

4 Générateur d'une variable aléatoire suivant une loi donnée

On remarque que si x est une variable aléatoire de densité $\text{rect}_{[0,1]}(x)$, $px + q$ est une v.a. de densité $\frac{1}{p}\text{rect}_{[q, p+q]}(x)$ (le facteur $\frac{1}{p}$ assurant la normalisation à 1). De façon plus générale, comment, à partir d'une v. a. de densité uniforme, construire une v. a. de densité quelconque ?

4.1 Méthode de l'inversion de la primitive

4.1.1 Principe

1) Soit x une v. a. de densité $f(x)$. Par définition de la densité, la probabilité dp pour que x soit dans l'intervalle $[x, x + dx]$ est :

$$dp = f(x) dx$$

Soit une fonction $y(x)$: quelle est la densité $g(y)$ de la v. a. y ? On a :

$$x \in [x, x + dx] \implies y \in [y, y + dy]$$

donc, par définition de la densité de y :

$$dp = g(y) dy = f(x) dx \implies g(y) = f(x) \frac{dx}{dy}$$

2) On envisage la réciproque du problème précédent : quelle doit être la fonction $y(x)$ pour que, x ayant la densité $f(x)$, y ait la densité $g(y)$? On a vu que :

$$g(y) \frac{dy}{dx} = f(x)$$

Soit G une primitive de g , F une primitive de f :

$$\frac{d}{dx} G(y) = g(y) \frac{dy}{dx} = f(x) = \frac{dF}{dx} \implies G(y) = F(x) + \text{cte}$$

On se place dans le cas où G^{-1} existe et est calculable⁵. On peut alors écrire :

$$y = G^{-1} [F(x) + \text{cte}]$$

5. Sinon il faudra employer la méthode de Von Neuman exposée ci-dessous

Il est normal de trouver une constante arbitraire puisqu'on a résolu une équation différentielle du premier ordre. Le choix de la valeur y_0 que l'on veut attribuer à y pour une valeur particulière de x notée x_0 détermine la constante qui vaut : $G(y_0) - F(x_0)$. Finalement :

$$y = G^{-1} [F(x) - F(x_0) + G(y_0)]$$

Cas particulier :

si $f(x) = \text{rect}_{[0,1]}(x)$ une primitive de $f(x)$ est $F(x) = x$ et :

$$y = G^{-1} [x - x_0 + G(y_0)]$$

4.1.2 Exemple

Prenons :

$$f(x) = \text{rect}_{[0,1]}(x)$$

$$g(y) = \frac{1}{\alpha} \exp\left(-\frac{y}{\alpha}\right) \text{ pour } y \geq 0 \text{ et } 0 \text{ ailleurs } (\alpha > 0)$$

$g(y)$ est la loi exponentielle que l'on rencontre fréquemment en physique.

α est un paramètre qui représente la valeur moyenne de y :

$$\langle y \rangle = \frac{1}{\alpha} \int_0^{+\infty} y \exp\left(-\frac{y}{\alpha}\right) dy = \alpha$$

Une primitive de $g(y)$ est $G(y) = -\exp\left(-\frac{y}{\alpha}\right)$ dont la fonction réciproque est $G^{-1}(z) = -\alpha \ln(-z)$. Donc :

$$y = G^{-1} [x - x_0 + G(y_0)] = -\alpha \ln \left[-x + x_0 + \exp\left(-\frac{y_0}{\alpha}\right) \right]$$

Choisissons par exemple $y(0) = 0$, on a alors $x_0 = 0$ et $y_0 = 0$ et :

$$y = -\alpha \ln(-x + 1)$$

$1 - x$ et x ayant la même distribution $\text{rect}_{[0,1]}(x)$ on peut écrire plus simplement :

$$y = -\alpha \ln(x)$$

On pourra donc, dans un programme, tirer des valeurs aléatoires dont la densité suit une loi exponentielle de paramètre α par l'instruction :

```
...
y = -alpha*log(alea());
...
```

4.2 Méthode du rejet de Von Neuman

Dans le cas où la primitive de la densité recherchée n'est pas inversible et calculable, on applique une méthode beaucoup plus générale.

Soit x une v. a. de loi $f(x)$ sur $[a, b]$ et soit M la valeur supérieure⁶ de $f(x)$ sur $[a, b]$.

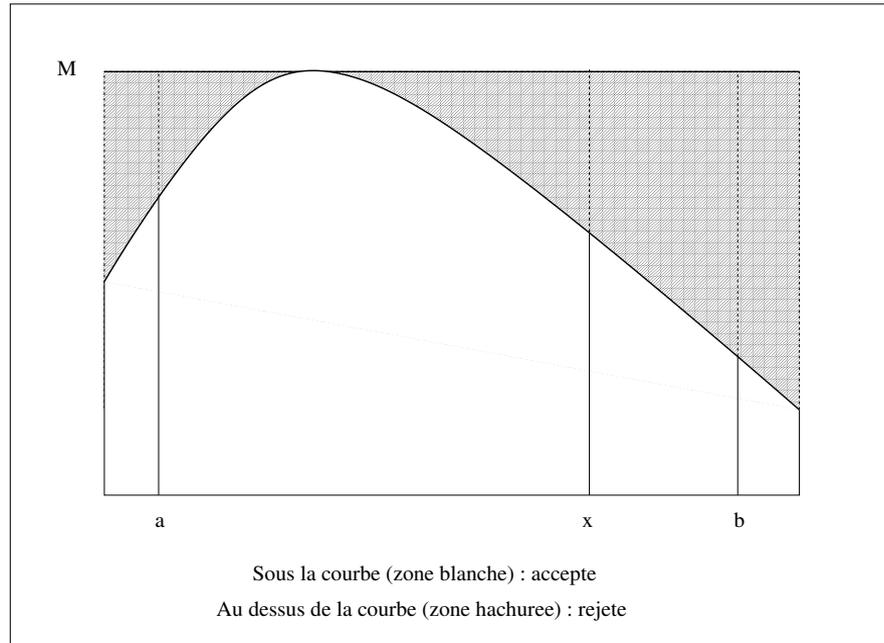


FIGURE 9 –

On tire une valeur aléatoire x avec une densité uniforme sur $[a, b]$, puis une valeur aléatoire y avec une densité uniforme sur $[0, M]$. Si le point de coordonnées (x, y) est au-dessus de la courbe de $f(x)$ on le rejette, sinon on l'accepte. Les valeurs de x ainsi sélectionnées suivent la loi de densité $f(x)$. En effet, supposons qu'on ait effectué N_0

6. supposée exister

doubles tirages (x et y). Le nombre de valeurs de x situées dans l'intervalle $[x, x + dx]$ est :

$$dN = N_0 \underbrace{\frac{dx}{b-a}}_{1^{\text{er}} \text{ tirage}} \underbrace{\frac{f(x)}{M}}_{2^{\text{ème}} \text{ tirage}}$$

Pour déduire de dN la densité de probabilité de x , il faut diviser dN par le nombre N de valeurs de x effectivement obtenues, qui est différent de N_0 puisqu'il y a des valeurs rejetées. Le nombre de valeurs de x retenues est :

$$N = \int_a^b dN = \frac{N_0}{(b-a)M} \int_a^b f(x) dx$$

$f(x)$ étant une densité de probabilité on a :

$$\int_a^b f(x) dx = 1$$

soit :

$$N = \frac{N_0}{(b-a)M}$$

Donc :

$$dp = \frac{dN}{N} = f(x) dx$$

La densité de probabilité de x est bien $f(x)$.

Remarque

En pratique si, au lieu d'utiliser $f(x)$ qui est normée à 1, on utilise $f'(x) = \lambda f(x)$ (λ quelconque $\in \mathbb{R}$), le résultat est le même puisque le maximum M' de $f'(x)$ est multiplié dans le même rapport λ et on a : $f(x)/M = f'(x)/M'$. Il n'est donc pas nécessaire de travailler avec une fonction normée à 1.

Complément :

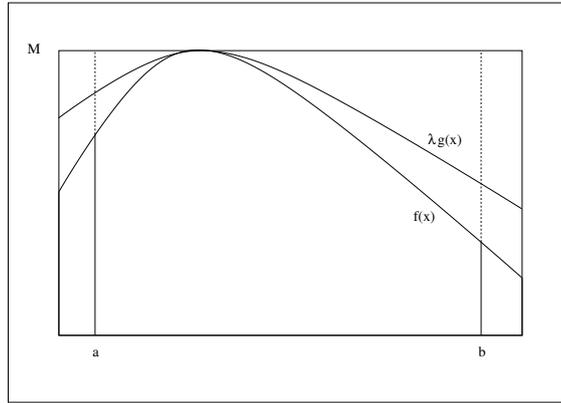


FIGURE 11 –

L'inconvénient de cette méthode est qu'il peut y avoir beaucoup de tirages inutiles. On peut y remédier si on connaît une densité $g(x)$ sur $[a, b]$ pour laquelle on sait générer des valeurs aléatoires et que cette densité satisfait à la propriété :

$$\exists \lambda \in \mathbf{R} \text{ tel que } \lambda g(x) \geq f(x) \forall x \in [a, b]$$

On procède encore à un double tirage mais :

au lieu de tirer x avec une densité uniforme sur $[a, b]$, on le tire avec la densité $g(x)$
on tire y avec une densité uniforme sur $[0, \lambda g(x)]$

puis, comme précédemment, on rejette le point s'il est au-dessus de la courbe de $f(x)$.

Pour N_0 doubles tirages le nombre de valeurs de x situées dans l'intervalle $[x, x + dx]$ est donc :

$$dN = N_0 \underbrace{g(x)dx}_{1^{\text{er}} \text{ tirage}} \underbrace{\frac{f(x)}{\lambda g(x)}}_{2^{\text{ème}} \text{ tirage}}$$

Le nombre de valeurs de x retenues est :

$$N = \int_a^b dN = \frac{N_0}{\lambda} \int_a^b f(x) dx = \frac{N_0}{\lambda}$$

Donc :

$$dp = \frac{dN}{N} = f(x) dx$$

La densité de probabilité de x est bien $f(x)$.

Remarque

On est obligé d'introduire le facteur λ car on ne pourrait avoir simultanément : $\int_a^b f(x) dx = 1$, $\int_a^b g(x) dx = 1$
et $g \geq f$ sauf si $f = g$.