TP5 : La méthode du rejet

11 Cas particuler de la méthode du rejet

On souhaite simuler une variable aléatoire réelle ayant une loi de densité $f : \mathbb{R} \to [0; K]$, où K est un réel fixé et f est une fonction mesurable positive s'annulant en dehors d'un intervalle $[a; b]$ avec $a < b$ (ainsi, $f(x) = 0$ si $x \notin [a, b]$).

Exercice 17. —

1. Écrire un programme `uniforme(a, b)` qui simule une variable aléatoire uniforme sur l'intervalle $[a; b]$, c'est-à-dire de densité $(b-a)^{-1}\,\mathbf{1}_{a<x<b}$.
2. Écrire un programme `points_uniformes(a, b, K)` qui renvoie un couple `(x,y)` aléatoire dont la loi est uniforme dans le rectangle $[a; b] \times [0; K]$.
3. On regarde la fonction $f(x) = cx^2\,\mathbf{1}_{1<x<2}$ sur l'intervalle $[a; b] = [1; 2]$. Que vaut c pour que f soit une densité de probabilité ? Quelle est la plus petite constante K telle que $f(x) \leqslant K$ pour tout x ? Quelle est la fonction de répartition de f ?
4. Écrire un programme `nuage_points(N, f, a, b, K)` qui :
   (a) Simule N points uniformes dans le rectangle $[a; b] \times [0; K]$.
   (b) Dessine dans le rectangle en bleu les points (x, y) obtenus tels que $f(x) < y$, et en rouge ceux tels que $f(x) > y$. On pourra utiliser `plt.scatter(x,y)` qui dessine un nuage de points dont les abscisses sont données par le tableau unidimensionnel x et les ordonnées par le tableau unidimensionnel y.
   (c) Superpose à ce nuage de points le graphe de la fonction f.

   Utiliser ce programme pour $N = 1000$ et a, b, f, K donnés dans la question précédente.
5. Soient $Z_1, Z_2, \ldots$ des points aléatoires indépendants de loi uniforme dans le rectangle $[a; b] \times [0; K]$. Ainsi, si $Z_i = (X_i, Y_i)$, alors toutes les variables X_i et Y_i sont indépendantes, chaque X_i est uniforme sur $[a; b]$ et chaque Y_i est uniforme sur $[0; K]$. On note

   $$M = \inf(\{i \in \mathbb{N}^* : f(X_i) > Y_i\}).$$

   (a) Quelle est la loi de M ? On pourra calculer $\mathbb{P}(M > n)$.
   (b) Quelle est la loi de X_M ? Calculer $\mathbb{P}(X_1 < t \text{ et } f(X_1) > Y_1)$ et en déduire $P(X_M < t)$.

   Remarque : on peut montrer que X_M et M sont indépendants.
6. Écrire un programme `point_rejet(f, a, b, K)` qui simule la variable aléatoire (X_M, Y_M). Autrement dit, on engendre des points (x, y) uniformes dans le rectangle $[a; b] \times [0; K]$ jusqu'à ce que l'un d'entre eux vérifie $f(x) > y$, et on renvoie ce premier couple admissible.
7. Écrire un programme `nuage_points_rejet(N, f, a, b, K)` qui :
   (a) Simule N points obtenus par le programme `point_rejet(f,a,b,K)`.
   (b) Dessine ces points dans le plan.

(c) Superpose à ce nuage de points le graphe de la fonction f.

Utiliser ce programme pour $N = 1000$ et les mêmes a, b, f, K que précédemment.

8. Écrire un programme `verification(N, f, F, a, b, K)` qui :

   (a) Simule un échantillon de variables aléatoires réelles indépendantes $(X_1, \dots, X_N)$, où chaque X_i a la loi de X_M.

   (b) Trace la fonction de répartition empirique de cet échantillon.

   (c) Lui superpose F, la fonction de répartition de la loi de densité f (question 3).

12 Méthode du rejet général pour les lois à densité sur $\mathbb{R}$

Théorème 5. *Soient f et g deux densités sur $\mathbb{R}$ telles que $f(x) \leqslant Kg(x)$ pour un certain $K > 0$ et pour tout x dans $\mathbb{R}$. Soient $X_1, X_2, \dots$ des variables aléatoires indépendantes de loi de densité g, et $U_1, U_2, \dots$ des variables aléatoires indépendantes de loi uniforme sur $[0; 1]$. On pose $Y_i = K\, g(X_i)\, U_i$. Alors :*

1. *Les points $Z_i = (X_i, Y_i)$ sont indépendants et de loi uniforme sur le domaine*
$$\{(x, y) \in \mathbb{R}^2 : 0 < y < Kg(x)\},$$
qui est d'aire finie K.

2. *Soit $M = \inf\{i \in \mathbb{N}^* : f(X_i) > Y_i\}$. La variable aléatoire M est indépendante de (X_M, Y_M), et a une loi géométrique de paramètre K^{-1}. Elle est indépendante de (X_M, Y_M), qui a la loi uniforme sur le domaine*
$$\{(x, y) \in \mathbb{R}^2 : 0 < y < f(x)\}.$$

3. *La variable X_M a la loi de densité f.*

Ainsi, si on connaît les fonctions f et g et si l'on sait simuler des variables aléatoires de loi g, alors on a une méthode pour de simuler des variables aléatoires de loi f.

Exercice 18. — On souhaite utiliser ce résultat théorique pour simuler une variable gaussienne, qui a une densité qui ne s'annule pas hors d'un compact.

1. Soit X une variable aléatoire de loi exponentielle $\mathcal{E}(1)$, et Z une variable aléatoire indépendante de X telle que $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$. Montrer que XZ a une loi de densité $g(x) = \frac{1}{2} \exp(-|x|)$ (appelée loi de Laplace). En déduire un programme `laplace()` qui simule une variable aléatoire de loi de Laplace.

2. On souhaite simuler une variable aléatoire gaussienne, donc de densité $f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$. Quel est le plus petit réel K tel que $f(x) \leqslant Kg(x)$ pour tout x ?

3. Écrire un programme `points_domaine(N, simg, g,K)` qui prend en paramètres un programme `simg` qui simule une variable aléatoire de loi g, et qui engendre N point aléatoires uniformes sur le domaine
$$\{(x, y) \in \mathbb{R}^2 : 0 < y < Kg(x)\}.$$
Tracer grâce à ce programme $N = 1000$ points avec K trouvé à la question précédente. Afficher en bleu les points tels que $f(x) < y$, en rouge ceux tels que $f(x) > y$, et superposer la densité de la gaussienne.

4. Écrire un programme `rejet(simg, g, K, f)` simule le point (X_M, Y_M) du théorème ; `simg` est un programme qui simule une variable aléatoire de loi g, et on suppose que $f \leqslant Kg$. Utiliser ce programme pour représenter $N = 1000$ points indépendants ayant la loi de (X_M, Y_M), et superposer la densité de la gaussienne.

5. Tracer sur un même graphique la fonction de répartition empirique d'un échantillon de taille $N = 1000$ de variables aléatoires ayant la loi de X_M, et la fonction de répartition de la gaussienne donnée par la commande `scipy.stats.norm.cdf`.