HEIG–VD
Simulation et Optimisation

Bastien Clément
11 mars 2016

**Rapport : Génération de variables aléatoires** **Travail personnel 1**

# Problème 1

a) Donner une définition par morceaux de $\tilde{f}(x)$.

$$\tilde{f}(x) = \begin{cases} 0 & \text{si } x < x_0 \\ m_0(x - x_0) + y_0 & \text{si } x_0 \leq x \leq x_1 \\ 0 & \text{si } x > x_1 \end{cases}$$

b) *On désire définir une variable aléatoire $X$ de densité $f(x)$ proportionnelle à $\tilde{f}(x)$. Déterminer la constante de proportionnalité, c.à-d. déterminer la constante $A_0$ telle que $f(x) = \frac{1}{A_0}\tilde{f}(x)$ soit une densité.*

Puisque $A_0$ est une constante et que $f(x)$ est une densité, nous savons que

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{x_0}^{x_1} \frac{1}{A_0}\tilde{f}(x)\,dx = \frac{1}{A_0}\int_{x_0}^{x_1} \tilde{f}(x)\,dx = 1$$

Il n'est pas nécessaire de considérer les cas $x < x_0$ et $x > x_1$ puisque la fonction est nulle en dehors de l'intervalle $[x_0; x_1]$. Il alors est évident que

$$A_0 = \int_{x_0}^{x_1} \tilde{f}(x)dx$$

c'est à dire l'aire sous la courbe de $\tilde{f}(x)$ sur l'intervalle $[x_0; x_1]$. Dans notre cas, le problème est simplifié puisque nous avons un trapèze rectangle dont l'aire est donnée par le produit de la base moyenne et de la hauteur :

$$A_0 = \frac{y_0 + y_0}{2} \cdot (x_1 - x_0).$$

Nous pouvons finalement définir

$$\begin{aligned} f(x) &= \frac{1}{A_0} \cdot \tilde{f}(x) \\ &= \frac{2}{(y_0 + y_1)(x_1 - x_0)} \cdot \tilde{f}(x) \end{aligned}$$

c) *Calculer l'espérance de la variable $X$, sans oublier de simplifier votre résultat.*

L'espérance mathématique d'une variable aléatoire est définie par

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x f(x)\,dx$$

Dans notre cas, nous pouvons à nouveau nous limiter à l'intervalle $[x_0; x_1]$ puisque en dehors de cet intervalle, $f(x) = 0$.

$$
\begin{aligned}
\mathbb{E}[x] &= \int_{x_0}^{x_1} x f(x)\, dx \\
&= \frac{1}{A_0} \int_{x_0}^{x_1} x \tilde{f}(x)\, dx \\
&= \frac{2}{(y_0 + y_1)(x_1 - x_0)} \cdot \int_{x_0}^{x_1} x(m_0 \cdot (x - x_0) + y_0)\, dx \\
&= \frac{2}{(y_0 + y_1)(x_1 - x_0)} \cdot \left( \frac{x^2\big(2x(y_0 - y_1) + 3x_0 y_1 - 3x_1 y_0\big)}{6(x_0 - x_1)} + c \Big|_{x_0}^{x_1} \right) \\
&= \frac{2}{(y_0 + y_1)(x_1 - x_0)} \cdot \frac{(x_1 - x_0)\big(x_0(2y_0 + y_1) + x_1(y_0 + 2y_1)\big)}{6} \\
&= \frac{x_0(2y_0 + y_1) + x_1(y_0 + 2y_1)}{3(y_0 + y_1)}
\end{aligned}
$$

d) *Vérifier que, dans le cas où $y_0 \neq y_1$, la fonction de répartition de la variable $X$ est*

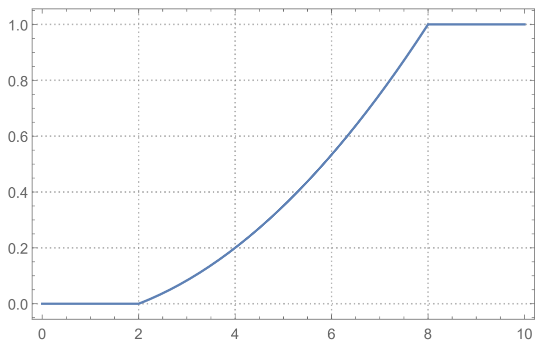
$$
F(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x < x_0 \\ \frac{1}{y_1^2 - y_0^2} \left( (m_0\,(x - x_0) + y_0)^2 - y_0^2 \right) & \text{si } x_0 \leq x \leq x_0 \\ 1 & \text{si } x > x_1 \end{cases}
$$

*Le développement du point d) n'est pas demandé.*

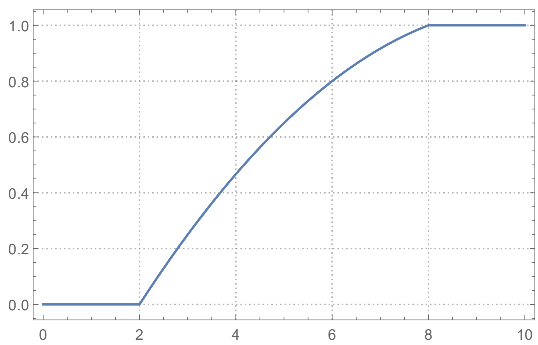
e) *Utiliser le résultat précédent pour proposer un algorithme permettant de générer des réalisations de la variable $X$ à l'aide de la méthode des fonctions inverses. Traiter séparément les cas $y_0 = y_1$ et $y_0 \neq y_1$.*

Notre fonction de répartition $F(x)$, liée à la fonction de densité $f(x)$, est définie comme une fonction $\mathbb{R} \to [0; 1]$ mais en pratique, le domaine intéressant se limite à $[x_0; x_1]$ puisque que

$$
\begin{aligned}
x < x_0 &\implies F(X) = 0 \\
x > x_1 &\implies F(X) = 1
\end{aligned}
$$




Exemple de graphe pour $F(x)$ avec $y_0 < y_1$




Exemple de graphe pour $F(x)$ avec $y_0 > y_1$

La méthode des fonctions inverses consiste à utiliser la réciproque $F^{-1}(y) : [0; 1] \to [x_0; x_1]$ et une variable $y \sim \mathcal{U}(0, 1)$ pour déterminer une réalisation $x = F^{-1}(y)$ conforme à la densité de probabilité désirée pour $X$.

Inversion de $y = F(x) \mid x \in [x_0, x_1], y_0 \neq y_1$ :

$$y = \frac{1}{y_1^2 - y_0^2} \cdot \left( (m_0 \cdot (x - x_0) + y_0)^2 - y_0^2 \right)$$

$$y \cdot (y_1^2 - y_0^2) + y_0^2 = (m_0 \cdot (x - x_0) + y_0)^2$$

$$\pm\sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2} = m_0 \cdot (x - x_0) + y_0$$

$$\frac{\pm\sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2} - y_0}{m_0} + x_0 = x$$

Dans le cas où $y_1 = y_0$, la solution est simple : la densité de probabilité définie par $f(x)$ est une uniforme sur l'intervalle $[x_0; x_1]$ que l'on peut générer très facilement à partir d'une uniforme sur l'intervalle $[0, 1]$.

$$u \sim \mathcal{U}(a, b) = a + (b - a)\big(v \sim \mathcal{U}(0, 1)\big)$$

Dans le cas contraire, nous devons déterminer le signe de la racine. En observant les graphes de la fonction $F(x)$ pour $y_0 < y_1$ et $y_0 > y_1$, on en déduit assez facilement que le signe dépend de la relation entre $y_0$ et $y_1$. Nous pouvons alors définir :

$$F^{-1}(y) : [0; 1] \to [x_0, x_1] = \begin{cases} x_0 + \frac{-y_0 + \sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2}}{m_0} & \text{si } y_0 < y_1 \\ x_0 + y \cdot (x_1 - x_0) & \text{si } y_0 = y_1 \\ x_0 + \frac{-y_0 - \sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2}}{m_0} & \text{si } y_0 > y_1 \end{cases}$$

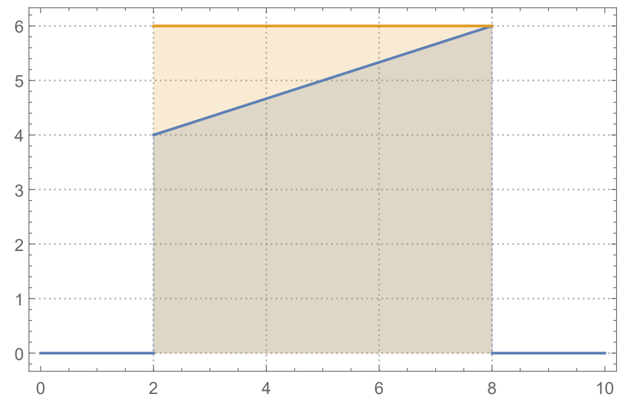
L'algorithme de génération de réalisations est alors trivial :

1 `Générer` $y \sim \mathcal{U}(0, 1)$

2 `Retourner` $F^{-1}(y)$

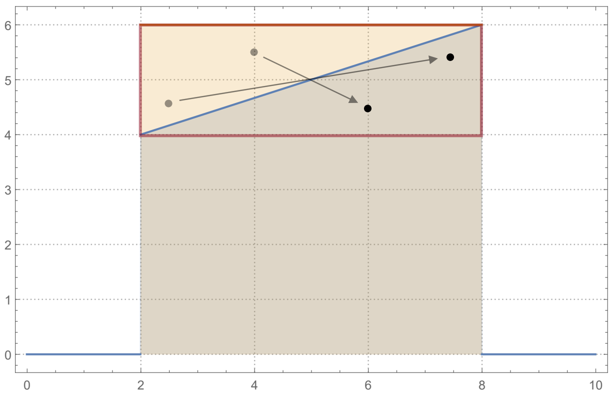
f) *Développer un algorithme permettant de générer des réalisations de la variable $X$ à l'aide d'une approche géométrique basée sur la méthode d'acceptation-rejet et en utilisant efficacement les symétries.*

La méthode d'acceptation-rejet consiste à *enfermer* notre fonction de densité $f(x)$ sous une fonction de densité $g(x)$ plus simple tel que $\forall x \in [x_0, x_1], f(x) \leq c \cdot g(x) \mid c \in \mathbb{R}$ pour laquelle nous savons générer des réalisations aléatoires. Nous retournerons alors des réalisations $y \sim g$ si $\mathcal{U}(0, 1) \leq \frac{f(y)}{cg(y)}$. Dans notre cas, il n'est même pas nécessaire de respecter strictement le concept de fonction de densité et nous pouvons utiliser directement $\tilde{f}(x)$, dont le maximum $y_m = \max(y_0, y_1)$ et générer des réalisations $(a, b) \mid a \sim \mathcal{U}(x_0, x_1), b \sim \mathcal{U}(0, y_m)$ qui correspondent à des points uniformément distribués dans le rectangle sous la courbe de la fonction $g(x) = y_m$ sur l'intervalle $[x_0, x_1]$.




Graphe de $\tilde{f}(x) = \begin{cases} \frac{1}{3}(x - 2) + 4 & \text{si } 2 \leq x \leq 8 \\ 0 & \text{sinon} \end{cases}$ et $g(x) = y_m = 6$

Nous retournerons alors $a$ si $b \leq \tilde{f}(a)$ est vérifié. Dans le cas contraire, de nouvelles réalisations de $a$ et $b$ devront être générées jusqu'à vérifier l'inégalité.
Cependant, puisque le segment $\{(x_0, y_0), (x_1, y_1)\}$ délimitant l'aire de $\tilde{f}$ est une diagonale du rectangle $\{(x_0, y_0), (x_1, y_0), (x_1, y_1), (x_0, y_1)\}$, il est possible de *rattraper le coup* lorsque le point $(a, b)$ est en dehors de l'aire de $\tilde{f}$ en effectuant une symétrie par rapport au centre du rectangle pour déplacer un point tombé dans le mauvais triangle dans le bon triangle.




Récupération de points dans le mauvais triangle

Nous avons donc l'algorithme suivant :

```
Définir $y_m = \max(y_0, y_1)$
Générer $a \sim \mathcal{U}(x_0, x_1)$ et $b \sim \mathcal{U}(0, y_m)$
Si $b \leq \tilde{f}(a)$ alors
  // Dans l'aire de $\tilde{f}(x)$
  Retourner $a$
Sinon
  // Hors de l'aire de $\tilde{f}(x)$, on prend la symétrie
  Retourner $x_1 - (a - x_0)$
Fin
```

## Problème 2