

On Brownian Motion, Ito Calculus and the Feynman-Kac Formula

21 février 2016

Dans cette partie nous décrirons une méthode probabiliste pour évaluer le noyau k_β que nous avons décrit dans la section précédente. Avant de commencer le calcul de k_β , nous allons introduire quelques notions utiles.

1 Definitions

Definition 1.1. Processus de Markov

Soit $(X_n)_{n \geq 0}$ un processus aléatoire discret sur un espace d'états dénombrable E . Le processus satisfait la propriété de Markov si pour toute collection d'états $(x_0, x_1, \dots, x_n) \in E$, nous avons

$$\mathbb{P}(X_{n+1} = y | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = y | X_n = x_n) \quad (1)$$

Il faut que les deux probabilités conditionnelles soient bien définies. Le processus $(X_n)_{n \geq 0}$ est alors appelé un processus de Markov.

Definition 1.2. Ergodicité

Un processus est ergodique s'il satisfait les conditions suivantes :

- φ est une mesure de probabilité invariante par le processus de Markov.
- **Condition d'accessibilité :**

$$\forall B \in \mathcal{B}(\mathbb{R}^d) \quad \text{t.q.} \quad \varphi(B) > 0, \quad \forall x \in \mathbb{R}^d, \quad \exists n \in \mathbb{N} \quad \text{t.q.} \quad \mathbb{P}(X_n \in B | X_0 = x) > 0 \quad (2)$$

Remark. 1. Nous avons défini la notion d'ergodicité sur \mathbb{R}^d , mais les mêmes définitions restent valables pour $\mathbb{L}\mathbb{T}^d$.

2. Au moment de simuler avec l'algorithme de Metropolis, il suffira de vérifier que la probabilité de transfert pour cet algorithme satisfait la condition d'accessibilité dans la définition d'ergodicité.

Definition 1.3. Chaînes de Markov

Une chaîne de Markov (discrète) est un processus de Markov défini par :

$$\begin{cases} X_{n+1}^{\Delta t} &= X_n^{\Delta t} + b(X_n^{\Delta t}) + \sqrt{\Delta t} \xi_n \\ X_0^{\Delta t} &\sim f(x) dx \end{cases} \quad (3)$$

avec $x \in \mathbb{R}^d$ et $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction régulière. Ici $\xi_j \sim \mathcal{N}(0, 1)_{\mathbb{R}^d}$, i.e. ξ_j est un vecteur gaussien de dimension d .

À partir de la chaîne de Markov défini ci-dessus nous pouvons construire un processus de Markov en temps continu en reliant les instants de temps t_n et t_{n+1} de manière affine. Dans ce cas, nous avons avec les mêmes conditions que ci-dessus :

$$\forall t \in [t_n, t_{n+1}], \quad \tilde{X}_t^{\Delta t} = X_n^{\Delta t} + (t - t_n) X_{n+1}^{\Delta t} \quad (4)$$

Nous avons de même

$$\tilde{X}_t^{\Delta t} \xrightarrow{\Delta t \rightarrow 0} X_t \quad (5)$$

où X_t suit l'équation différentielle stochastique donnée par :

$$dX_t = b(X_t)dt + dW_t \quad (6)$$

$$X_0 \sim f(x)dx \quad (7)$$

où W_t est le processus de Wiener.

2 Formule d'Itô

A la fin de la section précédente, nous avons abouti à une équation différentielle stochastique qui détermine le processus de Markov en temps continu. Pour faciliter la résolution de cette équation, nous mettrons en place des formules de calcul différentiel stochastique qui nous permettront d'aboutir à la formule de Feynman-Kac.

Considérons d'abord l'équation (6) sans le terme stochastique dW_t . Dans ce cas, l'équation devient déterministe et sa solution Y_t suit l'équation :

$$\frac{dY_t}{dt} = b(Y_t) \quad (8)$$

Pour rester dans un cadre simple, nous travaillerons dans \mathbb{R} , mais les résultats sont bien évidemment généralisables sur \mathbb{R}^d . Soit $h : t \times \mathbb{R}^d \rightarrow \mathbb{R}$. Alors nous avons :

$$\begin{aligned}
d(h(Y_t)) &= \frac{\partial h}{\partial t}(t, Y(t))dt + \nabla h(Y_t)dY_t \\
&= \frac{\partial h}{\partial t}(t, Y(t))dt + \nabla h(Y_t)b(Y_t)dt
\end{aligned}$$

Dans le cas de X_t , solution de l'équation différentielle stochastique (6), la différentielle de $h(t, X_t)$ contient un terme supplémentaire faisant intervenir les dérivées secondes de h . Nous obtenons finalement :

$$d(h(t, X_t)) = \left(\frac{\partial h}{\partial t}(t, X_t) + \nabla h(t, X_t)b(X_t) + \frac{1}{2}\Delta h(t, X_t) \right)dt + \nabla h(t, X_t)dW_t \quad (9)$$

3 Le Formule de Feynman-Kac

Munis des règles du calcul différentiel, nous sommes maintenant en mesure de dériver la formule de Feynman-Kac.

Pour ce faire, nous introduisons deux équations : les équations de Kolmogorov forward et backward

Definition 3.1. L'équation de Kolmogorov forward C'est une équation différentielle associée à (6) définie dans $]0, \beta] \times \Omega$ par :

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \frac{1}{2}\Delta u - \operatorname{div}(bu) - Vu \\
u(0) &= f(x)
\end{aligned} \quad (10)$$

Definition 3.2. L'équation de Kolmogorov backward

Dans $]0, \beta] \times \Omega$, nous définissons l'équation de Kolmogorov backward associée à (6)

$$\begin{aligned}
\frac{\partial v}{\partial t} + \frac{1}{2}\Delta v + b\nabla v - Vv &= 0 \\
v(\beta, x) &= g(x)
\end{aligned} \quad (11)$$

Nous considérons la dérivée totale $d(v(t, X_t)e^{-\int_0^t V(X_s)ds})$ où X_t est défini en (15). En utilisant l'équation (8), nous trouvons que :

$$\begin{aligned}
d(v(t, X_t)e^{-\int_0^t V(X_s)ds}) &= \left(\frac{\partial v}{\partial t} + b\nabla v + \frac{1}{2}\Delta v - Vv \right)e^{-\int_0^t V(X_s)ds} \\
&\quad + \nabla v(t, X_t)dW_te^{-\int_0^t V(X_s)ds}
\end{aligned} \quad (12)$$

et le premier terme est nul d'après (16). Donc :

$$d(v(t, X_t)e^{-\int_0^t V(X_s)ds}) = \nabla v(t, X_t)dW_t e^{-\int_0^t V(X_s)ds} \quad (13)$$

Enfin nous remarquons que l'espérance de $\nabla v(t, X_t)dW_t e^{-\int_0^t V(X_s)ds}$ est nulle, et en utilisant la commutation des différentes opérations avec l'espérance, nous avons ainsi que :

$$\mathbb{E}(v(t, X_t)e^{-\int_0^t V(X_s)ds}) = \text{cte} \quad (14)$$

En particulier :

$$\mathbb{E}(v(\beta, X_\beta)e^{-\int_0^\beta V(X_s)ds}) = \mathbb{E}(v(0, X_0)), \quad (15)$$

$$\mathbb{E}(g(x_\beta))e^{-\int_0^\beta V(X_s)ds} = \int_{\Omega} v(0, x)f(x)dx. \quad (16)$$

où nous avons utilisé la condition initiale de (16) et où l'intégration porte sur l'ouvert Ω considéré.

Considérons maintenant l'intégrale $\int_{\Omega} v(0, x)f(x)dx$. Par la condition initiale de (14), nous avons par conséquent :

$$\int_{\Omega} v(0, x)f(x)dx = \int_{\Omega} v(0, x)u(0, x)dx \quad (17)$$

D'où :

$$\begin{aligned} \int_{\Omega} v(0, x)u(0, x)dx &= \int_{\Omega} v(\beta, x)u(\beta, x)dx - \int_0^\beta \frac{d}{dt} \left(\int_{\Omega} v(t, x)u(t, x)dx \right) dt \\ &= \int_{\Omega} v(\beta, x)u(\beta, x)dx - \int_0^\beta \left(\int_{\Omega} \frac{\partial v}{\partial t} u dx + \frac{\partial u}{\partial v} dx \right) dt \end{aligned} \quad (18)$$

Nous remarquons alors que les problèmes (14) et (16) sont l'adjoint l'un de l'autre pour le produit scalaire usuelle défini pour l'espace L^2 . En d'autres termes, si L est un opérateur différentiel et L^* son opérateur adjoint, nous avons :

$$\begin{aligned} \frac{\partial u}{\partial t} &= Lu \\ \frac{\partial v}{\partial t} &= -L^*v \end{aligned} \quad (19)$$

En d'autres termes :

$$-\int_0^\beta \left(\int_\Omega \frac{\partial v}{\partial t} u dx + \frac{\partial u}{\partial v} dx \right) dt = -\int_0^\beta ((-L^*v, u)_{L^2} + (v, Lu)_{L^2}) dt = 0 \quad (20)$$

Ainsi :

$$\begin{aligned} \int_\Omega v(\beta, x) u(\beta, x) dx &= \int_\Omega v(0, x) f(x) dx \\ \int_\Omega g(x) u(\beta, x) dx &= \int_\Omega v(0, x) f(x) dx \end{aligned}$$

et on déduit la formule de Feynman-Kac :

$$\boxed{\int_\Omega g(x) u(\beta, x) dx = \mathbb{E} \left(g(X_\beta) e^{-\int_0^\beta V(X_s) ds} \right)} \quad (21)$$

où X_t est la solution de (6)

4 Application à notre problème

Revenons sur notre problème de départ, qui est de trouver le noyau de l'opérateur $e^{-\beta H}$. Nous rappelons que le noyau est une fonction k_β telle que :

$$(e^{-\beta H} f)(x) = \int_\Omega k_\beta(x, y) f(y) dy \quad (22)$$

Nous remarquons que, pour $b = 0$, $e^{-\beta H} f$ est la solution de (14). D'où en appliquant (26) :

$$\int_{\Omega \times \Omega} k_\beta(x, y) f(y) g(x) dx dy = \mathbb{E} \left(g(X_\beta) e^{-\int_0^\beta V(X_s) ds} \right) \quad (23)$$

où X_t est la solution de (6).

Cette dernière condition étant vraie pour tout f , nous prenons $f(y) = \delta(y)$, le delta de Dirac. Dans ce cas, nous avons pour le processus X_t :

$$X_\beta = y + W_\beta \quad (24)$$

D'où :

$$\begin{aligned}
\int_{\Omega} k_{\beta}(x, y) g(x) dx &= \mathbb{E} \left(g(y + W_{\beta}) e^{-\int_0^{\beta} V(y+W_s) ds} \right) \\
&= \int_{\Omega} g(x) \mathbb{E}_{y+W_{\beta}=x} \left(e^{-\int_0^{\beta} V(y+W_s) ds} \right) dx
\end{aligned} \tag{25}$$

Nous identifions alors le noyau comme étant :

$$\boxed{k_{\beta}(x, y) = \mathbb{E}_{y+W_{\beta}=x} \left(e^{-\int_0^{\beta} V(y+W_s) ds} \right)} \tag{26}$$

Dans notre cas, nous nous intéressons au calcul de la fonction de partition $Z = \text{Tr}(e^{-\beta H})$, qui d'après ce qui précède s'écrit :

$$\begin{aligned}
Z &= \text{Tr}(e^{-\beta H}) = \int_{\Omega} k_{\beta}(x, x) dx \\
&= \int_{\Omega} \mathbb{E}_{W_{\beta}=0} \left(e^{-\int_0^{\beta} V(x+W_s) ds} \right)
\end{aligned} \tag{27}$$

Remark. Nous devons interpréter les deux termes dans cette dernière relation. Prenons d'abord $\mathbb{E}_{W_{\beta}=0}$. C'est le terme d'espérance sur les ponts browniens, c'est-à-dire une espérance sur tous les chemins allant de x_0 au temps $t = 0$ jusqu'au un temps β où l'on revient au point de départ $x_{\beta} = x_0$.

Le deuxième terme en exponentiel représente la création ou annihilation des particules à un point $x + W_s$ donné. Nous formons donc une image de ce processus comme suivant :

1. La particule commence son trajet à $x_{t=0} = x_0 = 0$. Nous allons regarder seulement les trajets qui finissent à un temps $t = \beta$, de sorte que $x_{\beta} = 0$.
2. Pendant ce trajet, la particule peut être détruite ou créée selon le poids probabiliste $e^{-\int_0^{\beta} V(x+W_s) ds}$.
3. L'espérance sur tous ces trajets nous donne la fonction de partition $Z = \text{Tr}(e^{-\beta H})$.

5 Methodes numeriques

Dans cette partie nous allons décrire les méthodes numériques utilisés afin de simuler les systèmes bosoniques. Mais avant de parler des méthodes numériques, nous allons décrire, en quelques mots, les systèmes bosoniques.

5.1 Systemes Bosoniques

Les systèmes bosoniques sont des systèmes constitué par des bosons. Les bosons, en mécanique quantique, sont des particules portant un nombre quantique dit "nombre quantique de spin" (ou *Spin Quantum Number*) et ce nombre est toujours un entier. Dans la nature, il existe d'autres particules qui portent un spin qui est toujours un demi-entier : ceux sont des fermions. La différence entre les spins de bosons et fermions engendrent des grandes différences dans les comportements des systèmes constitués par des nombreuses particules de même type.

En mécanique quantique, l'état de chaque système est gouverné par un état $|\psi\rangle$, qui contient toute l'information du système. Cet état évolue selon la fameuse équation de Schroedinger. Le fait que les bosons ont un spin entier et les fermions un spin demi entier a des implication sur les propriétés de l'état $|\psi\rangle$. En particulier, l'état décrivant un système de N bosons est symétrique par échange de deux bosons tandis que l'état est anti-symétrique pour des fermions. C'est-à-dire que pour un système comprenant N bosons ou N fermions, nous avons les relations suivantes :

$$|n_1 n_2 \cdots n_N; B\rangle = \sqrt{\frac{\prod_n m_n!}{N!}} \sum_p |n_{p(1)}\rangle |n_{p(2)}\rangle \cdots |n_{p(N)}\rangle \quad (28)$$

$$|n_1 n_2 \cdots n_N; F\rangle = \frac{1}{\sqrt{N!}} \sum_p \text{sgn}(p) |n_{p(1)}\rangle |n_{p(2)}\rangle \cdots |n_{p(N)}\rangle \quad (29)$$

où B est pour les bosons, F pour les fermions. Ici la somme porte sur tous les états possible sous une permutation p de N particules. La racine est pris afin de normaliser l'état. Finalement, $\text{sgn}(p)$ est la signe de la permutation p (+1 si p contient un nombre paire des transpositions et -1 sinon).

Vu qu'un état complet d'un système bosonique contient tous les permutations possible de N particules, nous devons prendre en compte cette symétrie dans notre modélisation des systèmes bosoniques par un intégral de chemin. Concrètement, cela implique que pour un système des bosons, la matrice de densité est écrit de la manière suivante :

$$\rho_B(\mathbf{R}, \mathbf{R}'; \beta) = \frac{1}{N!} \sum_{\mathcal{P}} \rho_D(\mathbf{R}, \mathcal{P}\mathbf{R}'; \beta) \quad (30)$$

où ρ_D est la matrice de densité pour un système de N particules distinguables. Pour les particules distinguables comme les particules classiques, nous pouvons commencer à un ensemble des points \mathbf{R} donné et retourner à ce même point. Cependant, pour les particules comme des bosons(ou des fermions), nous pouvons arriver à un point $\mathcal{P}\mathbf{R}$, c'est-à-dire une permutation d'ensemble \mathbf{R} . Donc, en effectuant un calcul d'intégrale de chemin pour un système bosonique, nous devons sommer pas seulement sur tous les chemins allant de point

\mathbf{R} à \mathbf{R}' , mais sur tous les chemins partant de \mathbf{R} et aboutissant à $\mathcal{P}\mathbf{R}'$. Cela fait à la fois la richesse dans le comportement des bosons mais aussi un problème technique dont nous devons faire face.

5.2 L'algorithme de Metropolis

Un outil indispensable pour toutes les méthodes de Monte-Carlo est l'algorithme de Metropolis que nous allons décrire ici.

L'algorithme de Metropolis est une méthode qui nous permet d'échantillonner une distribution de probabilité $P(\mathbf{X})$ dans les cas où l'échantillonnage direct est difficile. En particulier, nous pouvons utiliser cet algorithme si nous connaissons la distribution de probabilité à une constante multiplicative près. Dans le cas de la physique statistique, et par extension pour notre cas, cela nous permet de contourner le calcul de la fonction de partition (une quantité souvent difficile à calculer sauf dans des cas simples) et échantillonner selon une loi de probabilité (souvent la distribution de Boltzmann).

Concrètement, l'algorithme de Metropolis utilise une chaîne de Markov qui atteint asymptotiquement une distribution de probabilité $\pi(x)$ telle que $\pi(x) = P(x)$ où $P(x)$ est la distribution de probabilité souhaitée. L'algorithme de Metropolis construit une chaîne de Markov qui satisfait les conditions d'ergodicité et la condition d'existence d'une distribution stationnaire asymptotique. La condition d'existence d'une distribution stationnaire est écrit comme une équation de "bilan détaillé" :

$$\pi(x)P(x \rightarrow x') = \pi(x')P(x' \rightarrow x) \quad (31)$$

où $\pi(x)$ est la distribution stationnaire et $P(x \rightarrow x')$ est la probabilité de passage de l'état x à l'état x' pour la chaîne de Markov.

Si la distribution stationnaire $\pi(x)$ est égale à $P(x)$, nous pouvons déduire la probabilité de passage entre deux états pour la chaîne de Markov comme suit :

1. Nous écrivons

$$P(x)P(x \rightarrow x') = P(x')P(x' \rightarrow x) \quad (32)$$

$$\frac{P(x \rightarrow x')}{P(x' \rightarrow x)} = \frac{P(x')}{P(x)} \quad (33)$$

2. Ensuite, nous décomposons la probabilité de passage en deux parties : une probabilité dite de "proposal" $g(x \rightarrow x')$, qui est la probabilité conditionnelle de proposer un état x' étant donné un état x et une probabilité d'acceptation $A(x \rightarrow x')$. Donc, la probabilité de passage s'écrit :

$$P(x \rightarrow x') = g(x \rightarrow x') A(x \rightarrow x') \quad (34)$$

$$\frac{A(x \rightarrow x')}{A(x' \rightarrow x)} = \frac{P(x') g(x' \rightarrow x)}{P(x) g(x \rightarrow x')} \quad (35)$$

3. Finalement nous devons choisir une probabilité d'acceptation qui est souvent prise comme :

$$A(x \rightarrow x') = \min \left(1, \frac{P(x') g(x' \rightarrow x)}{P(x) g(x \rightarrow x')} \right) \quad (36)$$

Dans le plupart des cas, la probabilité $g(x \rightarrow x')$ est pris uniforme. Ce qui nous amène à une probabilité d'acceptation :

$$\boxed{A(x \rightarrow x') = \min \left(1, \frac{P(x')}{P(x)} \right)} \quad (37)$$

Une fois la probabilité d'acceptation dérivée, nous pouvons échantillonner une probabilité de distribution $P(x)$ en utilisant l'algorithme suivant pour générer un nombre λ d'états du système :

1. Nous commençons avec un état x
2. Nous choisissons un nouvel état x' selon la loi uniforme
3. Nous acceptons l'état x' selon la loi $A(x \rightarrow x')$. Si l'état est accepté, le nouvel état du système est x' . Sinon, le système reste dans l'état x (et donc il n'y a pas de transition)
4. Nous répétons les étapes 2 et 3 jusqu'à la génération de λ états.

Dans la physique statistique, il est commode de prendre pour $P(x)$ la distribution de Boltzmann :

$$P(x) = \frac{e^{-\beta E_x}}{\mathcal{Z}} \quad (38)$$

$$\beta = \frac{1}{k_B T} \quad (39)$$

De même k_B est la constante de Boltzmann et T est la température du système. \mathcal{Z} est la fonction de partition et E_x est la énergie d'état x . On définit \mathcal{Z} comme :

$$\mathcal{Z} = \sum_{\{x\}} e^{-\beta E_x} \quad (40)$$

où la somme porte sur toutes les configurations possibles de x .

Nous voyons la grande utilité du méthode de Metropolis : dans la probabilité d'acceptation nous n'avons pas besoin de calculer \mathcal{Z} explicitement puisque ce qui nous intéresse sont les rapports de $P(x)$ et $P(x')$. Pour un système de N bosons en interaction, \mathcal{Z} est une quantité assez difficile à calculer dont l'importance des méthodes numériques fondé sur l'algorithme de Metropolis, alors que nous avons tout simplement :

$$\frac{P(x)}{P(x')} = e^{-\beta(E_x - E_{x'})} \quad (41)$$