
Sujet 2: variations autour de la méthode de Monte-Carlo

• **Méthode de la fonction de répartition inverse:**

On définit la fonction pseudo-inverse de F sur $[0, 1]$ par

$$F^{-1}(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\}$$

Proposition 1. *Si U suit une loi uniforme sur $[0, 1]$, alors $F^{-1}(U)$ a pour fonction de répartition F .*

Preuve: Commençons par montrer que $F^{-1}(u) \leq t$ ssi $u \leq F(t)$.

Soient $u \in [0, 1]$ et $t \in \mathbb{R}$ tels que $u \leq F(t)$. Par définition de la fonction de répartition inverse, on a alors $F^{-1}(u) \leq t$. Réciproquement, si $F^{-1}(u) \leq t$, alors pour tout $y > t$, $F(y) \geq u$ car F est croissante. Et puisque F est continue à droite, $F(t) \geq u$.

En utilisant ce résultat, on en déduit que

$$P(F^{-1}(U) \leq t) = P(U \leq F(t)) = F(t). \quad \square$$

Ainsi, dans le cas où F^{-1} est explicite, pour générer un échantillon X_1, \dots, X_n suivant la fonction de répartition F , on génère un échantillon (U_1, \dots, U_n) de variables uniformément distribuées sur $[0, 1]$ et on pose $X_i = F^{-1}(U_i)$.

• **Méthode de décomposition:**

On veut simuler des variables aléatoires de densité f par rapport à une mesure μ . On suppose que cette densité peut s'écrire sous la forme suivante (non unique):

$$f = \sum_{j \geq 0} p_j f_j$$

où $0 < p_j < 1$ pour tout $j \geq 0$ et $\sum_{j \geq 0} p_j = 1$ et où f_j est elle-même une densité par rapport à

la mesure μ . Soit W une variable aléatoire discrète de loi de probabilité $(p_j)_{j \geq 0}$. Soit $(Y_j)_{j \geq 0}$ une suite de variables indépendantes entre elles et indépendante de W , telle que $Y_j \sim f_j$ pour chaque $j \geq 0$. Alors $X := Y_W \sim f$.

En effet, d'après le théorème de Fubini, on a $\forall B \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}(X_W \in B) = \sum_{j \geq 0} \mathbb{P}(X_W \in B | W = j) \mathbb{P}(W = j) = \int_B \sum_{j \geq 0} p_j f_j(x) d\mu(x) = \int_B f(x) d\mu(x)$$

• **Algorithme d'acceptation-rejet (rejection-sampling):**

- On souhaite simuler un échantillon i.i.d. distribué selon une loi admettant une densité f . Considérons le cas où il est difficile voire impossible de simuler directement des variables selon la loi admettant une densité f . La méthode d'acceptation-rejet s'applique pourvu qu'il existe une densité g selon laquelle il est facile de simuler des variables et une constante M qui vérifient pour tout $x \in \text{supp}(f)$,

$$f(x) \leq Mg(x).$$

- Principe: simuler des Y_i selon une loi de proposition notée g (idéalement proche de f) selon laquelle il est facile de générer des échantillons aléatoires puis conserver seulement les Y_i qui ont une bonne chance de pouvoir être issus de f ...
- Comment mettre cela en oeuvre? Voyons comment générer une variable $X \sim f$:

Algorithme 1. 1. Générer $Y_1 \sim g$ et $U_1 \sim \mathcal{U}[0, 1]$

2. Accepter $X = Y_1$ ssi

$$U_1 \leq \frac{f(Y_1)}{Mg(Y_1)}.$$

3. Si on rejette Y_1 , recommencer en générant pour $k \geq 2$, $Y_k \sim g$ et $U_k \sim \mathcal{U}[0, 1]$ jusqu'à l'instant k_0 où $U_{k_0} \leq f(Y_{k_0})/(Mg(Y_{k_0}))$. On pose alors $X = Y_{k_0}$.

- Pourquoi ça marche?

Proposition 2. La variable aléatoire X simulée par l'algorithme 1 a pour densité f . La probabilité d'acceptation d'une proposition est égale à $1/M$.

Preuve: Calculons d'abord la probabilité d'acceptation p

$$\begin{aligned} p &= P\left(U \leq \frac{f(Y)}{Mg(Y)}\right) = \int \int \mathbb{1}_{\{u \leq \frac{f(y)}{Mg(y)}\}} f_U(u)g(y) du dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{f(y)}{Mg(y)}} f_U(u) du \cdot g(y) dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}\left(U \leq \frac{f(y)}{Mg(y)}\right) g(y) dy \\ &= \int_{-\infty}^{\infty} \frac{f(y)}{Mg(y)} \cdot g(y) dy \\ &= \frac{1}{M} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{M} \end{aligned}$$

Calculons maintenant la loi de X .

$$\begin{aligned}
 P(X \leq x) &= \sum_{n=1}^{\infty} \mathbb{P}(\{X = Y_n\} \cap \{X \leq x\}) \\
 &= \sum_{n=1}^{\infty} \mathbb{P} \left(\bigcap_{k=1}^{n-1} \left\{ U_k > \frac{f(Y_k)}{Mg(Y_k)} \right\} \cap \left\{ U_n \leq \frac{f(Y_n)}{Mg(Y_n)} \right\} \cap \{Y_n \leq x\} \right) \\
 &= \sum_{n=1}^{\infty} (1-p)^{n-1} \mathbb{P} \left(\left\{ U_1 \leq \frac{f(Y_1)}{Mg(Y_1)} \right\} \cap \{Y_1 \leq x\} \right) \\
 &= \frac{1}{p} \mathbb{P} \left(\left\{ U_1 \leq \frac{f(Y_1)}{Mg(Y_1)} \right\} \cap \{Y_1 \leq x\} \right) \\
 &= \frac{1}{p} \int_{-\infty}^x \mathbb{P} \left(U_1 \leq \frac{f(y)}{Mg(y)} \right) g(y) dy \\
 &= \frac{1}{p} \int_{-\infty}^x \frac{1}{M} f(y) dy = \int_{-\infty}^x f(y) dy. \quad \square
 \end{aligned}$$

- NB: le fait que, pour tout $x \in \text{supp}(f)$,

$$f(x) \leq Mg(x),$$

implique que les queues de f sont nécessairement plus légères que celles de g .

Exercice 1.

1. Simuler un échantillon selon la loi admettant la densité f par rapport à la mesure de Lebesgue:

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } 0 \leq x < 1, \\ \frac{1}{3} & \text{si } 1 \leq x < 2.5, \\ 0 & \text{sinon,} \end{cases}$$

en utilisant tour à tour

- (a) la méthode de la fonction de répartition inverse,
 - (b) la méthode d'acceptation-rejet,
 - (c) la méthode de décomposition.
2. Simuler un échantillon selon la loi admettant la densité f par rapport à la mesure de Lebesgue:

$$f(x) = \begin{cases} \frac{x}{2} & \text{si } 0 \leq x < 1, \\ \frac{1}{2} & \text{si } 1 \leq x < 2.5, \\ 0 & \text{sinon,} \end{cases}$$

- (a) la méthode de la fonction de répartition inverse,
 - (b) la méthode d'acceptation-rejet,
 - (c) la méthode de décomposition.
3. Estimer $I = \int \exp\left(-\frac{x^2}{2}\right) f(x) dx$ dans les deux cas précédents par méthode de Monte-Carlo. Quelle est la précision de l'approximation donnée?