Simulation stochastique et méthode de Monte Carlo

Afif Masmoudi

Laboratoire de probabilités et statistique
Université de Sfax
Journées de Statistique Mathématique et Data Science
(JSMDS 2019)

Plan

Plan

Plan

Plan

Plan

Plan

Méthode de Monte Carlo

Soit X une v.a. et soit $g : \mathbb{R} \to \mathbb{R}$ une fonction borélienne telle que $g(X) \in L^1$.
Problème: Calculer numériquement

$$\theta = \mathbb{E}(g(X)) = \int g(x)f(x)dx$$

avec f est la densité de probabilités de X.

Méthode de Monte Carlo

Exemples

- En finance: Prix d'une option d'achat.
  - Modèle de Black-Scholes:
    $$c = e^{-rT}\mathbb{E}\left[\left(S_0 e^{(r-\sigma^2/2)T+\sigma W_T} - K\right)^+\right] \text{ avec } W_T \sim \mathcal{N}(0,1).$$
- En assurance:
  - Prime pure: $\mathbb{E}(X)$;
  - Prime exponentielle: $\frac{1}{c}\log\left(\mathbb{E}(e^{cX})\right)$;
- Calcul de la Value At Risk en finance: $\mathbb{P}(X <$ seuil$)$.

Méthode de Monte Carlo

Méthode

Pour calculer $\theta = \mathbb{E}(g(X))$,

- Simuler N v.a. $(X_n)_{1 \leqslant n \leqslant N}$ i.i.d. de même loi que X,
- L'estimateur de θ :

$$\widehat{\theta} = \frac{1}{N} \sum_{i=1}^{N} g(X_i) = \frac{g(X_1) + ... + g(X_N)}{N}.$$

$$\widehat{\theta} \xrightarrow[N \to +\infty]{p-s} I \text{ : loi des grands nombre}$$

$\widehat{\theta} \simeq I$, pour N assez grand.

PROBLÈME: quelle est l'erreur d'estimation?

$$\text{T.C.L } \sqrt{N}(\widehat{\theta} - \theta) \xrightarrow[N \to +\infty]{d} \mathcal{N}(0, \sigma^2); \qquad \sigma^2 = Var(g(X)).$$

Méthode de Monte Carlo

Méthode

- Simuler N v.a. $(X_n)_{1 \leqslant n \leqslant N}$ i.i.d. de même loi que X,
- $$\widehat{\theta} = \frac{1}{N}\sum_{i=1}^{N} g(X_i) = \frac{g(X_1) + \ldots + g(X_N)}{N},$$
  $$\widehat{\sigma}_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(g(X_i) - \widehat{\theta})^2}.$$
- Intervalle de confiance :
  $$\theta \in \left(\widehat{\theta} \pm 1,96\frac{\widehat{\sigma}_N}{\sqrt{N}}\right).$$

Exemples.

Exemple 1 : Calculer $\theta_1 = \int_0^{\pi} e^x \sin(x)\, dx$

1. **la valeur théorique :** $\theta_1 = \frac{1}{2} e^{\pi} + 1 = 12.0703.$
2. **Valeur estimée :** $\hat{\theta}_1 = 12.1045$

Exemple 2 : $\theta_2 = \int\int_{x^2+y^2 \leq 1} \, dx\, dy$

1. **la valeur théorique :** $\theta_2 = \pi = 3.1415.$
2. **Valeur estimée :** $\hat{\theta}_2 = 3.144.$

Simulation stochastique

Théorème

$$(X_1, ..., X_d) \underset{en\,loi}{=} \Phi(U_1, ..., U_d)$$

avec

- $U_1, ..., U_d$ sont indépendantes et identiquement distribuées selon la loi uniforme sur $[0, 1]$,
- la fonction Φ est borélienne et a ses points de discontinuité dans un ensemble de mesure de Lebesgue nulle.

Simulation stochastique

Loi Gaussienne

Si $Y \sim \mathcal{N}(\mu, \sigma^2)$, alors $X = (Y - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Pour simuler X et Y

1. Simuler deux v.a U et V indépendantes et uniformes sur $[0, 1]$;
2. Poser

$$X = \sqrt{-2\, ln(U)} cos(2\pi V), \quad Y = \sqrt{-2\, ln(U)} sin(2\pi V).$$

X et Y sont deux v.a. indépendantes et de même loi normale centrées réduites.

Plan

Simulation stochastique

Simuler la loi uniforme consistera à produire par un algorithme des suites finies de nombres que nous pouvons considérer comme autant de réalisations indépendantes de variables aléatoires uniformes sur $[0, 1]$.

Méthode d'inversion

Définition

La fonction de répartition de X, notée F_X, est définie sur $\mathbb{R}$ par:

$$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}(X \leqslant x).$$

Proposition

Soit F_X la fonction de répartition d'une v.a.r. X. Alors:

- $F_X \in [0, 1]$.
- F_X est croissante.
- $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.
- F_X est continue à droite et a une limite à gauche en tout point.

Plan

Méthode d'inversion

Comme F_X est croissante, on peut définir la fonction pseudo-inverse q_X de F_X :

$$\forall u \in (0,1), \quad q_X(u) = \inf\left\{x \in \mathbb{R}, F_X(x) \geq u\right\}.$$

Théorème

Si U suit une loi uniforme sur $[0,1]$, alors $q_X(U)$ suit la même loi que X.

Méthode d'inversion

- $u \sim \mathcal{U}[0,1]$;
- $x = q_X(u)$. : réalisation de X

Méthode d'inversion

Si U suit la loi uniforme sur $[0, 1]$, alors

Loi uniforme sur [a,b]

$X = a + (b - a)U = q_X(u)$ suit la loi uniforme sur $[a, b]$.

Loi exponentielle

$X = -\frac{1}{\lambda} ln(1 - U) = q_X(u)$ suit la loi exponentielle de paramètre λ.

Loi de Cauchy

$X = c \tan(\pi(U - 1/2)) = q_X(u)$ suit la loi de Cauchy de paramètre c.

Loi de Bernoulli

si $U < 1 - p$, $X = 0$, sinon $X = 1$ suit la loi de Bernoulli de paramètre $p \in [0, 1]$

Méthode d'inversion

Soit $X = (X_1, X_2)$ un couple aléatoire absolument continu de densité $f(x, y)$.
Fonction de répartition de X_1:

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} \int_{\mathbb{R}} f(x, y) dy dx.$$

Fonction de répartition de $F_{X_2}^{X_1 = x_1}$ de la loi conditionnelle de X_2 sachant $X_1 = x_1$:

$$F_{X_2}^{X_1 = x_1}(x_2) = \frac{\displaystyle\int_{-\infty}^{x_2} f(x_1, y) dy}{\displaystyle\int_{-\infty}^{+\infty} f(x_1, y) dy}.$$

Méthode d'inversion

Si U_1 et U_2 sont deux v.a. uniformes sur $[0, 1]$ et indépendantes, alors

$$X_1 = q_1(U_1), \qquad X_2 = q_2(q_1(U_1), U_2).$$

Plan

Méthode de rejet

Soit f une densité de probabilité. On suppose qu' il existe une densité g simulable facilement et une constante $k \geq 1$ telles que

$$\forall x \in \mathbb{R}, f(x) \leq k\, g(x).$$

Soit $\alpha(x) = \frac{f(x)}{k\, g(x)}$.

Proposition

Soit $(X_i)_{i\geq 1}$ une suite de v.a. i.i.d de même densité g, et soit $(U_i)_{i\geq 1}$ une suite de v.a. i.i.d de même loi uniforme sur $[0, 1]$, indépendante de la suite $(X_i)_{i\geq 1}$. Si

$$T = inf \left\{ i \geq 1, U_i \leq \alpha(X_i) \right\}, \tag{1}$$

alors la variable aléatoire X_T a pour densité f.

Méthode de rejet

Méthode

Soit (X_1, U_1) un couple de v.a. indépendantes telles que X_1 a pour densité g et U_1 suit la loi uniforme sur $[0,1]$. Si $U_1 \leq \alpha(X_1)$, on choisit $X = X_1$.
Sinon on rejette X_1 et on recommence en générant une suite $(X_n, U_n)_{n\geq 2}$ de v.a. indépendantes de même loi que (X_1, U_1) jusqu'à l'instant p où $U_p \leq \alpha(X_p)$. On choisit alors $X = X_p$.

Remarques

1. On n'a pas besoin de connaitre F_X, ni F_X^{-1};
2. Elle s'étend à $\mathbb{R}^d$, à des lois discrètes, etc.

Méthode de rejet : Vitesse.

Calcul de la probabilité d'acceptation :

$$p = \mathbb{P}(U \leq \alpha(X)) = \frac{1}{k}.$$

Donc T suit la loi géométrique de paramètre p. En moyenne, on doit rejeter $k = 1/p$ fois avant d'accepter la valeur. Ainsi il faut choisir g telle que

$$k = max\left(\frac{f}{g}\right).$$

Exemple

Exemple plus concret. Calculer $\theta = \mathbb{E}((5e^{(Z-1/2)} - 3)^+)$ avec $Z \sim \mathcal{N}(0, 1)$

1. **Valeur exacte :** 2.71.
2. **Nombre de tirages :** $N = 1000$.
3. **Valeur estimée :** $\theta = 2.83$.
4. **Intervalle de confiance :** $\theta \in [2.42; 3.24]$.

Plan

Réduction de la variance.

BUT : Calculer $\theta = \mathbb{E}(Y) = \mathbb{E}(g(X))$.
Deux v.a Y_1 et Y_2 de même loi que Y.

1. $\mathbb{E}(Y) = \frac{1}{2}(\mathbb{E}(Y_1) + \mathbb{E}(Y_2)) = \mathbb{E}\left(\frac{Y_1+Y_2}{2}\right)$;
2. $\mathbb{V}ar\left(\frac{Y_1+Y_2}{2}\right) = \frac{\mathbb{V}ar(Y_1)+\mathbb{V}ar(Y_2)+2\mathbb{C}ov(Y_1,Y_2)}{4}$;
3. Si Y_1 et Y_2 dé-corrélées (ou indépendantes), alors $\mathbb{V}ar\left(\frac{Y_1+Y_2}{2}\right) = \frac{\mathbb{V}ar(Y)}{2}$;
4. Si $\mathbb{C}ov(Y_1, Y_2) < 0$, alors $\mathbb{V}ar\left(\frac{Y_1+Y_2}{2}\right) < \frac{\mathbb{V}ar(Y)}{2}$.

Réduction de la variance.

$$\theta = \int_0^1 g(x)\, dx = \mathbb{E}(g(X)),$$

avec $X \sim \mathcal{U}[0,1]$
$X_1, X_2, ..., X_n$: iid de même loi $\mathcal{U}[0,1]$.

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

Réduction de la variance.

$$\theta = \int_0^1 g(1-x)\,dx = \mathbb{E}(g(X))$$
$$= \int_0^1 \frac{g(x)+g(1-x)}{2}\,dx$$

$$\tilde{\theta} = \frac{1}{2n}\sum_{i=1}^{n}(g(X_i)+g(1-X_i))$$

$$Var(\tilde{\theta}) \leq Var(\hat{\theta})$$

Variables anthétiques uniformes.

Approcher $\theta = \mathbb{E}(g(U)) = \int_0^1 g(x)dx$ où U est uniforme sur $[0, 1]$.
Agorithme classique de Monte Carlo avec échantillon de taille $2n$.

Méthode

De $i = 1$ *à* $2n$
Générer U_i
Définir $Y_i = g(U_i)$
Fin
Définir $\hat{\theta} = \bar{Y}_{2n} = \frac{1}{2n}\sum_{i=1}^{2n} Y_i$ et $\hat{\sigma}^2 = \frac{1}{2n-1}\sum_{i=1}^{2n}(Y_i - \hat{\theta})^2$
Définir $I.C = [\hat{\theta} - 1.96\frac{\hat{\sigma}}{\sqrt{2n}}, \hat{\theta} + 1.96\frac{\hat{\sigma}}{\sqrt{2n}}]$

Variable anthétique uniforme.

Hypothèses: $Y = g(U)$ avec U uniforme sur $[0, 1]$.
Différents résultats:

- $\hat{\theta} = \bar{Y} = \dfrac{1}{2n} \sum_{i=1}^{2n} Y_i,$
- $\tilde{\theta} = \bar{Z}_n = \dfrac{1}{n} \sum_{i=1}^{n} Z_i$, avec $Z_i = Y_i + \tilde{Y}_i$, $\tilde{Y}_i = g(1 - U_i)$.

Réduction de la variance.

Cas général :

$$\theta = \int_{\Delta} g(x)f(x)\, dx,$$

où f est la densité de X.

- $x \to t(x) = y$ **: changement de variable (bijective) préservant la loi de X;**
- $Y = t(X)$ **a la même loi que X.**
- $$\theta = \int_{\Delta} g(t^{-1}(y))f(y)\, dy.$$

Réduction de la variance.

$$\theta = \frac{1}{2}\left\{\mathbb{E}(g(X)) + \mathbb{E}(g(t^{-1}(X)))\right\},$$

où f est la densité de X.

$$\tilde{\theta} = \frac{1}{2n}\sum_{i=1}^{n}\left(g(X_i) + g(t^{-1}(X_i))\right),$$

$$Var(\tilde{\theta}) \leq Var(\hat{\theta})$$

EXEMPLE : OPTION KNOCK-IN

CONTRAT FINANCIER de PAYOFF

$g(S_T) = max(0, S_T - K)\mathbf{1}_{S_T > B}$.

MODÉLE BLACK-SCHOLES

$S_T = S_0 \exp((r - \sigma^2/2)T + \sigma\sqrt{T}X)$, $X \sim \mathcal{N}(0,1)$.

PRIX DU CONTRAT

$C = \mathrm{e}^{-rT}\,\mathbb{E}(g(S_T)) = \mathrm{e}^{-rT}\,\mathbb{E}(max(0, S_T - K)\mathbf{1}_{S_T > B})$.

JEU DE PARAMÉTERS

$S_0 = 2, K = 1, B = 2,5, r = 0,1, \sigma = 0.3, T = 10, N = 5000$.

RÉSULTATS OBTENUS

- sans réduction de variance: $C = 1.5720$ (variance 5.4236),
- avec réduction de variance: $C = 1.5543$ (variance 1.5827).

Variable de contrôle

BUT: calculer $\theta = \mathbb{E}(Y) = \mathbb{E}(g(X))$.
AJOUT d'une variable Z

- facilement simulable,
- $\mathbb{E}(Z)$ connue ou facilement calculable (variance "petite").

Deux calculs possible de θ:

- par la méthode de Monte Carlo standard,
- en posant $W_c = Y + c(Z - \mathbb{E}(Z))$ et en calculant $\mathbb{E}(W_c)$ par **Monte Carlo**.

Question: $Var(W_c) < Var(Y)$?

Variable de contrôle

$$Var(W_c) = Var(Y) + c^2 Var(Z) + 2cCov(Y,Z)$$

Choix optimal de c: $Var(W_c) < Var(Y)$

- On pose $c^* = -2\dfrac{Cov(Y,Z)}{Var(Z)}$
- $c \in (0, c^*)$, $\hat{c} = \frac{c^*}{2}$
- $Var(W_{\hat{c}}) = Var(Y) - 3\dfrac{Cov(Y,Z)^2}{Var(Z)}$
- Z : variable de contrôle de Y.

Algorithme:

$$\hat{\theta}_{N,\hat{c}} = \frac{1}{N}\sum_{i=1}^{N}(Y_i + \hat{c}(Z_i - \mathbb{E}(Z))) \approx \theta.$$

Problème: Calcul de $\mathbb{E}(Z)$? De $Cov(Y,Z)$?

Plan

MC avec variable de contrôle.

Méthode

ÉTAPE 1: p petit $\longrightarrow$ déterminer une valeur approchée de $\hat{c}$.

- Simuler p v.a. $(Y_n)_{1\leq n\leq p}$ et p v.a.$(Z_n)_{1\leq n\leq p}$.
- Poser :

$$\hat{\mathbb{E}}(Z) = \frac{1}{p}\sum_{i=1}^{p} Z_i, \hat{\mathbb{E}}(Y) = \frac{1}{p}\sum_{i=1}^{p} Y_i$$

$$\hat{\mathbb{V}}ar(Z) = \frac{1}{p-1}\sum_{j=1}^{p}(Z_j - \hat{\mathbb{E}}(Z))^2,$$

$$\hat{\mathbb{C}}ov(Z,Y) = \frac{1}{p-1}\sum_{i,j=1}^{p}(Z_i - \hat{\mathbb{E}}(Z))(Y_j - \hat{\mathbb{E}}(Y)).$$

MC avec variable de contrôle.

Méthode

- Calculer $\hat{c} = -\frac{\hat{\mathbb{C}}ov(Y,Z)}{\hat{\mathbb{V}}ar(Z)}$.

ÉTAPE 2: N grand .

- Simuler N v.a. $(Y_n)_{1\leq n\leq N}$ et N v.a. $(Z_n)_{1\leq n\leq N}$.
- Poser :

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}(Y_i + \hat{c}(Z_i - \hat{\mathbb{E}}(Z))) \approx \theta.$$

Exemple

Énoncé

Calculer $\theta = \mathbb{E}(\mathrm{e}^{(U+V)^2})$ avec U et V sont de loi uniforme sur $[0, 1]$.

- $Y = \mathrm{e}^{(U+V)^2}$.
- Variables de contrôle : $Z_1 = U + V$, $Z_2 = (U + V)^2$ ou encore $Z_3 = \exp(U + V)$.

Exemple

X suit une loi normale de paramètres 0 et 1,

$$\theta = \mathbb{P}(X > 4) = \mathbb{E}(\mathbf{1}_{X>4}).$$

Monte-Carlo classique:

- Nombre de tirage N variant de 1000 à 10000000.
- Valeur estimée : 0.
- Conclusion erronée: $\theta = 0$.

Valeur exacte: $\theta = 7.6199 \times 10^{-24}$.
Nombre de tirages nécessaires de l'ordre de 10^{25}: impossible!

Plan

Échantillonnage préférentiel

Soit f la densité de X et soit ψ une autre densité telle que *Supp f* $\subset$ *Supp* ψ. Alors

$$\begin{aligned}
\theta &= \mathbb{E}(g(X)) = \int g(x) f(x)\, dx \\
&= \int g(x) \frac{f(x)}{\psi(x)} \psi(x)\, dx, \\
&= \mathbb{E}(h(Y)),
\end{aligned}$$

$$\text{avec } h(y) = \frac{g(y)f(y)}{\psi(y)} \text{ et } Y \sim \psi.$$

$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} h(Y_i)$ est l'estimateur de θ.

Définition: h est une fonction d'importance.
Question: comment choisir ψ?

Choix de la fonction d'importance.

Calcul de la variance:

$$Var(g(X)) = \int g^2(x) f(x)\, dx - \theta^2,$$

$$Var(h(Y)) = \int \frac{g^2(y) f^2(y)}{\psi(y)}\, dy - \theta^2.$$

Donc

$$Var(g(X)) - Var(h(Y)) = \int g^2(x) \left(1 - \frac{f(x)}{\psi(x)}\right) f(x)\, dx.$$

Choix de la fonction d'importance.

PISTE: Si $\psi(x) = \frac{g(x)f(x)}{\theta}$, alors $Var(h(Y)) = \int \theta \frac{g^2(y)f^2(y)}{g(y)f(y)}\, dy - \theta^2 = 0$!. Prendre ψ proche de $g.f$...

THÉORÈME (RUBINSTEIN)

La densité ψ qui minimise la variance est

$$\psi(x) = \frac{|g(x)|\, f(x)}{\int |g(x)|\, f(x) dx}.$$

Plan

Algorithme de Metropolis-Hastings

Principe

Points Communs avec la méthode d'acceptation-rejet

- f une *densité cible*, suivant laquelle on cherche à simuler.
- q une densité de proposition, selon laquelle on va échantillonner, en acceptant la proposition avec une probabilité donnée.

Différence

- q dépend de la valeur précédente de l'échantillon: $q() = q(. \mid x)$.
- Si la proposition à partir de x_n est refusée, on pose $x_{n+1} = x_n$: l'échantillon contiendra des valeurs répétées.
- L'échantillon correspond à une trajectoire d'une chaîne de Markov.

Algorithme de Metropolis-Hastings

Etant donné x_n,

1. Générer $y_n \sim q(y \mid x_n)$.
2. Choisir

$$x_{n+1} = \begin{cases} y_n & \text{avec probabilité } \rho(x_n, y_n), \\ x_n & \text{avec probabilité } 1 - \rho(x_n, y_n). \end{cases}$$

où

$$\rho(x, y) = \min\left(\frac{f(y)}{f(x)} \frac{q(x \mid y)}{q(y \mid x)}, 1\right)$$

Algorithme de Metropolis-Hastings

Théorème

Les (x_n) forment une chaîne de Markov. Si q est tel que cette chaîne est irréductible, sa distribution limite est f.

- Toute loi de proposition q rendant la chaîne irréductible convient.
- Contrairement à l'algorithme d'acceptation-rejet, on a plus besoin d'évaluer $\max \frac{f}{g}$.

Le choix de q n'influe pas sur le fait qu'il y a convergence, mais il influe sur la vitesse de celle-ci. Certains choix sont privilégiés.

Algorithme de Metropolis-Hastings

Algorithme de Metropolis-Hastings indépendant

- La proposition ne dépend pas de la valeur courante.

Etant donné x_n,

1. Générer $y_n \sim q(. \mid x_n)$.
2. Choisir

$$x_{n+1} = \begin{cases} y_n & \text{avec probabilité } \rho(x_n, y_n), \\ x_n & \text{avec probabilité } 1 - \rho(x_n, y_n). \end{cases}$$

où

$$\rho(x, y) = \min \left(\frac{f(y)}{f(x)} \frac{q(x \mid y)}{q(y \mid x)}, 1 \right)$$

Algorithme de Metropolis-Hastings

Algorithme de Metropolis-Hastings indépendant

- très ressemblant à l'acceptation-rejet.
- pas de besoin d'évaluer $\max \frac{f}{g}$, mais on perd l'indépendance entre les valeurs de l'échantillon.
- La convergence sera d'autant plus rapide que la distribution g est proche de f.

Algorithme de Metropolis-Hastings

Algorithme de Metropolis-Hastings par marche aléatoire

- La proposition suit une marche aléatoire symétrique.

1. Générer $y_n \sim g(. \mid x_n)$, g symétrique.
2. Choisir

$$x_{n+1} = \begin{cases} y_n & \text{avec probabilité } \rho(x_n, y_n), \\ x_n & \text{avec probabilité } 1 - \rho(x_n, y_n). \end{cases}$$

où

$$\rho(x, y) = \min\left(\frac{f(y)}{f(x)}, 1\right)$$

Algorithme de Metropolis-Hastings par marche aléatoire

- la probabilité d'acceptation ne dépend plus de g.
- approche très simple qui peut être utilisée quand f est très mal connue, contrairement à l'algorithme indépendant: espaces de grande dimension et /ou espaces d'objets discrets.