

Processus Stochastiques

Olivier DULCY

Chapitre 1

Théorème de Kolmogorov

$\mathcal{B}(\mathbb{R}^d)$ désigne la tribu borélienne de \mathbb{R}^d .

1.1 Rappels de probabilités

Définition 1.1.1 Une tribu B est un ensemble des sous-ensembles de E vérifiant :

- $E, \emptyset \in B$
- $\forall A \in B, A^C \in B$
- $\forall A_n \in B \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in B$

Définition 1.1.2 On appelle tribu engendrée par une famille d'ensembles la plus petite tribu au sens de l'inclusion contenant tous les éléments de cette famille d'ensembles.

1.2 Introduction

Définition 1.2.1 Un processus stochastique X est la donnée de l'ensemble défini par $\{X_t; 0 \leq t < +\infty\}$, où à t fixé, X_t est une variable aléatoire définie sur (Ω, \mathcal{F}) à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Remarque : On voit bien que t peut prendre un ensemble **infini** de valeurs.

Exemple 1.2.1 On pose $I = [0, 1]$. On considère que chaque X_t est à valeurs dans \mathbb{R} . Ainsi, le processus stochastique $X = (X_t)_{t \in [0, 1]}$ prend alors ses valeurs dans $\mathbb{R}^{[0, 1]}$. Ces éléments sont appelés « trajectoires » du processus.

L'idée est de définir une tribu B et une probabilité \mathbb{P} sur B telle que $\forall n \in \mathbb{N}, \forall 0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq 1$, on puisse trouver la probabilité de $(X_{t_1}, \dots, X_{t_n})$.

1.3 Le théorème de Kolmogorov

Soit I un ensemble quelconque et $(E_i, B_i, \mathbb{P}_i)_{i \in I}$ une famille d'espaces probabilisés. Soit I_f un sous-ensemble fini de I . On suppose qu'il existe une probabilité \mathbb{P}_{I_f} sur la tribu B_{I_f} . Ainsi, pour chaque sous-ensemble fini I_f de I , on a un espace probabilisé $(E_{I_f}, B_{I_f}, \mathbb{P}_{I_f})$.

Définition 1.3.1 On dit que l'ensemble des espaces probabilisés $(E_{I_f}, B_{I_f}, \mathbb{P}_{I_f})$ est un **système projectif** si pour tous I_1 et $I_2 \subset I_1$ **finis**, la probabilité \mathbb{P}_{I_2} est la loi marginale de \mathbb{P}_{I_1} .

Définition 1.3.2 On appelle tribu de KOLMOGOROV la tribu \tilde{B} engendrée par les tribus.

Chapitre 2

Mouvement brownien et calcul différentiel stochastique

2.1 Mouvement brownien

Définition : Soit X un mouvement brownien. On dit que le mouvement est de Markov si $\forall t_1, \dots, t_n$, $p(x_{t_n} | x_{t_1}, \dots, x_{t_n}) = p(x_n | x_{t_{n-1}})$.

Propriété : $p(b|a) \sim \mathcal{N}(m_b + C^T A^{-1}(a - m_a), B - C^T A^{-1}C)$ où $\begin{pmatrix} A & C \\ C^T & B \end{pmatrix}$

Question 1 : Montrer que $p(x_{t_n} | x_{t_{n-1}}, x_{t_{n-2}}) = p(x_{t_n} | x_{t_{n-1}})$

Remarque : Cela veut dire que $p(u|v, w) = p(u|v) \Leftrightarrow p(u, w|v) = p(u|v)p(w|v)$

Montrons alors que $p(x_{t_n}, x_{t_{n-2}} | x_{t_{n-1}}) = p(x_{t_n} | x_{t_{n-1}})p(x_{t_{n-2}} | x_{t_{n-1}})$

On remplit la matrice de covariance. Pour chaque coefficient, « coeff = inf(indice1, indice2) ». On place les coefficients de manière « intelligente » : on veut la matrice avec les entêtes $X_{t_{n-2}}$ et X_{t_n} . Ce qui donne (écrire $X_{t_{n-2}} X_{t_n} X_{t_{n-1}}$ au dessus de la matrice et sur le côté gauche) :

$$\begin{pmatrix} t_{n-2} & t_{n-2} & t_{n-2} \\ t_{n-2} & t_n & t_{n-1} \\ t_{n-2} & t_{n-1} & t_{n-1} \end{pmatrix}$$

Ainsi, $p(x_{t_n}, x_{t_{n-2}} | x_{t_{n-1}}) \sim \mathcal{N}(0,)$ (à terminer).

2.1.1 Loi de l'arrivée à un point

Considérons un point $a \in \mathbb{R}$ et un mouvement brownien $X(t)$ partant du point 0. Nous allons étudier la loi de la variable aléatoire associant à chaque trajectoire l'instant de son arrivée au point a . Soit τ_a l'instant de la première arrivée au point a de la trajectoire du processus partant du point 0.

Remarque : Il y a symétrie entre les variables aléatoires τ_a et $-\tau_a$. On supposera donc $a > 0$.

On recherche la fonction de répartition de τ_a . On remarque que $[X(t) > a] \subset [\tau_a \leq t]$. En effet, une trajectoire ne peut pas dépasser a sans l'avoir eu atteint.

Or, $\mathbb{P}(X_t > a | \tau_a < t) = \frac{1}{2}$ (par symétrie).

Ainsi,

$$\mathbb{P}(\tau_a < t) = 2\mathbb{P}(X_t > a) = 2 \int_a^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2t}} du$$

2.1.2 Loi du maximum

On cherche à connaître le comportement du maximum de la trajectoire, dans le cas du mouvement Brownien.

$$M_{[0,t]} = \max_{u \in [0,t]} X_u$$

Ici, on connaît la loi de τ_a . Donc,

$$\mathbb{P}(\max_{u \in [0, t]} X_u < b) = \mathbb{P}(\tau_b > t)$$

2.2 Intégrale et différentielle stochastiques

Formule de Itô : Soit $X_t = \varphi(t, \psi_t)$ un processus, où φ est une fonction de classe \mathcal{C}^2 allant de \mathbb{R}^2 dans \mathbb{R} , et ψ_t un mouvement brownien.

$$\text{On a alors : } dX_t = \left(\frac{\partial}{\partial t} \varphi(t, \psi_t) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \varphi(t, \psi_t) \right) dt + \frac{\partial}{\partial y} \varphi(t, \psi_t) d\psi_t$$

$$\text{Or, on cherche } \int \psi_t d\psi_t = \varphi(t, \psi_t), \text{ ce qui équivaut à } \frac{\partial \varphi}{\partial t} + \frac{1}{2} \frac{\partial^2 \varphi}{\partial y^2} = 0 \text{ et } \frac{\partial \varphi}{\partial y} = y.$$

$$\text{On trouve } \varphi(t, y) = \frac{1}{2}(y^2 - t)$$

$$\text{Trouver une solution de } dX_t = aX_t dt + bX_t d\psi_t.$$

On cherche une solution de la forme $x_t = ce^{at}$. Avec la formule d'Itô, si $X_t = \varphi(t, \psi_t)$, en identifiant les parties dt et $d\psi_t$, on a :

$$\frac{\partial}{\partial t} \varphi(t, \psi_t) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \varphi(t, \psi_t) = a\varphi(t, \psi_t)$$

et

$$\frac{\partial}{\partial y} \varphi(t, \psi_t) = b\varphi(t, \psi_t)$$

On trouve :

$$\varphi(t, \psi_t) = e^{(a - \frac{b^2}{2})t + b\psi_t}$$

$$\text{Or } m_t = \mathbb{E}[X_t]$$

Chapitre 3

Résultats utiles

3.1 Le théorème central limite

Soit X_1, X_2, \dots une suite de variables aléatoires réelles définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant la même loi D . Supposons que l'espérance μ et l'écart-type σ de D existent et soient finis avec $\sigma \neq 0$.

Considérons la somme $S_n = \sum_{k=1}^n X_k$. Alors

- l'espérance de S_n est $n\mu$ et
- l'écart-type de S_n est $\sigma\sqrt{n}$

De plus, quand n est « assez grand », la loi normale $\mathcal{N}(n\mu, n\sigma^2)$ est une bonne approximation de la loi de S_n .

On pose :

$$\bar{X}_n = \frac{S_n}{n}$$

et

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

La variable Z_n est centrée et réduite.

Théorème 3.1.1 (*Théorème Central Limite*) La suite de variables aléatoires Z_1, \dots, Z_n, \dots converge en loi vers une variable aléatoire Z , définie sur le même espace probabilisé, et de loi normale centrée réduite $\mathcal{N}(0, 1)$ lorsque n tend vers l'infini.

Cela signifie que si Φ est la fonction de répartition de $\mathcal{N}(0, 1)$, alors pour tout réel z :

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = \Phi(z)$$

ou, de façon équivalente :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$

Chapitre 4

Méthodes de simulation pour les équations différentielles stochastiques

4.1 Introduction (rappel) sur les équations différentielles stochastiques

4.1.1 Familles gaussiennes

Dans la suite, on se donne un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$

Définition 4.1.1 *Un vecteur $X = (X_1, \dots, X_n)$ de variable aléatoire réelles est un vecteur gaussien si et seulement si pour $a \in \mathbb{R}^n$, $\langle a, X \rangle$ est une gaussienne*

Si X est un vecteur gaussien, alors si $a \in \mathbb{R}^n$:

- $\mathbb{E}[\langle a, X \rangle] = \langle a, \mathbb{E}[X] \rangle = \langle a, \mu \rangle$
- $\mathbb{V}[\langle a, X \rangle] = \mathbb{V}[a^T X] = a^T \mathbb{V}[X] a = a^T \Sigma a$

Ainsi $\langle a, X \rangle \sim \mathcal{N}(\langle a, \mu \rangle, a^T \Sigma a)$

On en déduit que pour tout $a \in \mathbb{R}^n$:

$$\mathbb{E}[e^{i\langle a, X \rangle}] = e^{i\langle a, \mu \rangle - \frac{1}{2} a^T \Sigma a}$$

Propriété 4.1.1 *Soit X un vecteur gaussien de \mathbb{R}^n est (i_1, \dots, i_d) d indices distincts entre 1 et n . Si pour tout $(j_1, j_2) \in \{1, \dots, d\}^2$ tel que $j_1 \neq j_2$ $\text{cov}(X_{j_1}, X_{j_2}) = 0$ alors $(X_{i_1}, \dots, X_{i_d})$ sont indépendants.*

Preuve : Notons $\mu_i = \mathbb{E}[X_i]$ et $\sigma_i^2 = \mathbb{V}(X_i)$. Le vecteur $\begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_d} \end{pmatrix}$ est un vecteur gaussien de moyenne $\begin{pmatrix} \mu_{i_1} \\ \vdots \\ \mu_{i_d} \end{pmatrix}$

et de variance $\text{diag}(\sigma_1^2, \dots, \sigma_d^2)$. Si on introduit $\begin{pmatrix} \epsilon_{i_1} \\ \vdots \\ \epsilon_{i_d} \end{pmatrix}$ un vecteur gaussien avec des variables aléatoires i.i.d.

de loi normale centrée réduite alors $\begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_d} \end{pmatrix}$ a la même loi que $\begin{pmatrix} Z_{i_1} \\ \vdots \\ Z_{i_d} \end{pmatrix}$ où $Z_{i_j} = \mu_{i_j} + \sigma_{i_j} \epsilon_{i_j}$. Les z_{i_j} étant indépendants, on en déduit le résultat.

Ce résultat se généralise au cas suivant :

Propriété 4.1.2 *Si I et J sont deux sous-ensembles d'indices $\{1, \dots, n\}$ disjoints alors si pour tout $i \in I$ et tout $j \in J$ $\text{cov}(X_i, X_j) = 0$ on a $(X_i)_{i \in I}$ et $(X_j)_{j \in J}$ sont indépendants.*

Théorème 4.1.1 (Cochran) *Soit $X \simeq \mathcal{N}(0, I_n)$ un vecteur gaussien centré et réduit. Soit F un sous-espace vectoriel de \mathbb{R}^n de dimension $p \geq 1$ et F^\perp son orthogonal. Alors $P_F(X)$ et $P_{F^\perp}(X)$ sont des vecteurs gaussiens indépendants et $\|P_F(X)\|^2 \sim \chi^2(p)$ et $\|P_{F^\perp}(X)\|^2 \sim \chi^2(n-p)$.*

Preuve : Soit (u_1, \dots, u_n) une base orthonormée de \mathbb{R}^n adaptée à F et F^\perp . (u_1, \dots, u_p) est une base orthonormée de F et (u_{p+1}, \dots, u_n) est une base orthonormée de F^\perp .

On peut écrire $P_F = \sum_{i=1}^p \langle X, u_i \rangle u_i$ et $P_{F^\perp} = \sum_{i=p+1}^n \langle X, u_i \rangle u_i$. Si on considère $U_{(p)} = (u_1, \dots, u_p) \in \mathbb{R}^{n \times p}$ et $U_{(n-p)} = (u_{p+1}, \dots, u_n) \in \mathbb{R}^{n \times (n-p)}$. On a alors : $P_F(X) = U_{(p)} U_{(p)}^T X$ et $P_{F^\perp}(X) = U_{(n-p)} U_{(n-p)}^T X$.

On sait déjà que $P_F(X)$ et $P_{F^\perp}(X)$ sont des vecteurs gaussiens et

$$\begin{pmatrix} P_F(X) \\ P_{F^\perp}(X) \end{pmatrix} = \begin{pmatrix} U_{(p)} U_{(p)}^T \\ U_{(n-p)} U_{(n-p)}^T \end{pmatrix} X$$

$$\text{Or, } \mathbb{E}[P_F(X)] = 0 \text{ et } \mathbb{E}[P_{F^\perp}(X)] = 0 \text{ et } \mathbb{V} \left[\begin{pmatrix} P_F(X) \\ P_{F^\perp}(X) \end{pmatrix} \right] = \begin{pmatrix} U_{(p)} U_{(p)}^T \\ U_{(n-p)} U_{(n-p)}^T \end{pmatrix} \mathbb{V}[X] \begin{pmatrix} U_{(p)} U_{(p)}^T \\ U_{(n-p)} U_{(n-p)}^T \end{pmatrix}^T =$$

$$\begin{pmatrix} U_{(p)} U_{(p)}^T & 0 \\ 0 & U_{(n-p)} U_{(n-p)}^T \end{pmatrix}^T$$

Finalement, on a bien $P_F(X)$ et $P_{F^\perp}(X)$ sont indépendants.

4.1.2 Le mouvement brownien

Définition 4.1.2 Un processus à temps continu $(W_t)_{t \geq 0}$ est un mouvement brownien (issu de 0) si et seulement si

- $W_0 = 0$
- $(W_t)_{t \geq 0}$ est un processus gaussien
- Pour tout $(s, t) \in \mathbb{R}_+^2$, $W_t - W_s \sim \mathcal{N}(0, t - s)$
- Pour tout $(s, t) \in \mathbb{R}_+^2$, $s \leq t$, $W_t - W_s$ est indépendant de $\sigma((W_u)_{0 \leq u \leq s})$
- La trajectoire $t \mapsto W_t$ est continue

Si on n'insère pas la continuité dans la définition, on démontre tout de même que $t \mapsto W_t$ est continue presque sûrement.

Propriété 4.1.3 Un processus gaussien $(W_t)_{t \geq 0}$ à trajectoires continues et issu de 0 est un mouvement brownien si et seulement si

- $\forall t > 0, \mathbb{E}[W_t] = 0$
- $\forall (s, t) \in \mathbb{R}_+^2, \mathbb{E}[W_s W_t] = \min(s, t) = s \wedge t$

Preuve : \Rightarrow Supposons que $(W_t)_{t \geq 0}$ est un mouvement brownien.

- $\forall t \geq 0, \mathbb{E}[W_t] = \mathbb{E}[W_t - W_0] = 0$ car $W_t - W_s \sim \mathcal{N}(0, 1)$.
- Soit $(s, t) \in \mathbb{R}_+^2$, tel que $s \leq t$.

$$\begin{aligned} \mathbb{E}[W_s W_t] &= \mathbb{E}[W_s (W_s + W_t - W_s)] \\ &= \mathbb{E}[W_s^2] + \mathbb{E}[W_s (W_t - W_s)] \\ &= \mathbb{E}[(W_s - W_0)^2] + \mathbb{E}[W_s] \mathbb{E}[W_t - W_s] \text{ car } W_t - W_s \text{ est indépendant de } W_s \\ &= s + 0 \\ &= s \end{aligned}$$

\Leftarrow Supposons que $\mathbb{E}[W_t] = 0$ et pour tout $(s, t) \in \mathbb{R}_+^2, \mathbb{E}[W_s W_t] = s \wedge t$.

- Soit $(s, t) \in \mathbb{R}_+^2$ avec $s \leq t$ et on veut montrer que $W_t - W_s$ est indépendant de $\sigma((W_u)_{0 \leq u \leq s})$. Le processus étant gaussien, il faut (et il suffit) de montrer que pour tout $0 \leq u \leq s$, $\text{cov}(W_t - W_s, W_u) = 0$.

Or,

$$\begin{aligned}\text{cov}(W_t - W_s, W_u) &= \mathbb{E}[W_u W_t] - \mathbb{E}(W_u W_s) \\ &= u \wedge t - u \wedge s \\ &= u - u \\ &= 0\end{aligned}$$

Pour tout $(s, t) \in \mathbb{R}_+^2$, $s \leq t$, on sait que $W_t - W_s$ est une gaussienne centrée. De plus,

$$\begin{aligned}\mathbb{V}[W_t - W_s] &= \mathbb{E}[(W_t - W_s)^2] = \mathbb{E}[W_t^2] + \mathbb{E}[W_s^2] - 2\mathbb{E}[W_t W_s] \\ &= t + s - 2s \wedge t \\ &= t - s\end{aligned}$$

Corollaire 4.1.1.1 Soit $(W_t)_{t \geq 0}$ un mouvement brownien alors les processus suivants sont également des mouvements browniens

- $(W_{t+t_0} - W_{t_0})_{t \geq 0}$ pour $t_0 \in \mathbb{R}_+$
- $(tW_{1/t})_{t > 0}$
- $(\alpha W_{1/t})_{t > 0}$ pour $\alpha > 0$

Propriété 4.1.4 Soit $(W_t)_{t \geq 0}$ un mouvement brownien alors :

- $\limsup_{t \rightarrow +\infty} \frac{W_t}{\sqrt{t}} = +\infty$ (et donc $\limsup_{t \rightarrow +\infty} W_t = +\infty$) presque sûrement
- Le mouvement brownien prend presque sûrement toutes les valeurs réelles (et n'est dérivable nulle part).

Théorème 4.1.2 (Donsker) Soit $(U_k)_{k \geq 0}$ une suite de variables aléatoires centrées i.i.d. et de carré intégrable avec $\sigma^2 = \mathbb{V}[U_1]$. Pour tout $t \in [0, 1]$ et tout $n \leq 1$ on introduit $X_n(t) = \frac{1}{\sigma\sqrt{n}} \left(\sum_{k=1}^{\lfloor nt \rfloor} U_k + (nt - \lfloor nt \rfloor)U_{\lfloor nt \rfloor + 1} \right)$. Pour tout $n \leq 1$, $X_n = (X_n(t))_{0 \leq t \leq 1}$ est un élément de $\mathcal{C}([0, 1], \mathbb{R})$ que l'on munit de la topologie induite pour la norme inférieure. Alors,

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} (W_t)_{0 \leq t \leq 1}$$

4.1.3 Equations différentielles stochastiques

Pour construire une intégrale stochastique, on procède en mimant la construction de l'intégrale de RIE-MANN. Pour $T > 0$ fixé et une fonction $f : [0, T] \rightarrow \mathbb{R}$, on introduit la quantité

$$I_{n,T}(f) = \frac{T}{n} \sum_{i=0}^n f(t_i^n) (t_{i+1}^n - t_i^n)$$

où $(t_i^n)_{0 \leq i \leq n+1}$ est une subdivision de $[0, T]$, $t_0^n = 0 < t_1^n < \dots < t_{n+1}^n = T$.

Si $\sup_{0 \leq i \leq n+1} (t_{i+1}^n - t_i^n) \xrightarrow[n \rightarrow +\infty]{} 0$ et si f est continue alors $I_{n,T}(f)$ converge lorsque n tend vers l'infini vers une quantité notée $\int_0^T f(u) du$.

Construction

□ ETAPE 1 : On considère un processus étagé sur $[0, T]$:

$$\forall t \in [0, T], X_t = \sum_{i=0}^n X_{t_i^n} \mathbb{1}_{[t_i^n, t_{i+1}^n]}(t)$$

avec $X_{t_i^n}$ une variable aléatoire, $\mathcal{F}_{t_i^n}$ mesurable (où $\mathcal{F}_{t_i^n}$ est une tribu donnée) et $t_0^n = 0 < t_1^n < \dots < t_{n+1}^n = T$. Alors on introduit l'intégrale stochastique suivante :

$$\int_0^T X_s dW_s = \sum_{i=0}^n X_{t_i^n} (W_{t_{i+1}^n} - W_{t_i^n})$$

□ ETAPE 2 : Si $(X_t)_{0 \leq t \leq T}$ un processus continu et borné alors pour construire $\int_0^T X_s dW_s$ on procède de la façon suivante :

- on introduit pour $n \geq 1$, $X_t^n = \sum_{k=0}^{n+1} X_{\frac{kT}{n}} \mathbb{1}_{[\frac{kT}{n}, \frac{(k+1)T}{n}]}(t)$
- on construit $M_T^n = \int_0^T W_s dW_s$
- On montre ensuite que $(M_T^n)_{n \geq 0}$ converge dans \mathcal{L}^2 vers une variable aléatoire M_T :
 $\mathbb{E}[(M_T^n - M_T)^2] \xrightarrow{n \rightarrow +\infty} 0$. On note alors $M_T = \int_0^T X_s dW_s$

Dans la suite de ce cours nous allons étudier (d'un point de vue numérique) les solutions d'EDS de la forme :

$$dX_s = \alpha_\theta(X_s)ds + \sigma_\theta(X_s)dW_s$$

avec $\theta \in \mathbb{R}^q$ un paramètre (inconnu et à estimer)

- $\alpha_\theta : \mathbb{R} \rightarrow \mathbb{R}$ et $\sigma_\theta : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions continues.
- $(W_t)_{t \geq 0}$ un mouvement brownien (et sa filtration $F_t = \sigma((W_u)_{0 \leq u \leq t})$)

où être solution de (4.1.3) signifie :

$$X_t = X_0 + \int_0^t \alpha_\theta(X_s)ds + \int_0^t \sigma_\theta(X_s)dW_s$$

Exemple 4.1.1 (Ecologie du mouvement) $(X_s)_{s \geq 0}$ est la position à chaque instant d'un individu (animal par exemple) :

$$\begin{cases} dX_s = \nabla_{\sigma} A_{\sigma}(X_s)ds + \sigma dW_s \\ Y_t = X_{t_k} + \epsilon_k \text{ où } \epsilon_k \sim \mathcal{N}(0, \eta^2 I_2) \end{cases}$$

où $A_{\sigma} : \mathbb{R}^2 \rightarrow \mathbb{R}$ un « potentiel » (caractérise les zones attractives et répulsives).

Un autre exemple existe en neurosciences.

4.2 Simulation du mouvement brownien

4.2.1 Simulation d'un squelette de trajectoire

On a fixé l'horizon $T > 0$ et n instants (t_1, \dots, t_n) tels que $0 < t_1 < \dots < t_n < T$. Pour simuler $(W_{t_1}, \dots, W_{t_n})$ on propose l'algorithme suivant.

- On simule $(\varepsilon_1, \dots, \varepsilon_n)$ indépendantes avec $\varepsilon_1 \sim \mathcal{N}(0, t_1)$ et $\varepsilon_i \sim \mathcal{N}(0, t_i - t_{i-1})$ pour $1 < i \leq n$.
- On pose $X_{t_1} = \varepsilon_1$ et pour $i > 1$ $X_{t_i} = X_{t_{i-1}} + \varepsilon_i$.

Alors en posant $X_0 = 0$, $(X_{t_1}, \dots, X_{t_n}) \stackrel{\mathcal{L}}{=} (W_{t_1}, \dots, W_{t_n})$

En effet,

$$\begin{aligned} (X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}) &= (\varepsilon_1, \dots, \varepsilon_n) \\ &\stackrel{\mathcal{L}}{=} (W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}) \end{aligned}$$

Puisque $X_0 = W_0 = 0$, on :

$$(X_0, X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}) \stackrel{\mathcal{L}}{=} (W_0, W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}})$$

Par transformation linéaire et déterministe on en déduit que

$$(X_0, X_1, \dots, X_n) \stackrel{\mathcal{L}}{=} (W_0, W_1, \dots, W_n)$$

4.2.2 Complétion des trajectoires browniennes

Dans cette section, on suppose disponible une réalisation $(W_{t_1}, \dots, W_{t_n})$ d'un mouvement brownien aux instants prédéfinis (t_1, \dots, t_n) . On souhaite, conditionnellement à ce squelette, simuler les valeurs du mouvement brownien à d'autres instants.

Lemme 4.2.0.1 *On suppose que (X, Y, Z) est un vecteur gaussien, centré. Alors, la loi de Y conditionnellement à (X, Z) est une loi gaussienne de moyenne $P_{(X,Z)}(Y)$ et de variance $\|Y - P_{(X,Z)}(Y)\|^2$ où $P_{X,Z}$ est la projection orthogonale sur l'espace engendré par (X, Z) pour le produit scalaire $\langle U, V \rangle \mapsto \mathbb{E}[UV]$*

Preuve : Voir les notes du professeur.

On cherche à simuler W_u conditionnellement à $(W_{t_1}, \dots, W_{t_n})$ où $u \in (t_k, t_{k+1})$, $k \in \{1, \dots, n-1\}$. Cette loi est égale à la loi de W_u sachant $(W_{t_k}, W_{t_{k+1}})$. On sait que $(W_{t_k}, W_u, W_{t_{k+1}})$ est un vecteur gaussien centré, on peut donc appliquer le lemme 4.2.0.1.

Pour calculer la loi cherchée, on calcule donc $P_{(W_{t_k}, W_{t_{k+1}})}(W_u)$ et $\|W_u - P_{(W_{t_k}, W_{t_{k+1}})}(W_u)\|^2$. Ainsi, $\left(\frac{W_{t_k}}{\|W_{t_k}\|}, \frac{W_{t_{k+1}} - W_{t_k}}{\|W_{t_{k+1}} - W_{t_k}\|}\right)$ forme une base orthogonale de $\text{Vect}(W_{t_k}, W_{t_{k+1}})$.

Ainsi,

$$\begin{aligned} P_{(W_{t_k}, W_{t_{k+1}})}(W_u) &= \left\langle W_u, \frac{W_{t_k}}{\|W_{t_k}\|} \right\rangle \frac{W_{t_k}}{\|W_{t_k}\|} + \left\langle W_u, \frac{W_{t_{k+1}} - W_{t_k}}{\|W_{t_{k+1}} - W_{t_k}\|} \right\rangle \frac{W_{t_{k+1}} - W_{t_k}}{\|W_{t_{k+1}} - W_{t_k}\|} \\ &= \frac{\langle W_u, W_{t_k} \rangle}{\langle W_{t_k}, W_{t_k} \rangle} W_{t_k} + \frac{\langle W_u, W_{t_{k+1}} - W_{t_k} \rangle}{\langle W_{t_{k+1}} - W_{t_k}, W_{t_{k+1}} - W_{t_k} \rangle} (W_{t_{k+1}} - W_{t_k}) \end{aligned}$$

Or,

$$\begin{aligned} \langle W_{t_k}, W_{t_k} \rangle &= \mathbb{E}[W_{t_k}^2] \\ &= t_k \end{aligned}$$

Puis,

$$\begin{aligned} \langle W_{t_{k+1}} - W_{t_k}, W_{t_{k+1}} - W_{t_k} \rangle &= \mathbb{E}[(W_{t_{k+1}} - W_{t_k})^2] \\ &= t_{k+1} - t_k \end{aligned}$$

Enfin

$$\begin{aligned} \langle W_u, W_{t_k} \rangle &= \mathbb{E}[W_u W_{t_k}] \\ &= u \wedge t_k \\ &= t_k \end{aligned}$$

Et

$$\begin{aligned} \langle W_u, W_{t_{k+1}} - W_{t_k} \rangle &= \langle W_u, W_{t_{k+1}} \rangle - \langle W_u, W_{t_k} \rangle \\ &= \mathbb{E}[W_u W_{t_{k+1}}] - \mathbb{E}[W_u W_{t_k}] \\ &= u \wedge t_{k+1} - u \wedge t_k \\ &= u - t_k \end{aligned}$$

Conditionnellement à $(W_k, W_{t_{k+1}})$ la moyenne de W_u est $\mathbb{E}[W_u | W_{t_k}, W_{t_{k+1}}] = W_{t_k} + \frac{u-t_k}{t_{k+1}-t_k} (W_{t_{k+1}} - W_{t_k})$.

Soit

$$\mathbb{E}[W_u | W_{t_k}, W_{t_{k+1}}] = \frac{t_{k+1} - u}{t_{k+1} - t_k} W_{t_k} + \frac{u - t_k}{t_{k+1} - t_k} W_{t_{k+1}}$$

La variance est

$$\begin{aligned}
\mathbb{E}[(W_u - \mathbb{E}(W_u|W_{t_k}, W_{t_{k+1}}))^2] &= \left(\frac{t_{k+1} - u}{t_{k+1} - t_k}\right)^2 \underbrace{\mathbb{E}[(W_{t_k} - W_u)^2]}_{u-t_k} + \left(\frac{u - t_k}{t_{k+1} - t_k}\right)^2 \underbrace{\mathbb{E}[(W_u - W_{t_{k+1}})]}_{t_{k+1}-u} \\
&\quad \text{car } \mathbb{E}[(W_{t_k} - W_u)(W_{t_{k+1}} - W_u)] \underset{\text{indép}}{=} \mathbb{E}[W_{t_k} - W_u]\mathbb{E}[W_{t_{k+1}} - W_u] = 0 \\
&= \frac{(t_{k+1} - u)^2(u - t_k)}{(t_{k+1} - t_k)^2} + \frac{(u - t_k)^2(t_{k+1} - u)}{(t_{k+1} - t_k)^2} \\
&= \frac{(t_{k+1} - u)(u - t_k)}{t_{k+1} - t_k}
\end{aligned}$$

4.2.3 Utilisation pour simulations Monte Carlo

Dans le cas de l'équation différentielle stochastique suivante :

$$dS_t = \rho S_t dt + \sigma S_t dW_t$$

Soit $S_T - S_0 = \int_0^T \rho S_s ds + \int_0^T \sigma S_s dW_s$, on peut prouver dans ce cas que :

$$S_T = S_0 e^{(r - \frac{\sigma^2}{2})T + \sigma W_T}$$

En général, on s'intéresse ensuite à une quantité du type $\mathbb{E}[h(S_T)]$. Un choix standard : $h : x \mapsto (x - k)_+$ où $k > 0$. Une approche très classique pour estimer $\mathbb{E}[h(S_T)]$ est d'utiliser une méthode de MONTE CARLO. On simule $(S_t^i)_{1 \leq i \leq N}$ indépendants et de même loi pour $N \geq 1$ et on estime $\mathbb{E}[h(S_T)]$ par $\frac{1}{N} \sum_{i=1}^N h(S_t^i)$. Ici cela se résume à simuler $(W_T^i)_{1 \leq i \leq N}$ indépendants. Cette approche est justifiée par

- La loi des grands nombres : si h est borné, $\frac{1}{N} \sum_{i=1}^N h(S_T^i) \xrightarrow[N \rightarrow +\infty]{\text{p.s.}} \mathbb{E}[h(S_T)]$
- Le théorème central limite : si h est borné $\frac{\sqrt{N}}{\sqrt{\mathbb{V}[h(S_T)]}} \left(\frac{1}{N} \sum_{i=1}^N h(S_T^i) - \mathbb{E}[h(S_T)] \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$

4.3 Simulation approchée par discrétisation

Dans cette section, on s'intéresse à un processus $(X_t)_{0 \leq t \leq T}$ solution de l'EDS suivante :

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{4.1}$$

où $b : \mathbb{R} \rightarrow \mathbb{R}$ et $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ sont continues.

Tout d'abord, notons que si on suppose σ et b lipschitziennes, (4.1) admet une solution (forte) unique. S'il existe $K \in \mathbb{R}_+$ tel que pour tout $(x, y) \in \mathbb{R}^2$

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K|x - y|$$

alors il y a unicité de processus $(X_t)_{0 \leq t \leq T}$ s'écrivent :

$$X_t - X_0 = \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \text{ p.s.}$$

4.3.1 Approximations d'Euler

Pour écrire l'approximation d'EULER, on introduit une discrétisation de l'intervalle $[0, T]$: $(t_k^n = \frac{kT}{N})_{0 \leq k \leq n}$ et pour simuler $(Y_t)_{t \in [0, T]}$ aux instants $(t_k^n)_{0 \leq k \leq n}$, on « fige » les valeurs de σ et de b à ces instants.

Entre t_k^n et t_{k+1}^n pour $0 \leq k \leq n-1$, on approche (4.1) par $d\tilde{X}_t = b(X_{t_k^n})dt + \sigma(X_{t_k^n})dW_t$. On a alors

$$\begin{cases} \widetilde{X}_0 = X_0 \\ \forall k \in \{0, \dots, n-1\}, \widetilde{X_{t_{k+1}^n}} = \widetilde{X_{t_k^n}} + \frac{T}{n} b(\widetilde{X_{t_k^n}}) + \sigma(\widetilde{X_{t_k^n}}) \sqrt{\frac{T}{n}} \varepsilon_{k+1} \end{cases}$$

où $(\varepsilon_k)_{1 \leq k \leq n}$ sont i.i.d., de loi $\mathcal{N}(0, 1)$. Cette approximation de $(X_{t_k^n}, \dots, X_T)$ est appelée approximation d'EULER à pas constant.

On peut introduire une approximation d'EULER continue de la façon suivante :

$$\begin{cases} \widetilde{X}_0 = X_0 \\ \forall k \in 0, \dots, n-1, \forall t \in [t_k^n, t_{k+1}^n), \overline{X_t} - \overline{X_{t_k^n}} = b(\overline{X_{t_k^n}})(t - t_k^n) + \sigma(\overline{X_{t_k^n}})(W_t - W_{t_k^n}) \end{cases}$$

On s'intéresse à la quantification de l'erreur d'approximation lorsque l'on remplace $(x_t)_{0 \leq t \leq T}$ par $(\overline{X_t})_{0 \leq t \leq T}$. Avant d'étudier l'approximation de (4.1) par la méthode d'Euler, on étudie l'approximation du brownien.

□ ERREUR D'APPROXIMATION POUR LE BROWNIEN :

On cherche ici à quantifier :

$$\left\| \sup_{t \in [0, T]} |W_t - \overline{W_t}| \right\|_p \quad \text{pour } p \geq 2$$

□ BORNE INFÉRIEURE :

$$\begin{aligned} \left\| \sup_{t \in [0, T]} |W_t - \overline{W_t}| \right\|_p &= \left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [t_{k-1}^n, t_k^n)} |W_t - W_{t_k^n}| \right\|_p \\ &= \sqrt{\frac{T}{n}} \left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [t_{k-1}^n, t_k^n)} \sqrt{\frac{n}{T}} |W_t - W_{t_k^n}| \right\|_p \\ &= \sqrt{\frac{T}{n}} \left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [k-1, k)} \underbrace{\sqrt{\frac{n}{T}}}_{\alpha} |W_{t \frac{T}{n}} - W_{(t-1) \frac{T}{n}}| \right\|_p \\ &= \sqrt{\frac{T}{n}} \left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [k-1, k)} \sqrt{\frac{n}{T}} |W_t - W_{k-1}| \right\|_p \\ &\quad \text{car } (\alpha W_{\frac{t}{\alpha^2}})_{t \in [0, T]} \text{ est un mouvement brownien.} \\ &= \sqrt{\frac{T}{n}} \left\| \max_{k \in \{1, \dots, n\}} \varepsilon_k \right\|_p \quad \text{où } \varepsilon_k = \sup_{t \in [k-1, k)} |W_t - W_{k-1}| \end{aligned}$$

$$\forall k \in \{1, \dots, n\}, \varepsilon_k \geq |W_k - W_{k-1}| \geq 0.$$

Ainsi,

$$\begin{aligned} \left\| \sup_{t \in [0, T]} |W_t - \overline{W_t}| \right\|_p &\geq \sqrt{\frac{T}{N}} \left\| \max_{k \in \{1, \dots, n\}} |W_k - W_{k-1}| \right\|_p \\ &\geq \sqrt{\frac{T}{N}} \left\| \max_{k \in \{1, \dots, n\}} |W_k - W_{k-1}| \right\|_p \\ &\quad \text{où } (|W_k - W_{k-1}|)_{1 \leq k \leq n} \text{ sont iid avec } W_k - W_{k-1} \sim \mathcal{N}(0, 1) \\ &\geq \sqrt{\frac{T}{N}} \sqrt{\left\| \max_{k \in \{1, \dots, n\}} |W_k - W_{k-1}|^2 \right\|_2^{\frac{p}{2}}} \\ &\quad \text{car } \mathbb{E}[|Z|^p]^{\frac{1}{p}} = \sqrt{\mathbb{E}[|Z|^2]^{\frac{p}{2}}}]^{\frac{1}{p}} \\ &\geq \sqrt{\frac{T}{n}} c_p \sqrt{\log n} \end{aligned}$$

où c_p est une constante indépendante de n . Cette dernière inégalité est laissée en exercice.

□ BORNE SUPÉRIEURE

$$\left\| \sup_{t \in [0, T]} |W_t - \bar{W}_t| \right\|_p = \sqrt{\frac{T}{n}} \sqrt{\left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [k-1, k[} |W_t - W_{k-1}|^2 \right\|_{\frac{p}{2}}}$$

Par translation, $\sup_{t \in [k-1, k[} |W_t - W_{k-1}|^2 \stackrel{\mathcal{L}}{=} \sup_{t \in [0, 1[} |W_t|^2 = \varepsilon$

Si on est capable de prouver que $\mathbb{E}[e^{\lambda \varepsilon}] < +\infty$ pour $\lambda > 0$, alors,

$$\left\| \max_{k \in \{1, \dots, n\}} \sup_{t \in [k-1, k[} |W_t - W_{k-1}|^2 \right\|_{\frac{p}{2} \leq c_{p, \lambda}} \log(n+1)$$

Or, pour $\lambda > 0$,

$$\begin{aligned} \mathbb{E}[e^{\lambda \varepsilon}] &= \mathbb{E}[e^{\lambda \sup_{t \in [0, 1]} |W_t|^2}] \\ &= \mathbb{E}[e^{\lambda \max(\sup_{t \in [0, 1]} (W_t), \sup_{t \in [0, 1]} (-W_t))^2}] \\ &\leq \mathbb{E}[e^{\lambda \sup_{t \in [0, 1]} (W_t)} + e^{\lambda \sup_{t \in [0, 1]} (-W_t)}] \\ &\leq 2\mathbb{E}[e^{\lambda (\sup_{t \in [0, 1]} (W_t))^2}] \end{aligned}$$

car $(-W_t)_{0 \leq t \leq 1}$ a la même loi que $(W_t)_{0 \leq t \leq 1}$.

D'après le principe de réflexion $\sup_{t \in [0, 1]} W_t = |W_1|$ (cf TD). Ainsi,

$$\begin{aligned} \mathbb{E}[e^{\lambda \varepsilon}] &\leq 2\mathbb{E}[e^{\lambda W_1^2}] \\ &\leq 2 \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\lambda x^2} e^{-\frac{x^2}{2}} dx \text{ car } W_1 \sim \mathcal{N}(0, 1) \\ &\leq 2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(1-2\lambda)x^2} dx \\ &\leq \frac{2}{\sqrt{1-\lambda}} \underbrace{\frac{1}{\sqrt{2\pi}} \sqrt{1-2\lambda} \int_{\mathbb{R}} e^{\frac{-x^2}{2\left(\frac{1}{\sqrt{1-2\lambda}}\right)^2}} dx}_1 \\ &\leq \frac{2}{\sqrt{1-\lambda}} \end{aligned}$$

Si $\lambda \in [0, \frac{1}{2}]$, on a $\mathbb{E}[e^{\lambda \varepsilon}] < +\infty$. On a donc bien une borne supérieure de la forme $c_p \sqrt{\log(n+1)} \sqrt{\frac{T}{n}}$.

Contrôle de l'erreur dans le cas général

On veut contrôler ici $\left\| \sup_{t \in [0, T]} |X_t - \bar{X}_t| \right\|_p$. Pour cela on va utiliser le lemme intermédiaire suivant.

Lemme 4.3.0.1 (Gronwald) Soit $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction continue, localement bornée et $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ croissante. Si pour tout $t \geq 0$, pour $\alpha \geq 0$,

$$f(t) \leq \alpha \int_0^t f(s) ds + \psi(s)$$

Alors :

$$\sup_{s \in [0, t]} f(s) \leq e^{\alpha t} \psi(t)$$

Preuve : à écrire (cf cours de prépa)

Propriété 4.3.1 Si b et σ sont lipschitziennes alors pour tout $p \geq 2$, il existe une constante c_p telle que

$$\left\| \sup_{t \in [0, T]} |X_t - \bar{X}_t| \right\|_p \leq c_p \left(\sqrt{\frac{T}{n}} + 1 \right)$$

On effectue la preuve dans le cas $p = 2$.

Preuve : notons que $X_t - \bar{X}_t = \int_0^t (b(X_s) - b(\bar{X}_s))ds + \int_0^t (\sigma(X_s) - \sigma(\bar{X}_s))dW_s$

On écrit alors :

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in [0, T]} |X_t - \bar{X}_t|^2 \right] &= \mathbb{E} \left[\sup_{t \in [0, T]} \left(\int_0^t b(X_s) - b(\bar{X}_s)ds + \int_0^t \sigma(X_s) - \sigma(\bar{X}_s)dW_s \right)^2 \right] \\ &\leq 2\mathbb{E} \left[\sup_{t \in [0, T]} \left(\int_0^t (b(X_s) - b(\bar{X}_s))ds \right)^2 \right] + 2\mathbb{E} \left[\sup_{t \in [0, T]} \left(\int_0^t (\sigma(X_s) - \sigma(\bar{X}_s))dW_s \right)^2 \right] \\ &\text{en utilisant Cauchy Schwartz} \\ &\leq 2T \mathbb{E} \left[\underbrace{\int_0^T |b(X_s) - b(\bar{X}_s)|^2 ds}_{|b(x) - b(y)| \leq c_b |x - y| \text{ (Lipschitz)}} \right] + 2\mathbb{E} \left[\sup_{t \in [0, T]} \left(\int_0^t (\sigma(X_s) - \sigma(\bar{X}_s))dW_s \right)^2 \right] \\ &\leq 2T c_b^2 \mathbb{E} \left[\int_0^T |X_s - \bar{X}_s|^2 ds \right] + 2\mathbb{E} \left[\sup_{t \in [0, T]} \left(\int_0^t (\sigma(X_s) - \sigma(\bar{X}_s))dW_s \right)^2 \right] \\ &\text{par l'inégalité de Douk et par l'isométrie d'Ito, on obtient} \\ &\leq 2T c_b^2 \mathbb{E} \left[\int_0^T |X_s - \bar{X}_s|^2 ds \right] + 8\mathbb{E} \left[\int_0^T \underbrace{(\sigma(X_s) - \sigma(\bar{X}_s))^2}_{\leq c_\sigma^2 |X_s - \bar{X}_s|} ds \right] \\ &\leq C \mathbb{E} \left[\int_0^T |X_s - \bar{X}_s|^2 ds \right] \\ &\leq C \left(\mathbb{E} \left[\int_0^T |X_s - \bar{X}_s|^2 ds \right] + \mathbb{E} \left[\int_0^T |\bar{X}_s - \bar{X}_s|^2 ds \right] \right) \\ &\leq C \int_0^T \mathbb{E} \left[\sup_{0 \leq u \leq s} |X_u - \bar{X}_u|^2 ds \right] + \mathbb{E} \left[\int_0^T |\bar{X}_s - \bar{X}_s|^2 ds \right] \end{aligned}$$

Par la lemme de GRONWALD, sur la première partie de l'inégalité, on trouve le résultat.

4.3.2 Autres résultats (plus généraux) sur le schéma d'Euler

Dans cette section, nous proposons quelques résultats supplémentaires sur les performances du schéma d'EULER sous des hypothèses légèrement plus faibles. On considère l'EDS suivante :

$$\begin{cases} X_0 = x_0 \\ dX_t = b(X_t)dt + \sigma(X_t)dW_t \end{cases}$$

Propriété 4.3.2 Si b et σ sont à croissance polynomiale, (par exemple, pour tout x , $|b(x)| \leq C(1 + |x|)$ et $|\sigma(x)| \leq C(1 + |x|)$), on a pour tout $p > 0$

- $\| \sup_{0 \leq t \leq T} |X_t| \|_p \leq C_{p,T}(1 + \|X_0\|_p)$
- $\| \sup_{0 \leq t \leq T} |\bar{X}_t^n| \|_p \leq C_{p,T}(1 + \|X_0\|_p)$

On suppose dans le cas le plus général que b et σ ne sont plus homogènes et donc que

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$$

En ajoutant une hypothèse de régularité, on conserve la convergence du schéma de discrétisation. On suppose donc qu'il existe $\beta \in]0, 1[$ tel que

$$\forall (s, t) \in [0, T]^2, \forall x \in \mathbb{R}, |b(x, t) - b(x, s)| + |\sigma(x, t) - \sigma(x, s)| \leq C(1 + |x|)|t - s|^\beta$$

et que (comme dans le cas démontré),

$$\forall s \in [0, T], \forall (x, y) \in \mathbb{R}^2, |b(x, s) - b(y, s)| + |\sigma(x, s) - \sigma(y, s)| \leq C|x - y|$$

Alors, pour tout $p > 0$,

$$\|\sup |X_t - \overline{X}_t^n|\|_p \leq C \left(\frac{T}{n} \right)^{\min(\beta, \frac{1}{2})}$$

où $(X_t^n)_{0 \leq t \leq T}$ est le schéma d'EULER « continu » obtenu de la façon suivante :

$$\forall k \in \{1, \dots, n\}, \overline{X}_{t_{k+1}^n} = \overline{X}_{t_k^n} + b(\overline{X}_{t_k^n}, t_k^n) \frac{T}{n} + \sigma(\overline{X}_{t_k^n}, t_k^n) \underbrace{(W_{t_{k+1}^n} - W_{t_k^n})}_{\sqrt{\frac{T}{n}} \varepsilon_{k+1} \text{ où } \varepsilon_{k+1} \sim \mathcal{N}(0, 1)}$$

avec $t_k^n = k \frac{T}{n}$ et pour tout $k \in \{1, \dots, n\}$ et tout $t \in]t_k^n, t_{k+1}^n[$

$$\overline{X}_t = \overline{X}_{t_k^n} + b(\overline{X}_{t_k^n}, t_k^n)(t - t_k^n) + \sigma(\overline{X}_{t_k^n}, t_k^n)(W_t - W_{t_k^n})$$

Propriété 4.3.3 Dans le cas où on suppose $\sup_{0 \leq t \leq T, x \in \mathbb{R}} |\sigma(x, t)| < +\infty$, alors pour tout $n \in \mathbb{N}$ et pour tout $\lambda > 0$, pour toute fonction f régulière,

$$\mathbb{E}[e^{\lambda(f(\overline{X}_T^n) - \mathbb{E}[f(\overline{X}_T^n)])}] \leq e^{\frac{\lambda^2}{2} C_{f,T} K_n}$$

où pour tout $n \geq 0, K_n > 0$ et $(K_n)_{n \in \mathbb{N}}$ est décroissante.

Application à l'obtention d'intervalles de confiance

On simule $M \geq 1$ réalisations indépendantes de $\overline{X}_T^n : (X_T^{n,i})_{1 \leq i \leq M}$ et on estime $\mathbb{E}[f(\overline{X}_T^n)]$ par $\frac{1}{M} \sum_{i=1}^M f(\overline{X}_T^{n,i})$

On fixe alors $\varepsilon > 0$ et on veut majorer

$$\mathbb{P} \left(\frac{1}{M} \sum_{i=1}^M f(\overline{X}_T^{n,i}) - \mathbb{E}[f(\overline{X}_T^n)] > \varepsilon \right)$$

Pour tout $\lambda > 0$, en utilisant la croissance de $x \mapsto e^x$,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{M} \sum_{i=1}^M f(\overline{X}_T^{n,i}) - \mathbb{E}[f(\overline{X}_T^n)] \geq \varepsilon \right) &= \mathbb{P} \left(e^{\lambda \left(\frac{1}{M} \sum_{i=1}^M f(\overline{X}_T^{n,i}) - \mathbb{E}[f(\overline{X}_T^n)] \right)} \geq e^{\lambda \varepsilon} \right) \\ &= \mathbb{P} \left(e^{\lambda \left(\sum_{i=1}^M f(\overline{X}_T^{n,i}) - \mathbb{E}[f(\overline{X}_T^n)] \right)} \geq e^{\lambda \varepsilon M} \right) \\ &\leq e^{-\lambda \varepsilon M} \mathbb{E} \left[e^{\lambda \left(\sum_{i=1}^M f(\overline{X}_T^{n,i}) - \mathbb{E}[f(\overline{X}_T^n)] \right)} \right] \\ &\leq e^{-\lambda \varepsilon M} e^{\frac{\lambda^2 M}{2} C_{f,T} K_n} \text{ par indépendance} \end{aligned}$$

On minimise ensuite la fonction $\lambda \mapsto -\lambda \varepsilon M \frac{\lambda^2 M}{2} C_{f,T} K_n$. Ce minimum est

$$\lambda_0 = \frac{\varepsilon M}{2 \frac{M C_{f,T} K_n}{2}} = \frac{\varepsilon}{C_{f,T} K_n}$$

Finalement, avec ce choix de λ_0 ,

$$\begin{aligned} \mathbb{P} \left(\frac{1}{M} \sum_{i=1}^M f(\overline{X_T^{n,i}}) - \mathbb{E}[f(\overline{X_T^n})] \geq \varepsilon \right) &\leq e^{-\frac{\varepsilon^2}{CK_n} M + \frac{M}{2} \frac{\varepsilon^2}{(CK_n)^2} CK_n} \\ &\leq e^{-\frac{\varepsilon^2 M}{2CK_n}} \end{aligned}$$

Par symétrie, on obtient la même chose. Ainsi,

$$\begin{aligned} \mathbb{P} \left(\left| \frac{1}{M} \sum_{i=1}^M f(\overline{X_T^{n,i}}) - \mathbb{E}[f(\overline{X_T^n})] \right| \geq \varepsilon \right) &\leq 2e^{-\frac{\varepsilon^2}{CK_n} M + \frac{M}{2} \frac{\varepsilon^2}{(CK_n)^2} CK_n} \\ &\leq 2e^{-\frac{\varepsilon^2 M}{2CK_n}} \end{aligned}$$

En pratique, on fournit un seuil $\delta \in]0, 1[$, on fixe alors ε_δ tel que $\delta = 2e^{-\frac{\varepsilon_\delta^2 M}{2CK_n}}$ et on a, avec probabilité inférieure à δ ,

$$\left| \frac{1}{M} \sum_{i=1}^M f(\overline{X_T^{n,i}}) - \mathbb{E}[f(\overline{X_T^n})] \right| > \varepsilon_\delta$$

On sait donc quantifier l'erreur entre $\frac{1}{M} \sum_{i=1}^M f(\overline{X_T^{n,i}})$ et $\mathbb{E}[f(\overline{X_T^n})]$. Par ailleurs, pour f suffisamment régulière, on peut montrer que :

$$|\mathbb{E}[f(X_T)] - \mathbb{E}[f(\overline{X_T^n})]| \leq \frac{C}{n^\alpha} + o\left(\frac{1}{n}\right) \text{ pour un } \alpha \in \left]0, \frac{1}{2}\right[$$

où C est une constante dépendante de f et T .

4.4 Schémas d'ordre plus élevé et application à l'estimation

On peut définir des schémas d'approximation plus fins que le schéma d'Euler en approchant $\int_0^t b(X_s)ds$ et $\int_0^t \sigma(X_s)dW_s$ plus précisément. Un schéma très répandu est le schéma de MILSTEIN. On définit une discrétisation de $[0, T]$ par $t_k^n = k \frac{T}{n}$ pour $n \geq 1$ et $k \in \{0, \dots, n\}$.

On introduit alors

$$\begin{cases} \widetilde{X}_0 = x_0 \\ \forall k \in \{0, \dots, n-1\} \quad \widetilde{X_{t_{k+1}^n}} = \widetilde{X_{t_k^n}} + b(\widetilde{X_{t_k^n}}) \frac{T}{n} + \sigma(\widetilde{X_{t_k^n}}) \underbrace{\sqrt{\frac{T}{n}} \varepsilon_k}_{W_{t_{k+1}^n} - W_{t_k^n}} + \frac{\sigma^2(\widetilde{X_{t_k^n}})}{2} \frac{T}{n} (\varepsilon_k^2 - 1) \end{cases}$$

où $\varepsilon \sim \mathcal{N}(0, 1)$.

Une version continue de cette discrétisation pour $t \in]t_k^n, t_{k+1}^n[$, $k \in \{0, \dots, n-1\}$ se calcule de façon similaire au schéma d'EULER. Pour des coefficients de l'EDS réguliers (HÖLDER d'ordre $\beta \in]0, 1[$ par exemple), on obtient un résultat de la forme :

$$\| \sup_{0 \leq t \leq T} |X_t - \widetilde{X_t^n}|_p \leq C \left(\frac{T}{n} \right)^{\frac{\beta+1}{2}}$$

Estimation de paramètres

Dans ce cadre, $(X_t)_{0 \leq t \leq T}$ est solution de $dX_t = b_\theta(\overline{X}_t)dt + \sigma_\theta(\overline{X}_t)dW_t$ où $\theta \in \mathbb{R}^d$ est un paramètre à estimer. L'objectif en général est d'estimer θ à partir d'observations $(X_{t_k^n})_{1 \leq k \leq n}$ où $0 \leq t_1^n \leq \dots \leq t_n^n \leq T$. Une solution très répandue est de résoudre le problème

$$\hat{\theta}_n \in \arg \max \{ \theta \mapsto \log p_\theta(X_{t_1^n}, \dots, X_{t_n^n}) \}$$

On sait que pour tout $\theta \in \mathbb{R}^d$

$$\log P_\theta(X_{t_1^n}, \dots, X_{t_n^n}) = \sum_{k=1}^n \log P_\theta(X_{t_k^n} | X_{t_{k-1}^n}) \text{ avec } X_{t_0^n} = x_0$$

On se propose ensuite de maximiser $\theta \mapsto \frac{1}{n} \sum_{k=1}^n \log \widehat{P}_\theta(X_{t_k^n} | X_{t_{k-1}^n})$ où $\widehat{P}_\theta(X_{t_k^n} | X_{t_{k-1}^n})$ est une estimation de $P_\theta(X_{t_k^n} | X_{t_{k-1}^n})$.

Cas du schéma d'Euler

$\forall k \in \{1, \dots, n\}$, $\widehat{P}_\theta(\cdot | X_{t_{k-1}^n})$ est une densité gaussienne de moyenne $\mu_k^n = X_{t_{k-1}^n} + b(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)$ et de variance $(\sigma_k^n)^2(\theta) = \sigma_\theta^2(X_{t_{k-1}^n})(t_k^n - t_{k-1}^n)$.

\triangle Valide et efficace uniquement lorsque $\sup_{0 \leq t \leq T} |t_k^n - t_{k-1}^n| \xrightarrow{n \rightarrow +\infty} 0$.

4.5 Exercices

4.5.1 Brownian motion

Soit $(W_t)_{t \geq 0}$ un mouvement brownien issu de 0.

1. (a) On définit $Z_t = W_{t+t_0} - W_{t_0}$ pour tout $t \geq 0$ et un certain $t_0 \geq 0$. Montrons que Z_t est un mouvement brownien. On utilise la propriété 4.1.3. Utilisons cette équivalence :

$$\begin{aligned} \mathbb{E}[Z_t] &= \mathbb{E}[W_{t+t_0} - W_{t_0}] \\ &= \mathbb{E}[W_{t+t_0}] - \mathbb{E}[W_{t_0}] \\ &= 0 \end{aligned}$$

De façon immédiate, nous avons $t \mapsto Z_t$ qui est continue et nulle en 0. Par ailleurs, montrons que si $s \leq t$, $\mathbb{E}[Z_s Z_t] = s$.

$$\begin{aligned} \mathbb{E}[(W_{s+t_0} - W_{t_0})(W_{t+t_0} - W_{t_0})] &= \mathbb{E}[W_{s+t_0} W_{t+t_0} + W_{t_0}^2 - W_{s+t_0} W_{t_0} - W_{t_0} W_{t+t_0}] \\ &= \mathbb{E}[W_{s+t_0} W_{t+t_0}] + \mathbb{E}[W_{t_0}^2] - \mathbb{E}[W_{s+t_0} W_{t_0}] - \mathbb{E}[W_{t_0} W_{t+t_0}] \\ &= s + t_0 + t_0 - t_0 - t_0 \\ &= s \end{aligned}$$

Par linéarité, $(Z_t)_{t \geq 0}$ est un processus gaussien. On en déduit donc que $(Z_t)_{t \geq 0}$ est un mouvement brownien.

- (b) On définit $(\widetilde{Z}_t = \alpha W_{\frac{t}{\alpha^2}})_{t \geq 0}$. C'est un processus gaussien par linéarité, centré, nul en 0 et à trajectoires continues (par composition). Ensuite, pour tout $(s, t) \in \mathbb{R}^2$, tel que $s \leq t$,

$$\begin{aligned} \mathbb{E}[\widetilde{Z}_s \widetilde{Z}_t] &= \mathbb{E}\left[\left(\alpha W_{\frac{s}{\alpha^2}}\right)\left(\alpha W_{\frac{t}{\alpha^2}}\right)\right] \\ &= \alpha^2 \mathbb{E}\left[W_{\frac{s}{\alpha^2}} W_{\frac{t}{\alpha^2}}\right] \\ &= \alpha^2 \frac{s}{\alpha^2} = s = s \wedge t \end{aligned}$$

Ainsi $(\widetilde{Z}_t)_{t \geq 0}$ est donc bien un mouvement brownien.

2. Soit $0 \leq s \leq t$.

$$\begin{aligned}
\mathbb{E}[W_s W_t^2] &= \mathbb{E}[W_s (W_t - W_s + W_s)^2] \\
&= \mathbb{E}[W_s (W_t - W_s)^2] + 2\mathbb{E}[W_s^2 (W_t - W_s)] + \mathbb{E}[W_s^3] \\
&= \mathbb{E}[W_s] \mathbb{E}[(W_t - W_s)^2] + 2\mathbb{E}[W_s^2] \mathbb{E}[W_t - W_s] + \mathbb{E}[W_s^3] \text{ car } W_t - W_s \text{ est indépendant de } W_s \\
&= 0 + 0 + 0 \\
&= 0
\end{aligned}$$

Rappel : les moments impaire d'une gaussienne centré sont nuls.

Ensuite,

$$\begin{aligned}
\mathbb{E}[W_t | W_s] &= \mathbb{E}[W_t - W_s + W_s | W_s] \\
&= \mathbb{E}[W_t - W_s | W_s] + \mathbb{E}[W_s | W_s] \\
&= \mathbb{E}[W_t - W_s | W_s] + W_s \text{ car } W_t - W_s \text{ est indépendant de } W_s \\
&= W_s
\end{aligned}$$

Enfin, pour le dernier calcul,

$$\begin{aligned}
\mathbb{E}[(W_t - W_s)^2 Y] &= \mathbb{E}[\mathbb{E}[(W_t - W_s)^2 Y | \sigma(\{W_u\}_{0 \leq u \leq s})]] \\
&= \mathbb{E}[Y \mathbb{E}[(W_t - W_s)^2 | \sigma(\{W_u\}_{0 \leq u \leq s})]] \\
&= \mathbb{E}[Y \mathbb{E}[(W_t - W_s)^2]] \text{ car } W_t - W_s \text{ est indépendant de } \sigma(\{W_u\}_{0 \leq u \leq s}) \\
&= \mathbb{E}[Y(t - s)] \\
&= (t - s) \mathbb{E}[Y]
\end{aligned}$$

3. On dispose de deux browniens indépendants. $(Z_t)_{t \leq 0}$ est un processus gaussien, centré, à trajectoires continues, nul en 0, par combinaison linéaire de $(B_t)_{t \leq 0}$, $(W_t)_{t \leq 0}$ et de ρ qui est connu.

Soit $(s, t) \in \mathbb{R}^2$, tel que $s \leq t$. Alors

$$\begin{aligned}
\mathbb{E}[Z_s Z_t] &= \mathbb{E}[(\rho W_s + \sqrt{1 - \rho^2} B_s)(\rho W_t + \sqrt{1 - \rho^2} B_t)] \\
&= \rho^2 \mathbb{E}[W_s W_t] + (1 - \rho^2) \mathbb{E}[B_t B_s] + \rho \sqrt{1 - \rho^2} \mathbb{E}[W_s B_t + W_t B_s] \\
&= \rho^2 s + (1 - \rho^2)s + 0 \text{ car } W \text{ et } B \text{ sont indépendants} \\
&= s
\end{aligned}$$

4.5.2 Brownian bridge

Soit $(B_t)_{t \geq 0}$ un processus gaussien centré et tel que pour tout $(s, t) \in [0, 1]^2$, $\mathbb{E}[B_t B_s] = s \wedge t - st$. C'est la loi d'un brownien dont on sait qu'il se termine d'une certaine valeur en 1.

1. Soit $(\widetilde{B}_t)_{0 \leq t \leq 1}$ tel que $\forall 0 \leq t \leq 1$, $\widetilde{B}_t = B_{1-t}$. \widetilde{B} est un processus gaussien et centré. On sait donc que sa loi est caractérisé uniquement par sa moyenne et sa variance. On sait donc, en calculant ces valeurs, qu'il aura la même loi que B . Ainsi, pour tout $(s, t) \in [0, 1]^2$ tel que $s \leq t$,

$$\begin{aligned}
\mathbb{E}[\widetilde{B}_s \widetilde{B}_t] &= \mathbb{E}[B_{1-s} B_{1-t}] \\
&= (1 - t) \wedge (1 - s) - (1 - t)(1 - s) \\
&= 1 - t - (1 - t)(1 - s) \\
&= s(1 - t) \\
&= s - st \\
&= s \wedge t - st
\end{aligned}$$

2. **Remarque :** Il y a un Z dans l'énoncé mais non défini. Petite typo dans l'énoncé, on va voir où cela va nous amener :-)

Si $(W_t)_{0 \leq t \leq 1}$ est un mouvement brownien, on définit $(\widetilde{W}_t = W_t - tW_1)_{0 \leq t \leq 1}$. $(\widetilde{W}_t)_{0 \leq t \leq 1}$ est un processus gaussien, centré comme $(W_t)_{t \in [0,1]}$. Pour $(s, t) \in [0, 1]^2$, $s \leq t$.

$$\begin{aligned}\mathbb{E}[\widetilde{W}_t \widetilde{W}_s] &= \mathbb{E}[(W_t - tW_1)(W_s - sW_1)] \\ &= \mathbb{E}[W_t W_s] - s\mathbb{E}[W_t W_1] - t\mathbb{E}[W_1 W_s] + st\mathbb{E}[W_1^2] \\ &= s - st - st + st \text{ le dernier terme égal à 1 car } W_1 \sim \mathcal{N}(0, 1) \\ &= s - st \\ &= s \wedge t - st\end{aligned}$$

$(\widetilde{W}_t)_{0 \leq t \leq 1}$ a bien la même loi que $(B_{1-t})_{t \in [0,1]}$.

4.5.3 Reflection principle - simulation of a first passage time - Loi du temps d'attente

Soit $(W_t)_{t \geq 0}$ un mouvement brownien, $S_t = \sup_{0 \leq s \leq t} W_s$. Soient $a \leq b$, tel que $b > 0$ et on cherche à prouver que

$$\mathbb{P}(S_t \geq b; W_t \leq a) = \mathbb{P}(W_t \geq 2b - a)$$

On introduit $\tau_b = \inf\{s \geq 0; W_s \geq b\}$ et $\mathcal{F}_s = \sigma((W_u)_{0 \leq u \leq s})$. On peut alors écrire :

$$\begin{aligned}\mathbb{P}(S_t \geq b; W_t \leq a) &= \mathbb{P}(\tau_b \leq t; W_t \leq a) \\ &= \mathbb{P}(\tau_b \leq t; W_t - W_{\tau_b} \leq a - W_{\tau_b}) \\ &= \mathbb{P}(\tau_b \leq t; W_t - W_{\tau_b} \leq a - b)\end{aligned}$$

On utilise alors le fait que le mouvement brownien vérifie la propriété de MARKOV forte (c'est dire que la loi du brownien sachant tout son passé est égal à la loi du brownien son passé sachant son passé au dernier temps). On peut donc dire que le processus $(Z_t = W_{t+\tau_b} - W_{\tau_b})_{t \geq 0}$ est un processus gaussien indépendant de τ_b .

Ainsi,

$$\begin{aligned}\mathbb{P}(S_t \geq b; W_t \leq a) &= \mathbb{P}(\tau_b \leq t; Z_{t-\tau_b} \leq a - b) \\ &= \mathbb{P}(\tau_b \leq t; -Z_{t-\tau_b} \leq a - b) \text{ (car événements indép. et symétrie de la gaussienne centré)} \\ &= \mathbb{P}(\tau_b \leq t; -W_t \leq a - 2b) \\ &= \mathbb{P}(\tau_b \leq t; W_t \geq 2b - a) \\ &= \mathbb{P}(W_t \geq 2b - a) \text{ car } \{W_t \geq 2b - a\} \subset \{Z_b \leq t\} \text{ lorsque } a \leq b\end{aligned}$$

Ensuite, pour tout $b \geq 0$,

$$\begin{aligned}\mathbb{P}(S_t \geq b) &= \mathbb{P}(S_t \geq b; W_t \geq b) + \mathbb{P}(S_t \leq b; W_t \leq b) \\ &= \mathbb{P}(W_t \geq b) + \mathbb{P}(W_t \geq 2b - b) \\ &= \mathbb{P}(W_t \geq b) + \mathbb{P}(-W_t \geq b) \\ &= \mathbb{P}(|W_t| \geq b)\end{aligned}$$

On veut trouver la densité de la loi du temps d'arrêt $\tau_b = \inf\{t \geq 0; W_t \geq b\}$ avec $b > 0$.

Soit $t \geq 0$,

$$\begin{aligned}\mathbb{P}(\tau_b \leq t) &= \mathbb{P}(S_t \geq b) \\ &= \mathbb{P}(|W_t| \geq b) \\ &= \mathbb{P}(W_t^2 \geq b^2) \\ &= \mathbb{P}((W_t - W_0)^2 \geq b^2)\end{aligned}$$

Or, $W_t - W_0 \sim \mathcal{N}(0, t)$ donc $\sqrt{t}(W_1 - W_0) \stackrel{\mathcal{L}}{=} W_t - W_0$. D'où,

$$\begin{aligned}\mathbb{P}(\tau_b \leq t) &= \mathbb{P}(tW_1^2 \geq b^2) \\ &= \mathbb{P}\left(W_1^2 \geq \frac{b^2}{t}\right)\end{aligned}$$

On cherche à calculer $\frac{\partial}{\partial t} \left(\mathbb{P}(W_1^2 \geq \frac{b^2}{t}) \right)$.

$$\begin{aligned}\frac{\partial}{\partial t} \left(\mathbb{P}(W_1^2 \geq \frac{b^2}{t}) \right) &= \frac{\partial}{\partial t} \int_{\frac{b^2}{t}}^{+\infty} f_{W_1^2}(u) du \\ &= \frac{2}{\sqrt{2\pi}} \frac{\partial}{\partial t} \int_{\frac{b^2}{t}}^{+\infty} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{b^2}{2t}} b t^{-3/2}\end{aligned}$$

4.5.4 Simulation of the maximum of a Brownian motion - Loi du maximum du pont brownien

1. Soit $(x, y) \in \mathbb{R}^2$, tel que $y \geq \max(0, x)$. On cherche $\mathbb{P}\left(\sup_{0 \leq s \leq t} W_s \geq y | W_t = x\right)$. Grâce à l'exercice 3, on peut écrire que

$$\begin{aligned}\mathbb{P}\left(\sup_{0 \leq s \leq t} W_s \geq y | W_t = x\right) &= \lim_{\delta \rightarrow 0} \frac{(\mathbb{P}(W_t \geq 2y - x - \delta) - \mathbb{P}(W_t \geq 2y - x)) / \delta}{(\mathbb{P}(W_t \leq x + \delta) - \mathbb{P}(W_t \leq x)) / \delta} \\ &= \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2t}(2y-x)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2t}x^2}} \\ &= e^{-\frac{1}{2t}(4y^2 - 4xy)} \\ &= e^{-\frac{2y}{t}(y-x)}\end{aligned}$$

2. $\mathbb{P}(Z \geq y) = e^{-\frac{2y}{t}(y-x)}$ pour $y \geq 0 \forall x$. La fonction de répartition de Z est donnée par $F_Z(y) = \mathbb{P}(Z \leq y) = 1 - e^{-\frac{2y}{t}(y-x)}$. Alors si $U \sim \mathcal{U}([0, 1])$, $F_Z^{-1}(u)$ a la même loi que Z .

4.6 Exercices supplémentaires

Exercice 4.6.1 On considère l'EDS linéaire suivante :

$$dX_t = a_t X_t dt + b_t dt + c_t dW_t$$

avec a, b et c des fonctions continues de $[0, T]$ dans \mathbb{R} et W un mouvement brownien réel. On note $A_t = \int_0^t a_s ds$ et on introduit $Y_t = e^{-A_t} X_t$.

- Déterminez l'EDS vérifiée par $(Y_t)_{0 \leq t \leq T}$
- Calculer l'espérance de X_t
- Calculer la covariance $\forall (t_1, t_2) \geq 0 \text{ cov}(X_{t_1}, X_{t_2})$.

1. On introduit la fonction $f : (t, x) \mapsto e^{-A_t} x$ et on applique la formule d'Itô à $Y_t = f(t, X_t)$. Pour rappel :

Propriété 4.6.1 (Itô) Soit $X_t = \varphi(t, \psi_t)$ un processus, où φ est une fonction de classe \mathcal{C}^2 allant de \mathbb{R}^2 dans \mathbb{R} , et ψ_t un mouvement brownien. On a alors : $dX_t = \left(\frac{\partial}{\partial t} \varphi(t, \psi_t) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \varphi(t, \psi_t) \right) dt + \frac{\partial}{\partial y} \varphi(t, \psi_t) d\psi_t$

Alors, étant donné que $\partial_{xx}f(t, x) = 0$ et $\frac{\partial^2}{\partial x^2}f(t, x) = 0$

$$\begin{aligned} dY_t &= \frac{\partial}{\partial t}f(t, X_t)dt + \frac{\partial}{\partial x}f(t, X_t)dX_t \\ &= -a_te^{-A_t}X_tdt + e^{-A_t}dX_t \\ &= -a_te^{-A_t}X_tdt + e^{-A_t}(a_tX_tdt + b_tdt + c_tdW_t) \\ &= e^{-A_t}b_tdt + e^{-A_t}c_tdW_t \end{aligned}$$

Ainsi, pour tout $t \geq 0$,

$$Y_t = Y_0 + \int_0^t e^{-A_s}b_sds + \int_0^t e^{-A_s}c_sdW_s$$

et

$$\begin{aligned} X_t &= e^{A_t}Y_t \\ &= e^{A_t}\left(X_0 + \int_0^t e^{-A_s}b_sds + \int_0^t e^{-A_s}c_sdW_s\right) \end{aligned}$$

Application : Processus d'ORNSTEIN-ULHENBECK :

$$\begin{cases} a_t = -1 \\ b_t = \mu \in \mathbb{R} \\ c_t = \sigma \\ dX_t = (\mu - X_t)dt + \sigma dW_t \end{cases}$$

2. On a

$$\begin{aligned} \mathbb{E}[X_t] &= e^{A_t}\mathbb{E}[Y_t] \\ &= e^{A_t}\left(\mathbb{E}[X_0] + \underbrace{\mathbb{E}\left[\int_0^t e^{-A_s}b_sds\right]}_{\text{espérance d'une constante}} + \underbrace{\mathbb{E}\left[\int_0^t e^{-A_s}c_sdW_s\right]}_0\right) \\ &= e^{A_t}\left(\mathbb{E}[X_0] + \int_0^t e^{-A_s}b_sds\right) \end{aligned}$$

3. Sans perte de généralité, on pose $X_0 = 0$. $\forall (t_1, t_2) \geq 0$

$$\begin{aligned} \text{cov}(X_{t_1}, X_{t_2}) &= \text{cov}\left(\int_0^{t_1} e^{A_{t_1}-A_s}c_sdW_s, \int_0^{t_2} e^{A_{t_2}-A_s}c_sdW_s\right) \\ &= sqf d f d \end{aligned}$$

Chapitre 5

Chaînes de Markov cachées

On considère une variable aléatoire X à valeurs dans $\{\omega_1, \omega_2\}$, dont les valeurs sont cachées. On considère une autre variable aléatoire Y à valeurs dans \mathbb{R} , dont les valeurs sont observables.

5.1 Position du problème

On cherche à estimer la valeur cachée ω du paramètre. Etant donné que l'on souhaite estimer la valeur cachée de ω , nous allons avoir besoin d'un estimateur. Cet estimateur s'appellera « stratégie de classification » et les valeurs cachées les « classes ». D'une manière générale, nous utilisons l'estimateur du maximum de vraisemblance. Le principe est le suivant :

Soit une famille paramétrique de distributions de probabilités D_θ dont les éléments sont associés soit à une densité de probabilité connue (distribution continue), soit à une fonction de masse connue (distribution discrète), notée $f(x|\theta)$. On tire un échantillon de n valeurs x_1, x_2, \dots, x_n de la distribution, et l'on calcule la densité de probabilité associée aux données observées :

$$f_\theta(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i | \theta)$$

Ceci étant une fonction de θ avec x_1, \dots, x_n fixés, c'est une vraisemblance.

$$L(\theta) = f_\theta(x_1, \dots, x_n; \theta)$$

Lorsque θ n'est pas observable, la méthode du maximum de vraisemblance utilise les valeurs de θ qui maximisent $L(\theta)$ estimateur de θ : c'est l'estimateur du maximum de vraisemblance de θ noté $\hat{\theta}$. Par exemple dans le cas du produit discret, on effectue un tirage de n valeurs, il faut donc trouver le paramètre qui maximise la probabilité d'avoir tiré ce tirage. L'estimateur du maximum de vraisemblance est un estimateur (statistique) utilisé pour inférer les paramètres de la loi de probabilité d'un échantillon donné.

Définition 5.1.1 (EMV) Soit X une variable aléatoire réelle, de loi discrète ou continue, dont on veut estimer un paramètre θ . On note \mathcal{D}_θ cette famille de lois paramétriques. Alors on définit une fonction f telle que :

$$f(x; \theta) = \begin{cases} f_\theta(x) & \text{si } X \text{ est une v.a. continue} \\ P_\theta(X = x) & \text{si } X \text{ est une v.a. discrète} \end{cases}$$

$f_\theta(x)$ représente la densité de X (où θ apparaît) et $P_\theta(X = x)$ représente une probabilité discrète (où θ apparaît).

On appelle vraisemblance de θ au vu des observations $(x_1, \dots, x_i, \dots, x_n)$ d'un n -échantillon indépendamment et identiquement distribué selon la loi \mathcal{D}_θ , le nombre :

$$L(x_1, \dots, x_i, \dots, x_n; \theta) = f(x_1; \theta) \times f(x_2; \theta) \times \dots \times f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

On cherche à trouver le maximum de cette vraisemblance pour que les probabilités des réalisations observées soient aussi maximum.

Ici, $\mathbb{P}(\cdot|\omega_i)$ est une probabilité dite « a priori » tandis que $\mathbb{P}(\cdot|Y)$ est une probabilité « a posteriori »

5.1.1 Approche bayésienne

D'une manière générale, on parle d'une approche « bayésienne » si une connaissance « a priori » sur les paramètres est modélisée par une loi de probabilité sur l'ensemble des paramètres.

On cherche un lien entre ces deux variables. Il y a deux liens « extrêmes possibles » :

- les variables aléatoires sont indépendantes
- les variables aléatoires sont liées de manières déterministes

D'une manière générale, on étudie la loi jointe $\mathbb{P}_{(X,Y)}$. Notons $h(x,y)$ la densité de cette loi jointe. Faisons quelques rappels de calculs :

- $\mathbb{P}_X : f(x) = \int_{\mathbb{R}} h(x,y)dy$
- $\mathbb{P}_Y : f(y) = h(\omega_1,y) + h(\omega_2,y)$
- $\mathbb{P}(X|Y) = \frac{h(x,y)}{f(y)}$

Ainsi, $h(x,y) = \mathbb{P}(x)\mathbb{P}(Y|X)$.

Notons \hat{s} notre stratégie de classification (c'est notre estimateur). Pour évaluer notre estimateur, nous calculons l'erreur de notre estimation à travers une fonction de perte. En effet, notre estimateur peut nous donner une bonne réponse ou se tromper. On peut supposer que les différentes erreurs ne sont pas de gravité équivalente. Dans ce cas là, on modifie ces différences par la fonction de perte suivante :

$$L(\omega_i, \omega_j) = \begin{cases} 0 & \text{si } \omega_i \neq \omega_j \\ \lambda_{ij} & \text{sinon} \end{cases}$$

Remarque : les λ_{ij} sont définis de manière subjectif. Cela dépend de notre problème et de ce que nous définissons comme une erreur de classification.

Pour mesurer la qualité de notre estimateur (que l'on appelle aussi stratégie de classification) à l'aide de n observations, en calculant la perte moyenne $\mathbb{E}[L(\hat{s}(Y), X)]$. Cette valeur est approchée par :

$$\frac{L(\hat{s}(y_1), x_1) + \dots + L(\hat{s}(y_n), x_n)}{n}$$

Par la loi des grands nombres, ce dernier tend bien vers $\mathbb{E}[L(\hat{s}(Y), X)]$. Finalement on définit la stratégie bayésienne \hat{s}_B .

Définition 5.1.2 (Stratégie Bayésienne) *On appelle stratégie bayésienne la stratégie de classification \hat{s}_B qui vérifie*

$$\mathbb{E}[L(\hat{s}_B(Y), X)] = \min_s \mathbb{E}[L(\hat{s}(Y), X)]$$

5.2 Chaînes de Markov cachées

Une chaîne de MARKOV cachée est un processus à temps discret doublement stochastique, c'est à dire composé de deux processus $X = (X_n)_{n \in \mathbb{N}}$ et $Y = (Y_n)_{n \in \mathbb{N}}$. On suppose que X est une chaîne de MARKOV et nous supposons que Y est réel. Ici, on parle d'une chaîne de MARKOV « cachée » signifie que les réalisations de X sont inobservables. On cherche alors à estimer X à partir de la réalisation observée Y .

Définition 5.2.1 $\mathbb{P}_{(X,Y)}$ est une chaîne de MARKOV cachée si

$$h(x,y) = \mathbb{P}(x)\mathbb{P}(y|x) = \mathbb{P}(x_1)\mathbb{P}(x_2|x_1) \dots \mathbb{P}(x_n|x_{n-1})\mathbb{P}(y_1|x_1) \dots \mathbb{P}(y_n|x_n)$$

où

- X est une chaîne de MARKOV
- les (y_k) indépendants conditionnellement à $X = x_j$
- $\mathbb{P}(y_i|x_1, \dots, x_n) = \mathbb{P}(y_i|x_i)$

5.2.1 Représentation graphique des dépendances

On considère une famille finie de variables $U = (U_s)_{s \in S}$. On peut représenter les liens de dépendances entre ces différentes variables aléatoires $(U_s)_{s \in S}$. S'il existe une arête entre U_i et U_j on ne peut rien dire, S'il n'existe pas d'arête, alors U_i et U_j sont indépendantes.

5.2.2 Modèle de chaîne de Markov cachée

Petit calcul : sachant que l'on connaît $\mathbb{P}(x_i, y_1, \dots, y_n)$, peut-on trouver $\mathbb{P}(x_i | y_1, \dots, y_n)$, $\forall i \in \{1, \dots, n\}$. La réponse est oui. Montrons cela.

On note $\alpha_i(x_i) = \mathbb{P}(x_i, y_1, \dots, y_i)$ et $\beta_i(x_i) = \mathbb{P}(y_{i+1}, \dots, y_n | x_i)$. Montrons que $\mathbb{P}(x_i | y_1, \dots, y_n) = \alpha_i(x_i) \beta_i(x_i)$

On utilise le fait que $\mathbb{P}(a, b) = \mathbb{P}(a) \mathbb{P}(b | a) = \mathbb{P}(b) \mathbb{P}(a | b)$. On note sur le graphe suivant a en vert et b en rouge. Le tout correspond à a et b .

$$\begin{aligned} \mathbb{P}(\underbrace{x_i, y_1, \dots, y_i}_a, \underbrace{\dots, y_n}_b) &= \mathbb{P}(x_i, y_1, \dots, y_i) \mathbb{P}(y_{i+1}, \dots, y_n | x_i, y_1, \dots, y_i) \\ &= \alpha_i(x_i) \mathbb{P}(y_{i+1}, \dots, y_n | x_i, y_1, \dots, y_i) \\ &= \alpha_i(x_i) \mathbb{P}(y_{i+1}, \dots, y_n | x_i) \text{ par indépendance vis à vis des } y_1, \dots, y_i \\ &= \alpha_i(x_i) \beta_i(x_i) \end{aligned}$$

Maintenant, comment passer de $\alpha_i(x_i)$ à $\alpha_{i+1}(x_{i+1})$?

$$\begin{aligned} \alpha_{i+1}(x_{i+1}) &= \mathbb{P}(x_{i+1}, y_1, \dots, y_{i+1}) \\ &= \sum_{x_i \in \Omega} \mathbb{P}(x_i, x_{i+1}, y_1, \dots, y_{i+1}) \text{ (probabilité totale)} \\ &= \sum_{x_i \in \Omega} \mathbb{P}(x_i, y_1, \dots, y_n) \mathbb{P}(x_{i+1}, y_{i+1} | y_1, \dots, y_n, x_i) \\ &= \sum_{x_i \in \Omega} \alpha_i(x_i) \mathbb{P}(x_{i+1}, y_{i+1} | x_i) \text{ par indépendance conditionnelle des } y_k \\ &= \sum_{x_i \in \Omega} \alpha_i(x_i) \mathbb{P}(y_{i+1} | x_i, x_{i+1}) \mathbb{P}(x_{i+1} | x_i) \\ &= \sum_{x_i \in \Omega} \alpha_i(x_i) \mathbb{P}(y_{i+1} | x_{i+1}) \mathbb{P}(x_{i+1} | x_i) \end{aligned}$$

Enfin, on pose $\beta_n(x_n) = 1$. Comment passer de $\beta_{i+1}(x_{i+1})$ à $\beta_i(x_i)$?

$$\begin{aligned} \beta_i(x_i) &= \mathbb{P}(y_{i+1}, \dots, y_n | x_i) \\ &= \sum_{x_{i+1} \in \Omega} \mathbb{P}(x_{i+1}, y_{i+1}, \dots, y_n | x_i) \\ &= \sum_{x_{i+1} \in \Omega} \mathbb{P}(x_{i+1} | x_i) \mathbb{P}(y_{i+1}, \dots, y_n | x_i, x_{i+1}) \\ &= \sum_{x_{i+1} \in \Omega} \mathbb{P}(x_{i+1} | x_i) \mathbb{P}(y_{i+1}, \dots, y_n | x_{i+1}) \text{ 3ème propriété de la définition d'une CMC} \\ &= \sum_{x_{i+1} \in \Omega} \mathbb{P}(x_{i+1} | x_i) \mathbb{P}(y_{i+1} | x_{i+1}) \mathbb{P}(y_{i+2}, \dots, y_n | x_{i+1}) \\ &= \sum_{x_{i+1} \in \Omega} \mathbb{P}(x_{i+1} | x_i) \mathbb{P}(y_{i+1} | x_{i+1}) \beta_{i+1}(x_{i+1}) \end{aligned}$$

Nous souhaitons démontrer l'équivalence suivante :

$$\widehat{s}_B^b(y_1, \dots, y_n) = (\widehat{x}_1, \dots, \widehat{x}_n) \Leftrightarrow \widehat{x}_i = \arg \max_{x_i} \mathbb{P}(x_i | y_1, \dots, y_n)$$

On note pour cela $L(x^1, x^2) = \sum_{i=1}^n \mathbb{1}_{[x_i^1 \neq x_i^2]}$. On cherche déjà notre estimateur bayésien. Pour cela, on calcule l'espérance de la fonction de perte et on cherchera à minimiser cette espérance.

$$\begin{aligned}\mathbb{E}[L(\widehat{s(Y)}, X)] &= \mathbb{E}\left[\sum_{i=1}^n \mathbb{1}_{[\widehat{X}_i(Y) \neq X_i]}\right] \\ &= \sum_{i=1}^n \underbrace{\mathbb{E}\left[\mathbb{1}_{[\widehat{X}_i(Y) \neq X_i]}\right]}_{\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{[\widehat{X}_i(Y) \neq X_i]}\right] | Y=y\right]} \\ &= \sum_{i=1}^n \mathbb{P}\left[\widehat{X}_i(y) \neq X_i | Y = y\right]\end{aligned}$$

Chapitre 6

Mélanges de lois et classification non supervisée

Comme dans le chapitre précédent, on considère deux variables aléatoires X à valeurs discrètes $\{F, G\}$ et Y à valeurs dans \mathbb{R} . Le problème consiste à trouver une estimation \hat{x} que l'on note \hat{x} sachant $Y = y$. D'après ce que nous avons vu au chapitre précédent, c'est possible à l'aide de $h(x, y) = \mathbb{P}(x) \underbrace{\mathbb{P}(y|x)}_{\text{gaussienne}}$. On note la stratégie bayésienne :

$$\hat{s}(y) = \begin{cases} F & \text{si } \mathbb{P}(F)\mathbb{P}(Y|F) \geq \mathbb{P}(G)\mathbb{P}(Y|G) \\ G & \text{si } \mathbb{P}(F)\mathbb{P}(Y|F) \leq \mathbb{P}(G)\mathbb{P}(Y|G) \end{cases}$$

où $\mathbb{P}(F) = \Pi_F$ et $\mathbb{P}(G) = \Pi_G$. On considère ces observations correspondant à la taille d'un individu en mètre :

$$\begin{array}{cccc} 1.56 & 1.75 & 1.68 & 1.80 \\ y_1 & y_2 & y_3 & y_4 \end{array}$$

Il s'agit d'estimer les paramètres $\theta = (\Pi_F, \Pi_G, m_F, m_G, \sigma_F^2, \sigma_G^2)$. On souhaite savoir si, grâce à ces observations, on peut classer ces individus (est-ce une fille ou un garçon?). Sans ces paramètres, la tâche s'avère être impossible.

Pour estimer ces paramètres, la méthode des moments ainsi que la méthode du maximum de vraisemblance sont impossibles. Nous allons introduire de nouvelles méthodes pour estimer ces paramètres.

6.1 Méthode SEM

Si les x_1, \dots, x_n étaient observés : $\widehat{\Pi}_F(x_1, \dots, x_n, y_1, \dots, y_n) = \frac{\text{Nombre de filles}}{\text{Nombre de personne totale}}$. Il suffit juste de compter. Mathématiquement,

$$\widehat{\Pi}_F(x_1, \dots, x_n, y_1, \dots, y_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[x_i=F]}$$

De même, pour avoir la moyenne, nous faisons :

$$\widehat{m}_F(x_1, \dots, x_n, y_1, \dots, y_n) = \frac{\sum_{i=1}^n y_i \mathbb{1}_{[x_i=F]}}{\sum_{i=1}^n \mathbb{1}_{[x_i=F]}}$$

Enfin, pour la variance

$$\widehat{\sigma}_F^2(x_1, \dots, x_n, y_1, \dots, y_n) = \frac{\sum_{i=1}^n (y_i - \widehat{m}_F)^2 \mathbb{1}_{[x_i=F]}}{\sum_{i=1}^n \mathbb{1}_{[x_i=F]}}$$

Dans notre problème, nous n'avons la possibilité d'observer x_i . Nous utilisons alors la méthode SEM :

1. On dispose de $\widehat{\theta}(x, y)$.

2. On sait simuler $X = (X_1, \dots, X_n)$ selon $\mathbb{P}(X_1, \dots, X_n | Y_1, \dots, Y_n)$ c'est à dire selon la loi conditionnelle à ce qui est observé.
3. On initialise $\theta^0 = (\Pi_F^0, \Pi_G^0, m_F^0, m_G^0, (\sigma_F^0)^2, (\sigma_G^0)^2)$
4. Pour passer de θ^q à θ^{q+1} :
 - On tire $x_1^q, x_2^q, \dots, x_n^q$ selon $\mathbb{P}_\theta^q(x_1 | y_1), \dots, \mathbb{P}_\theta^q(x_n | y_n)$
 - On pose $\theta^{q+1} = \hat{\theta}(x^q, y)$

Exercice 6.1.1 Soit $X_1, \dots, X_n, Y_1, \dots, Y_n$ chaîne de MARKOV cachée, avec

- X_i sont à valeurs dans $\{F, G\}$ et Y_i à valeurs dans \mathbb{R}
 - $\mathbb{P}(X_k, X_{k+1})$ indépendante de k
 - $P(Y_k | X_k)$ gaussienne, indépendante de k
1. Donner les paramètres
 2. Montrer, en utilisant les graphes, que $\mathbb{P}(X|Y)$ est de MARKOV
 3. Peut-on appliquer le SEM ?

1. Tout d'abord, on a grâce à la première propriété $\mathbb{P}(X_k)$ indépendante de k . Ainsi, on en déduit que

$$\mathbb{P}(X_{k+1} | X_k) = \frac{\mathbb{P}(X_k, X_{k+1})}{\mathbb{P}(X_k)} = \frac{\mathbb{P}(X_1, X_2)}{\mathbb{P}(X_1)}$$

On a alors comme paramètre $\theta = (\Pi_{FF}, \Pi_{FG}, \Pi_{GF}, \Pi_{GG}, m_F, m_G, \sigma_F^2, \sigma_G^2)$

2. Si on conditionne par rapport aux y_k , on casse les segments partant des y (schéma y et x sous forme de peigne). On a alors uniquement un segment passant pour tous les x_k .
3. Oui, on peut trouver des estimateurs pour les paramètres. Par exemple, $\hat{\Pi}_{FF}(x_1, \dots, x_n, y_1, \dots, y_n) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{1}_{[x_i=F, x_{i+1}=F]}$ (remarque : il y a $n-1$ couples, d'où le $n-1$ au numérateur). Pour la deuxième propriété, il faut arriver à simuler la chaîne. Il nous faut la loi de X_1 conditionnellement aux Y_k puis les transitions conditionnellement aux Y_k . On a :

$$\mathbb{P}(X_1 | Y_1, \dots, Y_n) = \frac{\alpha_1(X_1) \beta_1(X_1)}{\sum_{x_i} \alpha_i(X_i) \beta_i(X_i)}$$

Ensuite, pour par exemple X_3 et X_4

$$\begin{aligned} \mathbb{P}(X_3, X_4, Y_1, \dots, Y_3, Y_4, \dots, Y_N) &= \mathbb{P}(X_4, Y_{4:n} | \text{Vert} X_3, Y_{1:3}) \mathbb{P}(X_3, Y_{1:3}) \\ &= \mathbb{P}(X_3, Y_{1:3}) \mathbb{P}(X_4, Y_{4:n} | X_3) \\ &= \mathbb{P}(X_3, Y_{1:3}) \mathbb{P}(X_4 | X_3) \mathbb{P}(Y_{4:n} | X_4) \end{aligned}$$