



Département GMM  
5ème année Spécialité MMS-IF

# Statistique des processus financiers

**Aldéric Joulin**

A. Joulin  
Bureau 115 - GMM  
ajoulin@insa-toulouse.fr

**Année universitaire 2017-2018**



# Table des matières

<b>1</b>	<b>Rappels sur le calcul stochastique</b>	<b>5</b>
1.1	Le mouvement brownien . . . . .	5
1.2	Martingales . . . . .	8
1.3	Intégration stochastique . . . . .	11
1.4	Formule d'Itô . . . . .	13
1.5	Théorème de Girsanov . . . . .	15
<b>2</b>	<b>Équations différentielles stochastiques</b>	<b>17</b>
2.1	Résultat d'existence et d'unicité de la solution . . . . .	17
2.2	Diffusions en finance . . . . .	18
2.2.1	Le mouvement brownien avec dérive . . . . .	19
2.2.2	Le modèle de Black-Scholes . . . . .	19
2.2.3	Le processus d'Ornstein-Uhlenbeck . . . . .	20
2.2.4	Le modèle de Vasicek . . . . .	21
2.2.5	Le modèle de Cox-Ingersoll-Ross . . . . .	22
2.3	Processus de Markov . . . . .	23
<b>3</b>	<b>Maximum de vraisemblance</b>	<b>27</b>
3.1	Rappels . . . . .	27
3.1.1	Introduction à la vraisemblance . . . . .	27
3.1.2	Exemples classiques . . . . .	29
3.1.3	Le cas d'une chaîne de Markov . . . . .	31
3.2	Le cas des diffusions . . . . .	32
3.2.1	Le cadre général . . . . .	32
3.2.2	Exemples apparaissant en finance . . . . .	35
3.2.3	Test de Neyman-Pearson . . . . .	41
3.3	Discrétisation des diffusions . . . . .	44
3.3.1	Diffusions observées à des instants discrets . . . . .	44
3.3.2	Discrétisation par schéma d'Euler . . . . .	47



# Chapitre 1

## Rappels sur le calcul stochastique

### 1.1 Le mouvement brownien

Dans tout ce cours, on va considérer des variables aléatoires (en abrégé v.a.) et processus stochastiques définis sur un espace de probabilité générique  $(\Omega, \mathcal{A}, \mathbb{P})$ . Avant de définir le mouvement brownien, commençons par introduire la notion de vecteurs gaussiens généralisant les v.a. gaussiennes unidimensionnelles. Si  $X$  est un vecteur aléatoire en dimension  $d$  (i.e. une v.a. à valeurs dans  $\mathbb{R}^d$  vue comme un vecteur colonne) tel que chacune de ses coordonnées  $X_i$ ,  $i \in \{1, \dots, d\}$ , soit de carré intégrable, alors on définit dans la suite son vecteur espérance et sa matrice de covariance par

$$\mathbb{E}[X] = m := \begin{pmatrix} \mathbb{E}[X_1] \\ \cdot \\ \cdot \\ \cdot \\ \mathbb{E}[X_d] \end{pmatrix} \quad \text{et} \quad \text{Var}(X) = \Gamma := \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,d} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,d} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{d,1} & \sigma_{d,2} & \cdots & \sigma_{d,d} \end{pmatrix},$$

où  $\sigma_{i,j}$  désigne la covariance entre les variables  $X_i$  et  $X_j$  :

$$\sigma_{i,j} = \text{Cov}(X_i, X_j) := \mathbb{E}[(X_i - m_i)(X_j - m_j)] = \mathbb{E}[X_i X_j] - m_i m_j,$$

où  $m_i := \mathbb{E}[X_i]$  pour tout  $i \in \{1, \dots, d\}$ . La matrice  $\Gamma$  est symétrique et semi-définie positive au sens où pour tout  $x \in \mathbb{R}^d$ , on a  $x^T \Gamma x \geq 0$ , le symbole  $^T$  désignant la transposition.

**Définition 1.1.1.** Soit  $X$  un vecteur aléatoire en dimension  $d$  et de matrice de covariance  $\Gamma$ , supposée inversible. Il est dit gaussien si la densité jointe est donnée par

$$f_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} \exp\left(-\frac{1}{2}(x - m)^T \Gamma^{-1}(x - m)\right), \quad x \in \mathbb{R}^d.$$

On note alors  $X \sim \mathcal{N}_d(m, \Gamma)$  (et  $\mathcal{N}(m, \Gamma)$  si  $d = 1$ ).

Ainsi, comme dans le cas unidimensionnel, la donnée du vecteur espérance et de la matrice de covariance caractérise la loi d'un vecteur gaussien. Rappelons quelques propriétés importantes des vecteurs gaussiens :

1 - La fonction caractéristique étant vue comme la transformée de Fourier de la densité, on a la caractérisation suivante de la loi d'un vecteur gaussien (on notera indifféremment  $\langle x, y \rangle$  ou  $x^T y$  le produit scalaire entre deux éléments de  $\mathbb{R}^d$ ) : un vecteur aléatoire  $X$  est gaussien d'espérance  $m$  et de matrice de covariance  $\Gamma$  si et seulement si sa fonction caractéristique est donnée par

$$\phi_X(\theta) := \mathbb{E}[e^{i\langle \theta, X \rangle}] = \exp\left(i\theta^T m - \frac{1}{2}\theta^T \Gamma \theta\right), \quad \theta \in \mathbb{R}^d.$$

2 - Un vecteur aléatoire  $X$  est gaussien si et seulement si  $\theta^T X$  est une v.a. gaussienne unidimensionnelle pour tout  $\theta \in \mathbb{R}^d$  différent du vecteur nul, i.e. toute combinaison linéaire non nulle de ses coordonnées est gaussienne (donc chacune des coordonnées d'un vecteur gaussien suit une loi gaussienne ; en revanche, la réciproque est fautive).

3 - Miracle gaussien : les coordonnées d'un vecteur gaussien  $X$  sont indépendantes si et seulement si sa matrice de covariance est diagonale.

4 - Transformation linéaire : soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . Si  $A$  est une matrice inversible  $d \times d$  et  $b$  un vecteur dans  $\mathbb{R}^d$ , alors le vecteur  $AX + b$  est gaussien d'espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ .

5 -  $X \sim \mathcal{N}_d(m, \Gamma)$  si et seulement si  $X = AU + m$ , où  $U \sim \mathcal{N}_d(0, I_d)$  et  $A$  est une matrice carrée  $d \times d$  inversible et vérifiant  $AA^T = \Gamma$ .

6 - Théorème Central Limite (TCL) multidimensionnel : soit  $(X_n)_{n \geq 1}$  une suite de vecteurs aléatoires i.i.d. à valeurs dans  $\mathbb{R}^d$ , et dont les coordonnées sont de carré intégrable. Notons  $m := \mathbb{E}[X_1]$  et  $\Gamma := \text{Var}(X_1)$ . Alors on a la convergence en loi suivante

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}_d(0, \Gamma).$$

À présent, introduisons la notion de processus gaussien, qui est une généralisation non-dénombrable des vecteurs gaussiens.

**Définition 1.1.2.** *Un processus (stochastique) est une famille de variables aléatoires réelles  $X_t$  indexées par le temps  $t \geq 0$  et dont les trajectoires  $t \rightarrow X_t$  sont continues. Il est dit gaussien si chacun des vecteurs extraits est un vecteur gaussien, i.e. pour tout  $d \in \mathbb{N}_*$  et tout  $d$ -uplet  $(t_1, \dots, t_d)$ , le vecteur  $d$ -dimensionnel  $(X_{t_1}, \dots, X_{t_d})$  est gaussien.*

Un processus gaussien  $X = (X_t)_{t \geq 0}$  est donc caractérisé en loi par sa "gaussianité" ainsi que par ses fonctions espérance et covariance :

$$t \longrightarrow \mathbb{E}[X_t] \quad \text{et} \quad (s, t) \longrightarrow \text{Cov}(X_s, X_t), \quad s, t \geq 0.$$

Nous sommes maintenant en mesure d'introduire le mouvement brownien, qui peut être construit comme objet limite de marches aléatoires renormalisées (théorème de Donsker) ou encore par le développement en série à base d'ondelettes de Haar (théorème de Lévy).

**Définition 1.1.3.** Soit  $B = (B_t)_{t \geq 0}$  un processus à valeurs dans  $\mathbb{R}$ . Il est appelé mouvement brownien si c'est un processus gaussien centré et de fonction de covariance donnée par

$$K(s, t) = \text{Cov}(B_s, B_t) = \min\{s, t\}, \quad s, t \geq 0.$$

Cette dénomination est essentiellement due à un botaniste anglais, Robert Brown, qui le décrit pour la première fois en 1827 en observant des mouvements de particules à l'intérieur de grains de pollen. Un siècle plus tard, en 1923, l'américain Norbert Wiener le construit rigoureusement, et c'est pour cela que l'on parle aussi de processus de Wiener.

**Proposition 1.1.4.** Soit  $B$  un mouvement brownien. Alors il vérifie les assertions suivantes :

- (i)  $B_0 = 0$  p.s.
- (ii) pour tous  $0 \leq s \leq t$ , la v.a.  $B_t - B_s$  a même loi que  $B_{t-s}$ , qui suit la loi normale centrée  $\mathcal{N}(0, t-s)$  : on dit que le mouvement brownien est à accroissements stationnaires.
- (iii) pour tous  $0 = t_0 < t_1 < \dots < t_d$ , les v.a.  $B_{t_i} - B_{t_{i-1}}$ ,  $i \in \{1, \dots, d\}$ , sont indépendantes : on dit que le mouvement brownien est à accroissements indépendants.

En tant que processus, le mouvement brownien peut être considéré comme une v.a. à valeurs dans  $\mathcal{C}$ , l'espace vectoriel des fonctions continues de  $[0, +\infty[$  dans  $\mathbb{R}$  muni de la tribu borélienne associée (la topologie sous-jacente est celle de la convergence uniforme sur tout compact). Notons  $P$  la loi de  $B$  sur  $\mathcal{C}$ , c'est-à-dire que l'on a pour tout ensemble borélien  $A$  de  $\mathcal{C}$ ,

$$P(A) := \mathbb{P}(B \in A).$$

Cette loi sur l'espace  $\mathcal{C}$ , appelée mesure de Wiener, est déterminée par les lois fini-dimensionnelles du mouvement brownien, c'est-à-dire par celles des vecteurs du type  $(B_{t_1}, \dots, B_{t_d})$  où  $d \in \mathbb{N}_*$  et  $t_1 < t_2 < \dots < t_d$ . Il résulte de la proposition précédente que si l'on se donne  $0 = t_0 < t_1 < t_2 < \dots < t_d$ , alors la densité jointe du vecteur  $(B_{t_1}, \dots, B_{t_d})$  est donnée par

$$\frac{1}{(2\pi)^{d/2} \sqrt{t_1(t_2 - t_1) \cdots (t_d - t_{d-1})}} \exp\left(-\sum_{k=1}^d \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}\right), \quad x \in \mathbb{R}^d,$$

où par convention  $x_0 = 0$ . Ainsi, pour montrer qu'un processus  $X$  est un mouvement brownien, il suffit de montrer que ses lois fini-dimensionnelles coïncident avec celles du mouvement brownien.

Bien qu'à trajectoires continues, le mouvement brownien est un objet très irrégulier, ce fait étant illustré par sa représentation graphique tout à fait singulière. En effet, non seulement les trajectoires du mouvement brownien ne sont pas dérivables, mais de surcroît

elles ne sont pas à variation bornée. On rappelle qu'une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  est dite à variation bornée sur l'intervalle  $[a, b]$  si

$$\sup \sum_i |f(t_{i+1}) - f(t_i)| < +\infty,$$

où le supremum est pris sur l'ensemble des subdivisions  $(t_i)$  de  $[a, b]$ . Notons que la plupart des fonctions que l'on rencontre en pratique sont à variation bornée (les fonctions de classe  $\mathcal{C}^1$ , les fonction monotones, etc).

## 1.2 Martingales

Introduisons à présent la notion de martingale.

**Définition 1.2.1.** Une famille  $(\mathcal{F}_t)_{t \geq 0}$  de sous-tribus de  $\mathcal{A}$  est une filtration de l'espace  $(\Omega, \mathcal{A}, \mathbb{P})$  si

$$\mathcal{F}_s \subset \mathcal{F}_t, \quad 0 \leq s \leq t.$$

L'espace  $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  est alors appelé un espace de probabilité filtré.

Un processus  $X$  est adapté à une filtration  $(\mathcal{F}_t)_{t \geq 0}$  si  $X_t$  est  $\mathcal{F}_t$ -mesurable pour tout  $t \geq 0$ .

**Définition 1.2.2.** Considérons un processus  $M$  adapté à une filtration  $(\mathcal{F}_t)_{t \geq 0}$ , et dont tous les éléments sont intégrables. On dit que  $M$  est une martingale pour  $(\mathcal{F}_t)_{t \geq 0}$  si

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s, \quad 0 \leq s \leq t.$$

En particulier, une martingale est d'espérance constante. Voici quelques exemples de martingales, toutes par rapport à la filtration engendrée par le mouvement brownien  $B$ , c'est-à-dire  $\mathcal{F}_t = \sigma(B_s : s \in [0, t])$ ,  $t \geq 0$  :

- le mouvement brownien lui-même ;
- le processus  $(B_t^2 - t)_{t \geq 0}$  ;
- le processus exponentiel  $(e^{\theta B_t - \theta^2 t/2})_{t \geq 0}$ , où  $\theta \in \mathbb{R}$ .

Dans la suite de ce cours, on suppose l'espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  filtré par une filtration générique  $(\mathcal{F}_t)_{t \geq 0}$ . On appelle alors mouvement brownien par rapport à la filtration  $(\mathcal{F}_t)_{t \geq 0}$  un mouvement brownien  $B$  adapté pour cette filtration et à accroissements indépendants au sens suivant : pour tous  $0 \leq s \leq t$ , la v.a.  $B_t - B_s$  est indépendante de  $\mathcal{F}_s$  (cette définition d'indépendance des accroissements coïncide avec celle de la Proposition 1.1.4 lorsque la filtration  $(\mathcal{F}_t)_{t \geq 0}$  est celle du mouvement brownien considéré).

Pour les martingales, quitte à remplacer  $M_t$  par  $M_t - M_0$ , on suppose à présent, et sauf indication du contraire, qu'elles sont toujours issues de 0. Enfin, pour ne pas s'embêter avec des notions techniques pas très importantes, nous supposons que tous les processus considérés ont de bonnes propriétés d'intégrabilité.

Rappelons le théorème principal de convergence en temps long des martingales.



**Théorème 1.2.3** (Convergence  $L^2$ ). *Soit  $M$  une martingale bornée dans  $L^2$ , c'est-à-dire*

$$\sup_{t \geq 0} \mathbb{E}[M_t^2] < +\infty.$$

*Alors  $M_t$  converge p.s. et dans  $L^2$  lorsque  $t \rightarrow +\infty$  vers une v.a.  $M_\infty$ .*

À présent, définissons la variation quadratique d'une martingale.

**Théorème 1.2.4.** *Soit  $M$  une martingale. Alors il existe un unique processus croissant adapté issu de 0, appelé la variation quadratique de  $M$  et noté  $[M, M] = ([M, M]_t)_{t \geq 0}$ , tel que  $M^2 - [M, M]$  soit une martingale. De plus, on a la convergence en probabilité suivante :*

$$[M, M]_t = \lim_{n \rightarrow +\infty} \sum_{i=1}^n (M_{it/n} - M_{(i-1)t/n})^2, \quad t > 0.$$

Pour le mouvement brownien, on a la convergence dans  $L^2$  (donc en probabilité) suivante :

$$[B, B]_t = \lim_{n \rightarrow +\infty} \sum_{i=1}^n (B_{it/n} - B_{(i-1)t/n})^2 = t, \quad t > 0.$$

Notons aussi que  $M^2 - [M, M]$  étant une martingale, elle est d'espérance constante, d'où

$$\mathbb{E}[M_t^2] = \mathbb{E}[[M, M]_t], \quad t \geq 0.$$

En particulier  $M$  est bornée dans  $L^2$  si et seulement si  $[M, M]_\infty := \lim_{t \rightarrow +\infty} [M, M]_t$  existe comme limite p.s. et est dans  $L^1$ , auquel cas

$$\sup_{t \geq 0} \mathbb{E}[M_t^2] = \lim_{t \rightarrow +\infty} \mathbb{E}[M_t^2] = \mathbb{E}[[M, M]_\infty].$$

Ainsi, pour utiliser le théorème de convergence des martingales, on peut calculer sa variation quadratique et voir si la v.a.  $[M, M]_\infty$ , lorsqu'elle existe, est intégrable. Notons également que la variation quadratique peut être généralisée à deux martingales  $M$  et  $N$  : c'est l'unique processus  $[M, N]$  adapté issu de 0 et à variation bornée tel que  $MN - [M, N]$  soit une martingale. On a aussi la convergence en probabilité

$$[M, N]_t = \lim_{n \rightarrow +\infty} \sum_{i=1}^n (M_{it/n} - M_{(i-1)t/n}) (N_{it/n} - N_{(i-1)t/n}).$$

En particulier, l'application  $(M, N) \mapsto [M, N]$  est bilinéaire.

Dans ce qui suit, nous allons énoncer deux résultats de convergence des martingales, la Loi des Grands Nombres (LGN) et le Théorème Central Limite (TCL), dont nous aurons besoin dans la partie Statistique. Bien que ces résultats soient aussi valables en temps continu, nous ne nous focaliserons que sur le cas du temps discret, afin de faire apparaître la variation quadratique d'une martingale à temps discret, et de faire le lien

avec les théorèmes associés aux sommes de variables i.i.d. Si  $M = (M_n)_{n \in \mathbb{N}}$  désigne une martingale à temps discret pour une filtration  $(\mathcal{F}_n)_{n \in \mathbb{N}}$ , on définit sa variation quadratique discrète par

$$[M, M]_n := \sum_{k=1}^n \mathbb{E} [(M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}], \quad [M, M]_0 := 0.$$

Comme dans le cas du temps continu, c'est l'unique processus croissant adapté (même prévisible, c'est-à-dire que  $[M, M]_n$  est  $\mathcal{F}_{n-1}$ -mesurable pour tout  $n \in \mathbb{N}_*$ ) et issu de 0 tel que le processus  $M^2 - [M, M]$  soit une martingale.

**Théorème 1.2.5** (LGN). *Soit  $M$  une martingale telle que p.s.  $[M, M]_\infty = +\infty$ . Alors on a la convergence p.s. suivante :*

$$\lim_{n \rightarrow +\infty} \frac{M_n}{[M, M]_n} = 0.$$

Rappelons que si  $X = (X_n)_{n \in \mathbb{N}}$  est une suite de variables i.i.d. centrées et de carré intégrable, alors la suite  $M$  donnée par

$$M_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N}_*,$$

est une martingale pour la filtration  $\mathcal{F}_n = \sigma(X_i : 1 \leq i \leq n)$ ,  $n \in \mathbb{N}_*$ . En appliquant le théorème précédent, on retrouve la LGN classique. Lorsque l'on remplace le temps discret par le temps continu, on obtient le même résultat et appliqué au mouvement brownien, il vient

$$\lim_{t \rightarrow +\infty} \frac{B_t}{t} = 0.$$

Pour le TCL, on a plusieurs résultats à notre disposition. Cependant, nous n'allons énoncer que celui dont on se sert le plus. Bien évidemment, on retrouve le TCL pour les variables i.i.d. centrées avec le choix de la suite  $a_n = n$ .

**Théorème 1.2.6** (TCL). *Soit  $M$  une martingale et soit  $(a_n)_{n \in \mathbb{N}}$  une suite positive tendant vers l'infini lorsque  $n$  tend vers l'infini. On suppose la convergence en probabilité suivante*

$$\lim_{n \rightarrow +\infty} \frac{[M, M]_n}{a_n} = \sigma^2 > 0.$$

Alors on a les convergences en loi

$$\frac{M_n}{\sqrt{a_n}} \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, \sigma^2) \quad \text{et} \quad \sqrt{a_n} \frac{M_n}{[M, M]_n} \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, \sigma^{-2}).$$

La première convergence en loi est le TCL proprement dit tandis que la seconde en est une conséquence grâce au point (iii) du lemme de Slutsky.

**Lemme 1.2.7** (Lemme de Slutsky). *Soient  $(X_n)_{n \in \mathbb{N}_*}$  et  $(Y_n)_{n \in \mathbb{N}_*}$  deux suites de v.a. convergeant en loi respectivement vers un nombre  $c \in \mathbb{R}$  et une v.a.  $Y$ . Alors*

- (i) *la somme  $X_n + Y_n$  converge en loi vers  $c + Y$ .*
- (ii) *le produit  $X_n Y_n$  converge en loi vers  $cY$ .*
- (iii) *le ratio  $Y_n/X_n$  converge en loi vers  $Y/c$  dès que  $c \neq 0$ .*

Dans l'énoncé, l'hypothèse selon laquelle  $X_n$  converge vers une constante est cruciale. En effet, si la limite était une v.a., le résultat ne serait plus valide et il faudrait une hypothèse plus forte comme la convergence en loi du couple  $(X_n, Y_n)$  pour que le résultat reste vrai. Par ailleurs, le lemme reste valide lorsque l'on remplace toutes les convergences en loi par des convergences en probabilité.

### 1.3 Intégration stochastique

Ce paragraphe constitue l'objet central d'un cours de calcul stochastique, le calcul d'Itô étant considéré comme l'un des thèmes les plus importants de la théorie des probabilités. Il s'agit de donner un sens à l'intégrale  $\int_0^t H_s dX_s$  en tant que processus, où  $X$  est une martingale et  $H$  un processus adapté suffisamment intégrable. Si  $X$  est un processus à variation bornée, la théorie classique de l'intégration pour les fonctions à variation bornée nous permet de donner un sens à cette intégrale, c'est-à-dire qu'elle peut être construite comme une limite p.s. Étant donné que la seule martingale à variation bornée est la martingale nulle, cette intégrale ne peut être définie comme ceci. Pour contourner cette difficulté, l'idée est de la définir sur une classe de processus simples et de l'étendre à une classe plus générale par un argument de densité-continuité dans  $L^2$ .

**Définition 1.3.1.** *Soit  $M$  une martingale et soit  $\mathcal{H}_0^2(M)$  l'espace formé des processus simples de la forme*

$$H_t = \sum_{k \geq 0} a_k 1_{(t_k, t_{k+1}]}(t), \quad t \geq 0,$$

*où la v.a.  $a_k$  est bornée et  $\mathcal{F}_{t_k}$ -mesurable, et  $0 = t_0 < t_1 < \dots < t_n < \dots$  est une suite croissant vers l'infini. On définit l'intégrale stochastique de  $H$  par rapport à  $M$  de la manière suivante :*

$$\int_0^{+\infty} H_t dM_t := \sum_{k \geq 0} a_k (M_{t_{k+1}} - M_{t_k}).$$

On peut ensuite démontrer que  $\mathcal{H}_0^2(M)$  est dense dans l'espace  $\mathcal{H}^2(M)$  constitué des processus adaptés  $H$  tels que

$$\|H\|_{\mathcal{H}^2(M)}^2 := \mathbb{E} \left[ \int_0^{+\infty} H_t^2 d[M, M]_t \right] < +\infty.$$

On rappelle que  $[M, M]$  étant à variation bornée (car croissante), cette intégrale est construite au sens classique. Ainsi, en définissant l'intégrale stochastique sur l'espace des processus simples  $\mathcal{H}_0^2(M)$ , on peut espérer l'étendre par densité aux processus dans  $\mathcal{H}^2(M)$ .

C'est l'objet du résultat suivant, qui justifie le mot "stochastique" dans l'expression "intégrale stochastique" : la limite est construite dans l'espace  $L^2$  et non au sens de la convergence p.s.

**Théorème 1.3.2.** *Soit  $H \in \mathcal{H}^2(M)$ . Alors l'intégrale stochastique  $\int_0^{+\infty} H_t dM_t$  déterminée par la limite dans  $L^2$  d'intégrales du type  $\int_0^{+\infty} H_t^n dM_t$ , où  $(H^n)_{n \in \mathbb{N}} \subset \mathcal{H}_0^2(M)$  est une suite de processus simples convergeant vers  $H$  pour la norme de  $\mathcal{H}^2(M)$ , est bien définie. De plus, on a l'isométrie dite d'Itô :*

$$\mathbb{E} \left[ \left( \int_0^{+\infty} H_t dM_t \right)^2 \right] = \mathbb{E} \left[ \int_0^{+\infty} H_t^2 d[M, M]_t \right].$$

Enfin, si  $X_t$  désigne la v.a. définie par  $X_t = \int_0^{+\infty} H_s 1_{[0,t]}(s) dM_s$ , où le processus  $H \in \mathcal{H}^2(M)$ , alors la famille  $(X_t)_{t \geq 0}$  est une martingale de carré intégrable. On note alors  $\int_0^t H_s dM_s$  la v.a.  $X_t$ .

Dans la suite, on ne précisera plus l'intégrabilité de  $H$  afin de simplifier les énoncés. Une martingale étant d'espérance constante et la valeur de  $\int_0^t H_s dM_s$  en 0 étant nulle par construction, on a

$$\mathbb{E} \left[ \int_0^t H_s dM_s \right] = 0, \quad t \geq 0.$$

Dans le cas brownien, l'isométrie d'Itô nous donne

$$\mathbb{E} \left[ \left( \int_0^t H_s dB_s \right)^2 \right] = \mathbb{E} \left[ \int_0^t H_s^2 ds \right], \quad t \geq 0.$$

Regardons à présent comment se comporte l'intégrale stochastique par rapport à la variation quadratique.

**Proposition 1.3.3.** *Soient  $M, \tilde{M}$  deux martingales. Alors on a l'identité suivante :*

$$\left[ \int_0^\cdot H_s dM_s, \int_0^\cdot K_s d\tilde{M}_s \right]_t = \int_0^t H_s K_s d[M, \tilde{M}]_s, \quad t \geq 0.$$

À présent, introduisons les semimartingales, qui généralisent la notion de martingale vue précédemment. Il s'agit de la classe de processus la plus générale pour laquelle nous allons donner un sens à l'intégrale stochastique.

**Définition 1.3.4.** *Un processus  $X$  est une semimartingale s'il s'écrit sous la forme*

$$X_t = M_t + A_t, \quad t \geq 0,$$

où  $M$  est une martingale et  $A$  un processus adapté et à variation bornée.

On peut montrer que cette décomposition est unique. Si  $Y$  est une autre semimartingale de décomposition  $Y_t = \tilde{M}_t + \tilde{A}_t$ , on définit la variation quadratique des semimartingales  $X$  et  $Y$  par celle de leur partie martingale,

$$[X, Y] := [M, \tilde{M}].$$

En particulier, si  $A$  et  $\tilde{A}$  sont deux processus à variation bornée et si  $M$  est une martingale, alors on a

$$[M, \tilde{A}] = [A, \tilde{A}] = 0.$$

Enfin, l'intégrale stochastique par rapport à une semimartingale est définie de la manière suivante.

**Définition 1.3.5.** Soit  $X = M + A$  une semimartingale et  $H$  un processus adapté. L'intégrale stochastique  $\int_0^t H_s dX_s$  est alors définie comme la semimartingale

$$\int_0^t H_s dX_s := \int_0^t H_s dM_s + \int_0^t H_s dA_s, \quad t \geq 0.$$

Pour conclure que  $X$  est une semimartingale, on utilise implicitement le fait que l'intégrale par rapport à un processus à variation bornée est elle-même un processus à variation bornée.

## 1.4 Formule d'Itô

À présent, nous allons énoncer l'un des résultats les plus importants de la théorie du calcul stochastique, la formule d'Itô. Ces travaux ont été publiés entre 1942 et 1950 par le mathématicien japonais Kiyoshi Itô. Elle montre qu'une fonction de classe  $\mathcal{C}^2$  de  $d$  semimartingales est encore une semimartingale, et exprime explicitement sa décomposition. Rappelons que dans le cadre classique de l'intégration (appliqué aux processus), un des théorèmes fondamentaux de cette théorie est le suivant : étant donné  $A$  un processus à variation bornée et une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  de classe  $\mathcal{C}^1$ , on a

$$f(A_t) - f(A_0) = \int_0^t f'(A_s) dA_s, \quad t \geq 0.$$

De même, si  $\tilde{A}$  est un autre processus à variation bornée, la formule d'intégration par parties est vérifiée :

$$A_t \tilde{A}_t - A_0 \tilde{A}_0 = \int_0^t A_s d\tilde{A}_s + \int_0^t \tilde{A}_s dA_s, \quad t \geq 0.$$

On va voir que ces formules ne sont plus valables dès que l'on sort du cadre des processus à variation bornée. Cependant, en reprenant le même type de démonstration via la formule de Taylor et en contrôlant de manière adéquate le reste quadratique (qui est négligeable dans le cas précédent), on est en mesure d'obtenir la fameuse formule d'Itô, faisant donc apparaître un terme supplémentaire : la variation quadratique.

**Théorème 1.4.1** (Formule d'Itô). *Soit  $X^1, \dots, X^d$  des semimartingales et soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction de classe  $\mathcal{C}^2$ . Alors pour tout  $t \geq 0$ ,*

$$\begin{aligned} f(X_t^1, \dots, X_t^d) &= f(X_0^1, \dots, X_0^d) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s^1, \dots, X_s^d) dX_s^i \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s^1, \dots, X_s^d) d[X^i, X^j]_s. \end{aligned} \quad (1.4.1)$$

Dans le cas unidimensionnel, si  $M$  est une martingale, alors pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  de classe  $\mathcal{C}^2$ ,

$$f(M_t) = f(M_0) + \int_0^t f'(M_s) dM_s + \frac{1}{2} \int_0^t f''(M_s) d[M, M]_s, \quad t \geq 0.$$

Par exemple, la formule d'Itô appliquée à la fonction  $f(x) = x^2$  entraîne que

$$M_t^2 - [M, M]_t = M_0^2 + 2 \int_0^t M_s dM_s, \quad t \geq 0.$$

Ainsi, non seulement on retrouve le fait que le processus  $M^2 - [M, M]$  est une martingale, mais de plus on donne sa valeur sous forme d'intégrale stochastique. Une autre application intéressante de la formule d'Itô est la formule d'intégration par parties, généralisant celle vue ci-dessus dans le cadre des processus à variation bornée.

**Corollaire 1.4.2.** *Si  $X$  et  $Y$  sont deux semimartingales, alors*

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_s dX_s + \int_0^t X_s dY_s + [X, Y]_t, \quad t \geq 0.$$

En appliquant la formule d'Itô bidimensionnelle au mouvement brownien  $B$  ainsi qu'à la semimartingale déterministe  $X_t := t$ , on obtient pour toute fonction  $f : (x, t) \in \mathbb{R} \times \mathbb{R}_+ \mapsto f(x, t) \in \mathbb{R}$  de classe  $\mathcal{C}^2$  en  $x$  et  $\mathcal{C}^1$  en  $t$ ,

$$f(B_t, t) = f(0, 0) + \int_0^t \frac{\partial f}{\partial x}(B_s, s) dB_s + \int_0^t \left( \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \right)(B_s, s) ds, \quad t \geq 0.$$

En effet, vu que  $X$  est à variation bornée, on a  $[B, X] = 0$ .

À présent, énonçons les versions des LGN et TCL pour les intégrales browniennes.

**Corollaire 1.4.3** (LGN pour les intégrales browniennes). *Soit  $H$  un processus adapté tel que p.s.  $\int_0^{+\infty} H_t^2 dt = +\infty$ . Alors on a le résultat de convergence p.s. suivant :*

$$\lim_{t \rightarrow +\infty} \frac{\int_0^t H_s dB_s}{\int_0^t H_s^2 ds} = 0.$$

Ci-dessous, le TCL pour les intégrales stochastiques browniennes est donné en toute généralité (i.e avec une fonction  $f$  générale), bien qu'en pratique la fonction identité convienne la plupart du temps.

**Corollaire 1.4.4** (TCL pour les intégrales browniennes). *Soit  $H$  un processus adapté et  $f : ]0, +\infty[ \rightarrow ]0, +\infty[$  une fonction tendant vers l'infini à l'infini. On suppose la convergence en probabilité suivante*

$$\lim_{t \rightarrow +\infty} \frac{1}{f(t)} \int_0^t H_s^2 ds = \sigma^2 > 0.$$

Alors on a les convergences en loi

$$\frac{1}{\sqrt{f(t)}} \int_0^t H_s dB_s \xrightarrow[t \rightarrow +\infty]{} \mathcal{N}(0, \sigma^2) \quad \text{et} \quad \sqrt{f(t)} \frac{\int_0^t H_s dB_s}{\int_0^t H_s^2 ds} \xrightarrow[t \rightarrow +\infty]{} \mathcal{N}(0, \sigma^{-2}).$$

## 1.5 Théorème de Girsanov

Terminons ce chapitre par le théorème de Girsanov, outil très utilisé en finance lorsqu'il s'agit de changer de probabilité pour faire apparaître de nouvelles martingales.

On rappelle qu'une probabilité  $\mathbb{Q}$  est dite absolument continue sur l'espace  $(\Omega, \mathcal{A})$  (ou sur  $\mathcal{A}$  par abus de langage) par rapport à la probabilité  $\mathbb{P}$  si pour tout  $A \in \mathcal{A}$ ,

$$\mathbb{P}(A) = 0 \implies \mathbb{Q}(A) = 0.$$

Ceci équivaut à la propriété suivante : il existe une unique v.a. (à égalité  $\mathbb{P}$ -p.s.)  $\mathbb{P}$ -intégrable et positive ou nulle, dite dérivée de Radon-Nykodym de  $\mathbb{Q}$  par rapport à  $\mathbb{P}$  et notée  $d\mathbb{Q}/d\mathbb{P}$ , telle que pour tout  $A \in \mathcal{A}$ ,

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}} \left[ \frac{d\mathbb{Q}}{d\mathbb{P}} 1_A \right].$$

On note dans la suite  $\mathcal{F}_{\infty} := \sigma(\mathcal{F}_t : t \geq 0)$ .

**Théorème 1.5.1** (Girsanov). *Soit  $H$  un processus adapté. Alors le processus donné par*

$$L_t := \exp \left( \int_0^t H_s dB_s - \frac{1}{2} \int_0^t H_s^2 ds \right), \quad t \geq 0,$$

*est une  $\mathbb{P}$ -martingale qui converge p.s. lorsque  $t \rightarrow +\infty$ . Soit  $\mathbb{Q}$  la probabilité équivalente à  $\mathbb{P}$  sur la tribu  $\mathcal{F}_{\infty}$ , dont la densité est donnée par la v.a. limite  $L_{\infty}$ . Alors le processus*

$$\widetilde{B}_t := B_t - \int_0^t H_s ds, \quad t \geq 0,$$

*est une martingale par rapport à  $\mathbb{Q}$ , et même un  $\mathbb{Q}$ -mouvement brownien.*





# Chapitre 2

## Équations différentielles stochastiques en finance

### 2.1 Résultat d'existence et d'unicité de la solution

La plupart des processus intervenant en finance satisfont une équation de la forme

$$X_t = x_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s,$$

ou sous une forme différentielle,

$$\begin{cases} dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t; \\ X_0 = x_0. \end{cases}$$

Ces équations, appelées Équations Différentielles Stochastiques (EDS), sont des équations différentielles perturbées par un bruit aléatoire, lequel est représenté par une partie brownienne. Remarquons que le sens donné à cette équation dépend de la théorie de l'intégrale stochastique introduite dans le chapitre précédent. La solution  $X$  de cette EDS est appelée processus de diffusion (ou seulement une diffusion), terme rappelant le lien étroit entre le mouvement brownien et l'EDP de la chaleur. La fonction  $b$  s'appelle la dérive (ou drift en anglais) car elle indique la tendance de la diffusion (lorsqu'on prend l'espérance, l'intégrale stochastique disparaît). *A contrario*, la fonction  $\sigma$  devant le mouvement brownien reflète l'intensité ou variabilité du bruit : on parle de volatilité (stochastique) en finance.

Depuis quelques années, on a incorporé à la modélisation des marchés financiers des processus qui ne sont pas des diffusions comme par exemple les solutions d'EDS dirigées par un processus de Poisson ou plus généralement par un processus de Lévy (à la place du mouvement brownien). Néanmoins, l'étude statistique de ce type de modèle est bien plus difficile, la différence principale résidant dans la présence des sauts : ces processus ne sont pas à trajectoires continues. En particulier, la théorie du calcul stochastique sur laquelle repose l'estimation statistique est différente pour ces processus. C'est pourquoi nous n'allons considérer dans la suite de ce cours que des processus de diffusion. Par

ailleurs, nous n'étudierons essentiellement que des modèles dits paramétriques, c'est-à-dire que l'on estimera statistiquement un (ou plusieurs) paramètre(s) inconnu(s) apparaissant dans ces EDS à travers les fonctions  $b$  et  $\sigma$ , plutôt qu'estimer les fonctions elles-mêmes si elles devaient nous être inconnues. Ce dernier cadre est celui de l'estimation dite non-paramétrique, théorie en progrès mais moins développée à ce jour que l'estimation paramétrique.

Tout d'abord, établissons un théorème général à propos de l'existence et de l'unicité de la solution d'une EDS. Il s'avère qu'il existe plusieurs notions d'existence et d'unicité. Cependant, nous avons pris le parti de passer sous silence ces différentes notions pour n'en retenir qu'une seule.

**Théorème 2.1.1.** *Étant donné un horizon fini fixé  $T > 0$ , considérons l'EDS suivante sur  $[0, T]$ :*

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

où  $X_0$  est une v.a.  $\mathcal{F}_0$ -mesurable, de carré intégrable et indépendante du mouvement brownien  $B$ . Supposons les coefficients  $b$  et  $\sigma$  continus en temps et lipschitziens en espace, i.e. pour tout  $t \in [0, T]$ ,

$$|b(t, x) - b(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \leq K_T |x - y|^2, \quad x, y \in \mathbb{R}. \quad (2.1.1)$$

Alors:

(i) *existence* : il existe une solution  $X$  sur  $[0, T]$  continue et adaptée, qui de plus est de carré intégrable.

(ii) *unicité* : si  $X$  et  $Y$  sont deux telles solutions de cette EDS (avec le même mouvement brownien et la même valeur initiale), alors elles sont égales p.s., i.e.

$$\mathbb{P}(X_t = Y_t \quad \forall t \in [0, T]) = 1.$$

La généralisation de ce résultat à la dimension supérieure est immédiate (le mouvement brownien multidimensionnel étant construit comme un vecteur de mouvements browniens indépendants). Par ailleurs, la conclusion de ce théorème reste valable dans le cadre d'hypothèses affaiblies que l'on rencontre souvent en pratique. Par exemple  $b$  et  $\sigma$  peuvent être supposées localement lipschitziennes en espace et satisfaisant une condition de croissance linéaire convenable pour éviter une explosion en temps fini.

## 2.2 Exemples classiques de diffusions en finance

À présent, introduisons quelques exemples de diffusions apparaissant très fréquemment en finance. Pour certaines, il existe une représentation explicite de la solution sous forme d'une fonction du mouvement brownien  $B$  (ou de sa trajectoire), ce qui va nous permettre de faire une étude statistique approfondie des paramètres associés aux coefficients  $b$  et  $\sigma$ .

### 2.2.1 Le mouvement brownien avec dérive

Il s'agit bien évidemment de la diffusion la plus simple que nous allons considérer. Les fonctions  $b$  et  $\sigma$  sont supposées constantes, à savoir  $b(x) = \mu \in \mathbb{R}$  et  $\sigma(x) = \sigma > 0$ . Le processus est donc le suivant :

$$X_t = X_0 + \mu t + \sigma B_t, \quad t \geq 0.$$

Cette diffusion a été étudiée par Bachelier dans sa thèse (soutenue en 1900), constituant ainsi les fondations des mathématiques financières modernes. Le processus  $X$  est évidemment gaussien, d'espérance  $\mathbb{E}[X_t] = \mathbb{E}[X_0] + \mu t$  et de covariance

$$\text{Cov}(X_s, X_t) = \text{Var}(X_0) + \sigma^2 \min\{s, t\}.$$

### 2.2.2 Le modèle de Black-Scholes

Considérons l'EDS linéaire très simple :

$$dX_t = \mu X_t dt + \sigma X_t dB_t,$$

où les constantes  $\mu$  et  $\sigma$  sont dans  $\mathbb{R}$  et  $]0, +\infty[$ , respectivement. On suppose aussi que p.s.  $X_0 > 0$ . On remarque que si  $X_t$  est différent de 0, l'EDS se réécrit

$$\frac{dX_t}{X_t} = \mu dt + \sigma dB_t,$$

c'est-à-dire que le terme de gauche, qui est l'intégrale stochastique par rapport à  $X$  de la dérivée du logarithme népérien, est simplement un mouvement brownien avec dérive. La formule d'Itô appliquée au logarithme népérien nous donne

$$d \log(X_t) = \left( \mu - \frac{\sigma^2}{2} \right) dt + \sigma dB_t,$$

et il en résulte alors que l'unique solution de l'EDS est

$$X_t = X_0 \exp \left( \sigma B_t + \left( \mu - \frac{\sigma^2}{2} \right) t \right).$$

Ce processus, appelé mouvement brownien géométrique par les probabilistes, est le fameux modèle de Black-Scholes (1973), dans lequel l'évolution du prix d'une action donnée est régie par un processus stochastique. Plus précisément, la célèbre formule de Black-Scholes permet de calculer la valeur théorique d'une option européenne à partir des données suivantes :

- $X_t$  (resp.  $x_0$ ) est la valeur au temps  $t$  (resp. valeur initiale, supposée déterministe) de l'action sous-jacente ;
- $T$  est l'échéance, ou maturité de l'option ;
- $K$  est le prix d'exercice fixé par l'option (strike) ;

- $\mu = r$  est le taux d'intérêt sans risque (sous la probabilité risque-neutre) ;
- $\sigma$  est la volatilité du prix de l'action.

Le prix théorique d'une option d'achat européenne (on parle alors de *call*), qui donne le droit mais pas l'obligation d'acheter l'actif  $X$  à la valeur  $K$  à la date  $T$ , est l'espérance du *payoff*  $(X_T - K)^+ = \max\{X_T - K, 0\}$  actualisé. La loi de l'actif sous-jacent étant log-normale (i.e. son logarithme suit une loi normale), le prix théorique peut être calculé explicitement :

$$C(x_0, K, r, \sigma, T) := e^{-rT} \mathbb{E} [(X_T - K)^+] = x_0 \varphi(d) - K e^{-rT} \varphi(d - \sigma\sqrt{T}),$$

où  $\varphi$  est la fonction de répartition de la loi gaussienne centrée et réduite et

$$d := \frac{\ln(x_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}.$$

De même, le prix théorique d'une option de vente européenne (on parle alors de *put*), qui donne le droit mais pas l'obligation de vendre l'actif  $X$  à la valeur  $K$  à la date  $T$ , est l'espérance du *payoff*  $(K - X_T)^+$  actualisé, et est donné par la formule

$$P(x_0, K, r, \sigma, T) := e^{-rT} \mathbb{E} [(K - X_T)^+] = -x_0 \varphi(-d) + K e^{-rT} \varphi(\sigma\sqrt{T} - d),$$

L'intérêt de modéliser les marchés financiers par la diffusion de Black et Scholes est que les calculs peuvent être faits de manière explicite, la v.a.  $X_t$  s'exprimant comme une fonction très simple de  $B_t$ . En revanche, ce modèle est limité au sens où il ne colle pas réellement à la réalité des marchés financiers. Par exemple, la formule de Black et Scholes n'est plus valable dès que le taux d'intérêt et la volatilité ne sont plus constants ou encore pour la prise en compte d'éventuels krachs boursiers (modélisés dans ce cas par des processus à sauts).

### 2.2.3 Le processus d'Ornstein-Uhlenbeck

La méthode par changement de variable que l'on vient de proposer dans le cas du modèle de Black-Scholes peut très bien se révéler inutile pour des cas très simples, comme celui que nous allons regarder maintenant. L'EDS que l'on va considérer est la suivante :

$$dX_t = -\mu X_t dt + \sigma dB_t,$$

où  $\mu \in \mathbb{R}_*$  et  $\sigma > 0$ . Ce modèle a été introduit au début des années 30 par les physiciens Ornstein et Uhlenbeck lorsqu'ils ont étudié la théorie cinétique des gaz, et plus précisément le comportement en vitesse de ces molécules. Notons tout de même que cette équation était écrite légèrement différemment, "à la physicienne", car le calcul stochastique introduit par Itô n'est né que 10 ans après. Concernant la modélisation financière, il s'agit d'un modèle décrivant l'évolution de taux d'intérêt.

Cette EDS peut être interprétée comme une perturbation aléatoire brownienne de l'équation différentielle ordinaire  $dx_t = -\mu x_t dt$  dont la solution est  $x_t = x_0 e^{-\mu t}$ . On s'attend donc à ce que ce terme exponentiel joue un rôle dans la résolution de l'EDS. Appliquons la formule d'Itô au processus  $(e^{\mu t} X_t)_{t \geq 0}$  :

$$\begin{aligned} d(e^{\mu t} X_t) &= \mu e^{\mu t} X_t dt + e^{\mu t} dX_t \\ &= \mu e^{\mu t} X_t dt - \mu e^{\mu t} X_t dt + \sigma e^{\mu t} dB_t \\ &= \sigma e^{\mu t} dB_t, \end{aligned}$$

d'où la solution de l'EDS,

$$X_t = X_0 e^{-\mu t} + \sigma \int_0^t e^{-\mu(t-s)} dB_s, \quad t \geq 0.$$

L'espérance est  $\mathbb{E}[X_t] = \mathbb{E}[X_0] e^{-\mu t}$  et en utilisant l'indépendance de  $X_0$  et du mouvement brownien  $B$ , ainsi que l'isométrie d'Itô, on obtient la covariance suivante : si  $0 \leq s \leq t$ ,

$$\text{Cov}(X_s, X_t) = e^{-\mu(s+t)} \left( \text{Var}(X_0) + \sigma^2 \frac{e^{2\mu s} - 1}{2\mu} \right).$$

Par exemple dans le cas où  $X_0 = 0$  p.s., le processus est centré et a pour variance  $\text{Var}(X_t) = \sigma^2(1 - e^{-2\mu t})/2\mu$ . Enfin, si l'on suppose  $X_0$  gaussienne, la présence d'une intégrale stochastique d'une fonction déterministe nous assure du caractère gaussien de ce processus. En effet, on peut démontrer que si  $f$  est une fonction continue alors la martingale de carré intégrable  $\left( \int_0^t f(s) dB_s \right)_{t \geq 0}$  est un processus gaussien. En particulier on a pour tout  $t > 0$  :

$$M_t \sim \mathcal{N} \left( 0, \int_0^t f(s)^2 ds \right).$$

### 2.2.4 Le modèle de Vasicek

Ce modèle est une généralisation du modèle de Black-Scholes, auquel on a ajouté une dérive. Le processus  $X$  est solution de l'EDS

$$dX_t = -(\mu X_t - \nu) dt + \sigma dB_t,$$

où  $\mu, \nu \in \mathbb{R}_*$  et  $\sigma > 0$ . En notant  $\tilde{X}$  le processus  $\tilde{X}_t = X_t - \nu/\mu$  qui est un processus d'Ornstein-Uhlenbeck, on trouve facilement la représentation explicite de  $X$  :

$$X_t = X_0 e^{-\mu t} + \sigma \int_0^t e^{-\mu(t-s)} dB_s + \frac{\nu}{\mu} (1 - e^{-\mu t}).$$

Tout comme le processus d'Ornstein-Uhlenbeck, l'inconvénient principal lorsque l'on utilise ce processus dans la modélisation financière est qu'il peut prendre des valeurs négatives. C'est pourquoi le prochain exemple est plus approprié, bien que plus délicat à étudier.

### 2.2.5 Le modèle de Cox-Ingersoll-Ross

Considérons l'EDS suivante :

$$dX_t = \mu(\nu - X_t) dt + \sigma \sqrt{X_t} dB_t,$$

où p.s.  $X_0 > 0$  et  $\mu, \nu \in \mathbb{R}$  et  $\sigma > 0$ . Cette équation admet une unique solution dès lors que le processus ne touche jamais 0 : une condition nécessaire et suffisante pour cela est que

$$\mu > 0 \quad \text{et} \quad 2\mu\nu > \sigma^2,$$

ce que l'on supposera dans la suite. Au contraire des processus précédents, la représentation de  $X$  n'est pas explicite, ce qui rend son étude plus délicate. En revanche, on peut donner sa loi pour certaines valeurs des paramètres  $\mu, \nu$  et  $\sigma$ . Supposons que la quantité  $d := 4\mu\nu/\sigma^2$  soit un nombre entier strictement supérieur à 2 et considérons un processus d'Ornstein-Uhlenbeck  $d$ -dimensionnel  $\tilde{X}$  solution de l'EDS

$$d\tilde{X}_t = -\frac{\mu}{2} \tilde{X}_t dt + \frac{\sigma}{2} d\tilde{B}_t, \quad \tilde{X}_0 = (0, 0, \dots, 0, \sqrt{X_0})^T,$$

où  $\tilde{B}$  est un mouvement brownien dans  $\mathbb{R}^d$  supposé indépendant de  $X_0$ , la condition initiale du modèle CIR. En appliquant la formule d'Itô multidimensionnelle au processus  $\tilde{X}$  et à la fonction  $f$  de classe  $\mathcal{C}^2$  sur  $\mathbb{R}^d$  définie par  $f(x) = \|x\|^2$ , où  $\|\cdot\|$  désigne la norme euclidienne, on a

$$f(\tilde{X}_t) - f(\tilde{X}_0) = \int_0^t \frac{\sigma}{2} \langle \nabla f(\tilde{X}_s), d\tilde{B}_s \rangle + \int_0^t \left( \frac{\sigma^2}{8} \Delta f(\tilde{X}_s) - \frac{\mu}{2} \langle \tilde{X}_s, \nabla f(\tilde{X}_s) \rangle \right) ds,$$

c'est-à-dire

$$\|\tilde{X}_t\|^2 - \|\tilde{X}_0\|^2 = \int_0^t \sigma \langle \tilde{X}_s, d\tilde{B}_s \rangle + \int_0^t \mu \left( \nu - \|\tilde{X}_s\|^2 \right) ds.$$

Par ailleurs on peut démontrer que l'intégrale stochastique ci-dessus a même loi que l'intégrale stochastique  $\int_0^t \sigma \|\tilde{X}_s\| dB_s$  où  $B$  est un mouvement brownien unidimensionnel indépendant de  $X_0$ . Ainsi, le processus  $\|\tilde{X}\|^2$  est solution de l'EDS unidimensionnelle du modèle CIR et par unicité de la solution, les processus  $X$  et  $\|\tilde{X}\|^2$  coïncident. Comme le mouvement brownien multidimensionnel  $\tilde{B}$  a ses coordonnées indépendantes par définition, le processus  $\tilde{X}$  a lui aussi ses  $d$  coordonnées indépendantes : la v.a.  $X_t$  suit donc la loi d'une somme de carrés de gaussiennes indépendantes, toutes centrées et de même variance (sauf la dernière), c'est-à-dire une loi dite du  $\chi^2$  non centrée à  $d$  degrés de liberté. Son espérance peut alors être calculée et l'on trouve

$$\mathbb{E}[X_t] = \mathbb{E}[X_0] e^{-\mu t} + \nu (1 - e^{-\mu t}),$$

résultat auquel nous nous attendions. En effet, il suffit de prendre l'espérance directement dans l'EDS (l'intégrale brownienne disparaît).

## 2.3 Processus de Markov

Dans cette partie, nous faisons un bref résumé de la théorie des processus de Markov, dont nous aurons besoin dans la partie Statistique pour établir la normalité asymptotique de l'estimateur de maximum de vraisemblance.

**Définition 2.3.1.** *Soit  $X$  un processus adapté à une filtration  $(\mathcal{F}_t)_{t \geq 0}$ . On dit que c'est un processus de Markov s'il vérifie la propriété suivante, dite propriété de Markov : pour toute fonction borélienne positive ou bornée  $f$  et tous  $0 \leq s \leq t$ , on a l'égalité p.s.*

$$\mathbb{E}[f(X_t) | \mathcal{F}_s] = \mathbb{E}[f(X_t) | X_s].$$

Autrement dit, le futur est indépendant du passé conditionnellement au présent. Si l'espérance conditionnelle, qui est une fonction borélienne de  $X_s$ , ne dépend que de  $t - s$ , l'écart entre  $t$  et  $s$ , alors le processus de Markov est dit homogène. On supposera cette propriété d'homogénéité vérifiée dans la suite pour simplifier la présentation.

À présent, introduisons les notions d'irréductibilité, de récurrence et de récurrence positive des processus de Markov comme on le fait dans le cas discret des chaînes de Markov à espace d'état dénombrable. On considère dans la suite

$$T_A := \inf\{t \geq 0 : X_t \in A\}, \quad \text{et} \quad V_A := \int_0^{+\infty} 1_A(X_t) dt,$$

où  $A$  désigne un ensemble borélien de mesure de Lebesgue strictement positive (on parle de borélien positif). Ces quantités sont respectivement le temps d'entrée dans  $A$  et le temps passé dans  $A$  par le processus.

(i) irréductibilité : partant de n'importe quel point  $x \in \mathbb{R}$ , le processus peut atteindre en temps fini n'importe quel borélien positif  $A$ . Ceci s'écrit  $\mathbb{P}_x(T_A < +\infty) > 0$  ou encore de manière équivalente,  $\mathbb{E}_x[V_A] > 0$ .

(ii) récurrence : non seulement le processus est irréductible, mais aussi partant de n'importe quel point  $x \in \mathbb{R}$ , le processus atteint en temps fini n'importe quel borélien positif  $A$ . Ceci s'écrit  $\mathbb{P}_x(T_A < +\infty) = 1$ . On peut montrer que ceci est équivalent à la propriété apparemment plus forte : partant de n'importe quel point  $x \in \mathbb{R}$ , la durée passée dans n'importe quel borélien positif  $A$  est infinie, c'est-à-dire  $\mathbb{P}_x(V_A = +\infty) = 1$ . Enfin ceci équivaut à la propriété apparemment plus faible :  $\mathbb{E}_x[V_A] = +\infty$ .

(iii) transience : le processus est dit transitoire s'il est irréductible et non récurrent, c'est-à-dire que pour tout  $x \in \mathbb{R}$  et tout borélien positif  $A$ , on a  $\mathbb{P}_x(T_A < +\infty) < 1$  ou encore que  $\mathbb{P}_x(V_A = +\infty) = 0$  ou encore que  $\mathbb{E}_x[V_A] < +\infty$ . C'est le cas des processus qui tendent p.s. vers l'infini ou vers une constante déterministe lorsque  $t$  tend vers l'infini.

(iv) récurrence positive, ou ergodicité : le processus est récurrent et partant de n'importe quel point  $x \in \mathbb{R}$ , le temps d'entrée dans n'importe quel borélien positif  $A$  est fini en moyenne. En d'autres termes, on a  $\mathbb{E}_x[T_A] < +\infty$ .

(v) récurrence nulle : le processus est récurrent et pour tout  $x \in \mathbb{R}$  et tout borélien positif  $A$ , on a  $\mathbb{E}_x[T_A] = +\infty$ .

Dans le cas ergodique, et seulement dans ce cas, il existe une unique probabilité invariante  $\pi$  pour le processus, c'est-à-dire que si  $X_0$  suit la loi  $\pi$  alors pour tout  $t > 0$ , la variable  $X_t$  la suit aussi. Lorsque  $X_0$  suit la loi invariante, le processus  $X$  est dit stationnaire. L'ergodicité est associée à une propriété de convergence en temps long, apparaissant aussi dans le cas des chaînes de Markov, et connu sous le nom de théorème ergodique (ou loi des grands nombres markovienne). Ce résultat est à la base de ce que nous allons faire dans la partie Statistique.

**Théorème 2.3.2** (Théorème ergodique, ou LGN markovienne). *Supposons que le processus  $X$  soit ergodique. Alors pour toute fonction  $f \in L^1(\pi)$ , on a la convergence p.s. suivante :*

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t f(X_s) ds = \pi(f),$$

où  $\pi(f)$  désigne l'intégrale de  $f$  sous  $\pi$ , i.e.  $\pi(f) := \int_{\mathbb{R}} f d\pi$ .

Maintenant, posons-nous la question suivante, en lien avec les EDS : les solutions d'EDS sont-elles des processus de Markov, et si oui, existe-il des critères sur les fonctions  $b$  et  $\sigma$  assurant les propriétés ci-dessus ? Le résultat suivant répond à ces questions.

**Théorème 2.3.3.** *La solution  $X$  d'une EDS est un processus de Markov, qui est homogène si les fonctions  $b$  et  $\sigma$  ne dépendent pas du temps. Supposons de plus que  $\sigma$  ne s'annule pas sur  $\mathbb{R}$  et notons  $U$ ,  $V$  et  $Z$  les fonctions et quantité suivantes : pour tout  $x \in \mathbb{R}$ ,*

$$U(x) = 2 \int_0^x \frac{b(u)}{\sigma(u)^2} du, \quad V(x) = \int_0^x e^{-U(y)} dy \quad \text{et} \quad Z = \int_{\mathbb{R}} \frac{e^{U(y)}}{\sigma(y)^2} dy.$$

Alors le processus  $X$  est :

- récurrent si et seulement si  $\lim_{x \rightarrow \pm\infty} V(x) = \pm\infty$ .
- récurrent nul si et seulement s'il est récurrent et  $Z = +\infty$ .
- récurrent positif (ergodique) si et seulement s'il est récurrent et  $Z < +\infty$ .

Dans ce dernier cas uniquement, la probabilité invariante existe et a pour densité par rapport à la mesure de Lebesgue

$$f_{\pi}(y) = \frac{1}{Z} \frac{e^{U(y)}}{\sigma(y)^2} \quad y \in \mathbb{R}.$$

Les critères précédents sont des conditions nécessaires et suffisantes pour obtenir la propriété désirée. Cependant elles peuvent ne pas être très explicites, en particulier lorsqu'il s'agit d'évaluer des intégrales. Dans ce qui suit, nous donnons une condition suffisante très simple assurant l'ergodicité. Notons  $\mathcal{P}$  l'ensemble des fonctions régulières à croissance au plus polynomiale en l'infini.

**Proposition 2.3.4.** *Supposons que la fonction  $1/\sigma \in \mathcal{P}$  et que*

$$\lim_{|x| \rightarrow +\infty} \text{sign}(x) \frac{b(x)}{\sigma(x)^2} < 0.$$



Alors le processus est ergodique et la probabilité invariante  $\pi$  admet des moments de tout ordre, c'est-à-dire que pour tout  $p > 0$ ,

$$\int_{\mathbb{R}} |x|^p f_{\pi}(x) dx < +\infty.$$

Ce jeu d'hypothèses sera appelé dans la suite l'hypothèse (H).

Tous les processus que l'on a vus précédemment sont des processus de Markov homogènes. On peut montrer qu'ils sont tous irréductibles sur  $\mathbb{R}$ , sauf les modèles de Black-Scholes et CIR qui le sont sur  $]0, +\infty[$  (qui est l'ensemble des valeurs possibles prises par ces deux processus). Plus précisément, nous avons pour :

- le mouvement brownien avec dérive : il est transitoire dans le cas  $\mu \neq 0$  car p.s.  $\lim_{t \rightarrow +\infty} X_t = \pm\infty$  (selon le signe de  $\mu$ ) et récurrent nul sinon (il s'agit alors du mouvement brownien).

- le modèle de Black-Scholes : il est transitoire sur  $]0, +\infty[$  dans le cas  $\mu \neq \sigma^2/2$  car il tend p.s. vers 0 ou  $+\infty$  selon que  $\mu < \sigma^2/2$  ou  $\mu > \sigma^2/2$ . Si  $\mu = \sigma^2/2$  alors il est récurrent nul, comme fonction croissante du mouvement brownien.

- le processus d'Ornstein-Uhlenbeck : il est ergodique dans le cas  $\mu > 0$  car il satisfait l'hypothèse (H) et l'unique probabilité invariante est la loi normale centrée et de variance  $\sigma^2/2\mu$ . Si  $\mu < 0$  alors il est transitoire et  $X_t e^{\mu t}$  converge p.s. et dans  $L^2$  vers la v.a. gaussienne

$$X_0 + \sigma \int_0^{+\infty} e^{\mu r} dB_r.$$

- le modèle de Vasicek : on a les mêmes conclusions que pour le processus d'Ornstein-Uhlenbeck (seule l'espérance pour la loi limite est modifiée).

- le modèle CIR : il est ergodique sur  $]0, +\infty[$  car il satisfait l'hypothèse (H) sur  $]0, +\infty[$  et l'unique probabilité invariante  $\pi$  est la loi Gamma de paramètres  $\alpha = d/2$  et  $\beta = \sigma^2/2\mu$ , i.e. de densité sur  $\mathbb{R}_+$  donnée par

$$f_{\pi}(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}.$$

Pour terminer avec les processus de Markov, regardons ce que donne la combinaison du théorème ergodique (la convergence p.s. entraînant la convergence en probabilité) avec le TCL pour les intégrales browniennes, lorsque le processus  $H$  est de la forme  $f(X)$ , où  $X$  est la solution d'une EDS.

**Théorème 2.3.5** (TCL markovien). *Supposons que les fonctions  $b$  et  $\sigma$  satisfassent l'hypothèse (H). Alors pour toute fonction  $h \in \mathcal{P}$ , on a les convergences en loi*

$$\frac{1}{\sqrt{t}} \int_0^t h(X_s) dB_s \xrightarrow[t \rightarrow +\infty]{} \mathcal{N}(0, \pi(h^2)) \quad \text{et} \quad \sqrt{t} \frac{\int_0^t h(X_s) dB_s}{\int_0^t h(X_s)^2 ds} \xrightarrow[t \rightarrow +\infty]{} \mathcal{N}(0, \pi(h^2)^{-1}).$$

Ainsi, le TCL markovien pourra s'appliquer aux processus d'Ornstein-Uhlenbeck, de Vasicek et au modèle CIR, comme nous allons le voir lors de l'estimation statistique par maximum de vraisemblance.



# Chapitre 3

## Estimation par maximum de vraisemblance

### 3.1 Rappels sur l'estimation par maximum de vraisemblance

Nous allons nous intéresser à la statistique paramétrique, où la loi d'une v.a.  $X$  peut être caractérisée par un paramètre  $\theta \in \Theta$  qui est un nombre ou un vecteur inconnu que l'on souhaite estimer. L'ensemble  $\Theta$ , quant à lui, est un sous-ensemble de  $\mathbb{R}$  ou de  $\mathbb{R}^d$  supposé implicitement compact ou borné (pour assurer des résultats de régularité sur lesquels nous n'insisterons pas). La méthode que l'on va étudier repose sur la notion de vraisemblance et est très souvent utilisée en pratique (nous n'aborderons pas d'autres types de procédures d'estimation comme les méthodes des moments, bayésienne ou de la distance minimale).

#### 3.1.1 Introduction à la vraisemblance

On se donne un  $n$ -échantillon  $(X_1, \dots, X_n)$  d'observations issues d'une v.a.  $X$  dépendant d'un paramètre inconnu  $\theta \in \Theta$ , c'est-à-dire une suite de variables i.i.d. de même loi que  $X$ , notée  $P_\theta$ . Rappelons que  $\mathbb{P}$  est une probabilité sur  $(\Omega, \mathcal{A})$  tandis que  $P_\theta$  est une probabilité sur l'ensemble des valeurs prises par  $X$  muni de sa tribu naturelle (la tribu borélienne dans le cas d'un espace continu et l'ensemble des parties dans le cas discret d'un ensemble fini ou dénombrable). On suppose aussi que la famille  $(P_\theta)_{\theta \in \Theta}$  vérifie la condition d'identifiabilité, c'est-à-dire que l'application  $\theta \mapsto P_\theta$  est injective : si deux lois de probabilités  $P_{\theta_1}$  et  $P_{\theta_2}$  sont égales alors  $\theta_1 = \theta_2$ .

**Définition 3.1.1.** On appelle vraisemblance de la loi  $P_\theta$  la fonction  $L_n : (x_1, \dots, x_n, \theta) \mapsto L_n(x_1, \dots, x_n, \theta)$  définie :

- dans le cas discret par  $L_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P_\theta(\{x_i\})$ ;
- dans le cas continu par la densité de la loi jointe du  $n$ -échantillon, c'est-à-dire  $L_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i)$ , où  $f_\theta$  désigne la densité de la loi marginale  $P_\theta$  par rapport

à la mesure de Lebesgue.

La v.a. obtenue en appliquant la fonction  $(x_1, \dots, x_n) \mapsto \operatorname{argmax}_{\theta \in \Theta} L_n(x_1, \dots, x_n, \theta)$  au  $n$ -échantillon  $(X_1, \dots, X_n)$  s'appelle l'estimateur du maximum de vraisemblance (EMV) du paramètre  $\theta$ .

Par exemple dans le cas discret, on effectue un tirage de  $n$  valeurs et il faut donc trouver le paramètre qui maximise la probabilité d'avoir tiré ce tirage. Notons que l'EMV peut être unique, ne pas être unique, ou même ne pas exister. Cependant nous n'étudierons que des cas où il y a existence et unicité de l'EMV. Par ailleurs, lorsque la vraisemblance est strictement positive et grâce à la croissance stricte du logarithme népérien (noté  $\log$  dans la suite), il est équivalent (et souvent plus simple en pratique) de maximiser la log-vraisemblance, c'est-à-dire le logarithme népérien de la vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver). La recherche de l'EMV est alors un problème d'optimisation de la log-vraisemblance : si c'est une fonction assez régulière de  $\theta$ , alors il s'agit :

- de trouver un point critique, i.e. un point pour lequel le gradient de la log-vraisemblance s'annule, puis
- de vérifier que la matrice hessienne en ce point est négative (il s'agit donc d'un maximum local) puis
- de montrer que ce maximum local est en fait global. Ce dernier point est en général assuré par l'éventuelle concavité de la log-vraisemblance.

Notons  $\partial^l$  la dérivée partielle d'ordre  $l \in \mathbb{N}_*$ . Rappelons quelques propriétés satisfaites par l'EMV, que l'on note dans la suite  $\hat{\theta}_n$ .

**Proposition 3.1.2.** *Soit  $\theta \in \Theta \subset \mathbb{R}^d$  un paramètre inconnu. L'EMV est un estimateur consistant de  $\theta$ , c'est-à-dire qu'il converge en probabilité vers  $\theta$  lorsque  $n$  tend vers l'infini, (i.e. la norme euclidienne de  $\hat{\theta}_n - \theta$  tend vers 0 en probabilité). De plus, il est asymptotiquement normal au sens de la convergence en loi vers un vecteur gaussien  $d$ -dimensionnel :*

$$\sqrt{n} \left( \hat{\theta}_n - \theta \right) \xrightarrow[n \rightarrow +\infty]{} U_\theta,$$

où  $U_\theta \sim \mathcal{N}_d(0, I(\theta)^{-1})$  et  $I(\theta)$  est la matrice (ou information) de Fisher  $d \times d$  formée des éléments

$$I(\theta)_{i,j} = \operatorname{Cov} \left( \partial_{\theta_i} \log f_\theta(X_1), \partial_{\theta_j} \log f_\theta(X_1) \right) = -\mathbb{E} \left[ \partial_{\theta_i, \theta_j}^2 \log f_\theta(X_1) \right],$$

que l'on suppose bien définie et inversible. En particulier l'EMV atteint à la limite la borne de Cramer-Rao : il est asymptotiquement sans biais, i.e.  $\lim_{n \rightarrow \infty} \mathbb{E}[\hat{\theta}_n] = \theta$ , et de variance minimale : on dit qu'il est asymptotiquement efficace.

Bien que la démonstration de ce résultat soit assez technique, donnons-en les grandes étapes pour la normalité asymptotique en dimension 1 (nous ne préciserons ni les hypothèses techniques ni les modes de convergence). Tout d'abord, notons  $H_n$  la fonction

(aléatoire) de  $\theta$  définie par

$$H_n(X_1, \dots, X_n, \theta) := \frac{1}{n} \partial_\theta \log L_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{i=1}^n \partial_\theta \log f_\theta(X_i),$$

qui est donc la somme de v.a. i.i.d. centrées et de variance  $I(\theta)$ , l'information de Fisher. Par le TCL, on a la convergence

$$\sqrt{n} H_n(X_1, \dots, X_n, \theta) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, I(\theta)).$$

À présent, appliquons le théorème des accroissements finis à  $\theta \mapsto H_n(X_1, \dots, X_n, \theta)$  :

$$0 = H_n(X_1, \dots, X_n, \hat{\theta}_n) \underset{n \rightarrow +\infty}{\approx} H_n(X_1, \dots, X_n, \theta) + (\hat{\theta}_n - \theta) K_n(X_1, \dots, X_n, \theta),$$

où  $K_n$  est la fonction (aléatoire) donnée par

$$K_n(X_1, \dots, X_n, \theta) = \partial_\theta H_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{i=1}^n \partial_\theta^2 \log f_\theta(X_i).$$

Par la LGN, on obtient que  $K_n(X_1, \dots, X_n, \theta)$  converge vers l'espérance de  $\partial_\theta^2 \log f_\theta(X_1)$ , qui n'est autre que  $-I(\theta)$ . Ainsi on obtient finalement que

$$\sqrt{n} (\hat{\theta}_n - \theta) \underset{n \rightarrow +\infty}{\approx} \frac{\sqrt{n} H_n(X_1, \dots, X_n, \theta)}{-K_n(X_1, \dots, X_n, \theta)},$$

qui suit la loi  $I(\theta)^{-1} \mathcal{N}_d(0, I(\theta))$ , c'est-à-dire la loi  $\mathcal{N}(0, I(\theta)^{-1})$ .

À présent, si l'on désire estimer non pas  $\theta$  mais plutôt  $g(\theta)$ , où  $g$  est une "bonne" fonction, on a à notre disposition un résultat très utile en pratique, connu sous le nom de "delta-method".

**Théorème 3.1.3.** *Soit  $g$  une fonction définie sur  $\Theta \subset \mathbb{R}^d$  et à valeurs dans  $\mathbb{R}^k$ , de classe  $\mathcal{C}^1$  et dont la matrice jacobienne  $d \times k$  au point  $\theta$ , notée  $Jac(g)(\theta)$ , est inversible. Alors  $g(\hat{\theta}_n)$  est l'EMV de  $g(\theta)$  et de plus, on a la convergence en loi suivante :*

$$\sqrt{n} \left( g(\hat{\theta}_n) - g(\theta) \right) \xrightarrow[n \rightarrow +\infty]{} Jac(g)(\theta)^T U_\theta,$$

où  $U_\theta \sim \mathcal{N}_d(0, I(\theta)^{-1})$ . Autrement dit, la loi limite est celle d'un vecteur gaussien  $k$ -dimensionnel centré et de matrice de covariance  $Jac(g)(\theta)^T I(\theta)^{-1} Jac(g)(\theta)$ .

### 3.1.2 Exemples classiques

Voici quelques exemples classiques pour lesquels nous sommes en mesure de déterminer l'EMV. En particulier, on reconnaît quelques estimateurs bien connus.

◦ cas d'une v.a. de Bernoulli de paramètre  $\theta \in \Theta \subset ]0, 1[$ . L'EMV est donné par la moyenne empirique

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i =: \bar{X}_n,$$

c'est-à-dire l'estimateur sans biais donné par la LGN puisque  $\mathbb{E}[X_1] = \theta$ . De plus, l'information de Fisher est

$$I(\theta) = \frac{1}{\theta(1-\theta)} = \frac{1}{\text{Var}(X_1)},$$

ce qui était attendu car d'après le TCL on a la convergence en loi

$$\sqrt{n} (\bar{X}_n - \theta) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, \text{Var}(X_1)).$$

◦ cas d'une v.a. de Poisson de paramètre  $\theta \in \Theta \subset ]0, \infty[$ . L'EMV est encore la moyenne empirique, comme dans le cas Bernoulli. L'information de Fisher, quant à elle, vaut  $1/\theta$ , c'est-à-dire l'inverse de la variance d'une v.a. de Poisson de paramètre  $\theta$ .

◦ cas d'une v.a. gaussienne de moyenne  $m$  et de variance  $\sigma^2$  : pour  $\theta = (m, \sigma^2)$  et  $\Theta \subset \mathbb{R} \times ]0, \infty[$ , l'EMV vaut

$$\hat{\theta}_n = (\bar{X}_n, S_n^2),$$

où  $S_n^2$  est la variance empirique

$$S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2.$$

Par la LGN, la variance empirique converge p.s. (donc en probabilité) vers la variance  $\sigma^2$ . En utilisant la théorie des vecteurs gaussiens et en particulier le théorème de Cochran, on peut montrer que les moyenne et variance empiriques sont indépendantes et suivent respectivement la loi  $\mathcal{N}(m, \sigma^2/n)$  et celle de la variable  $(\sigma^2/n)Y$ , où  $Y$  suit la loi du  $\chi^2$  à  $n-1$  degrés de liberté. De plus, notons que  $S_n^2$  est biaisée car  $\mathbb{E}[S_n^2] = (n-1)\sigma^2/n$  mais asymptotiquement sans biais. L'information de Fisher étant donnée par

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix},$$

on en déduit la matrice de covariance limite,

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

En particulier on a la convergence en loi suivante :

$$\sqrt{n} (S_n^2 - \sigma^2) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 2\sigma^4).$$

résultat auquel on s'attendait d'après le TCL appliqué à la variable  $Y$ .

◦ cas d'une v.a. uniforme sur  $[0, \theta]$  avec  $\theta \in \Theta \subset ]0, +\infty[$ . La vraisemblance vaut

$$L_n(x_1, \dots, x_n, \theta) = \frac{1}{\theta^n}, \quad 0 \leq x_1, \dots, x_n \leq \theta,$$

et alors l'EMV est donné par

$$\hat{\theta}_n = \max \{X_1, \dots, X_n\}.$$

On remarque que  $\hat{\theta}_n$  converge bien en probabilité vers  $\theta$  donc c'est un estimateur consistant, et que  $\mathbb{E}[\hat{\theta}_n] = n\theta/(n+1)$  : il est asymptotiquement sans biais. En revanche la normalité asymptotique n'est pas vérifiée, ce modèle n'étant pas régulier en un certain sens (on peut montrer par ailleurs que  $n(\theta - \hat{\theta}_n)$  converge en loi vers une v.a. exponentielle de paramètre  $1/\theta$ ).

### 3.1.3 Le cas d'une chaîne de Markov

Avant de passer au cas des diffusions, mentionnons que la vraisemblance peut être définie pour des v.a. dépendantes telles que les chaînes de Markov. Par exemple, soit  $(U_n)_{n \geq 1}$  une suite i.i.d. de loi commune  $\mathcal{N}(0, 1)$  et considérons la suite  $(X_n)_{n \geq 0}$  définie par récurrence par  $X_0 = x \in \mathbb{R}$  et

$$X_{n+1} = \theta X_n + U_{n+1}, \quad n \in \mathbb{N},$$

où  $\theta \in \Theta \subset \mathbb{R}$  est un paramètre inconnu que l'on souhaite estimer. La suite  $(X_n)_{n \geq 0}$  est une chaîne de Markov homogène qui est la version à temps discret d'un processus d'Ornstein-Uhlenbeck, comme on le verra à la fin de ce chapitre. Il n'est pas difficile de montrer que le vecteur  $(X_1, \dots, X_n)$  des observations est gaussien. Plus précisément on a la densité jointe suivante :

$$f_\theta(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta x_{i-1})^2\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

où par convention  $x_0 = x$ . Ainsi, en définissant la vraisemblance de la même manière que pour les v.a. i.i.d., c'est-à-dire

$$L_n(x_1, \dots, x_n, \theta) := f_\theta(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

l'EMV est égal à

$$\hat{\theta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2} = \theta + \frac{\sum_{i=1}^n U_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2} = \theta + \frac{M_n}{[M, M]_n},$$

où  $M$  est la martingale  $M_n = \sum_{i=1}^n U_i X_{i-1}$ . Ainsi, on verra que la consistance et la normalité asymptotique de l'EMV dépendent du comportement de la martingale  $M$ .

Cet exemple fait partie de la classe importante des processus auto-régressifs d'ordre 1, notés processus AR(1), intervenant comme modèle de régression pour des séries temporelles (dans lequel la série est expliquée par ses valeurs passées plutôt que par d'autres variables). Notons que l'on a pris  $X_0$  déterministe par simplicité, mais ceci reste valide dans le cas d'une variable initiale aléatoire, l'important étant que l'on connaisse sa loi. En effet, la valeur de la chaîne au temps 0 est observée en pratique et peut donc être considérée comme connue (sa loi ne dépendra pas du paramètre inconnu  $\theta$ ).

## 3.2 Maximum de vraisemblance pour des diffusions

### 3.2.1 Le cadre général

La notion de vraisemblance peut être définie dans un cadre bien plus général que ceux que nous avons vus précédemment, en particulier pour la loi des diffusions que nous avons introduites dans le chapitre 2. Soit  $T > 0$  un horizon fini et considérons sur l'intervalle  $[0, T]$  l'unique solution de l'EDS

$$X_t = X_0 + \int_0^t b_\theta(s, X_s) ds + B_t,$$

où la fonction  $b_\theta$  est continue en temps et lipschitzienne en espace, et où le paramètre  $\theta \in \Theta \subset \mathbb{R}$  est supposé inconnu. On suppose aussi que la v.a.  $X_0$  est indépendante du mouvement brownien  $B$  et que sa loi ne dépend pas de  $\theta$ . Du point de vue de l'estimation statistique, nous sommes dans la situation où toute la trajectoire sur  $[0, T]$  du processus  $X$  est observée. Évidemment, ce type d'observations continues est *stricto sensu* impossible à réaliser, sachant qu'en pratique le nombre d'observations est forcément fini. Cependant, cette idéalisation de la réalité a un intérêt mathématique indéniable, comme nous allons le voir ci-dessous, et peut être vue comme une "borne supérieure" des résultats que l'on peut obtenir par discrétisation.

En tant que processus,  $X$  peut être considéré comme une v.a. à valeurs dans  $\mathcal{C}_T$ , l'espace vectoriel des fonctions continues de  $[0, T]$  dans  $\mathbb{R}$  muni de la tribu borélienne associée (la topologie sous-jacente est celle de la convergence uniforme). Notons  $P_\theta$  la loi de  $X$  sur  $\mathcal{C}_T$  sous la probabilité  $\mathbb{P}$ , c'est-à-dire que l'on a pour tout ensemble borélien  $A$  de  $\mathcal{C}_T$ ,

$$P_\theta(A) := \mathbb{P}(X \in A).$$

On admet dans la suite que la fonction  $b_\theta$  est choisie de manière à ce que la loi  $P_\theta$  vérifie la condition d'identifiabilité. Pour définir une notion de vraisemblance, il nous faut considérer la loi  $P_\theta$  et montrer qu'elle est absolument continue sur  $\mathcal{C}_T$  par rapport à une mesure de référence, comme la loi d'une v.a. continue l'est par rapport à la mesure de Lebesgue. Pour ce faire, nous allons utiliser le théorème de Girsanov. Notons  $M^\theta$  le processus

$$M_t^\theta := \exp\left(-\int_0^t b_\theta(s, X_s) dB_s - \frac{1}{2} \int_0^t b_\theta(s, X_s)^2 ds\right), \quad t \in [0, T].$$



Le processus  $M^\theta$  est une  $\mathbb{P}$ -martingale sur  $[0, T]$  et si  $\mathbb{Q}$  désigne la probabilité absolument continue par rapport à  $\mathbb{P}$  sur l'espace  $(\Omega, \mathcal{F}_T)$  et de densité  $M_T^\theta$ , le théorème de Girsanov nous dit que le processus  $X$  est une  $\mathbb{Q}$ -martingale et même un  $\mathbb{Q}$ -mouvement brownien (issu de  $X_0$ ). Soit  $A$  un ensemble borélien de  $\mathcal{C}_T$ . On a

$$\begin{aligned} P_\theta(A) &= \mathbb{P}(X \in A) \\ &= \mathbb{E}_{\mathbb{P}} [1_{\{X \in A\}}] \\ &= \mathbb{E}_{\mathbb{Q}} \left[ 1_{\{X \in A\}} \frac{1}{M_T^\theta} \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[ 1_{\{X \in A\}} \exp \left( \int_0^T b_\theta(t, X_t) dB_t + \frac{1}{2} \int_0^T b_\theta(t, X_t)^2 dt \right) \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[ 1_{\{X \in A\}} \exp \left( \int_0^T b_\theta(t, X_t) dX_t - \frac{1}{2} \int_0^T b_\theta(t, X_t)^2 dt \right) \right]. \end{aligned}$$

Ainsi, si l'on note  $P$  la loi sur l'espace  $\mathcal{C}_T$  du  $\mathbb{Q}$ -mouvement brownien  $X$  issu de  $X_0$ , c'est-à-dire que pour tout ensemble borélien  $A$  de  $\mathcal{C}_T$ ,

$$P(A) := \mathbb{Q}(X \in A),$$

alors on en déduit que  $P_\theta$  est absolument continue sur  $\mathcal{C}_T$  par rapport à  $P$  et admet pour densité

$$\frac{dP_\theta}{dP}(x) = \exp \left( \int_0^T b_\theta(t, x_t) dx_t - \frac{1}{2} \int_0^T b_\theta(t, x_t)^2 dt \right), \quad x \in \mathcal{C}_T.$$

On peut d'ores et déjà définir la notion de vraisemblance dans ce cas. Cependant, nous allons la définir dans un cadre plus général. Considérons le processus de diffusion  $X$  solution de l'EDS suivante :

$$X_t = X_0 + \int_0^t b_\theta(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s, \quad t \in [0, T],$$

où les fonctions  $b_\theta$  et  $\sigma$  sont supposées continues en temps et lipschitziennes en espace, avec de surcroît  $\sigma > 0$ . De plus, la v.a.  $X_0$  est supposée indépendante du mouvement brownien  $B$  et de  $\theta$ . Donnons-nous une autre fonction  $b$  satisfaisant les mêmes hypothèses et notons  $h_\theta$  la fonction  $h_\theta := (b_\theta - b)/\sigma$  puis  $M^\theta$  le processus

$$M_t^\theta := \exp \left( - \int_0^t h_\theta(s, X_s) dB_s - \frac{1}{2} \int_0^t h_\theta(s, X_s)^2 ds \right), \quad t \in [0, T].$$

Soit  $\mathbb{Q}$  la probabilité absolument continue par rapport à  $\mathbb{P}$  sur l'espace  $(\Omega, \mathcal{F}_T)$  et de densité  $M_T^\theta$ . Alors on peut montrer comme précédemment que la loi de  $X$  sous  $\mathbb{Q}$  est celle d'un processus de diffusion solution de l'EDS

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) d\tilde{B}_s, \quad t \in [0, T],$$

où  $\tilde{B}$  est le  $\mathbb{Q}$ -mouvement brownien donné par

$$\tilde{B}_t = B_t + \int_0^t h_\theta(s, X_s) ds, \quad t \in [0, T].$$

Notons que si  $b \equiv 0$  et  $\sigma \equiv 1$ , on retrouve bien le  $\mathbb{Q}$ -mouvement brownien de la situation précédente. On en déduit que si  $P_\theta$  et  $P$  désignent respectivement la loi sur l'espace  $\mathcal{C}_T$  du processus  $X$  sous  $\mathbb{P}$  et sous  $\mathbb{Q}$ , la loi  $P_\theta$  est absolument continue sur  $\mathcal{C}_T$  par rapport à  $P$  et admet pour densité (après quelques lignes de calculs)

$$\frac{dP_\theta}{dP}(x) = \exp \left( \int_0^T \frac{b_\theta - b}{\sigma^2}(t, x_t) dx_t - \frac{1}{2} \int_0^T \frac{b_\theta^2 - b^2}{\sigma^2}(t, x_t) dt \right).$$

**Définition 3.2.1.** *On suppose que la loi  $P_\theta$  sur  $\mathcal{C}_T$  associée à la diffusion  $X$  sous  $\mathbb{P}$  vérifie la condition d'identifiabilité. La vraisemblance de  $P_\theta$  par rapport à  $P$  est la fonction  $L_T$ , définie sur  $\mathcal{C}_T \times \Theta$  et à valeurs strictement positives, donnée par la densité*

$$L_T(x, \theta) := \frac{dP_\theta}{dP}(x) = \exp \left( \int_0^T \frac{b_\theta - b}{\sigma^2}(t, x_t) dx_t - \frac{1}{2} \int_0^T \frac{b_\theta^2 - b^2}{\sigma^2}(t, x_t) dt \right), \quad x \in \mathcal{C}_T.$$

La v.a. obtenue en appliquant la fonction  $x \mapsto \operatorname{argmax}_{\theta \in \Theta} L_T(x, \theta)$  au processus observé  $X$  sur  $[0, T]$  est l'EMV du paramètre  $\theta$ , que l'on note  $\hat{\theta}_T$ .

Dans la suite de ce cours, on admettra que les modèles que l'on étudie satisfont la condition d'identifiabilité.

Ainsi, trouver l'EMV revient à déterminer le paramètre  $\theta \in \Theta$  pour lequel la densité est maximale. En particulier, l'EMV satisfait l'équation en  $\theta$  suivante :

$$\int_0^T \frac{\partial_\theta b_\theta}{\sigma^2}(t, x_t) (dx_t - b_\theta(t, x_t) dt) = 0.$$

Comme on a le choix sur la loi de référence  $P$ , on choisit souvent en pratique le cas le plus simple, c'est-à-dire  $b \equiv 0$ . En revanche, la volatilité du processus  $X$  est la même sous  $\mathbb{P}$  comme sous  $\mathbb{Q}$ , donc fixée par le modèle à étudier. En modifiant la fonction  $h_\theta$ , on aurait pu s'attendre à obtenir une volatilité différente de  $\sigma$  lorsque l'on considère le processus  $X$  sous  $\mathbb{Q}$ . Il s'avère que ce n'est pas possible car les lois  $P_\theta$  et  $P$  ne seraient plus absolument continues sur  $\mathcal{C}_T$ . D'après le calcul de la variation quadratique du mouvement brownien, on a pour tout  $T > 0$  la convergence p.s. suivante :

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n (B_{iT/n} - B_{(i-1)T/n})^2 = T.$$

Étant donné  $\beta > 0$ , notons  $A_\beta$  l'ensemble borélien de  $\mathcal{C}_T$  défini par

$$A_\beta := \left\{ x \in \mathcal{C}_T : \lim_{n \rightarrow +\infty} \sum_{i=1}^n (x_{iT/n} - x_{(i-1)T/n})^2 = \beta^2 T \right\}.$$

Si  $\alpha$  est un nombre strictement positif et différent de 1, les lois des processus  $B$  et  $\alpha B$  sont étrangères sur  $\mathcal{C}_T$ , c'est-à-dire qu'elles sont portées par deux ensembles disjoints, en l'occurrence  $A_1$  et  $A_\alpha$  :

$$\mathbb{P}(B \in A_1) = \mathbb{P}(\alpha B \in A_\alpha) = 1.$$

Autrement dit, elles ne peuvent être absolument continues l'une par rapport à l'autre.

Par ailleurs, on remarque que dans la définition de  $X$ , la volatilité ne dépend pas de  $\theta$ . En effet, si c'était le cas, alors en utilisant l'approximation en probabilité de la variation quadratique, on aurait

$$\lim_{n \rightarrow +\infty} \sum_{i=1}^n (X_{iT/n} - X_{(i-1)T/n})^2 = [X, X]_T = \int_0^T \sigma_\theta(s, X_s)^2 ds.$$

La connaissance de ces intégrales permettrait alors de trouver la vraie valeur de  $\theta$ , c'est-à-dire que le problème statistique serait complètement résolu. Par exemple soit  $X = \sigma B$  et  $\sigma^2 \in \Theta \subset ]0, +\infty[$  le paramètre inconnu à estimer. Notons pour tout  $n \in \mathbb{N}_*$ ,

$$V_T^n := \sum_{i=1}^n (X_{iT/n} - X_{(i-1)T/n})^2.$$

On sait que  $V_T^n$  tend vers  $\sigma^2 T$  p.s. et dans  $L^2$  lorsque  $n$  tend vers l'infini, et que  $\mathbb{E}[V_T^n] = \sigma^2 T$ . On montre aussi en utilisant le TCL pour les variables i.i.d. que la convergence en loi suivante est satisfaite :

$$\sqrt{n} \left( \frac{V_T^n}{T} - \sigma^2 \right) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 2\sigma^4).$$

En effet, la v.a.  $V_T^n - \sigma^2 T$  s'écrit comme la somme

$$V_T^n - \sigma^2 T = \sum_{i=1}^n Z_{i,n,T}, \quad \text{avec} \quad Z_{i,n,T} := (X_{iT/n} - X_{(i-1)T/n})^2 - \frac{\sigma^2 T}{n},$$

où les  $(Z_{i,n,T})_{i=1,\dots,n}$  sont i.i.d. centrées et de variance égale à  $2\sigma^4 T^2/n^2$ .

### 3.2.2 Exemples apparaissant en finance

On vient de voir comment était défini l'EMV dans le cas des diffusions. Cet estimateur possède-t-il de bonnes propriétés de convergence lorsque  $T$  tend vers l'infini ? Pour répondre à cette question, étudions un exemple très simple, où les fonctions apparaissant dans l'EDS ne dépendent pas de  $X$ . Considérons le processus de diffusion suivant :

$$dX_t = \theta b(t) dt + \sigma(t) dB_t, \quad t \in [0, T],$$

où  $b$  et  $\sigma$  sont deux fonctions continues sur  $[0, T]$  avec de plus  $\sigma > 0$ , et  $\theta \in \Theta \subset \mathbb{R}$  est un paramètre inconnu que l'on souhaite estimer. La variable  $X_0$ , quant à elle, est

indépendante du mouvement brownien  $B$  et sa loi ne dépend pas de  $\theta$ , comme dans tous les exemples qui vont suivre (auquel cas le processus n'est pas stationnaire, la probabilité invariante dépendant de  $\theta$ ). Dans ce cas, la log-vraisemblance est donnée par

$$\log L_T(x, \theta) = \theta \int_0^T \frac{b(t)}{\sigma(t)^2} dx_t - \frac{\theta^2}{2} \int_0^T \frac{b(t)^2}{\sigma(t)^2} dt, \quad x \in \mathcal{C}_T.$$

Cette fonction de  $\theta$  étant concave et comme il n'y a qu'un seul point critique, l'EMV est donné par

$$\begin{aligned} \hat{\theta}_T &= \frac{\int_0^T \frac{b(t)}{\sigma(t)^2} dX_t}{\int_0^T \frac{b(t)^2}{\sigma(t)^2} dt} \\ &= \theta + \frac{\int_0^T f(t) dB_t}{\int_0^T f(t)^2 dt}. \end{aligned}$$

où  $f$  est la fonction  $f := b/\sigma$ . Nous sommes dans l'un des rares cas où l'on connaît explicitement la loi de l'EMV, qui est la loi  $\mathcal{N}(\theta, 1/\int_0^T f(t)^2 dt)$ . Ainsi,  $\hat{\theta}_T$  est un estimateur sans biais dont le comportement asymptotique en temps long dépend de l'intégrabilité de la fonction  $f$ . Par exemple si  $f \notin L^2(\mathbb{R}_+)$  alors  $\hat{\theta}_T$  converge vers  $\theta$  dans  $L^2$ , donc en probabilité : l'EMV est consistant. On a même la convergence p.s. en utilisant la LGN pour les martingales à temps continu, en remarquant que l'EMV s'écrit

$$\hat{\theta}_T = \theta + \frac{M_T}{[M, M]_T},$$

où  $M$  est la martingale de carré intégrable  $M_T = \int_0^T f(t) dB_t$ . En revanche si  $f^2$  est intégrable à l'infini, alors cela signifie que la v.a.  $[M, M]_\infty$  est bien définie. Ainsi,  $M$  est bornée dans  $L^2$  et d'après le théorème de convergence des martingales,  $M_T$  admet lorsque  $T$  tend vers l'infini une limite p.s. et dans  $L^2$ , notée  $M_\infty$ . On obtient alors que p.s. et dans  $L^2$ ,

$$\begin{aligned} \lim_{T \rightarrow +\infty} \hat{\theta}_T &= \theta + \frac{M_\infty}{[M, M]_\infty} \\ &= \theta + \frac{\int_0^{+\infty} f(t) dB_t}{\int_0^{+\infty} f(t)^2 dt}. \end{aligned}$$

Ainsi, il n'y a pas de convergence en temps long et l'EMV n'est pas consistant.

Calculons à présent l'EMV et étudions ses principales propriétés pour les exemples classiques que nous avons introduits dans le chapitre précédent.

### Mouvement brownien avec dérive

Commençons par le mouvement brownien avec dérive. Soit  $X$  le processus donné par

$$X_t = X_0 + \theta t + \sigma B_t, \quad t \in [0, T],$$

où  $\sigma > 0$  est supposé connu tandis que  $\theta \in \Theta \subset \mathbb{R}$  est un paramètre inconnu que l'on va estimer. Ce processus est le cas le plus simple entrant dans la classe des processus dont on vient de calculer l'EMV. On a alors que

$$\hat{\theta}_T = \theta + \sigma \frac{B_T}{T},$$

qui tend p.s. et dans  $L^2$  vers le paramètre inconnu  $\theta$ . L'EMV est donc sans biais, consistant et a pour loi  $\mathcal{N}(\theta, \sigma^2/T)$ .

### Modèle de Black-Scholes

Considérons maintenant le modèle de Black-Scholes. L'EDS satisfaite est la suivante :

$$dX_t = \theta X_t dt + \sigma X_t dB_t, \quad t \in [0, T],$$

où  $X_0 > 0$  p.s.,  $\sigma > 0$  est supposé connu tandis que  $\theta \in \Theta \subset \mathbb{R}$  est le paramètre inconnu à estimer. La log-vraisemblance du modèle est donnée par

$$\log L_T(x, \theta) = \frac{\theta}{\sigma^2} \int_0^T \frac{dx_t}{x_t} - \frac{\theta^2 T}{2\sigma^2}, \quad x \in \mathcal{C}_T.$$

Cette fonction de  $\theta$  étant concave et comme il n'y a qu'un seul point critique, l'EMV est donné par

$$\begin{aligned} \hat{\theta}_T &= \frac{1}{T} \int_0^T \frac{dX_t}{X_t} \\ &= \theta + \sigma \frac{B_T}{T}, \end{aligned}$$

c'est-à-dire le même estimateur que dans le cas du mouvement brownien avec dérive, les conclusions étant alors les mêmes. Notons que ce résultat était attendu, l'EMV étant stable par bijection. En effet, on peut montrer que s'il existe une fonction bijective  $f$  nous permettant d'exprimer un processus  $X^2$  en fonction d'un processus  $X^1$  dépendant d'un paramètre inconnu  $\theta$ , alors l'EMV de  $\theta$  associé à  $X^2$  est le même que celui associé à  $X^1$ . Dans notre cas, le processus  $X^1$  est un mouvement brownien avec dérive,

$$dX_t^1 = \sigma dB_t + (\theta - \sigma^2/2) dt,$$

tandis que  $X^2 := e^{X^1}$  est le processus de Black-Scholes. Le lecteur attentif aura remarqué que la dérive de  $X^1$  fait apparaître le terme supplémentaire  $-\sigma^2/2$  qui ne dépend pas de  $\theta$ . Il s'agit donc tout d'abord d'estimer  $\theta^1 := \theta - \sigma^2/2$ , puis d'estimer  $\theta = g(\theta^1)$  où  $g(u) := u + \sigma^2/2$ . Le calcul précédent pour le mouvement brownien géométrique montre que l'EMV de  $\theta^1$  vaut  $\hat{\theta}_T^1 = \theta^1 + \sigma B_T/T$  puis que celui de  $\theta$  vaut  $g(\hat{\theta}_T^1) = \theta + \sigma B_T/T$  : nous sommes dans le cadre d'application de la "delta-method", qui est aussi valable pour les diffusions, comme mentionné ci-dessous pour les processus ergodiques.

### Processus ergodiques

Étudions à présent le cas des processus ergodiques, pour lesquels il existe un résultat général sur l'EMV multidimensionnel.

**Théorème 3.2.2.** *Soit  $\theta \in \Theta \subset \mathbb{R}^d$  un paramètre inconnu. Sous de bonnes hypothèses d'ergodicité du processus, par exemple si fonctions  $b_\theta$  et  $\sigma$  sont polynomiales et satisfont l'hypothèse (H), l'EMV est consistant, asymptotiquement sans biais et asymptotiquement normal au sens de la convergence en loi vers un vecteur gaussien :*

$$\sqrt{T} \left( \hat{\theta}_T - \theta \right) \xrightarrow[T \rightarrow +\infty]{} U_\theta,$$

où  $U_\theta \sim \mathcal{N}_d(0, I(\theta)^{-1})$ ,  $I(\theta)$  est la matrice de Fisher  $d \times d$  formée des éléments

$$I(\theta)_{i,j} := \int_{\mathbb{R}} \frac{\partial_{\theta_i} b_\theta \partial_{\theta_j} b_\theta}{\sigma^2} d\pi_\theta,$$

supposée inversible, et  $\pi_\theta$  est l'unique probabilité invariante du processus, qui dépend de  $\theta$ . De plus, il est asymptotiquement efficace.

*Delta-method* : si  $g : \Theta \rightarrow \mathbb{R}^k$  est une fonction de classe  $\mathcal{C}^1$  et dont la matrice jacobienne  $Jac(g)(\theta)$  en  $\theta$  est inversible, alors  $g(\hat{\theta}_T)$  est l'EMV de  $g(\theta)$  et l'on a la convergence en loi suivante :

$$\sqrt{T} \left( g(\hat{\theta}_T) - g(\theta) \right) \xrightarrow[T \rightarrow +\infty]{} Jac(g)(\theta)^T U_\theta.$$

Examinons à présent le cas des trois processus ergodiques que nous avons vus dans le chapitre précédent, qui satisfont l'hypothèse (H) et pour lesquels le théorème précédent va s'appliquer.

Processus d'Ornstein-Uhlenbeck. Commençons par le processus d'Ornstein-Uhlenbeck solution de l'EDS

$$dX_t = -\theta X_t dt + \sigma dB_t, \quad t \in [0, T],$$

où  $\sigma > 0$  est supposé connu et  $\theta \in \Theta \subset ]0, +\infty[$  est le paramètre inconnu, supposé strictement positif pour assurer l'ergodicité du processus. L'unique probabilité invariante  $\pi_\theta$  est la loi  $\mathcal{N}(0, \sigma^2/2\theta)$ . La log-vraisemblance pour ce modèle est

$$\log L_T(x, \theta) = -\frac{\theta}{\sigma^2} \int_0^T x_t dx_t - \frac{\theta^2}{2\sigma^2} \int_0^T x_t^2 dt, \quad x \in \mathcal{C}_T.$$

Cette fonction de  $\theta$  est concave et n'a qu'un seul point critique. Ainsi, l'EMV est donné par

$$\hat{\theta}_T = -\frac{\int_0^T X_t dX_t}{\int_0^T X_t^2 dt}$$

$$= \theta - \sigma \frac{\int_0^T X_t dB_t}{\int_0^T X_t^2 dt}.$$

En utilisant la LGN pour les martingales, on en déduit que l'EMV est consistant et asymptotiquement sans biais. En revanche, contrairement aux précédents modèles, la normalité asymptotique n'est pas immédiate à cause de la présence de l'intégrale stochastique. Cependant, le TCL markovien peut s'appliquer et l'on obtient la convergence en loi suivante :

$$\sqrt{T} (\hat{\theta}_T - \theta) \xrightarrow{T \rightarrow +\infty} \mathcal{N}(0, 2\theta),$$

résultat correspondant bien au théorème ci-dessus car l'information de Fisher est donnée par

$$I(\theta) = \frac{1}{\sigma^2} \int_{\mathbb{R}} x^2 \pi_{\theta}(dx) = \frac{1}{2\theta}.$$

Modèle de Vasicek. Focalisons-nous maintenant sur le modèle de Vasicek qui généralise le cas précédent, et pour lequel plusieurs situations sont possibles. Le processus  $X$  est solution de l'EDS

$$dX_t = -(\mu X_t - \nu) dt + \sigma dB_t,$$

où  $\mu > 0$ ,  $\nu \in \mathbb{R}_*$  et  $\sigma > 0$ . L'unique probabilité invariante  $\pi_{\theta}$  est la loi  $\mathcal{N}(\nu/\mu, \sigma^2/2\mu)$ . Notons dans la suite

$$Y_1 := \int_0^T X_t^2 dt, \quad Y_2 := \int_0^T X_t dt \quad \text{and} \quad Z := - \int_0^T X_t dX_t.$$

◦ Cas où  $\nu$  et  $\sigma$  sont supposés connus tandis que  $\theta = \mu \in \Theta \subset ]0, +\infty[$  est le paramètre inconnu à estimer. En procédant de la même manière que pour le processus d'Ornstein-Uhlenbeck, on obtient que

$$\hat{\theta}_T = \frac{\nu Y_2 + Z}{Y_1} = \theta - \sigma \frac{\int_0^T X_t dB_t}{\int_0^T X_t^2 dt},$$

et en combinant les TCL et LGN markoviens, l'EMV est consistant, asymptotiquement sans biais et l'on a la convergence en loi

$$\sqrt{T} (\hat{\theta}_T - \theta) \xrightarrow{T \rightarrow +\infty} \mathcal{N}\left(0, \frac{2\sigma^2\theta^2}{2\nu^2 + \sigma^2\theta}\right).$$

◦ Cas où  $\mu$  et  $\sigma$  sont supposés connus tandis que  $\theta = \nu \in \Theta \subset ]0, +\infty[$  est à estimer. On a alors que

$$\hat{\theta}_T = \frac{X_T - X_0 + \mu Y_2}{T} = \theta + \sigma \frac{B_T}{T},$$

donc que l'EMV est consistant, sans biais et a pour loi  $\mathcal{N}(\theta, \sigma^2/T)$ .

◦ Cas où seul  $\sigma$  est supposé connu, le paramètre inconnu à estimer étant  $\theta = (\mu, \nu) \in \Theta \subset ]0, +\infty[ \times \mathbb{R}_*$ . On trouve que l'EMV bidimensionnel vaut

$$\hat{\theta}_T = \left( \frac{TZ + Y_2(X_T - X_0)}{TY_1 - Y_2^2}, \frac{Y_2Z + Y_1(X_T - X_0)}{TY_1 - Y_2^2} \right).$$

D'après le théorème précédent, l'EMV est consistant, asymptotiquement sans biais et asymptotiquement normal au sens de la convergence en loi

$$\sqrt{T} \left( \hat{\theta}_T - \theta \right) \xrightarrow[T \rightarrow +\infty]{} \mathcal{N}_2 \left( 0, I(\theta)^{-1} \right),$$

où  $I(\theta)$  est la matrice de Fisher

$$I(\theta) = \frac{1}{2\mu^2\sigma^2} \begin{pmatrix} 2\nu^2 + \mu\sigma^2 & -2\mu\nu \\ -2\mu\nu & 2\mu^2 \end{pmatrix}.$$

Modèle CIR. Terminons par le cas du modèle CIR. Le processus  $X$  est solution de l'EDS

$$dX_t = \mu(\nu - X_t) dt + \sigma \sqrt{X_t} dB_t,$$

où p.s.  $X_0 > 0$  et les paramètres satisfont  $\mu > 0, \nu > 0, \sigma > 0$  avec  $2\mu\nu > \sigma^2$ . Réécrivons ce modèle légèrement différemment afin de dissocier les paramètres apparaissant dans la dérive. À présent, le processus  $X$  est solution de l'EDS

$$dX_t = (a - bX_t) dt + \sigma \sqrt{X_t} dB_t,$$

où p.s.  $X_0 > 0$  et les paramètres satisfont cette fois  $a > 0, b > 0, \sigma > 0$  avec  $2a > \sigma^2$ . L'unique probabilité invariante  $\pi_\theta$  est la loi Gamma de paramètres  $\alpha = 2a/\sigma^2$  et  $\beta = \sigma^2/2b$ . Notons dans la suite

$$Y_1 := \int_0^T X_t^{-1} dt, \quad Y_2 := \int_0^T X_t dt \quad \text{and} \quad Z := \int_0^T X_t^{-1} dX_t.$$

◦ Cas où  $b$  et  $\sigma$  sont supposés connus tandis que  $\theta = a \in \Theta \subset ]\sigma^2/2, +\infty[$  est le paramètre inconnu. On a alors que

$$\hat{\theta}_T = \frac{Z + bT}{Y_1} = \theta + \sigma \frac{\int_0^T X_t^{-1/2} dB_t}{\int_0^T X_t^{-1} dt},$$

et les TCL et LGN markoviens entraînent la convergence en loi

$$\sqrt{T} \left( \hat{\theta}_T - \theta \right) \xrightarrow[T \rightarrow +\infty]{} \mathcal{N} \left( 0, \frac{2a\sigma^2 - \sigma^4}{2b} \right),$$

car  $\int_{\mathbb{R}} x^{-1} \pi_\theta(dx) = 1/\beta(\alpha - 1) = 2b/(2a - \sigma^2)$ .



◦ Cas où  $a$  et  $\sigma$  sont supposés connus tandis que  $\theta = b \in \Theta \subset ]0, +\infty[$  est le paramètre inconnu à estimer. L'EMV est donné par

$$\hat{\theta}_T = \frac{aT - X_T + X_0}{Y_2} = \theta - \sigma \frac{\int_0^T \sqrt{X_t} dB_t}{\int_0^T X_t dt}.$$

En combinant les TCL et LGN markoviens, on obtient la convergence en loi

$$\sqrt{T} (\hat{\theta}_T - \theta) \xrightarrow{T \rightarrow +\infty} \mathcal{N} \left( 0, \frac{\sigma^2 b}{a} \right).$$

car  $\int_{\mathbb{R}} x \pi_{\theta}(dx) = \alpha\beta = a/b$ .

◦ Cas où seul  $\sigma$  est supposé connu, le paramètre inconnu à estimer étant  $\theta = (a, b) \in \Theta \subset ]\sigma^2/2, +\infty[ \times ]0, +\infty[$ . On trouve que l'EMV bi-dimensionnel vaut

$$\hat{\theta}_T = \left( \frac{ZY_2 - T(X_T - X_0)}{Y_1Y_2 - T^2}, \frac{ZT - Y_1(X_T - X_0)}{Y_1Y_2 - T^2} \right).$$

D'après le théorème précédent, l'EMV est consistant, asymptotiquement sans biais et asymptotiquement normal au sens de la convergence en loi

$$\sqrt{T} (\hat{\theta}_T - \theta) \xrightarrow{T \rightarrow +\infty} \mathcal{N}_2 (0, I(\theta)^{-1}),$$

où  $I(\theta)$  est la matrice de Fisher

$$I(\theta) = \frac{1}{\sigma^2} \begin{pmatrix} \frac{2b}{2a - \sigma^2} & -1 \\ -1 & \frac{a}{b} \end{pmatrix}.$$

### 3.2.3 Test de Neyman-Pearson

La propriété d'absolue continuité apparaissant dans la définition de la vraisemblance permet d'appliquer des procédures statistiques comme le test de Neyman-Pearson que l'on va voir ci-dessous. On a vu que la vraisemblance était donnée par la densité

$$L_T(x, \theta) = \frac{dP_{\theta}}{dP}(x) = \exp \left( \int_0^T \frac{b_{\theta} - b}{\sigma^2}(t, x_t) dx_t - \frac{1}{2} \int_0^T \frac{b_{\theta}^2 - b^2}{\sigma^2}(t, x_t) dt \right), \quad x \in \mathcal{C}_T,$$

où  $\theta \in \Theta$  est le paramètre inconnu à estimer,  $P_{\theta}$  est la loi sur  $\mathcal{C}_T$  sous  $\mathbb{P}$  associée à la diffusion  $X$  solution de l'EDS

$$dX_t = b_{\theta}(t, X_t) dt + \sigma(t, X_t) dB_t,$$

tandis que  $P$  est la loi de  $X$  sur  $\mathcal{C}_T$  (sous une autre probabilité  $\mathbb{Q}$ ) correspondant à celle d'une diffusion solution de l'EDS

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) d\tilde{B}_t,$$

où  $\tilde{B}$  est un  $\mathbb{Q}$ -mouvement brownien. Notons que si l'on fait dépendre  $b$  d'un paramètre  $\tilde{\theta}$  supposé connu, alors on définit le rapport de vraisemblance par

$$L_T(x, \theta, \tilde{\theta}) := \frac{dP_\theta}{dP_{\tilde{\theta}}}(x) = \exp \left( \int_0^T \frac{b_\theta - b_{\tilde{\theta}}}{\sigma^2}(t, x_t) dx_t - \frac{1}{2} \int_0^T \frac{b_\theta^2 - b_{\tilde{\theta}}^2}{\sigma^2}(t, x_t) dt \right), \quad x \in \mathcal{C}_T.$$

Rappelons brièvement quelques éléments à propos des tests statistiques. Un test d'hypothèse est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon (c'est-à-dire des observations), de faire un choix entre deux hypothèses statistiques, l'hypothèse nulle, notée  $(H_0)$ , et l'hypothèse alternative, notée  $(H_1)$ . Cette règle de décision va nous conduire à l'acceptation ou au rejet de  $(H_0)$ . Il ne s'agit pas de déterminer si elle est fondamentalement vraie ou non, mais plutôt de voir si c'est une hypothèse cohérente avec les observations. Cette décision étant fondée sur une information partielle, les observations, il est donc impossible de prendre la bonne décision à coup sûr. En pratique, on met en oeuvre une démarche qui nous permet de rejeter à tort  $(H_0)$  dans une faible proportion de cas. La conclusion déduite des résultats de l'échantillon aura un caractère probabiliste : on ne pourra prendre une décision qu'en ayant conscience qu'il y a un certain risque qu'elle soit erronée. Ce risque, dit de première espèce, nous est donné par le niveau du test, noté  $\alpha$ , consenti à l'avance. Plus précisément, il y a deux façons de se tromper lors d'un test statistique :

- rejeter à tort  $(H_0)$  alors qu'elle est vraie : c'est le risque de première espèce et l'on note  $\alpha$  la probabilité de se tromper dans ce sens, fixée à l'avance. On appelle parfois  $\alpha$  le risque de faux positif : en rejetant l'hypothèse nulle, on considère l'hypothèse à tester comme validée (positif) alors qu'elle ne l'est pas (faux).

- accepter  $(H_0)$  alors qu'elle est fautive. C'est le risque de deuxième espèce et l'on note  $\beta$  la probabilité de se tromper dans ce sens, correspondant aux faux négatifs : comme on accepte  $(H_0)$ , l'hypothèse à tester ne peut pas être validée (négatif) alors qu'elle est vraie (faux). La quantité  $1 - \beta$  est appelée puissance du test.

En pratique, il s'agit souvent d'effectuer un compromis entre ces deux types d'erreur car l'on ne diminue l'un des risques qu'en consentant à augmenter l'autre. Notons de surcroît que ces deux erreurs ne jouent pas un rôle symétrique. On contrôle uniquement le risque de première espèce  $\alpha$  : cela revient à considérer que le risque de rejeter  $(H_0)$  alors que cette hypothèse est vraie est beaucoup plus coûteux/dangereux que celui de la conserver à tort (ce dernier risque  $\beta$  n'étant pas maîtrisé).

Regardons à présent comment mettre en oeuvre le test de Neyman-Pearson dans un cas simple. Un radar est utilisé pour détecter le passage éventuel d'un avion correspondant à un signal  $m_t := \int_0^t f(s) ds$  connu, où  $t \in [0, T]$  et  $f : [0, T] \rightarrow \mathbb{R}$  est une fonction continue. Le radar présente un bruit d'enregistrement brownien avec un coefficient de diffusion  $\sigma > 0$  connu. Le signal enregistré  $x \in \mathcal{C}_T$  est la réalisation d'un processus  $X$  sur  $[0, T]$  pour lequel deux hypothèses distinctes sont possibles :

- hypothèse nulle  $(H_0)$  :  $X = m + \sigma B$ , correspondant à la présence d'un avion.
- hypothèse alternative  $(H_1)$  :  $X = \sigma B$ , traduisant l'absence d'avion.

Quels sont ici les risques de première et deuxième espèces ? Rejeter ( $H_0$ ) à tort revient à croire à l'absence d'avion alors qu'en réalité il y en a un, tandis qu'accepter ( $H_0$ ) à tort signifie que l'on pense qu'un avion passe alors qu'il n'en est rien. On voit ainsi que ces deux erreurs n'ont pas les mêmes conséquences et c'est pourquoi l'hypothèse nulle a été choisie de cette manière.

Notons  $\mathbb{P}_{H_0}, \mathbb{P}_{H_1}$  les probabilités définies sur l'espace  $(\Omega, \mathcal{F}_T)$  de la manière suivante : sous  $\mathbb{P}_{H_0}$  le processus  $X$  vaut  $m + \sigma B$  (on note  $P_0$  sa loi sur  $\mathcal{C}_T$ ) et sous  $\mathbb{P}_{H_1}$ , on a  $X = \sigma B$  (on note  $P_1$  sa loi sur  $\mathcal{C}_T$ ). Ainsi, le rapport de vraisemblance s'écrit

$$\frac{dP_1}{dP_0}(y) = \exp\left(-\frac{1}{\sigma^2} \int_0^T f(t) dy_t + \frac{1}{2\sigma^2} \int_0^T f(t)^2 dt\right), \quad y \in \mathcal{C}_T.$$

Soit  $\alpha \in ]0, 1[$ . Le test de Neyman-Pearson admet la règle de décision suivante : on rejette ( $H_0$ ) lorsque le rapport de vraisemblance est grand, c'est-à-dire lorsque

$$\int_0^T f(t) dx_t \leq C_\alpha,$$

où le seuil  $C_\alpha$  est déterminé selon le niveau  $\alpha$ ,

$$\mathbb{P}_{H_0}\left(\int_0^T f(t) dX_t \leq C_\alpha\right) = \alpha.$$

Autrement dit, la probabilité de rejeter ( $H_0$ ) à tort est égale à  $\alpha$  (risque de première espèce). Sous l'hypothèse ( $H_0$ ), on a  $X = m + \sigma B$  donc

$$\int_0^T f(t) dX_t = \sigma \int_0^T f(t) dB_t + \int_0^T f(t)^2 dt \stackrel{\text{loi}}{=} \sigma v_T Z + v_T^2,$$

où  $Z$  désigne une v.a. de loi  $\mathcal{N}(0, 1)$  et  $v_T^2 := \int_0^T f(t)^2 dt$ . D'où

$$\alpha = \mathbb{P}_{H_0}\left(\int_0^T f(t) dX_t \leq C_\alpha\right) = \mathbb{P}\left(Z \leq \frac{C_\alpha - v_T^2}{\sigma v_T}\right).$$

Ainsi, en notant  $q_\alpha$  le quantile d'ordre  $\alpha$  pour la loi normale centrée et réduite, on en déduit la valeur de  $C_\alpha$  :

$$C_\alpha = q_\alpha \sigma v_T + v_T^2.$$

Enfin, les paramètres  $\sigma$  et  $v_T$  étant connus, il reste à fixer le risque de première espèce  $\alpha$  (en général  $\alpha = 0,05$ ) puis conclure le test en utilisant la règle de décision en fonction de la réalisation  $x$  du processus  $X$  : si  $\int_0^T f(t) dx_t \leq C_\alpha$ , alors l'hypothèse ( $H_0$ ) est rejetée. Ou encore la probabilité de se tromper en rejetant ( $H_0$ ) est inférieure au seuil  $\alpha$  préalablement déterminé et le test est jugé significatif.

Au contraire, sous ( $H_1$ ), on a  $X = \sigma B$  et donc l'intégrale  $\int_0^T f(t) dX_t$  a même loi que la v.a.  $\sigma v_T Z$ . D'où le risque de seconde espèce  $\beta$  (ou probabilité d'accepter ( $H_0$ ) à tort) est

$$\beta = \mathbb{P}_{H_1}\left(\int_0^T f(t) dX_t > C_\alpha\right) = \mathbb{P}\left(Z > q_\alpha + \frac{v_T}{\sigma}\right).$$

Comme  $C_\alpha$  est croissant en  $\alpha$ , diminuer  $\alpha$  revient à diminuer  $C_\alpha$  et donc à augmenter  $\beta$  : il faut donc trouver un compromis entre les risques de première et de deuxième espèces. Notons par ailleurs que lorsque le rapport signal sur bruit  $v_T/\sigma$  grandit,  $\beta$  tend vers 0 très rapidement, à vitesse exponentielle car pour tout  $x > 0$ ,

$$\mathbb{P}(Z > x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}.$$

### 3.3 Discrétisation des diffusions

#### 3.3.1 Diffusions observées à des instants discrets

Comme nous l'avons déjà souligné, l'EMV défini pour les diffusions est purement théorique au sens où il faudrait que toute la trajectoire entre 0 et  $T$  du processus  $X$  soit observée, ce qui est impossible en pratique. En réalité on observe uniquement des variables correspondant à une discrétisation temporelle de la diffusion, c'est-à-dire une suite de variables  $X_{h_n}, X_{2h_n}, \dots, X_{nh_n}$  définie par

$$X_{ih_n} = X_{(i-1)h_n} + \int_{(i-1)h_n}^{ih_n} b(r, X_r) dr + \int_{(i-1)h_n}^{ih_n} \sigma(r, X_r) dB_r, \quad i \in \{1, \dots, n\},$$

avec  $T = nh_n$  et où l'on autorise le pas  $h_n$  des observations à dépendre du nombre d'observations  $n$ . Une diffusion étant un processus de Markov, la suite extraite ( $X_{ih_n}$ ) indiquée par  $i$  forme une chaîne de Markov. Plusieurs études asymptotiques sont alors envisageables, qui diffèrent suivant les modèles que l'on considère ou selon les observations dont on dispose :

- si  $h_n$  ne dépend pas de  $n$  alors faire tendre le nombre  $n$  d'observations vers l'infini revient à faire tendre  $T$  vers l'infini. Ce cadre est similaire à celui étudié précédemment en temps continu et c'est la convergence des martingales, voire les théorèmes limites pour les chaînes de Markov ergodiques, qui sont utilisés pour démontrer les éventuelles consistance et normalité asymptotique de l'EMV associé. C'est par exemple le cas de la chaîne de Markov correspondant au processus d'Ornstein-Uhlenbeck, dont nous avons calculé l'EMV en début de chapitre et sur lequel nous allons revenir ci-dessous.

- si  $h_n \rightarrow 0$  lorsque  $n \rightarrow +\infty$  tandis que le produit  $T = nh_n$  reste fixe (c'est le cas de la subdivision uniforme), on est dans le cas d'observations de plus en plus rapprochées à l'intérieur d'un intervalle fixe. Cette situation a été étudiée lors de l'estimation par la variation quadratique approchée d'un paramètre inconnu  $\theta$  apparaissant dans la volatilité  $\sigma_\theta$  (en particulier pour estimer le paramètre  $\sigma$  lui-même s'il devait être inconnu, dans le cas d'une volatilité constante). En revanche cette discrétisation n'est pas appropriée si le paramètre inconnu apparaît dans la fonction de dérive  $b$  car l'EMV n'est alors même pas consistant.

- si  $h_n \rightarrow 0$  et  $T \rightarrow +\infty$  lorsque  $n \rightarrow \infty$  (par exemple  $h_n = n^{-1/2}$ ), il s'agit d'observations de plus en plus rapprochées à l'intérieur d'un intervalle de plus en plus

grand. Deux régimes asymptotiques rentrent alors en concurrence et c'est pourquoi l'étude est en général plus délicate.

Remarquons que si l'on n'est pas dans le second cas, l'estimation de la volatilité  $\sigma$  comme paramètre inconnu devient intéressante car nous ne disposons plus de l'approximation par la variation quadratique approchée.

Illustrons dans ce cas discret la mise en oeuvre de l'estimation par maximum de vraisemblance sur deux exemples, le modèle de Black-Scholes et le processus d'Ornstein-Uhlenbeck. Comme on l'a vu précédemment, il est suffisant pour Black-Scholes de se consacrer à l'estimation des paramètres inconnus intervenant non pas dans ce modèle mais plutôt dans celui du mouvement brownien avec dérive. Ainsi, soit  $X = (X_{ih_n})_{i \in \{1, 2, \dots, n\}}$  la discrétisation du mouvement brownien avec dérive sur l'intervalle  $[0, T]$ , c'est-à-dire pour tout  $i \in \{1, 2, \dots, n\}$ ,

$$X_{ih_n} = X_{(i-1)h_n} + \mu h_n + \sigma (B_{ih_n} - B_{(i-1)h_n}), \quad X_0 = x \in \mathbb{R},$$

où  $\theta = (\mu, \sigma^2) \in \Theta \subset \mathbb{R} \times ]0, +\infty[$  est le paramètre bidimensionnel inconnu à estimer. La densité jointe  $f_\theta$  de la chaîne de Markov  $X$  est celle d'un vecteur gaussien  $n$ -dimensionnel et l'on a

$$f_\theta(x_1, \dots, x_n) = \frac{1}{(2\pi \sigma^2 h_n)^{n/2}} \exp\left(-\frac{1}{2\sigma^2 h_n} \sum_{i=1}^n (x_i - x_{i-1} - \mu h_n)^2\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

où par convention  $x_0 = x$ . D'où la log-vraisemblance est donnée pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$  par

$$\log L_n(x_1, \dots, x_n, \theta) = -\frac{n}{2} \log(2\pi \sigma^2 h_n) - \frac{1}{2\sigma^2 h_n} \sum_{i=1}^n (x_i - x_{i-1} - \mu h_n)^2,$$

et l'on en déduit l'EMV bidimensionnel

$$\hat{\theta}_n = \left( \frac{1}{h_n} \bar{Y}_{n, h_n}, \frac{1}{T} \sum_{i=1}^n (Y_{ih_n} - \bar{Y}_{n, h_n})^2 \right),$$

où  $Y$  est la suite définie par  $Y_{ih_n} := X_{ih_n} - X_{(i-1)h_n}$ ,  $i \in \{1, 2, \dots, n\}$ , et  $\bar{Y}_{n, h_n}$  est la moyenne empirique associée,

$$\bar{Y}_{n, h_n} := \frac{1}{n} \sum_{i=1}^n Y_{ih_n}.$$

Notons que grâce à l'indépendance et à la stationnarité des accroissements browniens, la suite  $Y$  est i.i.d. de loi  $\mathcal{N}(\mu h_n, \sigma^2 h_n)$  : on obtient alors les mêmes conclusions que dans le cas de variables gaussiennes i.i.d. vu en début de chapitre : les coordonnées  $\hat{\mu}_n$  et  $\hat{\sigma}_n^2$  de l'EMV sont indépendantes et suivent respectivement la loi  $\mathcal{N}(\mu, \sigma^2/T)$  et celle de la variable  $(\sigma^2/n)Y$ , où  $Y$  suit la loi du  $\chi^2$  à  $n-1$  degrés de liberté. Par ailleurs on peut réécrire l'EMV comme

$$\hat{\theta}_n = \left( \mu + \sigma \frac{B_T}{T}, \frac{\sigma^2}{n} \sum_{i=1}^n \left( \frac{B_{ih_n} - B_{(i-1)h_n}}{\sqrt{h_n}} \right)^2 - h_n \sigma^2 \left( \frac{B_T}{T} \right)^2 \right),$$

ce qui nous permet d'en déduire les convergences suivantes :

◦ si  $h_n$  ne dépend pas de  $n$  alors lorsque  $n \rightarrow +\infty$ , on a que  $T \rightarrow +\infty$ , d'où la consistance de l'EMV en utilisant la LGN pour les deux coordonnées (on sait que  $B_T/T$  tend vers 0 p.s. lorsque  $T \rightarrow +\infty$ ). De plus le TCL ainsi que le lemme de Slutsky entraînent la normalité asymptotique de l'estimateur  $\hat{\sigma}_n^2$  au sens de la convergence en loi suivante :

$$\sqrt{n} (\hat{\sigma}_n^2 - \sigma^2) \underset{n \rightarrow +\infty}{\Longrightarrow} \mathcal{N}(0, 2\sigma^4).$$

◦ si  $h_n = T/n$  alors  $T$  reste fixe lorsque  $n \rightarrow +\infty$  et aucune convergence n'a lieu pour l'estimateur  $\hat{\mu}_n$ , au contraire de  $\hat{\sigma}_n^2$  : bien que la quantité  $B_T/T$  reste fixe, la présence de  $h_n$  devant ce terme entraîne la convergence p.s. vers 0 lorsque  $n \rightarrow +\infty$ . Ainsi, le TCL et le lemme de Slutsky entraînent la même normalité asymptotique que précédemment.

◦ si  $h_n \rightarrow 0$  et  $T \rightarrow +\infty$  lorsque  $n \rightarrow +\infty$ , on obtient exactement les mêmes conclusions que dans le cas  $h_n$  constant.

À présent, étudions le cas du processus d'Ornstein-Uhlenbeck. On note  $X = (X_{ih_n})_{i \in \{1, 2, \dots, n\}}$  la discrétisation de ce processus sur l'intervalle  $[0, T]$ , c'est-à-dire pour tout  $i \in \{1, 2, \dots, n\}$ ,

$$X_{ih_n} = X_{(i-1)h_n} - \mu \int_{(i-1)h_n}^{ih_n} X_r dr + \sigma (B_{ih_n} - B_{(i-1)h_n}), \quad X_0 = x \in \mathbb{R},$$

où  $\theta = (\mu, \sigma^2) \in \Theta \subset ]0, +\infty[^2$  est le paramètre inconnu à estimer. Comme il est difficile de calculer les densités conditionnelles en utilisant la formule ci-dessus, tirons profit de l'expression explicite du processus, c'est-à-dire

$$\begin{aligned} X_{ih_n} &= e^{-\mu h_n} X_{(i-1)h_n} + \sigma e^{-\mu i h_n} \int_{(i-1)h_n}^{ih_n} e^{\mu r} dB_r \\ &\stackrel{\text{loi}}{=} e^{-\mu h_n} X_{(i-1)h_n} + \sigma \sqrt{\frac{1 - e^{-2\mu h_n}}{2\mu}} U, \end{aligned}$$

où  $U$  suit la loi  $\mathcal{N}(0, 1)$  et est indépendante de  $X_{(i-1)h_n}$  grâce à l'indépendance des accroissements browniens. Afin d'utiliser la delta-method, supposons que  $h_n = h$  ne dépende pas de  $n$  et posons

$$\tilde{\mu} := e^{-\mu h} \in ]0, 1[ \quad \text{et} \quad \tilde{\sigma} := \sigma \sqrt{\frac{1 - e^{-2\mu h}}{2\mu}} > 0,$$

de sorte que si l'on note respectivement  $\hat{\mu}_n$  et  $\hat{\sigma}_n^2$  les EMV des paramètres  $\tilde{\mu}$  et  $\tilde{\sigma}^2$ , alors le couple  $(\mu, \sigma^2)$  peut être estimé par l'EMV bi-dimensionnel

$$\left( -\frac{\log(\hat{\mu}_n)}{h}, \quad -\frac{2 \log \hat{\mu}_n}{h(1 - \hat{\mu}_n^2)} \hat{\sigma}_n^2 \right).$$

Ainsi, on peut se restreindre à l'estimation du paramètre  $\tilde{\theta} := (\tilde{\mu}, \tilde{\sigma}^2)$  dans le modèle AR(1) suivant :

$$Y_i = \tilde{\mu} Y_{i-1} + \tilde{\sigma} U_i,$$

où  $Y_i = X_{ih}$  et la suite  $(U_i)$  indicée par  $i \in \{1, \dots, n\}$  est i.i.d. de loi  $\mathcal{N}(0, 1)$ . Évidemment,  $Y$  est un vecteur gaussien et pour tout  $i \in \{1, \dots, n\}$ ,

$$Y_i \sim \mathcal{N}\left(\tilde{\mu}^i x, \tilde{\sigma}^2 \frac{1 - \tilde{\mu}^{2i}}{1 - \tilde{\mu}^2}\right).$$

En particulier, la probabilité invariante est la loi normale centrée et de variance  $\tilde{\sigma}^2/(1 - \tilde{\mu}^2)$ . Comme nous l'avons vue en début de chapitre, la densité jointe  $f_\theta$  de la chaîne de Markov  $Y$  est donnée par

$$f_\theta(y_1, \dots, y_n) = \frac{1}{(2\pi\tilde{\sigma}^2)^{n/2}} \exp\left(-\frac{1}{2\tilde{\sigma}^2} \sum_{i=1}^n (y_i - \tilde{\mu} y_{i-1})^2\right), \quad (y_1, \dots, y_n) \in \mathbb{R}^n,$$

où par convention  $y_0 = x$ . En calculant la log-vraisemblance, on en déduit l'EMV bidimensionnel  $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$  :

$$\begin{aligned} \hat{\theta}_n &= \left( \frac{\sum_{i=1}^n Y_{i-1} Y_i}{\sum_{i=1}^n Y_{i-1}^2}, \quad \frac{1}{n} \sum_{i=1}^n (Y_i - Y_{i-1} \hat{\mu}_n)^2 \right) \\ &= \left( \tilde{\mu} + \tilde{\sigma} \frac{\sum_{i=1}^n Y_{i-1} U_i}{\sum_{i=1}^n Y_{i-1}^2}, \quad \frac{\tilde{\sigma}^2}{n} \sum_{i=1}^n U_i^2 - \frac{\tilde{\sigma}^2}{n} \frac{(\sum_{i=1}^n Y_{i-1} U_i)^2}{\sum_{i=1}^n Y_{i-1}^2} \right). \end{aligned}$$

En notant  $M$  la martingale discrète  $M_n := \sum_{i=1}^n Y_{i-1} U_i$ , l'EMV se réécrit

$$\hat{\theta}_n = \left( \tilde{\mu} + \tilde{\sigma} \frac{M_n}{[M, M]_n}, \quad \frac{\tilde{\sigma}^2}{n} \sum_{i=1}^n U_i^2 - \frac{\tilde{\sigma}^2}{n} \frac{M_n^2}{[M, M]_n} \right),$$

et les théorèmes de convergence des martingales discrètes (LGN et TCL du premier chapitre) combinés à la LGN pour les chaînes de Markov (pour la convergence p.s. de  $[M, M]_n/n$  comme dans le cas du temps continu) et au lemme de Slutsky entraînent d'une part la consistance de l'EMV et d'autre part sa normalité asymptotique au sens de la convergence en loi. Plus précisément, on a les convergences en loi des marginales

$$\sqrt{n} (\hat{\mu}_n - \tilde{\mu}) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1 - \tilde{\mu}^2) \quad \text{et} \quad \sqrt{n} (\hat{\sigma}_n^2 - \tilde{\sigma}^2) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 2\tilde{\sigma}^4).$$

### 3.3.2 Discrétisation par schéma d'Euler

Lorsque l'on observe la diffusion à des instants discrets  $ih_n$  où  $i \in \{1, \dots, n\}$ , l'étude précédente n'est valable que lorsque l'on est capable de déterminer la loi jointe de la chaîne

de Markov  $(X_{ih_n})_{i \in \{1, \dots, n\}}$ , ce qui n'est pas toujours le cas (les densités conditionnelles de  $X_{ih_n}$  sachant  $X_{(i-1)h_n}$  nous sont inconnues en dehors de certains cas particuliers comme ceux que nous venons d'étudier). On met alors en oeuvre un schéma d'Euler, c'est-à-dire une discrétisation de l'EDS du type

$$\begin{aligned} X_{t_{i+1}} &= X_{t_i} + \int_{t_i}^{t_{i+1}} b(r, X_r) dr + \int_{t_i}^{t_{i+1}} \sigma(r, X_r) dB_r \\ &\approx X_{t_i} + b(t_i, X_{t_i}) (t_{i+1} - t_i) + \sigma(t_i, X_{t_i}) (B_{t_{i+1}} - B_{t_i}). \end{aligned}$$

Plus précisément, si  $0 = t_0^n < t_1^n < \dots < t_{p_n}^n = T$  est une suite de subdivisions de l'intervalle  $[0, T]$ , de pas tendant vers 0 lorsque  $n \rightarrow +\infty$  (en général on prend la subdivision uniforme donnée par  $t_i^n = iT/n$  pour  $i \in \{0, 1, \dots, n\}$  et  $p_n = n$ ), on construit les variables  $\hat{X}_{t_i^n}^n$  par le schéma récursif suivant :  $\hat{X}_0^n = X_0$  et

$$\hat{X}_{t_i^n}^n = \hat{X}_{t_{i-1}^n}^n + b_\theta(t_{i-1}^n, \hat{X}_{t_{i-1}^n}^n) (t_i^n - t_{i-1}^n) + \sigma(t_{i-1}^n, \hat{X}_{t_{i-1}^n}^n) (B_{t_i^n} - B_{t_{i-1}^n}).$$

La suite  $(\hat{X}_{t_i^n}^n)_{i=0, \dots, p_n}$  est une chaîne de Markov dont les densités conditionnelles sont gaussiennes. En effet, sachant que  $\hat{X}_{t_i^n}^n = x_i$ , on a

$$\hat{X}_{t_i^n}^n = x_i + b_\theta(t_{i-1}^n, x_i) (t_i^n - t_{i-1}^n) + \sigma(t_{i-1}^n, x_i) (B_{t_i^n} - B_{t_{i-1}^n}),$$

qui suit une loi gaussienne de paramètres

$$\mathbb{E}[\hat{X}_{t_i^n}^n] = x_i + b_\theta(t_{i-1}^n, x_i) (t_i^n - t_{i-1}^n) \quad \text{et} \quad \text{Var}(\hat{X}_{t_i^n}^n) = \sigma^2(t_{i-1}^n, x_i) (t_i^n - t_{i-1}^n).$$

On obtient alors la densité de la loi jointe et donc la log-vraisemblance associée, qui diffère de la log-vraisemblance pour la discrétisation de la diffusion elle-même (qui n'est pas accessible) : on parle alors de "pseudo log-vraisemblance".

Néanmoins, le schéma d'Euler défini ci-dessus converge-t-il vers la diffusion ? Pour répondre positivement à cette question, il nous faut trouver une diffusion associée à cette discrétisation, qui va converger en un certain sens vers la vraie diffusion. On considère alors le processus (encore noté  $\hat{X}^n$ ) passant par les points  $(t_i^n, \hat{X}_{t_i^n}^n)_{i=0, \dots, p_n}$  en les reliant de manière brownienne :

$$\hat{X}_t^n = \hat{X}_{t_{i-1}^n}^n + b_\theta(t_{i-1}^n, \hat{X}_{t_{i-1}^n}^n) (t - t_{i-1}^n) + \sigma(t_{i-1}^n, \hat{X}_{t_{i-1}^n}^n) (B_t - B_{t_{i-1}^n}), \quad t \in [t_{i-1}^n, t_i^n].$$

Notons que le processus  $\hat{X}^n$  ne coïncide pas avec la diffusion originelle  $X$ . On a le théorème d'approximation suivant.

**Théorème 3.3.1.** *On suppose que les fonctions  $b_\theta$  et  $\sigma$  sont continues sur  $[0, T]$  et lipschitziennes en espace pour assurer l'existence et l'unicité de la solution  $X$  de l'EDS. On suppose de plus qu'elles sont  $\alpha$ -höldériennes en temps pour un  $\alpha \in ]0, 1[$ , c'est-à-dire qu'il existe  $K_T > 0$  telle que pour tout  $x \in \mathbb{R}$ ,*

$$|b(t, x) - b(s, x)| + |\sigma(t, x) - \sigma(s, x)| \leq K_T (1 + |x|) |t - s|^\alpha, \quad s, t \in [0, T].$$



Alors pour tout  $p \geq 1$ , il existe  $C_p > 0$  telle que pour tout  $n \in \mathbb{N}_*$ , on ait

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |\hat{X}_t^n - X_t|^{2p} \right] \leq \frac{C_p}{n^{2\beta p}},$$

où  $\beta$  est le minimum entre  $\alpha$  et  $1/2$ . En particulier, si la diffusion est homogène, on obtient

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |\hat{X}_t^n - X_t|^{2p} \right] \leq \frac{C_p}{n^p}.$$

Ce résultat est suffisamment fort pour nous permettre d'obtenir un résultat de convergence p.s. en utilisant le lemme de Borel-Cantelli. En effet, par l'inégalité de Chebyshev, on a pour tout  $\varepsilon > 0$  et tout  $\gamma \in ]0, \beta[$ ,

$$\mathbb{P} \left( n^\gamma \sup_{t \in [0, T]} |\hat{X}_t^n - X_t| > \varepsilon \right) \leq \frac{\mathbb{E} \left[ \sup_{t \in [0, T]} |\hat{X}_t^n - X_t|^{2p} \right]}{\varepsilon^{2p} n^{-2\gamma p}} \leq \frac{C_p}{\varepsilon^{2p} n^{2p(\beta - \gamma)}}.$$

Ainsi, la probabilité précédente est le terme général d'une série convergente dès lors que  $2p(\beta - \gamma) > 1$ . Le paramètre  $p \geq 1$  étant arbitraire, choisissons-le de sorte que cette condition soit satisfaite. Par le lemme de Borel-Cantelli, pour tout  $\omega$  en dehors d'un ensemble négligeable, il existe  $N(\omega) \in \mathbb{N}_*$  tel que pour tout  $n \geq N(\omega)$ , on ait

$$n^\gamma \sup_{t \in [0, T]} |\hat{X}_t^n(\omega) - X_t(\omega)| \leq \varepsilon.$$

Autrement dit, pour tout  $\gamma \in ]0, \beta[$ , on a la convergence p.s. uniforme sur  $[0, T]$  de  $\hat{X}$  vers  $X$  à vitesse  $n^\gamma$  :

$$\lim_{n \rightarrow +\infty} n^\gamma \sup_{t \in [0, T]} |\hat{X}_t^n - X_t| = 0.$$

Dans le cas homogène on obtient la même convergence pour tout  $\gamma \in ]0, 1/2[$ .

Pour terminer l'étude statistique, donnons la pseudo log-vraisemblance obtenue par la discrétisation via le schéma d'Euler dans le cadre du modèle CIR ergodique, pour lequel la log-vraisemblance associée à la discrétisation temporelle de la diffusion n'est pas calculable. Le schéma d'Euler considéré est le suivant :

$$X_{ih} = X_{(i-1)h} - (b X_{(i-1)h} - a) h + \sigma \sqrt{X_{(i-1)h}} (B_{ih} - B_{(i-1)h}), \quad X_0 = x > 0,$$

où le paramètre inconnu est

$$\theta := (a, b, \sigma^2) \in \Theta \subset \{(u, v, w) \in \mathbb{R}_* \times ]0, +\infty[ \times ]0, +\infty[ : 2u > w\}.$$

Notons que  $h$  ne doit pas dépendre de  $n$  car dans le cas contraire, le modèle est modifié à chaque fois que le nombre  $n$  d'observations change, ce qui n'a pas de sens du point de vue

statistique. Après avoir calculé la densité jointe de la chaîne de Markov  $(X_{ih})_{i \in \{1, \dots, n\}}$ , on en déduit la log-vraisemblance : pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$\log L_n(x_1, \dots, x_n, \theta) = - \sum_{i=1}^n \left( \frac{1}{2} \log (2\pi \sigma^2 x_{i-1} h) + \frac{1}{2\sigma^2 x_{i-1} h} (x_i - x_{i-1} + h(bx_{i-1} - a))^2 \right).$$

Après des calculs pénibles, on en tire alors l'EMV tridimensionnel de  $\theta$  ainsi que les mêmes conclusions (pour la consistance et la normalité asymptotique au sens de la convergence en loi) que celles obtenues dans le cas du modèle CIR à temps continu.