Séance 2
Simulation d'une distribution quelconque

Durant la première séance d'exercices, nous avons appris à générer des nombres aléatoires suivant une distribution uniforme sur l'intervalle $[0,1[$. Nous allons à présent utiliser cela pour simuler d'autres distributions plus intéressantes, le but étant de pouvoir générer n'importe quelle distribution dont on a l'expression analytique.

Nous verrons deux méthodes de génération dans cette séance. La première, la *méthode de la transformée inverse*, est la plus efficace mais ne peut pas être appliquée dans tous les cas. La deuxième, la *méthode de réjection de von Neumann*, fonctionne à tous les coups mais peut être très inefficace. Le troisième exercice vous fera investiguer des pistes pour améliorer l'efficacité de cette méthode.

1. Méthode de la transformée inverse : Cette méthode est la seule qui garantit une efficacité de 100 %, mais elle nécessite que l'on soit capable de calculer la fonction cumulative de la fonction de distribution (et donc de connaître une primitive), et de calculer sa réciproque. C'est le cas pour des distributions exponentielle, (co)sinusoïdale ou polynomiale de degré plus élevé. Dans les autres cas, il est nécessaire d'employer la méthode de réjection, comme nous le verrons à l'exercice suivant.

Le principe de la méthode de la transformée inverse est le suivant. Soit $f(x)$ la fonction de densité de probabilité de x sur l'intervalle $[a,b]$. Puisqu'il s'agit d'une fonction de densité de probabilité, son intégrale vaut 1 :

$$\int_a^b f(x)\,\mathrm{d}x = 1.$$

On définit alors la *fonction cumulative* $F(x)$ comme l'intégrale de $f(x)$ jusqu'à la valeur x :

$$F(x) \equiv \int_a^x f(\xi)\,\mathrm{d}\xi.$$

Par construction, $F(x)$ est toujours comprise entre 0 et 1, et on a que $F(a) = 0$ et $F(b) = 1$.
Nous allons à présent prouver que si u suit une loi uniforme sur $[0,1[$, alors $x = F^{-1}(u)$ est distribuée selon $f(x)$.
<u>Démonstration :</u> par changement de variables et conservation de la probabilité, on a :

$$f(u)\,\mathrm{d}u = f(x)\,\mathrm{d}x \quad \Rightarrow \quad f(x) = f(u)\cdot\frac{\mathrm{d}u}{\mathrm{d}x} = 1\cdot\frac{\mathrm{d}u}{\mathrm{d}x}$$

puisque u suit une loi uniforme sur $[0,1[$. D'autre part, par définition :

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}.$$

Donc, $F(x) = u$, autrement dit $x = F^{-1}(u)$.
Pour générer x à partir d'une variable aléatoire distribuée selon une loi uniforme, il « suffit » donc de calculer la fonction inverse de la fonction cumulative de $f(x)$.

a) Générez une loi exponentielle de paramètre λ en utilisant la méthode de la transformée inverse, et fittez la distribution obtenue avec la forme attendue de la fonction.
b) [**Devoir**] Générez une loi en $\frac{1}{x}$ sur l'intervalle $[0{,}001; 10]$. N'oubliez pas de normaliser proprement la fonction de densité de probabilité.
c) [**Devoir**] Générez une loi de Cauchy en utilisant la méthode de la transformée inverse.

2. Méthode de réjection de von Neumann : Il s'agit de la méthode « par défaut », qui peut permettre de générer n'importe quelle distribution. Le problème est qu'elle n'a pas une efficacité de 100 % et est donc plus lente au niveau du temps de calcul.

On veut générer des variables aléatoires x qui suivent une fonction de densité de probabilité $f(x)$ sur l'intervalle $[a,b]$. La procédure à suivre est la suivante :

1. On détermine $f_{\max}$, la valeur maximale que peut prendre $f(x)$ sur l'intervalle $[a, b]$.
2. On génère x_i uniformément sur l'intervalle $[a, b]$.
3. On génère y_i uniformément sur l'intervalle $[0, f_{\max}]$.
4. Si $f(x_i) > y_i$, alors on garde x_i. Sinon, on rejette la valeur et on recommence.

Les nombres x_i acceptés suivent alors exactement la distribution de $f(x)$. On constate facilement que l'efficacité de la méthode (la fraction d'événements qui seront acceptés) est :

$$\mathcal{E} = \frac{\displaystyle\int_a^b f(x)\,\mathrm{d}x}{f_{\max} \cdot (b-a)}.$$

a) Utilisez la méthode de réjection pour simuler la distribution

$$f(x) = \frac{3}{8}\left(1 + x^2\right)$$

sur l'intervalle $[-1, 1]$. Fittez la distribution obtenue. Calculez l'efficacité attendue de la méthode de réjection et comparez-la avec l'efficacité observée.

b) **[Devoir]** Utilisez la méthode de réjection pour simuler la distribution bidimensionnelle

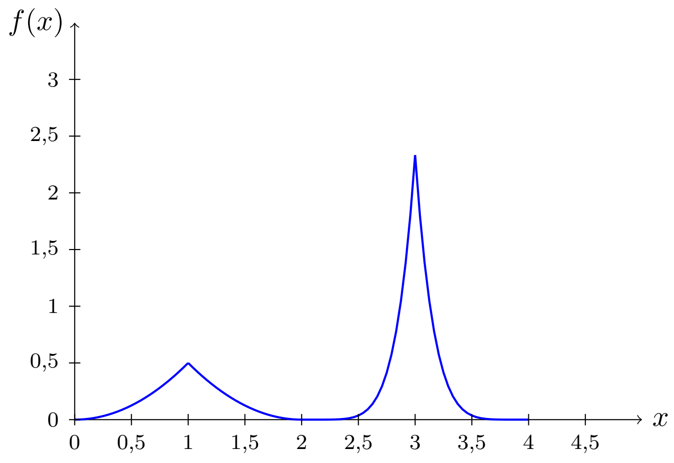
$$f(x, y) = \frac{1}{12}(3 - y)\left(1 + x^2\right),$$

pour $x \in [-1, 1]$ et $y \in [0, 3]$. Quelle est l'efficacité de la méthode ?

3. Limitations et améliorations de la méthode de réjection (gros devoir) : Dans cet exercice, nous allons rencontrer un cas où la méthode de réjection donne une très mauvaise efficacité. Nous allons voir comment améliorer la méthode pour ce genre de situations.

a) Utilisez la méthode de réjection pour simuler la distribution $f(x)$ définie sur l'intervalle $x \in [0, 4]$ par :

$$f(x) = \begin{cases} \frac{1}{2}x^2 & \text{si } 0 \leq x \leq 1 \\ \frac{1}{2}(x-2)^2 & \text{si } 1 < x \leq 2 \\ \frac{7}{3}(x-2)^6 & \text{si } 2 < x \leq 3 \\ \frac{7}{3}(x-4)^6 & \text{si } 3 < x \leq 4 \end{cases}$$



b) Estimez l'efficacité de la méthode à partir des simulations et de manière analytique.

c) Trouvez un moyen d'augmenter l'efficacité de la simulation à 3/17 (vous savez le faire).

d) Trouvez un moyen d'augmenter l'efficacité de la simulation à 6/17 (difficile).

e) Peut-on mieux faire avec la méthode de la transformée inverse ?