HEIG–VD
Simulation et Optimisation

Bastien Clément
15 mars 2016

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

Problème 1

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

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

$$\text{Avec } x_0 < x_1 \text{ ; } y_0, y_1 \in \mathbb{R}^+ \text{ et } m_0 = \frac{y_1 - y_0}{x_1 - x_0}$$

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 $[x_0; x_1]$. Nous en déduisons 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

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

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

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

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

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

Nous avons donc :

$$
\begin{aligned}
\mathbb{E}[x] &= \int_{x_0}^{x_1} x f(x)\, dx \\
&= \frac{1}{A_0} \int_{x_0}^{x_1} x \tilde{f}(x)\, dx \\
&= \frac{2}{(y_0+y_1)(x_1-x_0)} \cdot \int_{x_0}^{x_1} x(m_0 \cdot (x-x_0) + y_0)\, dx \\
&= \frac{2}{(y_0+y_1)(x_1-x_0)} \cdot \left[\frac{x^2\left(2x(y_0-y_1)+3x_0y_1-3x_1y_0\right)}{6(x_0-x_1)} + c \right]_{x_0}^{x_1} \\
&= \frac{2}{(y_0+y_1)(x_1-x_0)} \cdot \frac{(x_1-x_0)\left(x_0(2y_0+y_1)+x_1(y_0+2y_1)\right)}{6} \\
&= \frac{x_0(2y_0+y_1)+x_1(y_0+2y_1)}{3(y_0+y_1)}
\end{aligned}
$$

d) *Vérifier que, la fonction de répartition de la variable X est (le cas $y_0 = y_1$ a été ajouté)*

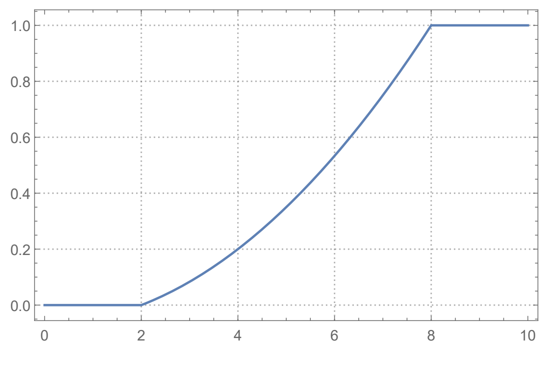
$$
F(x) = P(X \le x) = \begin{cases} 0 & \text{si } x < x_0 \\ \frac{1}{y_1^2-y_0^2}\left((m_0\,(x-x_0)+y_0)^2 - y_0^2\right) & \text{si } x_0 \le x \le x_0 \text{ et } y_0 \ne y_1 \\ \frac{1}{x_1-x_0}\,(x-x_0) & \text{si } x_0 \le x \le x_0 \text{ et } y_0 = y_1 \\ 1 & \text{si } x > x_1 \end{cases}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

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

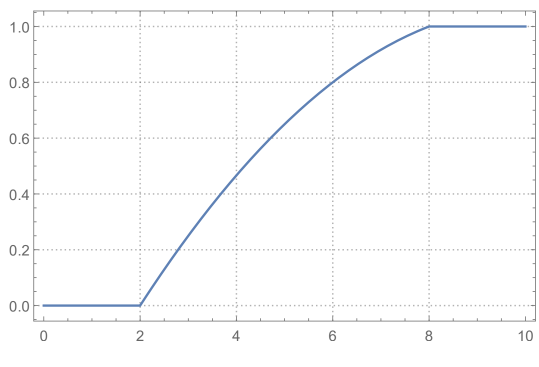
e) *Utiliser le résultat précédent pour proposer un algorithme permettant de générer des réalisations de la variable X à l'aide de la méthode des fonctions inverses.*

Notre fonction de répartition $F(x)$, liée à la fonction de densité $f(x)$, est définie comme une fonction $\mathbb{R} \to [0;1]$. En pratique cependant, 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.

Dans le cas où $y_1 = y_0$, la solution est simple : la densité de probabilité définie par f est une uniforme sur l'intervalle $[x_0; x_1]$ que l'on peut générer trivialement 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 inverser $y = F(x)$ lorsque $x \in [x_0, x_1]$ et que $y_0 \neq y_1$:

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

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

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

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

Dans le cas contraire, nous devons déterminer le signe de la racine. La fonction $F(x)$ étant croissante, son inverse $F^{-1}(x)$ l'est également. Une étude des fonctions

$$F_a^{-1}(y) = \frac{+\sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2} - y_0}{m_0} + x_0$$
$$F_b^{-1}(y) = \frac{-\sqrt{y \cdot (y_1^2 - y_0^2) + y_0^2} - y_0}{m_0} + x_0$$

révèle que la fonction F_a^{-1} est croissante, alors que la fonction F_b^{-1} est décroissante. Nous pouvons alors définir :

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

L'algorithme de génération de réalisations est alors trivial :

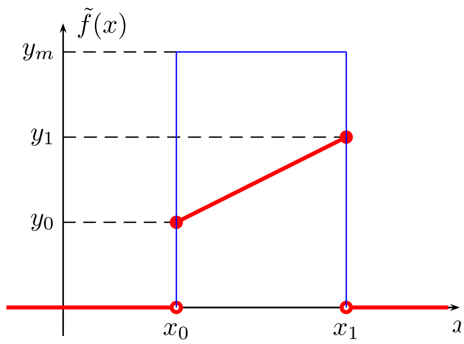
1 `Générer` $y \sim \mathcal{U}(0,1)$

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.

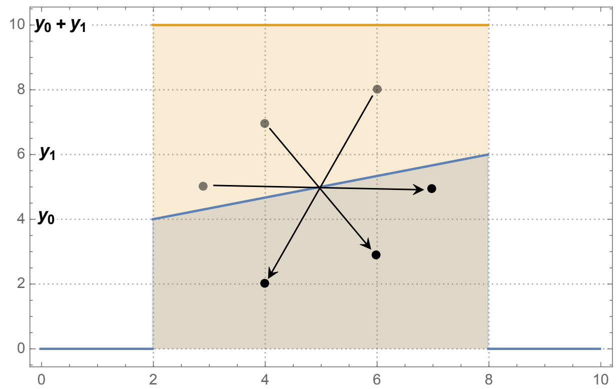
La méthode d'acceptation-rejet consiste à *enfermer* notre fonction de densité $f(x)$ sous une fonction de densité $g(x)$ plus simple tel que $\forall x \in [x_0, x_1], f(x) \leq c \cdot g(x) \mid c \in \mathbb{R}$ pour laquelle nous savons générer des réalisations aléatoires. Nous retournerons alors 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 $\tilde{f}$ et une borne supérieure $y_m = y_0 + y_1$, puis 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 $g(x) = y_0 + y_1$ sur l'intervalle $[x_0, x_1]$.



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 la fonction $\tilde{f}$ divise le rectangle $\{(x_0, 0), (x_1, 0), (x_1, y_0 + y_1), (x_0, y_0 + y_1)\}$ en deux parts égales, 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 la mauvaise part du rectangle.



Récupération de points dans la mauvaise part

Nous avons donc l'algorithme suivant, sans possibilité de rejet :

```
Générer a ~ U(x0, x1) et b ~ U(0, y0 + y1)
Si b ≤ f̃(a) alors
  // Dans l'aire de f̃(x)
  Retourner a
Sinon
  // Hors de l'aire de f̃(x), on prend la symétrie
  Retourner x1 − (a − x0)
Fin
```

Problème 2

Déterminer la constante A telle que $f = \frac{1}{A}\tilde{f}$ soit une densité

Idem que pour le premier problème, nous cherchons ici $A = \int_{-\infty}^{\infty} \tilde{f}(x)\,dx$, soit l'aire sous la courbe de notre fonction $\tilde{f}$ afin de réduire cette aire à 1. À nouveau, chaque tranche de la fonction est un trapèze rectangle dont l'aire peut être calculée facilement.

$$\begin{aligned}
A &= \int_a^b \tilde{f}(x)\,dx \\
&= \sum_{k=0}^{n-1} \int_{x_k}^{x_{k+1}} \tilde{f}(x)\,dx \\
&= \sum_{k=0}^{n-1} \frac{(y_{k+1} + y_k)}{2}(x_{k+1} - x_k) \\
&= \frac{1}{2}\sum_{k=0}^{n-1}(y_{k+1} + y_k)(x_{k+1} - x_k)
\end{aligned}$$

Développer un algorithme pour générer des réalisations de la variable X basé sur une application directe, "bête et méchante", de la méthode d'acceptation-rejet.

La première étape consiste à déterminer le maximum y_m de la fonction $\tilde{f}(x)$. Ce maximum correspond nécessairement à une valeur y_k pour $k \in \{0, 1, 2, \ldots, n\}$.

Nous générons ensuite deux variables aléatoires (x, y) indépendantes selon une loi uniforme, une fois sur l'intervalle $[a, b]$ et une fois sur l'intervalle $[0, y_m]$. Le couple (x, y) de chaque tirage correspond à un point aléatoire dans le rectangle enfermant le graphe de la fonction $\tilde{f}$.

La dernière étape est de vérifier que $b \leq \tilde{f}(a)$, c'est à dire que le point $P(x, y)$ se situe sous la courbe de $\tilde{f}$. Si ce n'est pas le cas, nous effectuons un nouveau tirage des variables x et y, jusqu'à ce que la condition soit vérifiée.

Algorithme 2.b

1 `Définir` $y_m = \max(y_0, y_1, \ldots, y_k)$

2 `Générer` $x \sim \mathcal{U}(a, b)$ `et` $y \sim \mathcal{U}(0, y_m)$

3 `Si` $y \leq \tilde{f}(x)$ `alors`

4   `Retourner` x

5 `Sinon`

6   `Sauter à la ligne 2`

7 `Fin`

Développer un algorithme pour générer des réalisations de la variable X basé sur la méthode des mélanges et sur la méthode développée au point f) du problème précédent.

Un des concepts de la méthode des mélanges est d'associer à chaque élément une probabilité p_k proportionnelle à la part d'impact de l'élément sur l'ensemble du mélange et $\sum_{k=0}^{n-1} p_k = 1$.

Dans notre cas, nous associons à la k-ième tranche de $\tilde{f}$ une probabilité p_k égale au rapport entre l'aire sous la courbe sur l'intervalle $[x_k, x_{k+1}]$ et l'aire totale sous la courbe de la fonction $\tilde{f}$.

Par la suite, nous générons un indice j obéissant à la loi discrète $P(j = k) = p_k \mid k = 0, \ldots, n-1$.

Une fois la tranche choisie, nous pouvons appliquer l'algorithme du point 1.f avec quelques substitutions pour correspondre au contexte de ce problème.

Algorithme 2.c

Pour $k = 0$ jusqu'à $n - 1$:

  Définir $\tilde{f}_k = \begin{cases} \tilde{f}(x) & \texttt{si } x_k \leq x \leq x_{k+1} \\ 0 & \texttt{sinon} \end{cases}$

  Définir $p_k = \dfrac{\int_{x_k}^{x_{k+1}} \tilde{f}_k(x)\,dx}{A} = \dfrac{(x_{k+1} - x_k)(y_{k+1} + y_k)}{2A}$

Définir $F_0 = p_0$

Pour $k = 1$ jusqu'à $n - 1$:

  Définir $F_k = F_{k-1} + p_k$

Générer $u \sim \mathcal{U}(0, 1)$ et définir $j = 0$, l'indice d'une section de $\tilde{f}$

Répéter

  Si $u \leq F_j$, définir $t = j$ et quitter la boucle

  Sinon, incrémenter j

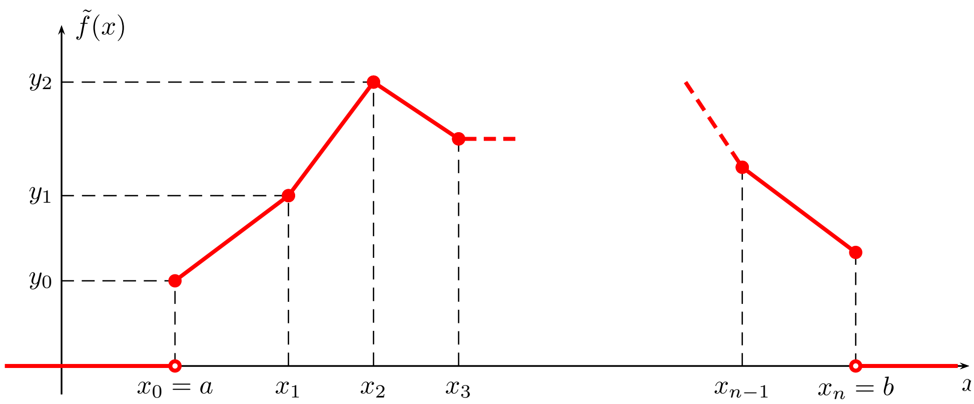
Appliquer l'algorithme 1.f sur la tranche $[x_j; x_{j+1}]$ de la fonction $\tilde{f}$, avec:

  // Substitutions dans l'algorithme du point 1.f...

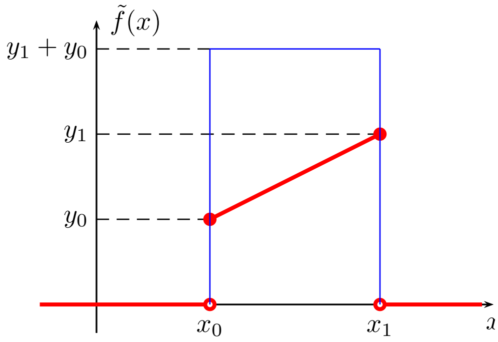
  $x_0,\, x_1 = x_j,\, x_{j+1}$

  $y_0,\, y_1 = y_j,\, y_{j+1}$

  $\tilde{f}(x) = \tilde{f}_j(x)$



foo



foo