Méthodes de Monte-Carlo et algorithmes stochastiques – TD/TP 2

Simulation de variables aléatoires et applications de Monte-Carlo

M2 Finance – Université Gustave Eiffel

Objectifs

Cette séance termine le chapitre 1. Elle met l'accent sur la simulation elle-même : inversion, acceptation–rejet, choix d'une loi de simulation, méthode delta et diagnostic numérique.

Exercice 1 – Inversion d'une loi mixte : un atome et une partie continue

On veut simuler une variable aléatoire positive qui vaut exactement zéro avec une probabilité non nulle, et qui, lorsqu'elle est strictement positive, se comporte comme une exponentielle. Plus précisément, soient $q \in (0,1)$ et $\lambda > 0$. On considère la loi définie par

$$\mathbb{P}(X = 0) = q, \qquad \mathbb{P}(X > x \mid X > 0) = \mathrm{e}^{-\lambda x}, \qquad x \geq 0.$$

1) Déterminer la fonction de répartition F de X. Faire apparaître explicitement le saut de F en 0.

2) On rappelle que l'inverse généralisée est

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \qquad 0 < u < 1.$$

Déterminer $F^{-1}(u)$ en distinguant les cas $0 < u \leq q$ et $q < u < 1$.

3) En déduire un algorithme de simulation de X à partir d'une seule variable $U \sim \mathcal{U}([0,1])$. Vérifier directement que l'algorithme produit bien une masse q en zéro et la bonne loi conditionnelle sur $(0, +\infty)$.

4) Calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.

Idée à retenir

L'inverse généralisée ne sert pas seulement lorsque la fonction de répartition est continue et strictement croissante. Un intervalle de valeurs de u peut être envoyé sur un même point : c'est exactement ainsi qu'un atome apparaît dans la loi simulée.

Exercice 2 – Acceptation–rejet : comparer deux lois de proposition

On veut simuler la densité

$$f(x) = 30x(1-x)^4 \mathbf{1}_{\{0 \leq x \leq 1\}},$$

qui est la densité d'une loi Beta$(2,5)$. On va comparer deux choix possibles pour la loi utilisée pour proposer les valeurs candidates dans l'algorithme d'acceptation–rejet.

Premier choix : la loi uniforme sur $[0,1]$, de densité

$$g_0(x) = \mathbf{1}_{\{0 \leq x \leq 1\}}.$$

1) Vérifier que f est bien une densité. Déterminer la plus petite constante c_0 telle que

$$f(x) \le c_0 g_0(x), \qquad x \in \mathbb{R}.$$

En déduire le taux d'acceptation théorique et le nombre moyen de propositions nécessaires pour obtenir une valeur acceptée.

Deuxième choix : la densité

$$g_1(x) = 4(1-x)^3 \mathbf{1}_{\{0 \le x \le 1\}}.$$

2) Vérifier que g_1 est une densité. Calculer sa fonction de répartition G_1, puis donner une formule permettant de simuler $Y \sim g_1$ par inversion à partir d'une uniforme.

3) Calculer le rapport $f(x)/g_1(x)$ sur $(0,1)$, puis déterminer la plus petite constante c_1 telle que $f \le c_1 g_1$. En déduire le taux d'acceptation théorique associé.

4) Comparer quantitativement les deux lois de proposition g_0 et g_1. Laquelle est la plus efficace ? Donner à la fois le taux d'acceptation et le nombre moyen de propositions par valeur acceptée.

Idée à retenir

Deux algorithmes de rejet peuvent être tous les deux exacts mais avoir des coûts très différents. Le choix de la loi de proposition consiste à rendre l'enveloppe cg aussi proche que possible de la densité cible dans les régions où celle-ci est importante.

Exercice 3 – Simulation gaussienne par la méthode polaire

On considère l'algorithme suivant. On tire U, V indépendantes uniformes sur $[-1,1]$ jusqu'à obtenir

$$0 < S = U^2 + V^2 < 1.$$

On note

$$R = \sqrt{U^2+V^2}, \qquad \Theta = \operatorname{Arg}(U + \mathrm{i}V) \in [0, 2\pi),$$

où Θ désigne l'angle polaire du point accepté. On pose ensuite

$$A = \sqrt{\frac{-2\log S}{S}}, \qquad X = AU, \qquad Y = AV.$$

1) Quelle est la loi du point (U,V) conditionnellement à son acceptation dans le disque unité ?

2) À l'aide du changement de variables polaires

$$(u,v) = (r\cos\theta, r\sin\theta),$$

déterminer la densité jointe conditionnelle de (R,Θ) sachant que le point est accepté. En déduire les lois marginales de R et de Θ, puis montrer que R et Θ sont indépendants après acceptation.

3) Montrer que $S = R^2$ est uniforme sur $[0,1]$ après acceptation. Poser

$$\rho = \sqrt{-2\log S}$$

et déterminer la fonction de répartition puis la densité de ρ.

4) Montrer que

$$X = \rho\cos\Theta, \qquad Y = \rho\sin\Theta.$$

En effectuant le changement de variables inverse des coordonnées polaires, déterminer la densité jointe de (X,Y). En déduire que X et Y sont indépendantes et de loi $\mathcal{N}(0,1)$.

5) Implémenter la méthode et vérifier numériquement les moyennes, variances et la covariance de X et Y.

Exercice 4 – Choisir une loi de simulation : un exemple calculable

On considère

$$J = \int_0^{+\infty} \sqrt{x}\, \mathrm{e}^{-x}\, dx.$$

1) Soit $X \sim \mathcal{E}(1)$. Montrer que $J = \mathbb{E}\left(\sqrt{X}\right)$. En déduire un estimateur $\widehat{J}_n^{(1)}$.

2) Calculer J explicitement à l'aide de la fonction Gamma. On pourra utiliser

$$\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

3) Calculer la variance de la variable $\sqrt{X}$, où $X \sim \mathcal{E}(1)$.

4) Soit maintenant $X_\lambda \sim \mathcal{E}(\lambda)$, avec $\lambda > 0$. Montrer que

$$J = \mathbb{E}\left[\frac{1}{\lambda}\sqrt{X_\lambda}\, \mathrm{e}^{-(1-\lambda)X_\lambda}\right].$$

5) Montrer que la variance de ce nouvel estimateur est finie si et seulement si $0 < \lambda < 2$, et que son second moment vaut

$$\mathbb{E}\left[\left(\frac{1}{\lambda}\sqrt{X_\lambda}\, \mathrm{e}^{-(1-\lambda)X_\lambda}\right)^2\right] = \frac{1}{\lambda(2-\lambda)^2}.$$

6) Déterminer la valeur de $\lambda \in (0, 2)$ qui minimise ce second moment. Commenter le résultat.

Exercice 5 – Estimation d'un paramètre exponentiel et méthode delta

Soient $X_1, \ldots, X_n$ i.i.d. de loi exponentielle de paramètre inconnu $\lambda > 0$, de densité

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}.$$

On pose

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad \widehat{\lambda}_n = \frac{1}{\overline{X}_n}.$$

1) Calculer $\mathbb{E}(X_1)$ et $\mathrm{Var}(X_1)$. En déduire, par le théorème central limite, la limite en loi de

$$\sqrt{n}\left(\overline{X}_n - \frac{1}{\lambda}\right).$$

2) En appliquant la méthode delta à la fonction $\varphi(x) = 1/x$, montrer que

$$\sqrt{n}(\widehat{\lambda}_n - \lambda) \Longrightarrow \mathcal{N}(0, \lambda^2).$$

En déduire un intervalle de confiance asymptotique de niveau 95% pour λ, ne faisant intervenir que $\widehat{\lambda}_n$ et n.

3) L'estimateur $\widehat{\lambda}_n$ est-il exactement sans biais ? On admet que

$$T_n = \sum_{k=1}^{n} X_k$$

a pour densité

$$f_{T_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \mathbf{1}_{\{t>0\}},$$

où, pour un entier $n \geq 1$, $\Gamma(n) = (n-1)!$. Pour $n \geq 2$, calculer $\mathbb{E}(1/T_n)$, puis $\mathbb{E}(\widehat{\lambda}_n)$.

4) Construire à partir de $\widehat{\lambda}_n$ un estimateur $\widetilde{\lambda}_n$ exactement sans biais de λ.

5) Pour $n \geq 3$, montrer que
$$\mathbb{E}\left(\frac{1}{T_n^2}\right) = \frac{\lambda^2}{(n-1)(n-2)}.$$
En déduire $\mathrm{Var}(\widetilde{\lambda}_n)$, puis comparer son comportement pour n grand avec la variance asymptotique suggérée par la méthode delta.

Idée à retenir

Cet exercice distingue trois notions qu'il ne faut pas confondre : convergence, comportement asymptotique et biais à taille finie. Un estimateur peut être consistant et asymptotiquement normal tout en restant légèrement biaisé pour tout n fini.

Exercice 6 – TP Python : maximum d'exponentielles et simulation en grande dimension

Soient $X_1, \ldots, X_d$ indépendantes de loi exponentielle de paramètre 1, et
$$M_d = \max(X_1, \ldots, X_d).$$
Cet exercice compare deux façons exactes de simuler M_d : simuler les d exponentielles puis prendre leur maximum, ou simuler directement la loi de M_d par inversion.

Étape 1 : loi exacte de M_d

1) Montrer que, pour $x \geq 0$,
$$\mathbb{P}(M_d \leq x) = \left(1 - e^{-x}\right)^d.$$
En déduire une formule permettant de simuler directement M_d à partir d'une seule variable $U \sim \mathcal{U}([0,1])$.

2) En utilisant l'identité
$$\mathbb{E}(M_d) = \int_0^{+\infty} \mathbb{P}(M_d > x)\,\mathrm{d}x,$$
montrer que
$$\mathbb{E}(M_d) = H_d := \sum_{k=1}^{d} \frac{1}{k}.$$
On pourra effectuer le changement de variable $t = e^{-x}$ puis utiliser
$$\frac{1-(1-t)^d}{t} = \sum_{j=0}^{d-1} (1-t)^j.$$

Étape 2 : deux algorithmes de simulation

À programmer vous-même

Programmer les deux méthodes suivantes.

— **Méthode A :** simuler d exponentielles par inversion, puis prendre leur maximum.

— **Méthode B :** utiliser directement l'inverse de la fonction de répartition de M_d, donc une seule uniforme par réalisation.

```python
import numpy as np

rng = np.random.default_rng(321)


def sample_max_method_A(n, d):
    # TODO : simuler un tableau de taille (n,d) d'exponentielles E(1)
    # TODO : prendre le maximum sur chaque ligne
    pass


def sample_max_method_B(n, d):
    # TODO : simuler n uniformes puis utiliser l'inversion de F_{M_d}
    pass


def harmonic_number(d):
    return np.sum(1.0 / np.arange(1, d + 1))
```

3) Pour $d = 10$ puis $d = 100$, simuler $n = 50\,000$ réalisations avec les deux méthodes. Comparer les moyennes empiriques à H_d.

4) Pour chaque méthode, construire un intervalle de confiance asymptotique à 95% pour $\mathbb{E}(M_d)$ à partir de la variance empirique des valeurs simulées. Vérifier si la valeur exacte H_d appartient aux intervalles obtenus.

5) Comparer le coût des deux méthodes lorsque d augmente. Expliquer mathématiquement pourquoi la méthode B évite une dépendance linéaire en d dans le nombre de variables aléatoires à générer pour chaque réalisation de M_d.

6) Répéter toute l'expérience $m = 500$ fois avec $d = 20$ et $n = 1000$, en utilisant la méthode B. Estimer la fréquence de couverture de l'intervalle à 95% pour la valeur exacte H_{20}. Commenter.

Idée à retenir

La dimension d peut intervenir fortement dans le coût de simulation d'une variable, même si le taux statistique de la moyenne de Monte-Carlo reste en $1/\sqrt{n}$. Une représentation plus intelligente de la loi simulée peut donc réduire fortement le coût sans modifier la quantité estimée.