ECOLE SUPÉRIEURE DE LA STATISTIQUE
ET DE L'ANALYSE DE L'INFORMATION

Méthodes de simulation
2023-2024(P3)
2$^{\text{ème}}$ année

Chapitre 1: Introduction aux méthodes de simulation

Tasnime Hamdeni

January 17, 2024

Sommaire

1- Principe de la méthode Monte Carlo

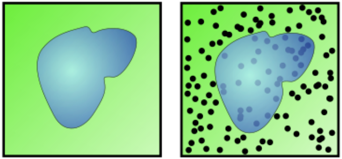
1- Principe de la méthode
1-1- Introduction

Plongez dans le monde fascinant des méthodes de **Monte Carlo**, où le hasard rencontre les mathématiques pour résoudre des **problèmes complexes**.

De la finance à la physique, ces méthodes transforment notre façon de modéliser et prédire des phénomènes variés, offrant des solutions innovantes là où **les approches traditionnelles échouent**.

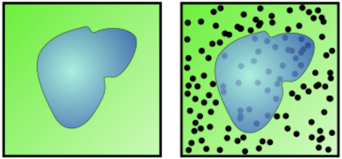
Technique puissante, véritable clé universelle pour comprendre et interagir avec un monde complexe.

1- Principe de la méthode
1-2- Vulgarisation: Détermination de la superficie d'un lac



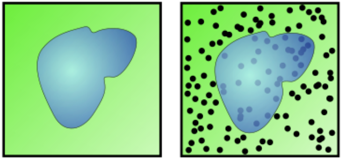
- On tire X coups de canon de manière aléatoire sur cette zone.

1- Principe de la méthode
1-2- Vulgarisation: Détermination de la superficie d'un lac



- On tire X coups de canon de manière aléatoire sur cette zone.
- On compte ensuite le nombre N de boulets qui sont restés sur le terrain.

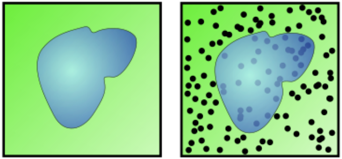
1- Principe de la méthode
1-2- Vulgarisation: Détermination de la superficie d'un lac



- On tire X coups de canon de manière aléatoire sur cette zone.
- On compte ensuite le nombre N de boulets qui sont restés sur le terrain.

  $\Longrightarrow$ on peut ainsi déterminer le nombre de boulets qui sont tombés dans le lac : $X - N$.

1- Principe de la méthode
1-2- Vulgarisation: Détermination de la superficie d'un lac



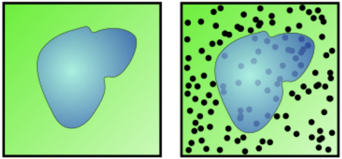
- On tire X coups de canon de manière aléatoire sur cette zone.
- On compte ensuite le nombre N de boulets qui sont restés sur le terrain.

  $\Longrightarrow$ on peut ainsi déterminer le nombre de boulets qui sont tombés dans le lac : $X - N$.

Il suffit ensuite d'établir un rapport entre les valeurs :

$$\frac{\text{superficie}_{\text{terrain}}}{\text{superficie}_{\text{lac}}} = \frac{X}{X - N}$$

1- Principe de la méthode
1-2- Vulgarisation: Détermination de la superficie d'un lac



- On tire X coups de canon de manière aléatoire sur cette zone.
- On compte ensuite le nombre N de boulets qui sont restés sur le terrain.

  $\Longrightarrow$ on peut ainsi déterminer le nombre de boulets qui sont tombés dans le lac : $X - N$.

Il suffit ensuite d'établir un rapport entre les valeurs :

$$\frac{\text{superficie}_{\text{terrain}}}{\text{superficie}_{\text{lac}}} = \frac{X}{X - N}$$

On pourrait ainsi déduire la superficie du lac.

1- Principe de la méthode
1-3- Estimation avec Monte Carlo

Supposons que l'on souhaite connaître la valeur d'une certaine quantité δ. La première étape de la méthode consiste à écrire le problème sous la forme d'une espérance.

Soient une variable aléatoire $X = (X_1, \ldots, X_d)$ de loi μ sur $\mathbb{R}^d$ (on abrègera cela par $X \sim \mu$) et une fonction $h : \mathbb{R}^d \to \mathbb{R}$. Le problème traité par les méthodes de Monte Carlo est l'estimation de

$$\delta = \mathbb{E}_\mu[h(X)] = \int_{\mathbb{R}^d} h(x)\mu(dx). \quad (1.1)$$

La solution standard à ce problème est de simuler une suite $(X_n)_{n\geq 1} = (X_{1,n}, \ldots, X_{d,n})_{n\geq 1}$ de variables aléatoires indépendantes identiquement distribuées (i.i.d.) suivant la loi μ, puis d'estimer l'espérance $\mathbb{E}_\mu[h(X)]$ par la moyenne empirique, i.e.,

$$\overline{h_n} = \frac{1}{n}\sum_{k=1}^{n} h(X_k) \quad (1.2)$$

Remarque.

Lorsqu'il n'y a pas d'ambiguïté sur la loi de X, on omettra μ dans les notations.

1- Principe de la méthode
1-3- Estimation avec Monte Carlo

Exemple : Calcul d'une intégrale

Soit $h : [a, b]^d \to \mathbb{R}$. On cherche à calculer

$$\delta = \int_{[a,b]^d} h(x_1, \ldots, x_d)\, dx_1 \ldots dx_d.$$

On peut réécrire I sous la forme

$$\delta = (b-a)^d \int_{\mathbb{R}^d} h(x_1, \ldots, x_d) \frac{1}{(b-a)^d}\, dx_1 \ldots dx_d.$$

Si l'on pose $X = (X_1, \ldots, X_d)$ un d-uplet de variables i.i.d. suivant la loi uniforme sur $[a, b]$, on a alors

$$\delta = (b-a)^d \mathbb{E}[h(X)].$$

1- Principe de la méthode
1-3- Estimation avec Monte Carlo

Exemple : Calcul d'une intégrale (suite)

De façon générale, si $\delta = \int_{\mathbb{R}^d} h(x)f(x)\,dx$, avec

- f une densité de probabilité sur $\mathbb{R}^d$
- h une fonction borélienne

alors on peut écrire, sous les hypothèses d'existence de δ, $\delta = E[h(X)]$ et l'estimer.

Exemple:

Donner l'approximation de l'intégrale $\int_0^2 e^{-x^2}\,dx$ par une méthode de Monte-Carlo.

2- Validité et comportement de la méthode

2- Validité et comportement de la méthode
2-1- Convergence de la méthode Monte-Carlo

La convergence de la méthode est assurée par la loi des grands nombres, sous l'hypothèse que h est intégrable par rapport à la mesure μ, c'est-à-dire que $E_{\mu}[|h(X)|]$ existe.

Rappel: (Loi faible et loi forte des grands nombres)

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles i.i.d. de loi μ telle que $h(X_1)$ soit μ-intégrable. Alors, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\overline{h_n} - \mathbb{E}[h(X_1)]\right| \leq \varepsilon\right] \to 1 \quad \text{(loi faible)},$$
$$\mathbb{P}\left[\left|\overline{h}_n - \mathbb{E}[h(X_1)]\right| \to 0\right] = 1 \quad \text{(loi forte)}.$$

2- Validité et comportement de la méthode
2-1- Convergence de Monte-Carlo

Remarque

La loi des grands nombres renseigne donc de deux façons sur l'erreur que l'on commet en estimant δ par la moyenne empirique $\bar{h}_n$.

Selon la **version faible**, il y a une probabilité négligeable de commettre une erreur plus grande que ε, tandis que la **version forte** dit qu'une fois que l'erreur d'estimation est inférieure à ε, alors cette erreur ne peut que décroître.

2- Validité et comportement de la méthode
2-2- Points forts de Monte Carlo

- **Applicabilité Générale :** Monte Carlo peut être appliqué à une large gamme de problèmes mathématiques.

2- Validité et comportement de la méthode
2-2- Points forts de Monte Carlo

- **Applicabilité Générale :** Monte Carlo peut être appliqué à une large gamme de problèmes mathématiques.
- **Traitement de Problèmes Stochastiques :** Efficace pour les problèmes avec des éléments aléatoires.

2- Validité et comportement de la méthode
2-2- Points forts de Monte Carlo

- **Applicabilité Générale :** Monte Carlo peut être appliqué à une large gamme de problèmes mathématiques.
- **Traitement de Problèmes Stochastiques :** Efficace pour les problèmes avec des éléments aléatoires.
- **Pas de Contraintes sur la Dimension :** Peut traiter des espaces de grande dimension.

2- Validité et comportement de la méthode
2-2- Points forts de Monte Carlo

- **Applicabilité Générale :** Monte Carlo peut être appliqué à une large gamme de problèmes mathématiques.
- **Traitement de Problèmes Stochastiques :** Efficace pour les problèmes avec des éléments aléatoires.
- **Pas de Contraintes sur la Dimension :** Peut traiter des espaces de grande dimension.
- **Estimation d'Incertitudes :** Fournit des estimations d'incertitudes naturelles.

2- Validité et comportement de la méthode
2-2- Points forts de Monte Carlo

- **Applicabilité Générale :** Monte Carlo peut être appliqué à une large gamme de problèmes mathématiques.
- **Traitement de Problèmes Stochastiques :** Efficace pour les problèmes avec des éléments aléatoires.
- **Pas de Contraintes sur la Dimension :** Peut traiter des espaces de grande dimension.
- **Estimation d'Incertitudes :** Fournit des estimations d'incertitudes naturelles.
- **Facilité d'Implémentation :** Relativement simple à mettre en œuvre.

2- Validité et comportement de la méthode
2-3- Points faibles de Monte Carlo

- **Convergence Lente :** Peut converger lentement, en particulier avec des variations rapides.

2- Validité et comportement de la méthode
2-3- Points faibles de Monte Carlo

- **Convergence Lente :** Peut converger lentement, en particulier avec des variations rapides.
- **Dépendance aux Échantillons Aléatoires :** Les résultats dépendent de la qualité des échantillons aléatoires.

2- Validité et comportement de la méthode
2-3- Points faibles de Monte Carlo

- **Convergence Lente :** Peut converger lentement, en particulier avec des variations rapides.
- **Dépendance aux Échantillons Aléatoires :** Les résultats dépendent de la qualité des échantillons aléatoires.
- **Calcul Intensif :** Peut nécessiter des ressources informatiques importantes.

2- Validité et comportement de la méthode
2-3- Points faibles de Monte Carlo

- **Convergence Lente :** Peut converger lentement, en particulier avec des variations rapides.
- **Dépendance aux Échantillons Aléatoires :** Les résultats dépendent de la qualité des échantillons aléatoires.
- **Calcul Intensif :** Peut nécessiter des ressources informatiques importantes.
- **Non-Déterministe :** Les résultats sont eux-mêmes aléatoires.

2- Validité et comportement de la méthode
2-3- Points faibles de Monte Carlo

- **Convergence Lente :** Peut converger lentement, en particulier avec des variations rapides.
- **Dépendance aux Échantillons Aléatoires :** Les résultats dépendent de la qualité des échantillons aléatoires.
- **Calcul Intensif :** Peut nécessiter des ressources informatiques importantes.
- **Non-Déterministe :** Les résultats sont eux-mêmes aléatoires.
- **Difficulté avec les Dimensions Élevées :** Rencontrer des difficultés dans des espaces de très grande dimension.

Il est possible de préciser davantage le comportement de la méthode lorsque h est de carré intégrable par rapport à la mesure μ, i.e., $\mathbb{E}_{\mu}\left[h(X)^2\right] < \infty$. [2] [3][1]

$\overline{h_n}$ étant un estimateur sans biais de $\delta = \mathbb{E}_{\mu}[h(X)]$, i.e., $\mathbb{E}_{\mu}\left[h_n\right] = \delta$, son erreur quadratique moyenne (EQM) est:

$$\mathbb{E}\left[\left(\overline{h_n} - \mathbb{E}[h(X)]\right)^2\right] = \frac{\sigma^2}{n}, \tag{1}$$

où $\sigma^2 = \mathrm{Var}[h(X)]$.

2- Validité et comportement de la méthode
2-4- Vitesse de convergence

théorème:

La vitesse de convergence de la méthode de Monte Carlo classique est en $o(n^{-1/2})$.

Pour doubler la précision de l'estimateur, que faut-il faire ?

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Lorsque la variance de $h(X)$ est finie, le théorème Central-Limite permet d'établir que $\overline{h_n} - \mathbb{E}_{\mu}[h(X)]$ suit asymptotiquement une loi normale.

Théorème Central-Limite (TCL)

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles i.i.d. de loi μ telle que $h(X_1)^2$ soit μ-intégrable. Alors, pour n suffisamment grand, alors

$$\sqrt{n}\left(\overline{h_n} - \mathbb{E}[h(X_1)]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

$$\Longleftrightarrow$$

$$\forall z \in \mathbb{R}, \qquad \mathbb{P}\left[\sqrt{\frac{n}{\sigma^2}}\left(\overline{h_n} - \mathbb{E}[h(X_1)]\right) \leq z\right] \xrightarrow[n\to+\infty]{} \phi(z)$$

où ϕ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

Le TCL permet de construire un intervalle de confiance autour de $\overline{h_n}$ pour quantifier l'erreur de la méthode de Monte Carlo.

Proposition 1:

L'erreur commise par la méthode de Monte Carlo est aléatoire, mais l'intervalle de confiance suivant fournit une quantification asymptotique de cette incertitude.

$\forall q \in \mathbb{R}$, on a :

$$\mathbb{P}\left[\sqrt{\frac{n}{\sigma^2}}\left|\overline{h_n} - \mathbb{E}[h(X)]\right| \leq q\right] \xrightarrow[n\to+\infty]{} \int_{-q}^{q} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = 2\phi(q) - 1$$

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Proposition 2 :

Sous les hypothèses du TCL, pour tout q réel et un niveau de confiance $1-\alpha$, l'intervalle de confiance bilatéral symétrique est donné par

$$IC_{1-\alpha} = \left[\overline{h_n} - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{h_n} + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

où $\overline{h_n}$ est la moyenne empirique, σ est l'écart-type empirique, $q_{1-\alpha/2}$ est le quantile de la loi normale standard associé à $1-\alpha/2$ $(\phi(q_{1-\alpha/2}) = 1-\alpha/2)$, et n est la taille de l'échantillon.

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Proposition 2 :

Sous les hypothèses du TCL, pour tout q réel et un niveau de confiance $1-\alpha$, l'intervalle de confiance bilatéral symétrique est donné par

$$IC_{1-\alpha} = \left[\overline{h_n} - q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{h_n} + q_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

où $\overline{h_n}$ est la moyenne empirique, σ est l'écart-type empirique, $q_{1-\alpha/2}$ est le quantile de la loi normale standard associé à $1-\alpha/2$ ($\phi(q_{1-\alpha/2}) = 1-\alpha/2$), et n est la taille de l'échantillon.

Remarque:

Le niveau de confiance usuel est $1-\alpha = 0.95$. Alors $q_{1-\alpha/2} = \phi^{-1}(0.975) \approx 1.96$.

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Remarque:

L'erreur quadratique moyenne et l'intervalle de confiance asymptotique dépendent tous deux de la variance σ^2, inconnue en pratique.
Lorsque la variance est finie, il est possible d'utiliser l'échantillon $(X_1, ..., X_n)$ pour obtenir un estimateur de σ^2, noté $\hat{\sigma}_n^2$ dans la suite.

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Remarque:

L'erreur quadratique moyenne et l'intervalle de confiance asymptotique dépendent tous deux de la variance σ^2, inconnue en pratique.
Lorsque la variance est finie, il est possible d'utiliser l'échantillon $(X_1, ..., X_n)$ pour obtenir un estimateur de σ^2, noté $\hat{\sigma}_n^2$ dans la suite.

Estimateurs de la variance

1. $\hat{\sigma}_n^2 \triangleq \frac{1}{n}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2 \triangleq \frac{1}{n}\sum_{k=1}^{n} Y_k^2 - (\bar{Y}_n)^2.$
2. $\hat{\sigma}_{n-1}^2 \triangleq \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y}_n)^2 \triangleq \frac{1}{n-1}\sum_{k=1}^{n} Y_k^2 - \frac{n}{n-1}(\bar{Y}_n)^2.$
3. $\hat{\sigma}_{2n}^2 \triangleq \frac{1}{2n}\sum_{k=1}^{n}(Y_{2k-1} - Y_{2k})^2.$

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

En substituant σ^2 par $\hat{\sigma}_n^2$, on obtient une approximation de l'EQM mais également de l'IC.

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

En substituant σ^2 par $\hat{\sigma}_n^2$, on obtient une approximation de l'EQM mais également de l'IC.

Rappel : Théorème de Slutsky

Soient $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ deux suites de variables aléatoires. S'il existe une variable aléatoire Y telle que $(Y_n)_{n\geq 1}$ converge en loi vers Y, et une constante c telle que $(Z_n)_{n\geq 1}$ converge en probabilité vers c, alors $(Y_n, Z_n)_{n\geq 1}$ converge en loi vers (Y, c). En particulier, $Z_n Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} cY$

En substituant σ^2 par $\hat{\sigma}_n^2$, on obtient une approximation de l'EQM mais également de l'IC.

Rappel : Théorème de Slutsky

Soient $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ deux suites de variables aléatoires. S'il existe une variable aléatoire Y telle que $(Y_n)_{n\geq 1}$ converge en loi vers Y, et une constante c telle que $(Z_n)_{n\geq 1}$ converge en probabilité vers c, alors $(Y_n, Z_n)_{n\geq 1}$ converge en loi vers (Y, c). En particulier, $Z_n Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} cY$

On en déduit que la convergence en loi du théorème TCL est préservée lorsque la variance est remplacée par un estimateur asymptotique sans biais.

2- Validité et comportement de la méthode
2-5- Estimation de l'erreur de Monte Carlo

Lemme: Intervalle de confiance

Sous les hypothèses du théorème TCL, on a

$$\sqrt{\frac{n}{\hat{\sigma}_n^2}}\left(\overline{h}_n - \mathbb{E}[h(X_1)]\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

et l'intervalle de confiance bilatéral symétrique au niveau de confiance asymptotique $1-\alpha$ est

$$\left[\overline{h}_n - q_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}_n^2}{n}},\ \overline{h}_n + q_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}_n^2}{n}}\right]$$

3- Conclusion

3- Conclusion

Si les méthodes déterministes sont efficaces pour les problèmes réguliers de très petites dimensions, les méthodes de Monte Carlo les surpassent et sont très compétitives pour les problèmes non-réguliers en grande dimension.

3- Conclusion

Si les méthodes déterministes sont efficaces pour les problèmes réguliers de très petites dimensions, les méthodes de Monte Carlo les surpassent et sont très compétitives pour les problèmes non-réguliers en grande dimension.

Les méthodes de Monte Carlo nécessitent uniquement de savoir formuler le problème sous la forme d'une espérance par rapport à une loi μ. Le choix de μ est guidé par le fait que l'on sache simuler suivant cette mesure (Chapitre 2) mais également par le fait que l'on puisse contrôler la variance de la méthode.

3- Conclusion

Si les méthodes déterministes sont efficaces pour les problèmes réguliers de très petites dimensions, les méthodes de Monte Carlo les surpassent et sont très compétitives pour les problèmes non-réguliers en grande dimension.

Les méthodes de Monte Carlo nécessitent uniquement de savoir formuler le problème sous la forme d'une espérance par rapport à une loi μ. Le choix de μ est guidé par le fait que l'on sache simuler suivant cette mesure (Chapitre 2) mais également par le fait que l'on puisse contrôler la variance de la méthode.

En effet, l'équation (1) montre que les deux seuls facteurs influant sur les performances de la méthode sont n et σ^2 (ou $\hat{\sigma}_n^2$). On peut donc réduire l'erreur en augmentant n au prix d'un coût de simulation (temps de calcul) plus important ou en réduisant σ^2.

Références

J. Geweke.
Monte carlo simulation and numerical integration.
Handbook of computational economics, 1:731–800, 1996.

D. P. Kroese and R. Y. Rubinstein.
Monte carlo methods.
Wiley Interdisciplinary Reviews: Computational Statistics, 4(1):48–58, 2012.

J. STOEHR.
Methodes de monte carlo.
Département MIDO, Université Paris Dauphine, 2022-2023.