $A\,x = b$, où b est connu, x inconnu et A est définie positive. Cela revient à résoudre $L\,L^t\,x = b$. On pose alors $y = L^t\,x$ et on résout d'abord $Ly = b$, ce qui est très rapide puisque L est triangulaire inférieure (on commence par $y_1 = b_1/L_{11}$, puis $y_2 = (b_2 - L_{21}y_1)/L_{22}$, etc. en descendant). On résout ensuite $L^t\,x = y$, ce qui est aussi très rapide pour la même raison (on commence par $x_n = y_n/L_{nn}$ puis on remonte).

en dehors de la zone d'intérêt $]2, +\infty[$. En passant par le complémentaire, on peut réécrire p comme

$$p = \frac{1}{2} - \int_0^2 \frac{1}{\pi(1+x^2)} dx,$$

dont le second terme peut être vu comme l'espérance de $h(U) = \frac{2}{\pi(1+U^2)}$ avec $U \sim \mathcal{U}_{[0,2]}$. Tirant un échantillon $(U_1, \ldots, U_n)$ i.i.d. de loi uniforme sur $[0, 2]$, on obtient un troisième estimateur :

$$\widehat{p}_3 = \frac{1}{2} - \frac{1}{n} \sum_{i=1}^{n} \frac{2}{\pi(1+U_i^2)},$$

dont la variance vaut $\mathbb{V}(\widehat{p}_3) = (\mathbb{E}(h(X)^2) - \mathbb{E}(h(U))^2)/n = 0.0285/n$ (par intégration par parties). Enfin, on peut encore réécrire (voir Ripley (1987))

$$p = \int_0^{1/2} \frac{y^{-2}}{\pi(1+y^{-2})} dy,$$

qui peut être vue comme $\mathbb{E}\left(\frac{V^{-2}}{2\pi(1+V^{-2})}\right)$ avec $V \sim \mathcal{U}_{]0,1/2[}$. L'estimateur formé à partir de cette représentation et d'un échantillon $(V_1, \ldots, V_n)$ i.i.d. de loi uniforme sur $[0, 1/2]$ a une variance de $0.95\, 10^{-4}/n$. Il est donc bien plus efficace que $\widehat{p}_1$ puisqu'il nécessite environ $\sqrt{10^3} \approx 32$ fois moins de simulations pour atteindre la même précision.

On a ainsi vu sur ce cas particulier que l'estimation d'une intégrale de la forme

$$\mathcal{I} = \mathbb{E}\left(h(X)\right) = \int_{\mathbb{R}^d} h(x) f(x) dx,$$

peut s'écrire de différentes manières, en faisant varier h et f. Par conséquent, un estimateur "optimal" devrait tenir compte de l'ensemble de ces possibilités. C'est justement l'idée développée dans la méthode d'échantillonnage d'importance dont le principe est décrit dans la définition suivante.

Définition – Echantillonnage d'importance

La méthode d'*échantillonnage d'importance* est une évaluation de $\mathcal{I}$ basée sur la simulation d'un échantillon $X_1, \ldots, X_n$ de loi de densité g et approximant :

$$\mathbb{E}_f\left(h(X)\right) \approx \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} h(X_i),$$

où la notation $\mathbb{E}_f$ signifie que l'espérance est calculée avec $X \sim f$. On appelle souvent les ratios $\frac{f(X_i)}{g(X_i)}$ les *poids d'importance* que l'on note w_i.

Cette méthode est basée sur la représentation suivante de $\mathcal{I}$:

$$\mathbb{E}_f\left[h(X)\right] = \int h(x) \frac{f(x)}{g(x)} g(x)\, dx$$

que l'on appelle l'*identité fondamentale de l'échantillonnage d'importance* et l'estimateur converge du fait de la loi forte des grands nombres.

Cette identité indique qu'une intégrale du type $\mathcal{I}$ n'est pas intrinsèquement associée à une loi donnée. L'intérêt de l'échantillonnage d'importance repose sur le fait qu'il n'y a aucune restriction sur le choix de la densité g, dite *instrumentale*, que l'on peut donc choisir parmi les densités des lois que l'on sait simuler aisément. Il y a bien évidemment des choix qui sont meilleurs que d'autres. Remarquons tout d'abord que bien que l'estimateur proposé dans la définition ci-dessus (p. 13) converge presque sûrement, sa variance est finie si

$$\mathbb{E}_g\left(h^2(X)\frac{f^2(X)}{g^2(X)}\right) = \mathbb{E}_f\left(h^2(X)\frac{f(X)}{g(X)}\right) = \int h^2(x)\frac{f^2(x)}{g(x)}\,dx < \infty.$$

On préconise alors l'usage de densités instrumentales g dont la queue de distribution est plus épaisse que celle de f pour éviter que cette variance puisse être infinie (on notera que cela dépend aussi de la fonction h à intégrer). En pratique, on utilise généralement l'estimateur suivant, de variance finie et qui donne des résultats plus stables numériquement que celui de la définition :

$$\frac{\sum_{i=1}^n w_i h(X_i)}{\sum_{i=1}^n w_i}$$

où on a remplacé n par la somme des poids d'importance. Puisque $\frac{1}{n}\sum_{i=1^n} w_i = \frac{1}{n}\sum_{i=1}^n \frac{f(x_i)}{g(x_i)}$ tend vers 1 quand $n \to \infty$, cet estimateur converge presque sûrement vers $\mathbb{E}_f\,(h(X))$ par la loi forte des grands nombres (voir C. P. Robert and G. Casella (2004)).

Parmi les densités g qui fournissent des estimateurs de variance finie, il est possible d'exhiber la densité optimale (au sens de la variance de l'estimateur associé) pour une fonction h et une densité f données.

Théorème – Densité optimale

Le choix de g qui minimise la variance de l'estimateur donné dans la définition ci-dessus (p. 13) est

$$g^*(x) = \frac{|h(x)|f(x)}{\int |h(x)|f(x)dx}.$$

Démonstration Notons d'abord que

$$\mathbb{V}_g\left(\frac{h(X)f(X)}{g(X)}\right) = \mathbb{E}_g\left(h^2(X)\frac{f^2(X)}{g^2(X)}\right) - \mathbb{E}_g\left(h(X)\frac{f(X)}{g(X)}\right)^2$$

et le second terme ne dépend pas de g. On minimise donc le premier terme. D'après l'inégalité de Jensen, on a

$$\mathbb{E}_g\left(h^2(X)\frac{f^2(X)}{g^2(X)}\right) \geq \mathbb{E}_g\left(|h(X)|\frac{f(X)}{g(X)}\right)^2 = \left(\int |h(x)|f(x)dx\right)^2,$$

qui nous donne une borne inférieure indépendante du choix de g. Elle est atteinte en prenant $g = g^*$. ∎

Ce résultat formel n'a que peu d'intérêt pratique : le choix optimal de g fait intervenir $\int |h(x)|f(x)dx$, qui est à une valeur absolue près la quantité que l'on souhaite estimer ! Il suggère néanmoins de considérer des densités g telles que $|h|f/g$ est quasi constante et de variance finie. On se reportera au chapitre 3 de C. P. Robert and G. Casella (2004) pour des exemples où un bon choix de g permet des améliorations considérables par rapport à des estimateurs de Monte-Carlo plus naïfs.

Exercices

Loi uniforme dans un domaine

Ecrire un algorithme pour simuler un point uniforme dans les trois domaines représentés ci-dessous.

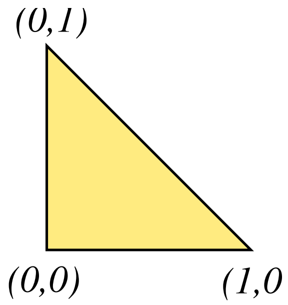

Figure 2 – Domaine A

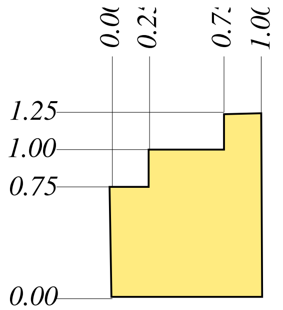

Figure 3 – Domaine B

Simulation selon la loi géométrique

Cette loi correspond à la loi du nombre d'essais avant d'obtenir un évènement de probabilité p

$$X \sim \mathcal{G}(p) \quad \mathbb{P}(X = n) = p(1-p)^{n-1} \quad n = 1, 2, \ldots$$

Question 1 Calculer l'espérance et la variance de X (Solution p. 20.)

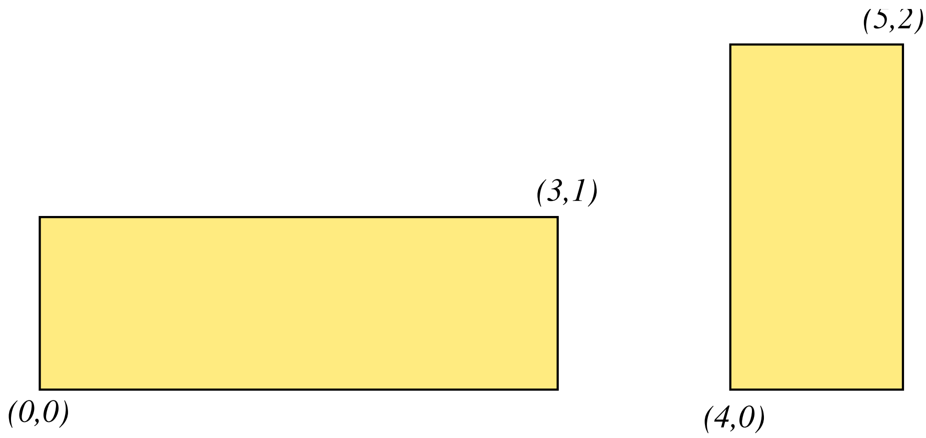

FIGURE 4 – Domaine C

Question 2 Générer un échantillon de taille 1000 pour $p = 0.6$, en utilisant une approche génétique. Calculer la moyenne et la variance. Comparer avec les valeurs théoriques. (Solution p. 22.)

Question 3 Montrer que si $X \sim \mathcal{E}(\lambda)$, alors $\lceil X \rceil \sim \mathcal{G}(p)$. On précisera la valeur de λ. (Solution p. 23.)

Question 4 Donner, implémenter et tester un algorithme pour simuler $\mathcal{G}(p)$ directement (Solution p. 23.)

Simulation de la loi gaussienne par la méthode de rejet

X est une variable aléatoire de loi gaussienne d'espérance $m = 0$ et de variance $\sigma^2 = 1$, $\mathcal{N}(0,1)$. Sa densité est

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \ x \in \mathbb{R}$$

Question 1 Montrer que $f(x) \leq C \exp(-|x|)$ où $C = \sqrt{\frac{e}{2\pi}}$ (Solution p. 23.)

Question 2 On considère $g(x) = \frac{1}{2}\exp(-|x|)\,,\ x \in \mathbb{R}$. Montrer que g est une densité de probabilité. (Solution p. 23.)

Question 3 Proposer un algorithme pour simuler selon g. (Solution p. 23.)

Question 4 Implémenter l'algorithme de rejet pour simuler selon f. (Solution p. 23.)

Simulation selon la loi de Wigner

La loi de Wigner (ou du demi-cercle) est la loi de support $[-2, 2]$ et de densité $f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}$. Simuler un grand nombre de variables aléatoires de loi de

Wigner, représenter l'histogramme associé et le comparer à la densité. (Solution p. 23.)

Loi des grands nombres et théorème central limite

Question 1 Afin d'illustrer la Loi des Grands Nombres, visualiser la suite $S_n = \frac{X_1+...+X_n}{n}$ pour $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes de loi uniforme sur [-1, 1].

Indication : pour $x = [x_1, \ldots, x_n]$ vecteur, `np.cumsum(x)` est le vecteur $[x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots, x_1 + \ldots + x_n]$ des sommes cumulées des coordonnées de x. (Solution p. 23.)

Question 2 Faire de même avec des variables aléatoires Y_i de loi de Cauchy (c'est-à-dire de densité $\frac{1}{\pi(1+x^2)}$). La suite $\frac{Y_1+...+Y_n}{n}$ semble-t-elle converger ? Pourquoi ? (Solution p. 23.)

Question 3 Calculer la moyenne μ et l'écart-type σ des X_i et vérifier, afin d'illustrer le Théorème Central Limite, que $\sqrt{n}(S_n - \mu)/\sigma$ converge en loi, lorsque $n \to \infty$, vers une variable de loi $\mathcal{N}(0,1)$: on se fixera une grande valeur de n, on simulera un grand nombre de fois $\sqrt{n}(S_n - \mu)/\sigma$, on en tracera l'histogramme et on le comparera à la densité $(2\pi)^-1/2e^{-x^2}$ la loi $\mathcal{N}(0,1)$. (Solution p. 23.)

Simulation d'un mélange de gaussiennes

On souhaite simuler selon la loi de mélange introduite dans l'exercice Mélange de loi du chapitre 3.

Question 1 Soit K une v.a.r. de loi :

$$\mathbb{P}(K = x) = \begin{cases} 1/2 & \text{si } x = 1 \\ 1/4 & \text{si } x = 2 \\ 1/4 & \text{si } x = 3 \\ 0 & \text{sinon} \end{cases}$$

Implémenter la méthode d'inversion pour simuler K (Solution p. 23.)

Question 2 Soit X_1, X_2 et X_3 des variables aléatoires de lois respectives $\mathcal{N}(0,1)$, $\mathcal{N}(5,1/2)$ et $\mathcal{N}(8,4)$. Implémenter un algorithme de simulation de X_K. Comparer l'histogramme obtenu pour un échantillon de taille 1000 avec la densité de X_K. (Solution p. 24.)

Echantillonnage d'importance

On cherche à évaluer l'espérance d'une variable aléatoire X gaussienne centrée réduite (de densité f_X et de f.d.r. F_X) sachant qu'elle dépasse la valeur 3.

Question 1 Exprimer la densité conditionnelle de $X|X > 3$. (Solution p. 24.)

Question 2 Calculer la quantité souhaitée à partir d'un échantillon de taille 1000 généré selon F_X. Quelle est la valeur du taux de rejet ? (Solution p. 24.)

Question 3 Implémenter l'algorithme d'échantillonnage d'importance en prenant comme loi instrumentale la loi gaussienne d'espérance 3 et de variance 1. Quelle est la proportion de poids nuls ?

(Solution p. 24.)

Solutions

Calcul numérique d'une probabilité cf. notebook

Exemples d'application 1 On suppose que $U \sim \mathcal{U}_{]0,1[}$

— Uniforme sur un intervalle $I \subset \mathbb{R}$:
Si $I =]a, b[$, $a < b$, alors la fonction de répartition de $X \sim \mathcal{U}_I$ est $F_X(x) = \frac{x-a}{b-a} 1_{]a,b[}(x) + 1_{[b,+\infty[}(x)$, d'où par inversion $X = a + (b-a)U$

— Exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$:
la fonction de répartition de $X \sim \mathcal{E}(\lambda)$ est $F_X(x) = 1 - \exp(-\lambda x)$, d'où par inversion $X = -\frac{\ln(1-U)}{\lambda}$. Pour simplifier, on remarquera que si $U \sim \mathcal{U}_{]0,1[}$, alors $1 - U \sim \mathcal{U}_{]0,1[}$, d'où $X = -\frac{\ln(U)}{\lambda}$

— de Cauchy, de densité $x \in \mathbb{R} \mapsto \left(\pi\left(1 + x^2\right)\right)^{-1}$,
la loi de Cauchy admet la fonction de répartition $F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}$, d'où $X = \tan(\pi U - 1/2)$

— de Laplace de paramètres $\mu \in \mathbb{R}$ et $s \in \mathbb{R}_+^*$, de densité $x \in \mathbb{R} \mapsto \frac{1}{2s}\exp\left\{-\frac{|x-\mu|}{s}\right\}$,
la loi de Laplace admet la fonction de répartition

$$F(x) = \begin{cases} \frac{1}{2}\exp(x) & \text{si } x < 0, \\ 1 - \frac{1}{2}\exp(-x) & \text{si } x \geq 0, \end{cases} = \frac{1}{2}\left(1 + \operatorname{sgn}(x)(1 - \exp(-|x|))\right)$$

d'où $X = \operatorname{sgn}(U - 1/2)\ln(1 - 2|U - 1/2|)$, où sgn est la fonction signe.

— Logistique de paramètres $\mu \in \mathbb{R}$ et $s \in \mathbb{R}_+^*$, de fonction de répartition $x \in \mathbb{R} \mapsto \left(1 + \exp\left\{-\frac{x-\mu}{s}\right\}\right)^{-1}$:
par inversion, $X = s\left(-\ln\left(\frac{1}{U} - 1\right) + \mu\right)$

Pile ou face Il suffit d'associer à "pile" un événement portant sur $U \sim \mathcal{U}_{]0,1[}$ de la bonne probabilité, $p \in]0, 1[$, par exemple $\{U \leq p\}$. On obtient ainsi l'algorithme :

1. Générer u selon $\mathcal{U}_{]0,1[}$
2. Renvoyer "pile" si $u \leq p$, "face" sinon

Exemples d'application 2

— Binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in]0,1[$,
La fonction de répartition de la loi binomiale est donnée par

$$F(x) = \begin{cases} 0 & \text{si } \ x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } \ 0 \leq x < n \\ 1 & \text{si } x \geq n \end{cases}$$

où $\lfloor x \rfloor$ est la partie entière de x.
On peut alors découper $]0,1[$ en intervalles de la forme $] \sum_{k=0}^{m} \binom{n}{k} p^k (1-p)^{n-k}, \sum_{k=0}^{m+1} \binom{n}{k} p^k (1-p)^{n-k}[$ pour $m \in \{0, n-1\}$ et renvoyer $m+1$ si un u généré selon $\mathcal{U}_{]0,1[}$ appartient à l'un de ces intervalles, 0 sinon.
Il est cependant beaucoup plus simple de remarquer qu'une variable aléatoire de loi $\mathcal{B}(n,p)$ s'obtient comme la somme de n variables de Bernoulli que l'on peut simuler comme ci-dessus (p. 18).

— Uniforme sur l'union de deux segments non vides et disjoints $[a,b], [c,d] \subset \mathbb{R}$, de densité $x \in \mathbb{R} \mapsto (b-a+d-c)^{-1} \, 1_{[a,b] \cup [c,d]}(x)$
La fonction de répartition d'une telle variable est

$$F(x) = \frac{x-a}{b-a+d-c} 1_{]a,b[}(x) + \frac{b-a}{b-a+d-c} 1_{[b,c]}(x) + \frac{x-c+b-a}{b-a+d-c} 1_{]c,d[}(x) + 1_{[d,+\infty[}(x)$$

Pour simuler cette variable, on peut donc utiliser l'algorithme suivant :

1. Générer u selon $\mathcal{U}_{]0,1[}$
2. Retourner

$$\begin{cases} a + u(b-a+d-c) & \text{si } u \in]0, \frac{b-a}{b-a+d-c}[\\ c + (u(b-a+d-c) - (b-a)) & \text{sinon} \end{cases}$$

Loi uniforme sur un pavé On note $[a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$. On peut voir que la densité d'une v.a. uniforme sur un tel pavé s'écrit $f(x) = \frac{1}{b_1-a_1} \cdots \frac{1}{b_d-a_d} 1_{[a_1,b_1] \times \cdots \times [a_d,b_d]}(x)$ soit comme le produit des densités de ses coordonnées qui sont donc indépendantes. On peut donc simuler la variable d'intérêt avec cet algorithme :

1. Générer $u_1, \ldots, u_d$ indépendamment selon $\mathcal{U}_{]0,1[}$
2. Retourner $(a_1 + u_1(b_1 - a_1), \ldots, a_d + u_d(b_d - a_d))$

Démonstration Soit $C \in \mathcal{B}(B)$ (considérer les boréliens de B suffit à caractériser la loi conditionnelle à l'événement $U \in B$ puisque si $C \cap B = \varnothing$, alors nécessairement, $\mathbb{P}(U \in C | U \in B) = 0$). On a

$$\begin{aligned} \mathbb{P}(U \in C | U \in B) &= \frac{\mathbb{P}(U \in C, U \in B)}{\mathbb{P}(U \in B)} = \frac{\mathbb{P}(U \in C)}{\mathbb{P}(U \in B)} \\ &= \frac{\int_C \frac{dx}{\lambda(A)}}{\int_B \frac{dx}{\lambda(A)}} \text{ puisque } U \sim \mathcal{U}_A \\ &= \frac{\lambda(C)}{\lambda(B)} = \int_C \frac{1}{\lambda(B)} dx \end{aligned}$$

on reconnaît bien la loi uniforme sur B.

Simulation On prend un pavé R tel que $B \subset R$ puis on applique l'algorithme suivant :

1. générer u selon $\mathcal{U}_R$
2. Si $u \notin R$, retour en 1.
3. retourner u

Taux de rejet Puisque le couple $(Y, Uf_Y(Y))$ est uniforme sur A_f, alors le couple $(Y, aUf_Y(Y))$ est uniforme sur A_{af}, d'où

$$\begin{aligned}
\mathbb{P}(\{aUf_Y(Y) > f_X(Y)\}) &= \mathbb{E}(1_{\{aUf_Y(Y) > f_X(Y)\}}) \\
&= \mathbb{E}(\mathbb{E}(1_{\{aUf_Y(y) > f_X(y)\}} | Y = y)) \\
&= \int_R \int_0^{af_Y(y)} 1_{\{auf_Y(Y) > f_X(Y)\}} \frac{1}{af_Y(y)} du f_Y(y) dy \\
&= \int_R \left(1 - \frac{f_X(y)}{af_Y(y)}\right) f_Y(y) dy \\
&= 1 - \frac{1}{a}
\end{aligned}$$

Loi uniforme dans un domaine

cf. notebook

Simulation selon la loi géométrique

Question 1 $\mathbb{P}(X=n)=p(1-p)^{n-1}=(1-q)q^{n-1}$, où p est la probabilité de succès, $n=1,2,\dots$

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{\infty} i\mathbb{P}(X=i) \\
&= \sum_{i=1}^{\infty} i(1-q)q^{i-1} \\
&= (1-q)\sum_{i=1}^{\infty} iq^{i-1} \\
&= (1-q)\sum_{i=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}q}q^{i} \\
&= (1-q)\frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{i=1}^{\infty} q^{i}\right) \\
&= (1-q)\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right), \text{ car } \sum_{i=1}^{\infty} q^{i} = q\sum_{i=0}^{\infty} q^{i} = \frac{q}{1-q} \\
&= (1-q)\frac{1}{(1-q)^2} \\
&= \frac{1}{(1-q)} \\
&= \frac{1}{p}
\end{aligned}
$$

Pour la variance $\mathbb{V}[X]=\mathbb{E}[X^2]-\mathbb{E}[X]$

$$
\begin{aligned}
\mathbb{E}[X^2] &= \sum_{i=1}^{\infty} i^2\mathbb{P}(X=i) \\
&= \sum_{i=1}^{\infty} i^2(1-q)q^{i-1} \\
&= (1-q)\sum_{i=1}^{\infty} i^2q^{i-1} \\
&= (1-q)\sum_{i=1}^{\infty} \left(i\left((i+1)q^{i-1}-q^{i-1}\right)\right) \\
&= (1-q)\sum_{i=1}^{\infty} \left(i(i+1)q^{i-1}-iq^{i-1}\right) \\
&= (1-q)\sum_{i=1}^{\infty} \left(\frac{\mathrm{d}^2}{\mathrm{d}q^2}q^{i+1}-\frac{\mathrm{d}}{\mathrm{d}q}q^{i}\right) \\
&= (1-q)\left(\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\sum_{i=1}^{\infty} q^{i+1}\right)-\frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{i=1}^{\infty} q^{i}\right)\right)
\end{aligned}
$$

On a déjà calculé le second terme ci-dessus.

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}q^2}\sum_{i=1}^{\infty} q^{i+1} &= \frac{\mathrm{d}^2}{\mathrm{d}q^2} q \sum_{i=1}^{\infty} q^i \\ &= \frac{\mathrm{d}^2}{\mathrm{d}q^2}\frac{q^2}{1-q}\end{aligned}$$

Calcul des dérivées :

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}q}\frac{q^2}{1-q} &= \frac{2q(1-q)+q^2}{(1-q)^2} \\ &= \frac{2q-q^2}{(1-q)^2}\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}q}\frac{2q-q^2}{(1-q)^2} &= \frac{2(1-q)(1-q)^2+2(1-q)(2q-q^2)}{(1-q)^4} \\ &= \frac{2(1-q)^2+2(2q-q^2)}{(1-q)^3} \\ &= \frac{2-4q+2q^2+4q-2q^2}{(1-q)^3} \\ &= \frac{2}{(1-q)^3}\end{aligned}$$

On en déduit

$$\begin{aligned}\mathbb{E}[X^2] &= (1-q)\left(\frac{2}{(1-q)^3}-\frac{1}{(1-q)^2}\right) \\ &= \frac{2}{(1-q)^2}-\frac{1}{(1-q)} \\ &= \frac{2-(1-q)}{(1-q)^2} \\ &= \frac{1+q}{(1-q)^2}\end{aligned}$$

Et finalement

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[X^2]-\mathbb{E}[X] \\ &= \frac{1+q}{(1-q)^2}-\frac{1}{(1-q)^2} \\ &= \frac{q}{(1-q)^2} \\ &= \frac{1-p}{p^2}\end{aligned}$$

Question 2 cf. notebook

Question 3

$$\begin{aligned}\mathbb{P}(\lceil X \rceil = n) &= \mathbb{P}(n-1 < X \leq n) \\ &= F_X(n) - F_X(n-1) \\ &= e^{-\lambda(n-1)}(1 - e^{-\lambda})\end{aligned}$$

d'où $\lambda = -\ln(1-p)$

Question 4 cf. notebook

Simulation de la loi gaussienne par la méthode de rejet

X est une variable aléatoire de loi gaussienne d'espérance $m = 0$ et de variance $\sigma^2 = 1$, $\mathcal{N}(0, 1)$. Sa densité est

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right),\ x \in \mathbb{R}$$

Question 1 Soit $x \in \mathbb{R}$

$$\begin{gathered}x^2 \geq 2|x| - 1 \\ \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \leq \sqrt{\frac{e}{2\pi}} e^{-|x|}\end{gathered}$$

Question 2 g est bien positive et son intégrale sur $\mathbb{R}$ vaut

$$\int_{\mathbb{R}} g(x)dx = \int_{\mathbb{R}} \frac{1}{2} e^{-|x|} dx = \int_{\mathbb{R}^+} e^{-x} dx = 1$$

Question 3 voir dans exemples d'application 1 (p. 6).

Question 4 cf. notebook

Simulation de la loi de Wigner

cf. notebook

Loi des grands nombres et théorème central limite

Question 1 cf. notebook

Question 2 cf. notebook

Question 3 cf. notebook

Simulation d'un mélange de gaussiennes

Question 1 cf. notebook

Question 2 cf. notebook

Echantillonnage d'importance

Question 1 On a $F_{X|X>3}(x) = \mathbb{P}(X \leq x|X > 3) = \frac{\mathbb{P}(X\leq x, X>3)}{1-F_X(3)} = \frac{F_X(x)-F_X(3)}{1-F_X(3)} 1_{]3,+\infty[}(x)$ d'où $f_{X|X>3}(x) = \frac{f_X(x)}{1-F_X(3)} 1_{]3,+\infty[}(x)$

Question 2 cf. notebook

Question 3 cf. notebook

Annexe

Preuve de la méthode d'inversion

Pour pouvoir démontrer le théorème de la méthode d'inversion (p. 7), il faut d'abord établir un certain nombre de propriétés de la réciproque généralisée d'une fonction de répartition. Elle peuvent être visualisées sur la figure ci-dessous.

Proposition – Proposition

Soit F une fonction de répartition. Alors sa réciproque généralisée F^- satisfait les propriétés suivantes.

1. F^- est croissante.
2. $\forall\, x \in \mathbb{R} : F^- \circ F(x) \leq x$.
3. $\forall\, u \in]0,1[\, : F \circ F^-(u) \geq u$ avec égalité si $u \in F(\mathbb{R})$.
4. $\forall\, (u,x) \ \in]0,1[\times \mathbb{R} \, : \{F(x) \geq u\} \ \Leftrightarrow \ \{x \geq F^-(u)\}$ et $\{F(x) < u\} \ \Rightarrow \{x \leq F^-(u)\}$.

Démonstration Pour tout $u \in]0,1[$ on note $F^{-1}([u,1]) := \{x \in \mathbb{R} : F(x) \geq u\}$ l'image réciproque de $[u,1]$ par F.

1. Soit $(u,v) \in]0,1[^2$. Si $u < v$ alors $F^{-1}([v,1]) \subset F^{-1}([u,1])$ d'où $F^-(u) = \inf F^{-1}([u,1]) \leq \inf F^{-1}([v,1]) = F^-(v)$. La fonction F^- est donc bien croissante.
2. Soit $x \in \mathbb{R}$, alors $F^- \circ F(x) = \inf\big\{z \in \mathbb{R} : F(z) \geq F(x)\big\} = \inf F^{-1}([F(x),1])$. Comme F est croissante sur $\mathbb{R}$ on a $[x,+\infty[\subseteq F^{-1}([F(x),1])$ donc $F^- \circ F(x) = \inf F^{-1}([F(x),1]) \leq \inf[x,+\infty[= x$.
3. Soit $u \in]0,1[$.

— Puisque $F^{-1}([u,1])$ est nécessairement non vide, il existe une suite décroissante $(x_n)_{n\in\mathbb{N}} \subseteq F^{-1}([u,1])$ convergeant vers $\inf F^{-1}([u,1]) = F^-(u)$. La croissance de F implique que la suite $\big(F(x_n)\big)_{n\in\mathbb{N}}$ est elle aussi décroissante, minorée par u car $(x_n)_{n\in\mathbb{N}} \subseteq F^{-1}([u,1])$ donc convergente. Sa limite est de même supérieure ou égale à u. Comme F est continue à droite, cette dernière n'est autre que

$$\lim_{n\to+\infty} F(x_n) = F\left(\lim_{n\to+\infty} x_n\right) = F \circ F^-(u).$$

Nous avons donc bien $F \circ F^{-}(u) \geq u$.

— Supposons maintenant que $u \in F(\mathbb{R})$. Alors $F^{-1}\left(\{u\}\right) := \{x \in \mathbb{R} : F(x) = u\} \neq \emptyset$. Il existe donc une suite décroissante $(x_n)_{n\in\mathbb{N}} \subseteq F^{-1}\left(\{u\}\right)$ convergeant vers $\inf F^{-1}\left(\{u\}\right) = F^{-}(u)$ par croissance de F. Comme F est continue à droite en tout point de $\mathbb{R}$ et $F(x_n) = u$ pour tout $n \in \mathbb{N}$, on a bien

$$u = \lim_{n\to+\infty} F(x_n) = F\left(\lim_{n\to+\infty} x_n\right) = F \circ F^{-}(u).$$

4. Soit $(u, x) \in]0, 1[\times\mathbb{R}$.

— **Equivalence.** Supposons $F(x) \geq u$. On a $F^{-} \circ F(x) \geq F^{-}(u)$ par croissance de F^{-} (propriété 1.) et $F^{-} \circ F(x) \leq x$ d'après la propriété 2. Réciproquement, supposons que $x \geq F^{-}(u)$. Alors par croissance de F on a $F(x) \geq F \circ F^{-}(u)$ puis $F \circ F^{-}(u) \geq u$ d'après la propriété 3.

— **Implication.** Supposons $F(x) < u$. Tout $z \in F^{-1}\left([u, 1]\right)$ vérifie $F(z) \geq u > F(x)$. Comme F est croissante sur $\mathbb{R}$ on a donc $z \geq x$, ce qui implique $x \leq \inf F^{-1}\left([u, 1]\right) = F^{-}(y)$.

■

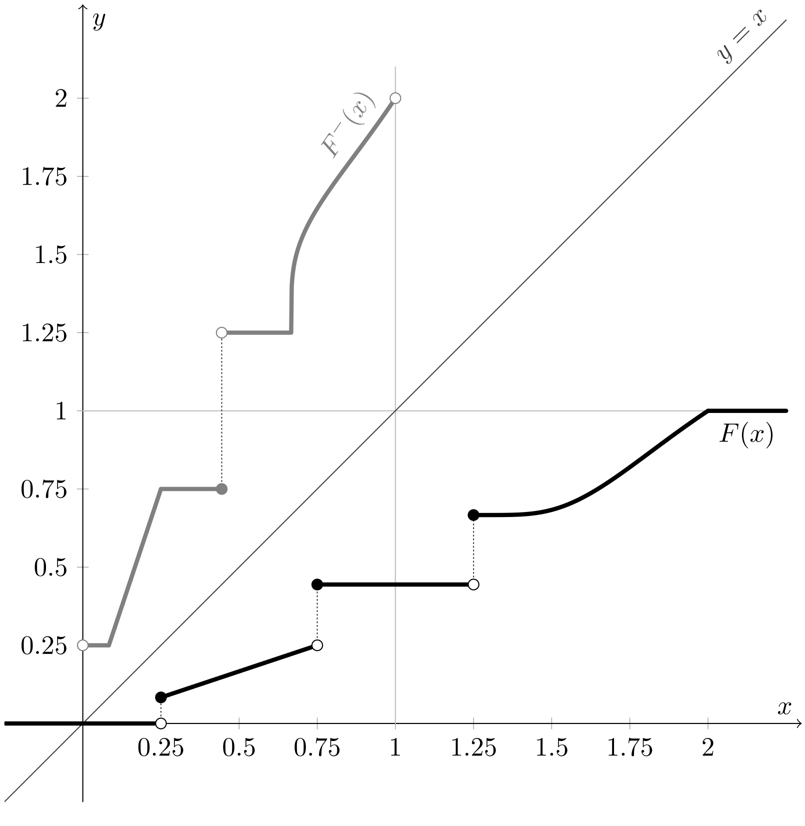

Figure 5 – Réciproque généralisée d'une fonction de répartition

Nous pouvons maintenant établir la preuve du théorème de la méthode d'inversion (p. 7).

Démonstration – Méthode d'inversion Soient $x \in \mathbb{R}$ et $(\Omega, \mathcal{A}, \mathbb{P})$ l'espace probabilisé sur lequel sont définies U et X. D'après la propriété 4 ci-dessus, on a $\left\{\omega \in \Omega : F_X^-(U(\omega)) \leq x\right\} = \{\omega \in \Omega : U(\omega) \leq F(x)\}$, d'où

$$\mathbb{P}\left(F_X^-(U) \leq x\right) = \mathbb{P}\left(U \leq F_X(x)\right) = F_X(x).$$

■

Références

C. P. Robert, and G. Casella. 2004. *Monte-Carlo Statistical Methods*. 2nd edition, Berlin: Springer.

G.E.P. Box, and M.E. Muller. 1958. "A Note on the Generation of Random Normal Deviates." *Ann. Math. Stat.* 29: 610–11.

Lehmer, Derrick H. 1951. "Mathematical Methods in Large-Scale Computing Units." *Annu. Comput. Lab. Harvard Univ.* 26: 141–46.

Ripley, Brian D. 1987. *Stochastic Simulation*. New York: Wiley.

Von Neumann, John. 1951. "Various Techniques Used in Connection with Random Digits." *U.S. Nat. Bur. Stand. Math. Ser.* 12: 36–38.