

Simulation

Nicolas Baskiotis - **Pierre-Henri Willemin**

Licence Informatique – Sorbonne Université

Méthode de Monte Carlo

➡ Définition (Simulation)

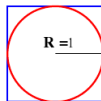
*“step by step the probabilities of separate events are merged into a composite picture which gives an **approximate** but **workable** answer to the problem”*

The Monte Carlo Method, D.D. McCracken, Scientific American, 1955

Monte Carlo est le nom d'un projet secret de John von Neumann et Stanislas Ulam (Los Alamos Scientific Laboratory) consistant à utiliser des nombres aléatoires pour simuler des séquences complexes d'évènements : Simulation de la diffusion des neutrons dans un matériau fissile.

roulette : méthode bien connue de génération de nombres aléatoires.

Un exemple : approximation de π



Méthode : on jette des cailloux dans le carré.

Hypothèse : $NbJets \propto Surface$

d'où

$$\frac{NbJets_{Cercle}}{NbJets_{Total}} = \frac{\pi \cdot R^2}{(2 \cdot R)^2} = \frac{\pi}{4}$$

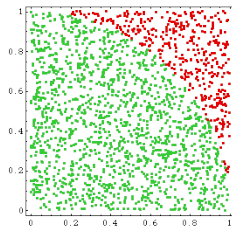
6 jets dans le cercle sur 10 jets en tout $\Rightarrow \hat{\pi}_{10} = 2.4$

89 jets dans le cercle sur 100 jets en tout $\Rightarrow \hat{\pi}_{100} = 3.57$

750 jets dans le cercle sur 1000 jets en tout $\Rightarrow \hat{\pi}_{1000} = 3$

7852 jets dans le cercle sur 10000 jets en tout $\Rightarrow \hat{\pi}_{10000} = 3.1408$

Autrement dit, on choisit **aléatoirement** un point du carré, et on vérifie (par $distance \leq 1$) si il est dans le cercle. Puis on itère.



simulation avec 4000 jets

Un exemple : approximation de π - formalisation

En notant $(x_i, y_i)_{i \leq N}$ les positions des N jets successifs.

Soit la fonction indicatrice du disque dans le carré : $1_{\bigcirc}(x, y) = \begin{cases} 1 & \text{si } x^2 + y^2 \leq 1 \\ 0 & \text{sinon} \end{cases}$

Le calcul précédent revient donc à calculer :

$$\frac{\sum_{i \leq N} 1_{\bigcirc}(x_i, y_i)}{N}$$

Qu'est-on en train d'estimer par une telle méthode ?

La solution **idéale** du problème serait de tester **tous les points du carré** pour connaître exactement la fraction de ceux-ci appartenant au cercle :

- Le “nombre” de point du carré : $\int_{(x,y) \in \square} dx dy$
- Le “nombre” de point du cercle : $\int_{(x,y) \in \square} 1_{\bigcirc}(x, y) dx dy$
- En introduisant une loi p uniforme sur \square (changement de mesure) :
 $\int_{(x,y) \in \square} p(x, y) dx dy = 1$ et $\int_{(x,y) \in \square} 1_{\bigcirc}(x, y) p(x, y) dx dy$

On estime donc $\int_{(x,y) \in \square} 1_{\bigcirc}(x, y) p(x, y) dx dy$ par $\frac{\sum_i 1_{\bigcirc}(x_i, y_i)}{N}$.

Examples d'applications

- finance : option pricing
- genetics and microarrays
- state space models : epidemiology and meteorology
- time series analysis
- biology - physics - chemistry
- mixture models for cluster analysis : astronomy, population studies
- operational research : traffic control, quality control, production optimization

Monte Carlo en statistique bayésienne

- Les méthodes de Monte Carlo proposent donc une **simulation stochastique** pour le calcul d'intégrales (ou d'équations différentielles).
- Il s'avère que l'intégration (ou l'équivalent discret : **la somme**) est une opération fondamentale dans les statistiques (et particulièrement dans la statistique bayésienne) à partir de :

$$posterior \propto L \times P$$

- Calculer la constante de normalisation : $\int L \times P$ car $posterior = \frac{L \times P}{\int L \times P}$
- Marginaliser une distribution jointe : $P(x_2) = \int P(x_1, x_2) dx_1$
- Statistiques sur une distribution : $E_P(f) = \int f(x) P(x) dx$
 - Moyenne de P : $f(x) = x$
 - Moment d'ordre 2 de P : $f(x) = x^2$
 - $P(A) : f(X) = 1_A$

Synthèse et résultats théoriques sur Monte Carlo

Supposons que nous voulions calculer $\mu = E_P(f) = \int f(x)P(x)dx$.

S'il n'y a pas de résultats analytiques, la méthode de Monte Carlo propose d'utiliser une suite $(X_i)_{i \leq N}$ d'observations de variables aléatoires, **i.i.d.**, **suivant la loi P** et d'estimer μ par :

$$\hat{\mu}_N = \frac{1}{N} \sum_{i \leq N} f(X_i)$$

Loi des grands nombres

Si $E(|X|) < \infty$ alors
(presque sûrement)

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i \leq N} X_i = E(X)$$

Théorème de la Limite Centrale

Soit $S_n = \sum_{i=1}^n X_i$, avec les X_i v.a. indépendante, à variance finie. Alors

$$S_n \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(n \cdot \mu, \sigma \cdot \sqrt{n})$$

D'où

Propriétés Monte Carlo

$$\hat{\mu}_N \xrightarrow[N \rightarrow \infty]{} \mu$$

$$\hat{\sigma}_N^2 = \frac{1}{N} \frac{1}{N-1} \sum_{i \leq N} [f(X_i) - \hat{\mu}_N]^2$$

$$\frac{\hat{\mu}_N - \mu}{\hat{\sigma}_N} \sim \mathcal{N}(0, 1)$$

Application de Monte Carlo

Dans de nombreuses applications (par exemple en génétique statistique), il est nécessaire de calculer la probabilité $P(Y = y)$ d'une variable Y observée comme la marginalisation d'une loi jointe $P(Y, X)$ où X représente des variables cachées. On utilise donc

$$P(Y = y) = \sum_{x \in \mathcal{X}} P(X = x, Y = y)$$

Mais, très souvent, \mathcal{X} est de taille bien trop grande pour pouvoir être décrit totalement. On ne peut pas calculer cette somme. Toutefois

$$P(Y = y) = \sum_{x \in \mathcal{X}} P(X = x, Y = y) = \sum_{x \in \mathcal{X}} P(Y = y | X = x)P(X = x) = \int_{x \in \mathcal{X}} f(x)P(x)dx$$

On se retrouve dans le cadre de l'utilisation de Monte Carlo avec $f(x) = P(Y = y | X = x)$
On peut donc estimer $P(Y = y)$ par

$$P(Y = y) \approx \frac{1}{N} \sum_{i \leq N} P(Y = y | X = x_i) \text{ où les } x_i \text{ suivent } P(X).$$

Monte Carlo : contre-exemple

- Soit E l'ensemble de permutations de $\{1, \dots, N\}$ ($N = 100$ par exemple),
- par exemple : E paramétrise le choix d'un tour dans le problème du voyageur de commerce,
- Soit $U(x), x \in E$ une fonction sur E , positive. Par exemple, U représente la longueur d'un tour dans le problème du voyageur de commerce.
- On veut simuler selon la loi : $\mu(x) = C \cdot \exp(-\beta \cdot U(x))$.
- C n'est pas connue et ne peut pas être calculée ! $\text{card}(E) \approx 10^{159}$
> nbre de particules de l'univers.

On ne peut pas toujours utiliser Monte Carlo !

Limites de Monte Carlo et solution

Pour estimer $\mu = \int f(x)P(x)dx$, la méthode de Monte Carlo propose d'utiliser une suite $(X_i)_{i \leq N}$ d'observations de v.a., **i.i.d, suivant la loi π** et d'estimer μ par $\hat{\mu}_N = \frac{1}{N} \sum_{i \leq N} f(X_i)$.

Que faire si il n'est pas possible d'obtenir des variables aléatoires i.i.d suivant la loi π ?

Principe de MCMC : Monte Carlo Markov Chain

Il s'agirait de construire une suite (X_i) de variables aléatoires qui seraient (presque) i.i.d, suivant π . Il serait alors possible d'utiliser la méthode de Monte Carlo pour estimer μ .

Une **Chaîne de Markov (à temps discret)**, de loi stationnaire π ($\pi = \pi \cdot P$) est un processus stochastique qui permet de générer une telle suite (X_i) .

PS : il faut un "certain nombre" d'itérations pour qu'une chaîne de Markov s'approche de la convergence (c'est à dire $P(X_t) \approx \pi$). Cette période (ou **burn-in**) passée, on peut considérer que $X_t \perp\!\!\!\perp X_{t+1}$, puisque les 2 v.a. suivent la même loi π .

MCMC : Monte Carlo Markov Chain

Changement de point de vue

- Quand on étudie les MC(TD), à partir d'une matrice de transition P , il s'agit de trouver la distribution stationnaire π .
- Pour les MCMC, étant donnée une loi π , il s'agit de **construire une MC(TD) convergent vers cette loi π** .

Comment construire cette chaîne de Markov ?

➡ Définition (Algorithme de Metropolis-Hastings – 1953)

Soit les lois $q(X | Y)$ **lois candidates ou instrumentales**, on construit alors

$$\alpha(X, Y) = \min \left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)} \right)$$

- Soit x_t la position courante du processus stochastique
- Itérations :
 - ➊ Proposer un candidat y suivant la loi $q(\cdot | x_t)$
 - ➋ Calculer $\alpha(x_t, y)$
 - ➌ Avec la probabilité $\alpha(x_t, y)$, $x_{t+1} = y$, sinon $x_{t+1} = x_t$
- les $(x_t)_{m \leq t \leq N}$ forment une suite de v.a. i.i.d utilisables pour une approximation MC.

MCMC - suite

- 1 Proposer un candidat y suivant la loi $q(\cdot | x_t)$
- 2 Calculer $\alpha(x_t, y)$
- 3 Avec la probabilité $\alpha(x_t, y)$, $x_{t+1} = y$: **acceptation**, sinon $x_{t+1} = x_t$: **rejet**

Les étapes 1 et 3 sont indépendantes, on peut donc calculer la probabilité de transition par :

$$\begin{cases} P(X_{t+1} = y | X_t = x) = q(y | x) \cdot \alpha(x, y) & \forall x \neq y \\ P(X_{t+1} = x | X_t = x) = 1 - \sum_{y \neq x} P(X_{t+1} = y | X_t = x) \end{cases}$$

Cette chaîne de Markov est irréductible et apériodique en fonction de $q(x | y)$ et $\alpha(x, y)$. Quelle est son point fixe ?

$$\alpha(X, Y) = \min \left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)} \right)$$

$$\Rightarrow \pi(X_t)q(X_{t+1} | X_t)\alpha(X_t, X_{t+1}) = \pi(X_{t+1})q(X_t | X_{t+1})\alpha(X_{t+1}, X_t)$$

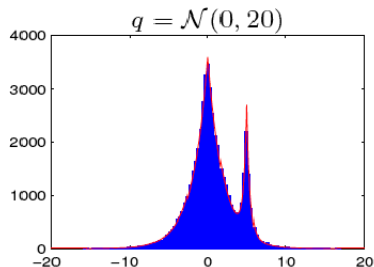
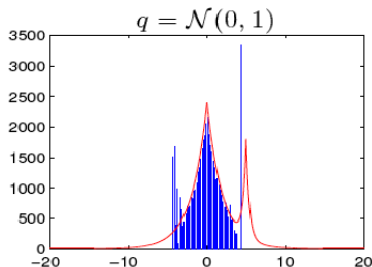
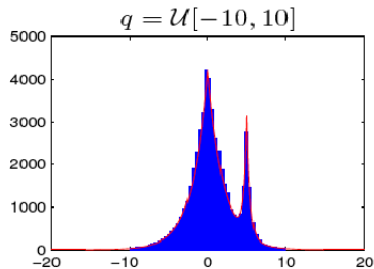
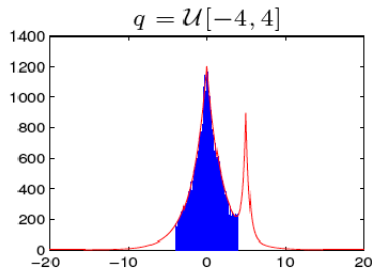
$$\Rightarrow \pi(X_t)P(X_{t+1} | X_t) = \pi(X_{t+1})P(X_t | X_{t+1})$$

$$\text{En sommant sur } X_t : \sum_x \pi(X_t = x)P(X_{t+1} | X_t = x) = \pi(X_{t+1}) \Rightarrow \pi \cdot P = \pi$$

Si cette CM(TD) converge, c'est vers π

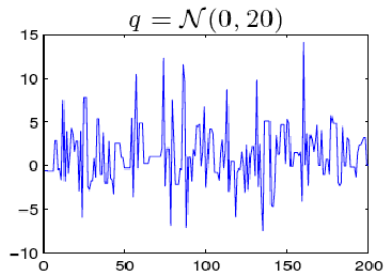
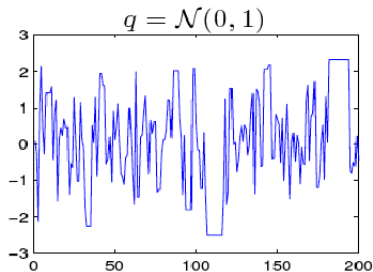
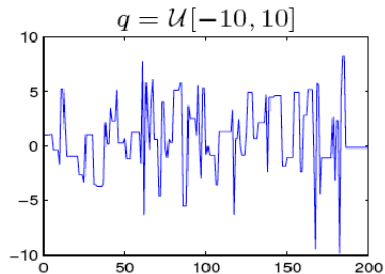
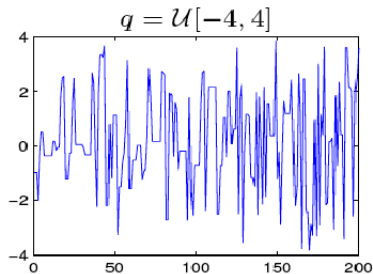


Influence de $q(x | y)$



Simulation

Influence de $q(x | y)$ – suite



Choix des lois candidates : $q(x | y)$

- q doit être **simulable rapidement** (\mathcal{N} ou U).
- q peut être connu analytiquement (à une constante près)
- autre exemple : $q(x | y) = q(y | x)$

La forme de q caractérise l'algorithme :

➡ Définition (formes de q)

- **Independence sampler** : $q(x | y) = q(x)$
- **Metropolis** : $q(x | y) = q(y | x)$ d'où $\alpha(x, y) = \min(1, \frac{\pi(y)}{\pi(x)})$