

Exposé: la technique de simulation MONTE-CARLO

Présenté par :

- Elmalki Hajar Alla Taoufiq
- Bourkkadi Salmane Benabdenbi Ilham

•Encadré par : Prof. Mohamed El Merouani

Le plan

- **Introduction**
- Définition
- Approche historique
- Domaines d'application
- Méthodologie d'exécution
- Caractéristiques et avantages
- Limites
- Composantes d'un algorithme MC
- Applications
- Le problème Monte Carlo
- **Conclusion**

Introduction

Définition

- Le terme **méthode de Monte-Carlo**, ou **méthode Monte-Carlo**, désigne toute méthode visant à calculer une valeur numérique en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. Le nom de ces méthodes, qui fait allusion aux jeux de hasard pratiqués à Monte-Carlo a été inventé en 1947 par Nicholas Metropolis, et publié pour la première fois en 1949 dans un article coécrit avec Stanislaw Ulam.

- Les méthodes de Monte-Carlo sont particulièrement utilisées pour calculer des intégrales en dimensions plus grandes que 1 (en particulier, pour calculer des surfaces et des volumes). Elles sont également couramment utilisées en physique des particules, où des simulations probabilistes permettent d'estimer la forme d'un signal ou la sensibilité d'un détecteur. La comparaison des données mesurées à ces simulations peut permettre de mettre en évidence des caractéristiques inattendues, par exemple de nouvelles particules.

Exemple: Détermination de la superficie d'un lac

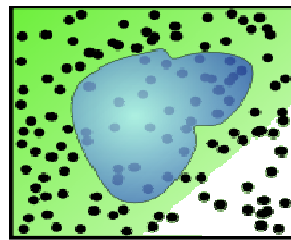
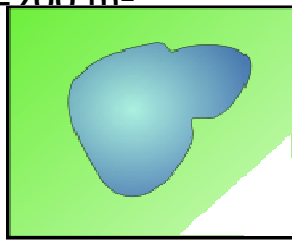
- Cet exemple est un classique en vulgarisation de la méthode de Monte-Carlo. Soit une zone rectangulaire ou carrée dont les côtés sont de longueur connue. Au sein de cette aire se trouve un lac dont la superficie est inconnue. Grâce aux mesures des côtés de la zone, on connaît l'aire du rectangle. Pour trouver l'aire du lac, on demande à une armée de tirer X coups de canon de manière aléatoire sur cette zone.

- On compte ensuite le nombre N de boulets qui sont restés sur le terrain ; 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 :

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

$$\Rightarrow \text{superficie}_{\text{lac}} = \frac{(X - N)}{X} \times \text{superficie}_{\text{terrain}}$$

- Par exemple, si le terrain fait 1 000 m², que l'armée tire 500 boulets et que 100 projectiles sont tombés dans le lac, alors une estimation de la superficie du plan d'eau est de : $1000 \cdot 100 / 500 = 200 \text{ m}^2$



- La qualité de l'estimation s'améliore en augmentant le nombre de tirs et en s'assurant que les artilleurs ne visent pas toujours le même endroit mais couvrent bien la zone. Cette dernière remarque est à mettre en parallèle avec la qualité du générateur aléatoire qui est primordiale pour avoir de bons résultats dans la méthode de Monte-Carlo. Un générateur biaisé est comme un canon qui tire toujours au même endroit : les informations qu'il apporte sont réduites.

Approche historique

- Les méthodes de monte-carlo ont été à l'origine pratiquées sous plusieurs noms génériques tel que <l'échantillonnage statistique>.
- Le nom « MONTE-CARLO » a été popularisé par les chercheurs en physique Stanislaw Ulam, Enrico Fermi, John von Neumann, et Nicholas Metropolis en référence à un célèbre casino de Monaco.

- En effet l'utilisation de l'aléatoire et le caractère répétitif du processus sont analogues à ceux menées dans un casino.
- Le véritable développement des méthodes de MONTE-CARLO s'est produit lors de la seconde guerre mondiale, lors des recherches sur la fabrication de la bombe atomique ainsi ces méthodes probabilistes ont été utilisées pour résoudre des équations aux dérivées partielles.

Domaines d'application

- On a recours à une simulation de Monte-Carlo lorsque le problème :
- Est trop complexe pour qu'une résolution par voie purement mathématique soit envisageable.
- Est trop volumineux (en particulier, contient un trop grand nombre de variables) pour que les techniques d'approximation numérique puissent conduire à un résultat précis dans un temps acceptable.

- Cette méthode permet aussi de résoudre d'autres problèmes tel que :
- Le problème du voyageur de commerce (c'est le problème qui consiste à trouver le plus court chemin qui relie toutes les villes, étant donné un ensemble de villes séparées par des distances données.
- Détermination de la valeur de π (π)
- Détermination de la superficie d'un lac

Méthodologie d'exécution

- La simulation de Monte-Carlo est une méthode habituelle pour l'évaluation d'un modèle déterministe utilisant un ensemble de nombres aléatoires comme intrants. Cette méthode est souvent utilisée lorsque le modèle est complexe, non linéaire, ou implique quelques paramètres incertains.

- Elle peut généralement faire intervenir plus de 10000 évaluations du modèle, tâche qui dans le passé était seulement pratique en utilisant super calculateurs.

- Les étapes à suivre sont les suivants :
- 1. Créer un modèle paramétrique $y = f(X_1, X_2, X_3, \dots, X_i)$
- 2. Générer un ensemble de données aléatoires $X_{i1}, X_{i2}, X_{i3}, \dots, X_{iq}$
- 3. Evaluer le modèle y
- 4. Répéter l'expérience n fois
- 5. Analyser les résultats à l'aide des histogrammes

Caractéristiques et avantages

- La résolution de nombreux problèmes scientifiques nécessite de calculer des sommes, des intégrales, ou encore de résoudre des équations ou des problèmes d'optimisation. Les techniques de calcul direct, encore appelées techniques analytiques, sont très vite dépassées par la complexité des modèles :

- elles nécessitent souvent des hypothèses trop fortes, de sorte qu'on ne peut pas les appliquer, ou alors, comme dans le cas de calcul de sommes, le nombre d'opérations requises peut être trop important pour être réalisé en un temps raisonnable.

- On doit alors nécessairement faire appel à des méthodes d'approximation.
- Les méthodes de simulation de Monte Carlo peuvent être vues comme des méthodes d'approximation, même s'il s'agit d'approximations au sens statistique du terme. Comme nous le verrons, ces méthodes sont moins exigeantes en termes d'hypothèses sur le modèle.

- Depuis son incorporation par Boyle (1977) dans la panoplie des outils de la finance computationnelle, la simulation de Monte Carlo n'a eu de cesse de gagner en popularité comme méthode de calcul de prix d'options de plus en plus sophistiquées et comme instrument de gestion des risques.

- Elle se caractérise par sa très grande souplesse et par sa capacité de traiter un problème en plusieurs dimensions.
- Dans certains cas, pour des fonctions à plusieurs dizaines de variables, une méthode de Monte Carlo devient le seul outil capable de donner une réponse en un temps raisonnable.

- Les méthodes de Monte Carlo sont aujourd'hui indispensables dans des domaines aussi variés et différents que la finance, les télécommunications, en ingénierie ou en physique, en biologie, en sciences sociales, etc. Par exemple, en chimie, en physique, ou même en biologie, de nombreux problèmes exigent l'analyse des propriétés dynamiques d'un nombre tellement grand d'objets, que ceci ne peut se faire que par des techniques de type Monte Carlo.

- La méthode de Monte-Carlo présente deux énormes avantages sur les méthodes par éléments finis.
- En procédant par échantillonnage, le nombre de calculs élémentaires nécessaire pour obtenir un résultat donné est considérablement plus petit.
- Il est extrêmement simple d'utiliser des processeurs en parallèle, par exemple un réseau d'ordinateurs, car chaque « réalisation » se fait par un calcul identique et peut se faire simultanément.

Limites

- Le problème de base est que l'hypothèse principale utilisée dans la simulation de Monte Carlo suppose une distribution normale et des coefficients de corrélation nuls, or aucun de ces deux suppositions n'est vraie sur le marché financier. Ces hypothèses peuvent engendrer des problèmes lors de l'analyse.

- En effet, la simulation MC est appropriée si :
- Il est impossible ou très coûteux d'obtenir des informations ;
- Le système observé est très complexe ;
- La solution analytique est difficile à obtenir ;
- Il est impossible ou coûteux de valider les modèles mathématiques ;

- Les variables de Monté Carlo supposent que les processus étudiés sont indépendants les uns des autres et que chaque valeurs est un tirage aléatoire d'une distribution.

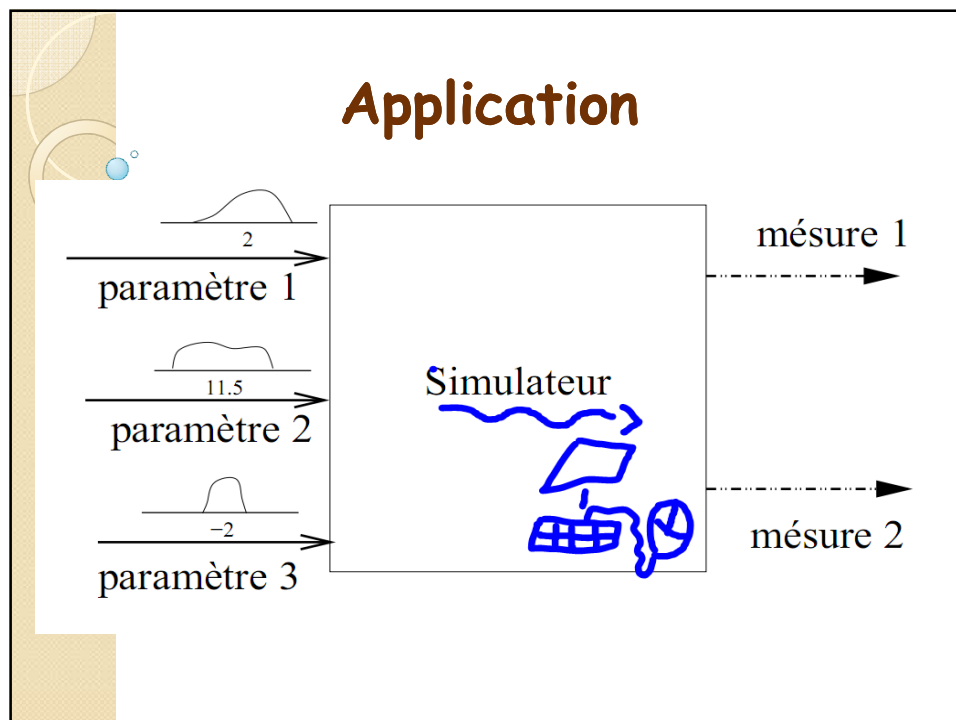
- Donc, les programmes utilisés par les ordinateurs peuvent manipuler les relations dépendantes entre les variables exogènes. Cependant, le problème est que la relation entre deux ou plusieurs variables est généralement compliquées et il est difficile de déterminer la vraie relation et les distributions des variables.

Composantes d'un algorithme MC

- Description probabiliste: un modèle stochastique du problème.
- Générateur uniforme de nombres aléatoires: un générateur de nombres aléatoires uniformément distribués sur l'intervalle $[0, 1]$.
- Loi d'échantillonnage: une technique pour échantillonner une distribution de probabilité générique.

- Simulateur: un simulateur déterministe qui renvoie l'output quand tous les paramètres sont connus.
- Collecteur des outputs: structure des données pour stocker tous les outputs de la simulation.
- Analyseur de l'output: ensemble de techniques statistiques qui permettent de tirer conclusions à partir des données générées par le simulateur.

- Estimateur d'erreur: ceci permet d'associer à chaque quantité estimée à part de l'output une indication sur l'erreur ou sur la confiance (par exemple en fonction du nombre de répétitions).

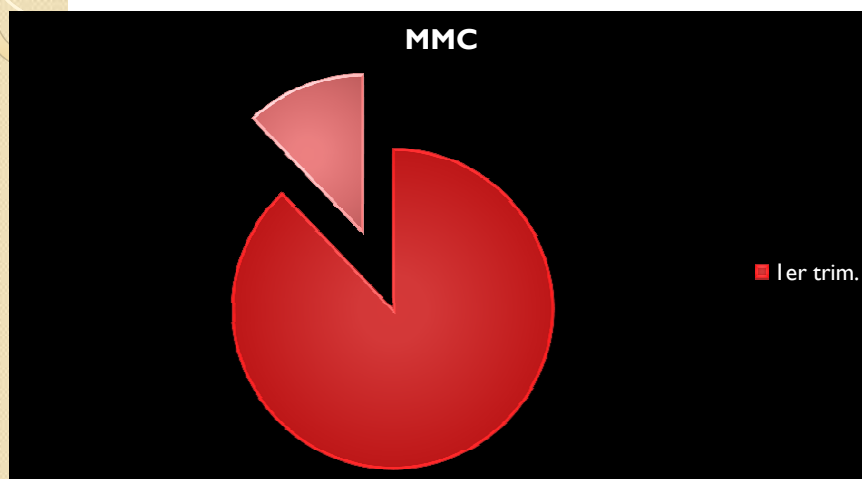


❖ Calcul surface d'un lac:

- Considérons une zone rectangulaire ou carrée dont la surface A_{terrain} est connue. Au sein de cette aire se trouve un lac dont la superficie A_{lac} est inconnue.
- Pour trouver l'aire du lac, on demande à une armée de tirer N coups de canon de manière aléatoire sur cette zone. On compte ensuite le nombre N_0 de boulets qui sont restés sur le terrain, on peut ainsi déterminer le nombre de boulets qui sont tombés dans le lac: $N - N_0$.
- Il suffit ensuite d'établir un rapport entre les valeurs :

$$\frac{A_{\text{terrain}}}{A_{\text{lac}}} = \frac{N}{N - N_0} \implies A_{\text{lac}} = \frac{(N - N_0)}{N} \cdot A_{\text{terrain}}$$

TS



Le problème Monte Carlo

Considérons une variable aléatoire $x \in \mathbb{R}^n$ de dimensionnalité n . Les méthodes Monte Carlo essaient de résoudre les deux problèmes suivants

Echantillonnage :

comment générer une série d'échantillons indépendantes

$\{x_i\}, i = 1, \dots, N$, qui soient tous tirés à partir de la même densité de probabilité $P_x(x)$.

Estimation :

étant donnée la fonction $g(x)$ et la variable aléatoire x , estimer l'espérance de la variable aléatoire $g(x)$

$$\theta = E[g(x)] = \int g(x)p(x)dx$$

θ est la moyenne de la variable aléatoire $g(x)$ qui représente la sortie du système.

La solution du premier problème rend possible la solution du deuxième. L'estimation de θ est obtenue à partir des N échantillons x_i par

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N g(x_i)$$

La loi des grands nombres nous garantit que

$$\lim_{N \rightarrow \infty} \hat{\theta} = \theta$$

Il est possible aussi de montrer que cet estimateur a des propriétés intéressantes d'un point de vue statistique. Il est non polarisé (en anglais unbiased), c.-à-d.

$$E[\hat{\theta}] = \theta$$

et sa variance est

$$\text{Var}[\hat{\theta}] = \frac{\sigma^2}{N}$$

où σ^2 est la variance de $g(x)$.

- La racine de la variance est une mesure de l'erreur moyenne commise une fois que nous estimons θ avec $\hat{\theta}$.
- Nous remarquons que la variance de $\hat{\theta}$ est indépendante de la dimensionnalité n .
- La propriété la plus importante des méthodes MC est donc que la précision de l'estimation renvoyée par Monte Carlo est indépendante de la dimensionnalité n de la variable x en entrée.
- Ceci signifie, que indépendamment de la dimensionnalité de x quelque dizaine d'échantillons de x peut être suffisante pour estimer θ d'une manière satisfaisante.

Conclusion