

## Correctif de la séance 2

### Simulation d'une distribution quelconque

#### 1. Méthode de la transformée inverse :

a) La fonction de densité de probabilité pour une exponentielle est

$$f(x) = \lambda e^{-\lambda x},$$

pour  $x \in [0, +\infty[$ . Calculons la fonction cumulative :

$$\begin{aligned} F(x) &= \lambda \int_0^x e^{-\lambda \xi} d\xi \\ &= \left[ -e^{-\lambda x} \right]_0^x \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

Employons à présent la méthode de la transformée inverse : on sait que

$$u = 1 - e^{-\lambda x}$$

est distribué suivant une loi uniforme. On a donc :

$$\begin{aligned} 1 - u &= e^{-\lambda x} \\ \Rightarrow -\lambda x &= \ln(1 - u) \\ \Rightarrow x &= \frac{-1}{\lambda} \ln(1 - u). \end{aligned}$$

Il suffit donc de générer une valeur de  $u$  en suivant une loi uniforme sur  $[0, 1[$ , puis de calculer  $\frac{-1}{\lambda} \ln(1 - u)$  pour avoir une distribution exponentielle de paramètre  $\lambda$ . Notez que comme  $u < 1$ , on n'a pas de problème de divergence pour le logarithme.

b) La loi est de la forme :

$$f(x) = \frac{N}{x},$$

où il faut déterminer la constante  $N$  pour que l'intégrale sur tout l'intervalle  $[0,001; 10]$  soit 1. On a donc :

$$N \cdot \int_{0,001}^{10} \frac{1}{\xi} d\xi = N \cdot \ln\left(\frac{10}{0,001}\right) = N \cdot 9,210 = 1 \quad \Rightarrow \quad N = 0,109.$$

La fonction cumulative vaut :

$$F(x) = N \cdot \int_{0,001}^x \frac{1}{\xi} d\xi = N \cdot (\ln x - \ln 0,001) = 0,109 \ln x + \frac{3}{4}.$$

En employant la méthode de la transformée inverse :

$$\begin{aligned} u &= 0,109 \ln x + \frac{3}{4} \\ \Rightarrow \frac{u - \frac{3}{4}}{0,109} &= \ln x \\ \Rightarrow x &= \exp\left(\frac{u - \frac{3}{4}}{0,109}\right). \end{aligned}$$

En générant cette loi, on doit obtenir une distribution en  $N \cdot \frac{1}{x}$ . Notez qu'il peut être important de bien choisir le binning et la valeur initiale des paramètres pour l'ajustement, sous peine que le fit ne converge pas.

c) La fonction de densité de probabilité d'une loi de Cauchy (ou de Lorentz) est :

$$f(x) = \frac{1}{\pi} \left[ \frac{a}{(x - x_0)^2 + a^2} \right].$$

Cette distribution est notamment utilisée en spectroscopie pour modéliser des raies d'émissions. Le quotient de deux variables aléatoires normales suit une loi de Cauchy.

Calculons la fonction de distribution. Le domaine de la fonction est  $x \in ]-\infty, +\infty[$ . On a donc :

$$F(x) = \int_{-\infty}^x \frac{1}{\pi} \left[ \frac{a}{(x - x_0)^2 + a^2} \right] dx.$$

En posant le changement de variables  $y \equiv \frac{x-x_0}{a}$ , on a :

$$\begin{aligned} F(x) &= \int_{-\infty}^{\frac{x-x_0}{a}} \frac{1}{\pi} \left[ \frac{a}{(ay)^2 + a^2} \right] a dy \\ &= \frac{1}{\pi} \int_{-\infty}^{\frac{x-x_0}{a}} \frac{dy}{y^2 + 1} \\ &= \frac{1}{\pi} \left[ \arctan y \right]_{-\infty}^{\frac{x-x_0}{a}} \\ &= \frac{1}{\pi} \arctan \left( \frac{x - x_0}{a} \right) + \frac{1}{2}. \end{aligned}$$

Notez que  $F(x) \rightarrow 1$  lorsque  $x \rightarrow +\infty$ , comme il se doit. On peut maintenant employer la méthode de la transformée inverse :

$$\begin{aligned} u &= \frac{1}{\pi} \arctan \left( \frac{x - x_0}{a} \right) + \frac{1}{2} \\ \Rightarrow \pi \left( u - \frac{1}{2} \right) &= \arctan \left( \frac{x - x_0}{a} \right) \\ \Rightarrow x &= x_0 + a \cdot \tan \left( \pi \cdot \left( u - \frac{1}{2} \right) \right). \end{aligned}$$

On sait donc générer une loi de Cauchy par méthode de la transformée inverse.

## 2. Méthode de réjection de von Neumann :

a) Commençons par vérifier la normalisation : on a bien

$$\int_{-1}^1 \frac{3}{8} (1 + x^2) dx = \left[ \frac{3}{8} \left( x + \frac{x^3}{3} \right) \right]_{-1}^1 = \frac{3}{8} \left( 1 + \frac{1}{3} \right) - \frac{3}{8} \left( -1 - \frac{1}{3} \right) = 1.$$

On peut aussi se baser sur ce calcul pour voir s'il y a moyen d'employer la méthode de la transformée inverse sur cette fonction : cela reviendrait à inverser l'équation

$$u = \frac{x^3}{8} + \frac{3}{8}x + \frac{1}{2}$$

pour isoler  $x$ . C'est faisable (il existe une méthode générale pour trouver la solution d'une équation du troisième degré), mais nous allons préférer ici employer la méthode de réjection, qui est plus "naturelle".

La première étape pour appliquer la méthode de réjection est de trouver la valeur maximale atteinte par la fonction. Dans notre cas, cette valeur est atteinte aux bornes, en  $x = \pm 1$ . On a donc  $f_{\max} = \frac{3}{8}(1 + 1) = \frac{3}{4}$ . Pour appliquer la méthode de réjection, on doit donc générer  $x$  uniformément sur  $[-1, 1]$  et  $y$  sur  $\left[0, \frac{3}{4}\right]$ .

L'efficacité attendue de la méthode de réjection est donnée par l'intégrale de la fonction divisée par la taille de la "boîte" qu'on emploie, autrement dit dans notre cas :

$$\mathcal{E} = \frac{1}{\frac{3}{4} \cdot (1 + 1)} = 67\%.$$

b) L'idée pour générer une distribution bidimensionnelle est la même que pour une unidimensionnelle ; on génère simplement 3 variables au lieu de 2.

La valeur maximale de la fonction est atteinte en  $x = \pm 1$  et  $y = 0$  ; on trouve ainsi  $f_{\max} = \frac{1}{12} \cdot 3 \cdot 2 = \frac{1}{2}$ .

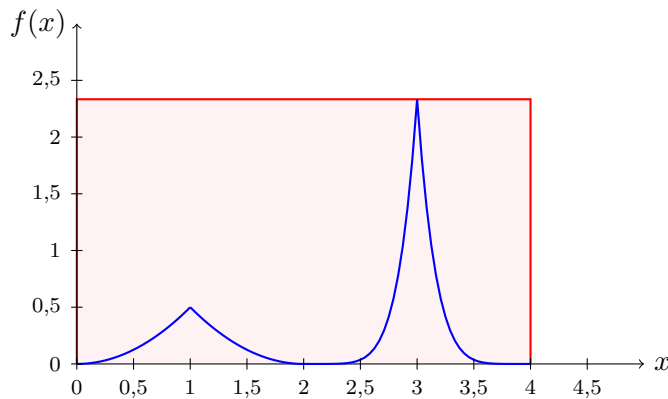
On doit donc générer 3 variables aléatoires qui suivent une distribution uniforme :  $x$  sur  $[-1, 1]$ ,  $y$  sur  $[0, 3]$  et  $z$  sur  $[0, \frac{1}{2}]$ . On garde le point généré uniquement si  $z < f(x, y)$ , comme pour la méthode à une dimension.

On peut vérifier facilement que la distribution est bien normalisée à 1 sur le domaine donné. L'efficacité attendue est donc de :

$$\mathcal{E} = \frac{1}{\frac{1}{2} \cdot 3 \cdot 2} = 33 \%$$

### 3. Limitations et améliorations de la méthode de réjection :

a) On peut se convaincre facilement que la fonction atteint son maximum en  $x = 3$  ; on a donc  $f_{\max} = \frac{7}{3}$ . Pour employer la méthode de réjection, il suffit donc de générer  $x$  uniformément sur  $[0, 4]$  et  $y$  uniformément sur  $[0, \frac{7}{3}]$ . Graphiquement, cela correspond à prendre ceci :



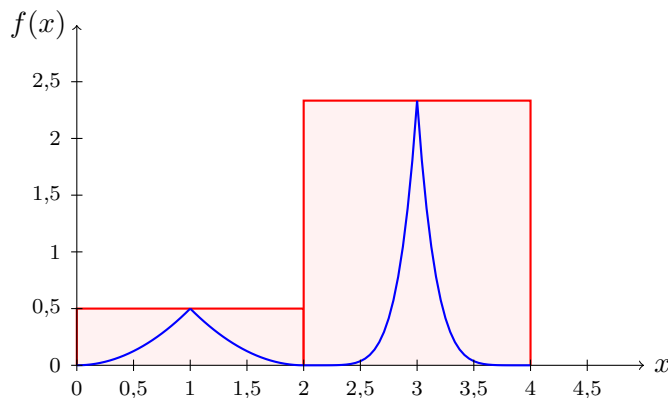
b) On peut vérifier facilement que la distribution est bien normalisée à 1 sur l'intervalle considéré. L'efficacité attendue est donc :

$$\mathcal{E} = \frac{1}{\frac{7}{3} \cdot 4} = \frac{3}{28} = 10,7 \%$$

Notre méthode de réjection est donc très inefficace !

c) Il est clair que l'on a deux régions distinctes et qu'il est inutile de générer  $y$  au-dessus de la hauteur du premier "pic" ( $\frac{1}{2}$ ) sur la région  $x \in [0, 2]$ . Une première méthode assez simple et naturelle pour améliorer l'efficacité est de générer non pas une, mais deux "boîtes" différentes, comme montré sur la figure ci-dessous :

- Première "boîte" :  $x \in [0, 2]$ ,  $y \in [0, \frac{1}{2}]$  ;
- Deuxième "boîte" :  $x \in [2, 4]$ ,  $y \in [0, \frac{7}{3}]$ .

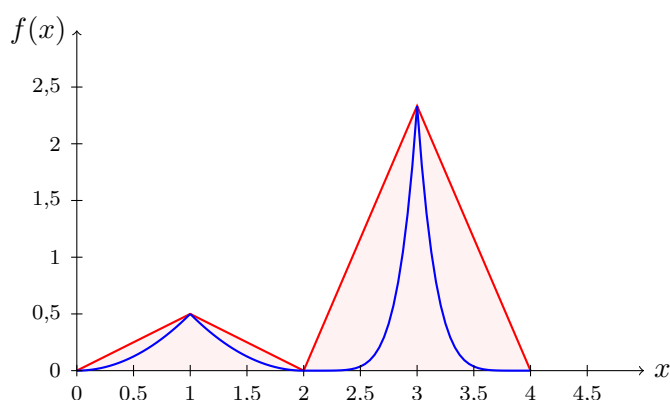


Il reste encore la question de décider dans quelle boîte on se place. Pour ce faire, il faut constater que l'aire totale des deux rectangles qu'on emploie est de  $\frac{1}{2} \times 2 + \frac{7}{3} \times 2 = \frac{17}{3}$ . On est donc dans la première boîte dans  $1/\frac{17}{3} = \frac{3}{17}$  des cas, et dans la deuxième boîte le reste du temps. Il suffit donc de générer un troisième nombre aléatoire  $z$  entre 0 et 1, qui va déterminer si on se place dans la première boîte (si  $z < \frac{3}{17}$ ) ou dans la deuxième (le reste du temps).

De cette manière, l'efficacité attendue devient :

$$\mathcal{E} = \frac{3}{17} \cdot \frac{1/3}{1} + \frac{14}{17} \cdot \frac{2/3}{14/3} = \frac{3}{17} = 17,6\%.$$

- d) En poursuivant l'idée de la sous-question précédente, qui est d'utiliser des boîtes qui suivent davantage la forme de la fonction qu'on veut générer, il est clair qu'employer des boîtes rectangulaires n'est pas la solution optimale : au vu du graphe de la fonction, il est évident que l'on gagnerait beaucoup en efficacité en tirant les points sur des *triangles* plutôt que sur des rectangles, ce qui correspondrait aux boîtes en rouge sur la figure ci-dessous.



La difficulté est dès lors de générer une variable aléatoire qui suit une loi triangulaire (et plus uniforme). Pour ce faire, il faut savoir qu'une distribution triangulaire est un cas particulier de loi de Irwin-Hall de paramètre 2, et donc qu'une somme de deux variables aléatoires indépendantes distribuées uniformément sur  $[0, 1[$  suit une loi triangulaire.

(Il est laissé à titre d'exercice de vérifier à partir de la forme générale de la loi de Irwin-Hall qu'on arrive bien à  $f(x) = 1 - |1 - x|$ , ce qui correspond bien à un triangle, en prenant comme paramètre 2.)

Le reste de la méthode est identique au point précédent : il suffit d'utiliser encore un autre nombre aléatoire pour savoir dans quel triangle on se place. Notez qu'il faut utiliser l'expression de la fonction triangulaire pour déterminer  $f_{\max}$ , qui dépend évidemment de la valeur de  $x$ .

Puisque l'aire du triangle est la moitié de l'aire du rectangle, l'efficacité de la méthode est multipliée par 2, soit  $\frac{6}{17} = 35,3\%$ .

- e) La fonction étant polynomiale (d'ordre 7 pour la dernière partie de la fonction), il y a toujours moyen de l'inverser même si c'est de manière numérique et pas analytique. Il est donc possible d'employer la méthode de la transformée inverse et d'atteindre une efficacité de 100 %.