

Modèles aléatoires

Mohamed Boutahar
Aix-Marseille Université

6 septembre 2021

Table des matières

I	Calcul stochastique	5
1	Espérance conditionnelle	7
1.1	Définitions	7
1.1.1	Tribu	7
1.1.2	Mesurabilité	7
1.1.3	Probabilité conditionnelle	8
1.1.4	Espérance conditionnelle dans L^2	8
1.1.5	Espérance conditionnelle dans L^1	8
1.1.6	Cas des variables continues.	9
1.2	Propriétés de l'espérance conditionnelle	10
1.3	Théorème de Radon-Nikodym	11
1.4	Exercices	12
2	Processus stochastiques	13
2.1	Définitions	13
2.2	Martingales	15
2.3	Le mouvement Brownien	16
2.3.1	Le mouvement Brownien multidimensionnel	16
2.4	Processus de Levy	17
2.5	Exercices	18
3	Intégrale stochastique	21
3.1	Intégrale d'Itô	21
3.1.1	Cas de processus étagé	21
3.1.2	Cas général	22
3.2	Formule d'Itô	26
3.2.1	Formule d'Itô multidimensionnel	27
3.3	Intégration par partie	27

3.4	Exercices	28
4	Equations différentielles stochastiques	31
4.0.1	Equations différentielles homogènes	32
4.0.2	Solution faible	32
4.0.3	Propriété de Markov	33
4.0.4	Théorème de Girsanov	34
4.1	Quelques équations du monde de la finance	34
4.1.1	Modèle de Black & Scholes	34
4.1.2	Modèle de Vasicek	35
4.1.3	Modèle CIR	35
4.2	Exercices	35
II	Partie II : Séries temporelles	39
5	Discrétisation des équations différentielles	41
5.1	Discrétisation exacte	41
5.2	Discrétisation approchée	44
5.2.1	Schéma d'Euler	44
5.2.2	Schéma de Milstein	46
5.3	Convergence des Schémas de discrétisation	49
6	Estimation des équations différentielles	51
6.0.1	Estimation du Mouvement Brownien géométrique	51
6.0.2	Estimation du processus d'Orstein-Uhlenbeck	52
7	Processus autorégressif-moyenne mobile (ARMA)	55
7.1	Un exemple réel	66
8	Processus non stationnaires	79
8.1	les modèles ARIMA	79
8.1.1	Un exemple réel, modélisation des indices boursiers (bourse de Paris).	86
8.2	Les modèles SARIMA	88
8.2.1	Deux exemples réels	90

Première partie
Calcul stochastique

Chapitre 1

Espérance conditionnelle

1.1 Définitions

1.1.1 Tribu

Soit Ω un espace (ou ensemble), soit \mathcal{F} une famille de sous-ensembles de Ω .

DÉFINITION 1 On dit que \mathcal{F} est une **tribu** (ou **σ -algèbre**) sur Ω si les conditions suivantes sont satisfaites :

- i. \emptyset et Ω sont dans \mathcal{F} ,
- ii. Si A est dans \mathcal{F} , alors A^c est dans \mathcal{F} , où A^c est le complémentaire de A dans Ω .
- iii. Si $(A_n, n \geq 1)$ est dans \mathcal{F} , alors $\bigcup_{n=1}^{\infty} A_n$ est dans \mathcal{F} .

1.1.2 Mesurabilité

DÉFINITION 2 Une variable aléatoire (v.a.) réelle X est mesurable par rapport à une tribu \mathcal{F} si $X^{-1}(\mathcal{B}) = \{A, X(A) \in \mathcal{B}\} \subset \mathcal{F}$, où \mathcal{B} est la tribu borélienne contenant les ouverts de \mathbb{R} .

On dit aussi que X est \mathcal{F} -mesurable.

DÉFINITION 3 Soit Y une variable aléatoire $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$. La tribu engendrée par la v.a. Y est

$$\sigma(Y) = Y^{-1}(\mathcal{B}^n) = \{A \subset \Omega : \exists B \in \mathcal{B}^n, Y(A) = B\}$$

THÉORÈME 1 Soit Y un vecteur aléatoire à valeurs \mathbb{R}^p . La variable X est mesurable par rapport à $\sigma(Y)$ ssi il existe une fonction borélienne (i.e. mesurable) g sur \mathbb{R}^p telle que $X = g(Y)$.

1.1.3 Probabilité conditionnelle

DÉFINITION 4 Soit A et B deux événements de \mathcal{F} , avec $P(B) \neq 0$. La **probabilité conditionnelle** de A sachant B est définie par

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

1.1.4 Espérance conditionnelle dans L^2

Soit $L^2(\Omega, \mathcal{F}, P)$ l'espace de Hilbert (c'est à dire l'espace des variables aléatoires de carré intégrable) muni du produit scalaire :

$$\langle X, Y \rangle = E(XY).$$

DÉFINITION 5 Soit $X \in L^2(\Omega, \mathcal{F}, P)$. L'espérance conditionnelle de X sachant Y , noté $E(X | Y)$, est l'unique élément $\hat{X} \in L^2(\Omega, \sigma(Y), P)$ tel que

$$E(XZ) = E(\hat{X}Z) \quad \forall Z \in L^2(\Omega, \sigma(Y), P)$$

$L^2(\Omega, \sigma(Y), P)$ est le sous espace de $L^2(\Omega, \mathcal{F}, P)$ formé par les variables aléatoires de carré intégrable et $\sigma(Y)$ -mesurable.

DÉFINITION 6 Soit $X \in L^2(\Omega, \mathcal{F}, P)$ et \mathcal{G} une tribu. L'espérance conditionnelle de X sachant \mathcal{G} , noté $E(X | \mathcal{G})$, est l'unique élément $\hat{X} \in L^2(\Omega, \mathcal{G}, P)$, \mathcal{G} -mesurable et tel que

$$E(XZ) = E(\hat{X}Z) \quad \forall Z \in L^2(\Omega, \mathcal{G}, P)$$

1.1.5 Espérance conditionnelle dans L^1 .

DÉFINITION 7 Soit $X \in L^1(\Omega, \mathcal{F}, P)$ et \mathcal{G} une tribu. L'espérance conditionnelle de X sachant \mathcal{G} , noté $E(X | \mathcal{G})$, est l'unique élément $\hat{X} \in L^1(\Omega, \mathcal{G}, P)$ tel que

$$E(XZ) = E(\hat{X}Z) \quad \text{pour toute v.a. } Z \text{ bornée } \mathcal{G}\text{-mesurable.}$$

1.1.6 Cas des variables continues.

DÉFINITION 8 Soient X et Y deux vecteurs aléatoires à valeurs dans \mathbb{R}^p et \mathbb{R}^q tels que $Z = (X, Y)$ admette la densité $f_{X,Y}(x, y)$. La densité conditionnelle de X sachant $Y = y$ est définie par

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ 0 & \text{si } f_Y(y) = 0. \end{cases}, \quad (1.1)$$

où $f_Y(y) = \int_{\mathbb{R}^p} f_{X,Y}(x, y) dx$ est la densité marginale de Y .

Posons

$$h(y) = \int_{\mathbb{R}^p} g(x, y) f_{X|Y}(x | y) dx, \quad (1.2)$$

$h(y)$ est l'espérance conditionnelle de $g(X, Y)$ sachant que $Y = y$.
L'espérance conditionnelle de $g(X, Y)$ sachant Y est

$$E(g(X, Y) | Y) = h(Y).$$

Exemple . On suppose que (X, Y) suit une loi uniforme sur le domaine

$$D = \{(x, y) \in \mathbb{R}^2, 0 \leq x \leq 1, 0 \leq y \leq 1, y \geq x\}.$$

On a $f_{(X,Y)}(x, y) = K = 2$, car $\int_D f_{(X,Y)}(x, y) dx dy = 1$.
 $f_Y(y) = \int_0^y 2 dx = 2y$, $f_X(x) = \int_x^1 2 dy = 2(1 - x)$.

$$f_{X|Y}(x | y) = \begin{cases} \frac{1}{y} & \text{si } x \in [0, y] \\ 0 & \text{si } x \notin [0, y]. \end{cases} \quad (1.3)$$

$$h(y) = E(X | Y = y) = \int_0^y \frac{x}{y} dx = y/2.$$

On a $E(X | Y) = Y/2$, $E(X) = 1/3$;

$$E(X | Y = 1/2) = 1/4, \quad E(X | Y = 1) = 1/2.$$

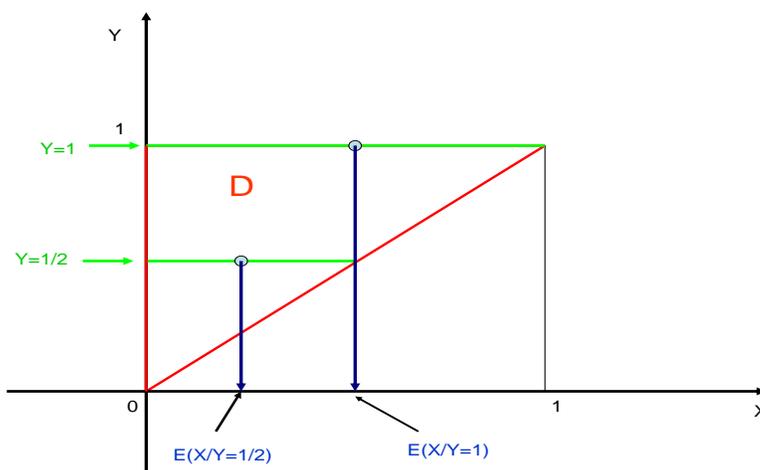


FIGURE 1.1 –

1.2 Propriétés de l'espérance conditionnelle

- 1) Linéarité. Soit a et b deux constantes. $E(aX + bY \mid \mathcal{G}) = aE(X \mid \mathcal{G}) + bE(Y \mid \mathcal{G})$.
- 2) Croissance. Soit X et Y deux v. a. telles que $X \leq Y$. Alors $E(X \mid \mathcal{G}) \leq E(Y \mid \mathcal{G})$.
- 3) Espérance imbriquée. $E[E(X \mid \mathcal{G})] = E(X)$.
- 4) Si X est \mathcal{G} -mesurable, $E(X \mid \mathcal{G}) = X$.
- 5) Si Y est \mathcal{G} -mesurable, $E(XY \mid \mathcal{G}) = YE(X \mid \mathcal{G})$.
- 6) Si X est indépendante de \mathcal{G} , $E(X \mid \mathcal{G}) = E(X)$.
- 7) Si \mathcal{G} et \mathcal{H} sont deux tribus telles que $\mathcal{G} \subset \mathcal{H}$ alors $E(X \mid \mathcal{G}) = E(E(X \mid \mathcal{H}) \mid \mathcal{G}) = E(E(X \mid \mathcal{G}) \mid \mathcal{H})$.
- 8) Si ϕ est une fonction convexe, alors $\phi(E(X \mid \mathcal{G})) \leq E(\phi(X) \mid \mathcal{G})$.

1.3 Théorème de Radon-Nikodym

DÉFINITION 9 Soit P_1 et P_2 deux mesures de probabilités. On dit que P_1 domine P_2 si P_2 est absolument continue par rapport à P_1 , noté $P_2 \ll P_1$, c'est à dire que pour tout $A \in \mathcal{F}$ tels que $P_1(A) = 0$, on a $P_2(A) = 0$,

THÉORÈME 2 Soit P une mesure de probabilité et μ une mesure finie. Si $P \ll \mu$, alors il existe une v.a. non négative f (presque sûrement unique) telle que pour tout $A \in \mathcal{F}$

$$P(A) = \int_A f(\omega) d\mu(\omega),$$

On note $f = \frac{dP}{d\mu}$ (f étant la densité de P par rapport à μ).

Exemple. P est la mesure de Gauss centrée réduite sur \mathbb{R} , c'est à dire que pour tout intervalle $]a, b[$

$$P(]a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt,$$

et μ est la mesure de Lebesgue, c'est à dire que $d\mu(t) = dt$.

THÉORÈME 3 (Autre version) Soient P et Q deux probabilités définies sur le même espace (Ω, \mathcal{F}) . Alors $P \ll Q$ si et seulement si il existe une variable $Z \geq 0$ \mathcal{F} -mesurable et d'espérance 1 sous Q telle que pour tout $A \in \mathcal{F}$,

$$P(A) = E_Q(Z \mathbf{1}_A) \tag{1.4}$$

$$= \int_A Z(\omega) dQ(\omega) \tag{1.5}$$

On appelle Z la densité de Radon-Nikodym de P par rapport à Q et on écrit $dP = Z dQ$ ou encore $Z = \frac{dP}{dQ}$.

$\mathbf{1}_A$ est la fonction indicatrice de A c'est à dire que

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

1.4 Exercices

Exercice 1. Montrer que si $X \in L^2$, $E(X | \mathcal{G}) = Y$ et $E(X^2 | \mathcal{G}) = Y^2$ alors $X = Y$ p.s.

Exercice 2. Soit X, Y deux v.a. telles que la v.a. $X - Y$ est indépendante de \mathcal{G} , d'espérance m et de variance σ^2 . On suppose que Y est \mathcal{G} -mesurable. Calculer $E(X - Y | \mathcal{G})$. En déduire $E(X | \mathcal{G})$. Calculer $E((X - Y)^2 | \mathcal{G})$. En déduire $E(X^2 | \mathcal{G})$.

Exercice 3. Soit $X = X_1 + X_2$. On suppose que X_1 est indépendante de \mathcal{G} , que X_2 est \mathcal{G} -mesurable et que X_1 est gaussienne.

1. Calculer $E(X | \mathcal{G})$ et $var(X | \mathcal{G})$.

2. Calculer $E(e^{\lambda X} | \mathcal{G})$.

Exercice 4 Soit Z_1, Z_2 deux variables aléatoires de carré intégrable. On définit

$$Cov(Z_1, Z_2 | \mathcal{G}) = E(Z_1 Z_2 | \mathcal{G}) - E(Z_1 | \mathcal{G})E(Z_2 | \mathcal{G}).$$

Montrer que

$$Cov(Z_1, Z_2 | \mathcal{G}) = E[(Z_1 - E(Z_1 | \mathcal{G}))Z_2 | \mathcal{G}].$$

Exercice 5 Soit \mathcal{G}_1 et \mathcal{G}_2 deux tribus indépendantes, $\mathcal{G} = \mathcal{G}_1 \vee \mathcal{G}_2$ et $(X_i, i = 1, 2)$ deux variables aléatoires bornées telles que X_i est \mathcal{G}_i -mesurable. Montrer que $E(X_1 X_2 | \mathcal{G}) = X_1 X_2$.

Exercice 6 Soit f et g deux densités strictement positives sur \mathbb{R} . Soit X une v.a. de densité f sur un espace (Ω, P) . Montrer qu'il existe une probabilité Q sur cet espace telle que X soit de densité g .

Chapitre 2

Processus stochastiques

2.1 Définitions

DÉFINITION 10 *Un processus stochastique est une famille de variables aléatoires $\{X_t, t \in \mathbb{T}\}$,. Dans le cas où $\mathbb{T} = \mathbb{R}_+$ ou \mathbb{R} on dit que le processus est à temps continu. Si $\mathbb{T} = \mathbb{N}$ ou \mathbb{Z} on parle alors de processus à temps discret. Dans ce dernier cas, on dit que X_t est une série temporelle ou série chronologique.*

Remarque. Un processus aléatoire modélise l'évolution au cours du temps d'une quantité aléatoire : température, indice boursier,..

DÉFINITION 11 *Soit (Ω, \mathcal{F}, P) un espace de probabilité et $\mathbb{T} = \mathbb{N}, \mathbb{Z}, \mathbb{R}_+$ ou \mathbb{R} On appelle filtration toute suite croissante de tribus $\{\mathcal{F}_t, t \in \mathbb{T}\}$ incluses dans \mathcal{F}*

DÉFINITION 12 *On appelle filtration naturelle du processus X_t la suite croissante de tribus*

$$\mathcal{F}_t = \sigma(X_s, s \leq t).$$

DÉFINITION 13 *$(\Omega, \{\mathcal{F}_t, t \in \mathbb{T}\}, \mathcal{F}, P)$ un espace de probabilité filtré si (Ω, \mathcal{F}, P) est un espace de probabilité et $\{\mathcal{F}_t, t \in \mathbb{T}\}$ est une filtration.*

Loi du processus

La loi du processus est la donnée des lois jointes de $(X_{t_1}, \dots, X_{t_k})$ pour toutes les suites finies d'instant t_1, \dots, t_k . Elle nous permet, par exemple de calculer

$$P(X_{t_1} \in B_1, \dots, X_{t_k} \in B_k), E(X_{t_1}), cov(X_{t_1}, X_{t_2}) \dots$$

Processus adapté

DÉFINITION 14 *On dit que le processus X_t est adapté à une filtration $(\mathcal{F}_t, t \in \mathbb{T})$ si $\forall t, X_t$ est \mathcal{F}_t -mesurable.*

Trajectoire du processus

Un processus stochastique peut être défini comme une fonction aléatoire qui à chaque $\omega \in \Omega$ associe la fonction, appelée aussi trajectoire,

$$t \longmapsto X_t(\omega)$$

Temps d'arrêt

Dans un jeu de hasard, un temps d'arrêt est un temps lors duquel le joueur décide d'arrêter de jouer, selon un critère ne dépendant que du passé et du présent. Il peut par exemple décider d'arrêter de jouer dès qu'il a gagné une certaine somme. Le joueur doit à tout moment pouvoir décider s'il arrête ou non.

DÉFINITION 15 *Soit $(X_t)_{t \in \mathbb{T}}$ un processus stochastique et $\mathcal{F}_t = \sigma(X_s, s \leq t)$ la filtration naturelle. Une variable aléatoire τ à valeurs dans $\mathbb{T} \cup \{+\infty\}$ est un temps d'arrêt si*

$$\forall t < \infty, \{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

Sa tribu associée est donnée par

$$\mathcal{F}_\tau = \{A \in \mathcal{F}, \forall t \in \mathbb{T}, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}.$$

$\{\tau \leq t\} \in \mathcal{F}_t$ signifie qu'à l'instant t , on sait si on s'est arrêté ou non. Mais on ne sait pas encore si on va s'arrêter avant une date ultérieure $t + s$: $\{\tau \leq t + s\} \notin \mathcal{F}_t$.

Exemple. Supposons que X_t soit réel, et soit B un intervalle de \mathbb{R} . Alors le temps de première atteinte de B

$$\tau_B = \inf\{t \in \mathbb{T} : X_t \in B\}$$

est un temps d'arrêt.

Processus arrêté .

Soit $(\mathcal{F}_t)_t$ une filtration, (X_t) un processus adapté à la filtration et τ un temps d'arrêt. On appelle processus arrêté en τ le processus $X_t^{(\tau)}$ défini par (qu'on note aussi $X_{\tau \wedge t}$) $X_t^{(\tau)} = X_{\tau \wedge t}$; avec $\tau \wedge t = \min(\tau, t)$. Si par exemple τ est le temps de première atteinte d'un ensemble B , alors $X_{\tau \wedge t}$ est le processus obtenu en gelant X à l'endroit où il atteint B pour la première fois.

2.2 Martingales

Le nom martingale est synonyme de jeu équitable, c'est-à-dire d'un jeu où le gain que l'on peut espérer faire en tout temps ultérieur est égal à la somme gagnée au moment présent.

DÉFINITION 16 Soit $X = (X_t)_{t \in \mathbb{T}}$ un processus adapté à la filtration (\mathcal{F}_t) et intégrable pour tout $t \in \mathbb{T}$.

1. X est une martingale si $E(X_t | \mathcal{F}_s) = X_s \forall s \leq t$, si $\mathbb{T} = \mathbb{N}$ l'égalité devient $E(X_{n+1} | \mathcal{F}_n) = X_n \forall n \in \mathbb{N}$.
2. X est une sur-martingale si $E(X_t | \mathcal{F}_s) \leq X_s \forall s \leq t$, ($E(X_{n+1} | \mathcal{F}_n) \leq X_n$ dans le cas discret).
3. X est une sous-martingale si $E(X_t | \mathcal{F}_s) \geq X_s \forall s \leq t$, ($E(X_{n+1} | \mathcal{F}_n) \geq X_n$ dans le cas discret)

DÉFINITION 17 Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus adapté à la filtration (\mathcal{F}_n) . On dit que le processus est prévisible si

$$X_0 = 0, X_n \text{ est } \mathcal{F}_{n-1} \text{ - mesurable.}$$

Exemple. Urne de Polya : On considère une urne contenant r boules rouges et v boules vertes. De manière répétée, on tire une boule de l'urne, puis on la remet en ajoutant un nombre fixé c de boules de la même couleur. Soit r_n le nombre de boules rouges après le nième tirage, v_n le nombre de boules vertes, et $X_n = r_n/(r_n + v_n)$ la proportion de boules rouges. On a

$$X_{n+1} = \begin{cases} \frac{r_n+c}{r_n+v_n+c} & \text{avec probabilité } \frac{r_n}{r_n+v_n} = X_n \\ \frac{r_n}{r_n+v_n+c} & \text{avec probabilité } \frac{v_n}{r_n+v_n} = 1 - X_n \end{cases}$$

$$E(X_{n+1}|X_n) = \frac{r_n+c}{(r_n+v_n+c)} \frac{r_n}{(r_n+v_n)} + \frac{r_n}{(r_n+v_n+c)} \frac{v_n}{(r_n+v_n)} = X_n.$$

La suite (X_n) est donc une martingale (par rapport à la filtration naturelle).

THÉORÈME 4 Soit $(X_n)_n$ une surmartingale et H_n un processus prévisible, non négatif et borné pour tout n . Alors $(H.X)_n = \sum_{k=1}^n H_k(X_k - X_{k-1})$ est une surmartingale.

THÉORÈME 5 Soit $(X_n)_n$ une martingale par rapport à \mathcal{F}_n , et soit ϕ une fonction convexe. Alors $\phi(X_n)$ est une sous-martingale par rapport à \mathcal{F}_n .

THÉORÈME 6 *Décomposition de Doob. Soit $X = (X_n)_n$ un processus adapté et intégrable. Alors X écrit d'une manière unique comme $X_n = M_n + A_n$, où M_n est une martingale et A_n est un processus prévisible avec $A_0 = 0$. Le processus A_n est non-décroissant (i.e $A_n \leq A_{n+1}$) ssi X est une sous-martingale.*

2.3 Le mouvement Brownien

DÉFINITION 18 *un processus stochastique $(X_t)_{t \in \mathbb{T}}$ est dit gaussien, si pour toute suite d'instants t_1, \dots, t_r dans \mathbb{T} le vecteur aléatoire $(X_{t_1}, \dots, X_{t_r})$ est gaussien.*

DÉFINITION 19 *Le mouvement Brownien standard ou processus de Wiener standard est le processus stochastique $(B_t)_{t > 0}$ satisfaisant :*

1. $B_0 = 0$ presque sûrement ;
2. Accroissements indépendants et stationnaires, pour tout $t > s > 0$, $B_t - B_s$ est indépendant de $(B_u)_{u \leq s}$;
3. Accroissements gaussiens : pour tout $t > s > 0$, $B_t - B_s$ suit une loi normale $N(0, t - s)$ de moyenne 0 et de variance $t - s$.

Remarque. X_t est dit stationnaire si pour toute suite d'instants $t_1, \dots, t_r \in \mathbb{T}$ et $h \in \mathbb{T}$, les vecteurs $(X_{t_1}, \dots, X_{t_r})$ et $(X_{t_1+h}, \dots, X_{t_r+h})$ ont la même loi.

2.3.1 Le mouvement Brownien multidimensionnel

DÉFINITION 20 *On dit que le processus $B_t = (B_t^1, \dots, B_t^d)$ est un mouvement Brownien d -dimensionnel si les processus B_t^1, \dots, B_t^d sont des mouvements Browniens standards indépendants.*

Continuité d'un processus stochastique

Soit I un intervalle de \mathbb{R}

DÉFINITION 21 *On dit que le processus stochastique $(X_t)_{t \in I}$ est continu à droite (resp. à gauche) si l'application $t \rightarrow X_t$ est continue à droite (resp. à gauche) presque sûrement, c'est à dire que pour tout $\omega \notin N$ avec $P(N) = 0$, $t \rightarrow X_t(\omega)$ est continue à droite (resp. à gauche)*

DÉFINITION 22 *On dit que le processus stochastique $(X_t)_{t \in I}$ est presque sûrement continu si l'application $t \rightarrow X_t$ est continue presque sûrement.*

DÉFINITION 23 *On dit que le processus stochastique $(X_t)_{t \in I}$ est continu en probabilité si*

$$\forall \epsilon > 0 \lim_{h \rightarrow 0} P(|X_{t+h} - X_t| > \epsilon) = 0.$$

DÉFINITION 24 *On dit que le processus stochastique $(X_t)_{t \in I}$ est continu en moyenne quadratique si*

$$\lim_{h \rightarrow 0} E(|X_{t+h} - X_t|^2) = 0.$$

DÉFINITION 25 *Soient X et Y deux processus stochastiques.*

On dit que X et Y sont des versions (on dit aussi modifications) l'un de l'autre si

$$P(X_t = Y_t) = 1, \quad \forall t \in I.$$

On dit que X et Y sont indistinguables si

$$P(\cap_{t \in I} \{X_t = Y_t\}) = 1.$$

THÉORÈME 7 *Soit X un processus stochastique. On suppose qu'il existe $\gamma, \delta, K > 0$ telles que*

$$E(|X_{t+h} - X_t|^\gamma) \leq Kh^{1+\delta} \quad \forall t, h,$$

alors X admet une version presque sûrement continue.

2.4 Processus de Levy

DÉFINITION 26 *Un processus stochastique X est un processus de Levy si*

1. $X_0 = 0$ presque sûrement;

2. Pour tout $n \in \mathbb{N}^*$, $0 \leq t_0 < \dots < t_n$, les variables aléatoires $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes.

3. Pour tout $t, h \geq 0$, on a

$$X_{t+h} - X_t \stackrel{\mathcal{L}}{=} X_h - X_0.$$

4. $\forall t \geq 0, \epsilon > 0$,

$$\lim_{s \rightarrow t} P(|X_s - X_t| > \epsilon) = 0.$$

5. X est un processus càdlàg c'est à dire que les trajectoires $(X_t(\omega), t \in \mathbb{R}^+)$ sont continues à droite et admettent une limite à gauche, pour tout $\omega \notin N$, $P(N) = 0$.

2.5 Exercices

Exercice 1. Démontrer les théorèmes 4 et 5.

Exercice 2 Soit X une v.a. intégrable et $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ est une filtration. Montrer que $(E(X | \mathcal{F}_t), t \geq 0)$ est une martingale.

Exercice 3. Surmartingale.

Montrer que si M est une martingale et A un processus croissant adapté ($A_s \leq A_t, \forall s \leq t$) alors $M - A$ est une surmartingale.

Exercice 4 Soit $(X_t, t \in [0, T])$ une martingale telle que $X_T = \zeta$. Exprimer X_t en fonction de ζ pour $t < T$.

Exercice 5 Soit $(M_t, t \geq 0)$ une \mathcal{F}_t -martingale de carré intégrable. Montrer que

1. $E((M_t - M_s)^2 | \mathcal{F}_s) = E(M_t^2 | \mathcal{F}_s) - M_s^2$ pour $t > s$.
2. $E((M_t - M_s)^2) = E(M_t^2) - E(M_s^2)$ pour $t > s$.
3. La fonction Φ définie par $\Phi(t) = E(M_t^2)$ est croissante.

Exercice 6 Soit X un processus stationnaire à accroissements indépendants, c'est-à-dire tel que, pour $t > s$, la v.a. $X_t - X_s$ est indépendante de $\sigma(X_u, u \leq s)$.

Montrer que si pour tout t la v.a. X_t est intégrable, X est une martingale et que si X est de carré intégrable, $X_t^2 - E(X_t^2)$ est une martingale.

Montrer que, si $e^{\lambda X_t}$ est intégrable, $M_t = e^{\lambda X_t} / E(e^{\lambda X_t})$ est une martingale.

Exercice 7 Tribu associée à un temps d'arrêt. Soit τ un temps d'arrêt. Montrer que \mathcal{F}_τ est une tribu.

Exercice 8 Soit τ un temps d'arrêt et ν une variable aléatoire \mathcal{F}_τ -mesurable, vérifiant $\nu \geq \tau$. Montrer que ν est un temps d'arrêt.

Exercice 9 Comparaison de tribus. Soient τ_1 et τ_2 deux temps d'arrêt tels que $\tau_1 \leq \tau_2$. Montrer que $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

Exercice 10 Montrer que le min (resp. le max) de deux temps d'arrêt est un temps d'arrêt.

Exercice 11 1. Calculer pour tout couple (s, t) les quantités $E(B_s B_t^2), E(B_t | \mathcal{F}_s), E(B_t | B_s)$ et $E(e^{\lambda B_t} | \mathcal{F}_s)$.

2 On admet que si $Z \rightsquigarrow N(0, \sigma^2)$ on a $E(Z^4) = 3\sigma^4$; calculer $E(B_t^2 B_s^2)$.

3. Quelle est la loi de $B_t + B_s$?

4. Soit θ_s une variable aléatoire bornée \mathcal{F}_s -mesurable. Calculer pour $t \geq s, E(\theta_s(B_t - B_s))$ et $E[\theta_s(B_t - B_s)^2]$.

Exercice 12 Soit B_t un mouvement Brownien.

1. Montrer que $B_t^2 - t$ est une martingale.

2. Soit $\gamma \in \mathbb{R}$, montrer que $\exp(\gamma B_t - \gamma^2 t / 2)$ est une martingale.

Exercice 13 Le pont Brownien. On définit le pont Brownien par

$$Z_t = B_t - tB_1, 0 \leq t \leq 1,$$

avec B_t un mouvement Brownien standard.

1. Montrer que Z est un processus Gaussien indépendant de B_1 . Préciser sa loi.
2. Montrer que le processus $\tilde{Z}_t = Z_{1-t}$ a même loi que Z_t .
3. Montrer que le processus Y défini par $Y_t = (1-t)B_{\frac{t}{1-t}}, 0 < t < 1$ a même loi que Z .

Chapitre 3

Intégrale stochastique

3.1 Intégrale d'Itô

3.1.1 Cas de processus étagé

On dit qu'un processus Θ est étagé (ou élémentaire) sur $[0, T]$ s'il existe une suite de réels $t_j; 0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ et une suite de variables aléatoires θ_j telles que θ_j soit \mathcal{F}_{t_j} -mesurable, appartienne à $L^2(\Omega)$ et que $\Theta_t = \theta_j$ pour tout $t \in [t_j, t_{j+1}[$, soit

$$\Theta_s(\omega) = \sum_{j=0}^{n-1} \theta_j(\omega) \mathbf{1}_{[t_j, t_{j+1}[}(s). \quad (3.1)$$

On définit alors

$$\int_0^T \Theta_s dB_s = \sum_{j=0}^{n-1} \theta_j (B_{t_{j+1}} - B_{t_j}).$$

On a

$$E \left(\int_0^T \Theta_s dB_s \right) = 0,$$

3.1.2 Cas général

Idée générale. Pour une fonction $f(t, \omega)$, $t \in [a, b]$, on définit la partition

$$t_j = t_j^{(n)} = \begin{cases} j/2^n & \text{si } a \leq j/2^n \leq b \\ a & \text{si } j/2^n < a \\ b & \text{si } j/2^n > b \end{cases} \quad (3.2)$$

La fonction $f(t, \omega)$ peut alors être approchée par

$$\sum_{j \geq 0} f(t_j^*, \omega) \mathbf{1}_{[t_j, t_{j+1}[}(t),$$

et l'intégrale $\int_a^b f(t, \omega) dB_t(\omega)$ est définie par

$$\int_a^b f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} f(t_j^*, \omega) (B_{t_{j+1}}(\omega) - B_{t_j}(\omega)).$$

Remarque. En général $t_j^* \in [t_j, t_{j+1}[$

1. Si $t_j^* = t_j$ on obtient alors l'intégrale d'Itô (c'est ce choix que nous allons considérer).
2. Si $t_j^* = (t_j + t_{j+1})/2$ on obtient alors l'intégrale de Stratonovich.
3. Le choix $t_j^* = t_{j+1}$ ne convient pas, car on obtient des intégrales non centrées. En effet

$$E \left(\sum_{j=0}^{n-1} B_{t_{j+1}} (B_{t_{j+1}} - B_{t_j}) \right) = b.$$

Soit $\mathcal{F}_t = \sigma(B_s, 0 \leq s \leq t)$ la tribu engendrée par le mouvement Brownien (information que nous possédons en observant la trajectoire du mouvement sur l'intervalle $[0, t]$).

On considère la classe $\mathbf{A}[a, b]$ des fonctions f définies sur $[a, b] \times \Omega$ vérifiant :

1. f est $\mathcal{B}([a, b]) \times \mathcal{F}$ -mesurable.
2. $f(t) = f(t, \cdot)$ est \mathcal{F}_t -mesurable.
3. $E \left(\int_a^b f^2(t) dt \right) < \infty$.

Remarque. Pour $t \in [a, b]$ fixé la variable aléatoire $f(t) = f(t, \cdot)$ est définie par $f(t)(\omega) = f(t, \omega)$.

Si $f(t, \omega)$ est étagée donnée par (3.1) alors $f \in \mathbf{A}(a, b)$ et donc l'intégrale d'Itô est donnée par

$$I(f) = \sum_{j=0}^{n-1} \theta_j (B_{t_{j+1}} - B_{t_j}).$$

THÉORÈME 8 *Si $f(t, \omega)$ est étagée et bornée alors*

$$E \left(\left(\int_a^b f(s) dB_s \right)^2 \right) = E \left(\int_a^b f^2(s) ds \right).$$

Pour une fonction quelconque $f \in \mathbf{A}(a, b)$, on peut trouver une suite de fonction élémentaires $g_n(t, \omega) = g_n(t)(\omega)$ (voir Annexe) telle que

$$\lim_{n \rightarrow \infty} E \left(\int_a^b (f - g_n)^2(t) dt \right) = 0.$$

On pose alors

$$I(f) = \int_a^b f(t) dB_t = \lim_{n \rightarrow +\infty} \int_a^b g_n(t) dB_t.$$

La limite est prise dans $L^2(\Omega)$.

Le théorème garantit la convergence dans $L^2(\Omega)$ de $\int_a^b g_n(t) dB_t$, (car c'est une suite de Cauchy et $L^2(\Omega)$ est complet).

THÉORÈME 9 *(Isométrie) Pour deux fonctions $f, g \in \mathbf{A}(a, b)$ on a*

$$E \left(\left(\int_a^b f(s) dB_s \right) \left(\int_a^b g(s) dB_s \right) \right) = E \left(\int_a^b f(s) g(s) ds \right).$$

En particulier pour $f \in \mathbf{A}(a, b)$ on a

$$E \left(\left(\int_a^b f(s) dB_s \right)^2 \right) = E \left(\int_a^b f^2(s) ds \right).$$

Quelques propriétés.

Soient α et β deux constantes f_1 et f_2 deux fonctions $\in \mathbf{A}(0, T)$. Alors

$$\alpha \int_0^T f_1(s)dB_s + \beta \int_0^T f_2(s)dB_s = \int_0^T (\alpha f_1(s) + \beta f_2(s))dB_s;$$

$$\int_0^T f_1(s)dB_s = \int_0^t f_1(s)dB_s + \int_t^T f_1(s)dB_s \text{ pour } t \in [0, T].$$

PROPOSITION 1 Pour tout $f \in \mathbf{A}(a, b)$ on a

$$E \left(\int_a^b f(s)dB_s \right) = 0$$

La variable aléatoire $\int_0^T f(s)dB_s$ est \mathcal{F}_T -mesurable

PROPOSITION 2 Pour tout $f \in \mathbf{A}(0, T)$ Alors $M_t = \int_0^t f(s)dB_s$ est une martingale par rapport à la filtration du mouvement Brownien \mathcal{F}_t .

THÉORÈME 10 Soit M_t une martingale par rapport à la filtration du mouvement Brownien \mathcal{F}_t de carré intégrable. Alors il existe une unique fonction $m \in \mathbf{A}(0, t)$ pour tout $t \geq 0$ telle que

$$M_t = E(M_0) + \int_0^t m(s)dB_s.$$

THÉORÈME 11 Soit $T > 0$, et $Z \in L^2(\mathcal{F}_T, P)$. Il existe un unique processus $z \in \mathbf{A}(0, T)$ tel que

$$Z = E(Z) + \int_0^T z(s)dB_s.$$

Exemple. Calcul de $\int_0^t B_s dB_s$.

Posons

$$g_n(s, \omega) = g_n(s)(\omega) = \sum_{j \geq 0} B_{t_j}(\omega) \mathbf{1}_{[t_j, t_{j+1}[}(s),$$

où les t_j sont données par (3.2) avec $a = 0$ et $b = t$

$$\begin{aligned}
E \left(\int_0^t (g_n(s) - B_s)^2 ds \right) &= \sum_{j \geq 0} E \left(\int_{t_j}^{t_{j+1}} (B_{t_j} - B_s)^2 ds \right) \\
&= \sum_{j \geq 0} \int_{t_j}^{t_{j+1}} (s - t_j)^2 ds \\
&= \frac{1}{2} \sum_{j \geq 0} (t_{j+1} - t_j)^2 \\
&\leq t/2^n,
\end{aligned}$$

donc

$$\begin{aligned}
\int_0^t B_s dB_s &= \lim_{n \rightarrow \infty} \int_0^t g_n(s) dB_s \\
&= \lim_{n \rightarrow \infty} \sum_{j \geq 0} B_{t_j} (B_{t_{j+1}} - B_{t_j}).
\end{aligned}$$

Or

$$\begin{aligned}
2 \sum_{j \geq 0} B_{t_j} (B_{t_{j+1}} - B_{t_j}) &= \sum_{j \geq 0} B_{t_{j+1}}^2 - B_{t_j}^2 - (B_{t_{j+1}} - B_{t_j})^2 \\
&= B_t^2 - \sum_{j \geq 0} (B_{t_{j+1}} - B_{t_j})^2
\end{aligned}$$

de plus

$$\sum_{j \geq 0} (B_{t_{j+1}} - B_{t_j})^2 \xrightarrow{L^2} t,$$

donc

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2}. \quad (3.3)$$

Si F est une fonction déterministe avec $F(0) = 0$. Comme $d(F^2(s))/ds = 2F(s)dF(s)/ds$ on déduit que l'intégrale de Riemann-Stieljes est donnée par

$$\int_0^t F(s) dF(s) = \frac{F^2(t)}{2}.$$

Dans le cas stochastique le terme supplémentaire $-t/2$ provient du fait que les trajectoires du mouvement Brownien sont irrégulières et nulle part différentiables.

3.2 Formule d'Itô

DÉFINITION 27 *Un processus X est un processus d'Itô si*

$$X_t = X_0 + \int_0^t \mu_s dB_s + \int_0^t \sigma_s dB_s,$$

que l'on écrit souvent

$$dX_t = \mu_t dt + \sigma_t dB_t. \quad (3.4)$$

Posons $\mathbf{A}_d = \cap_{T>0} \mathbf{A}_d(0, T)$ où $f \in \mathbf{A}_d(a, b)$ si

1. f est $\mathcal{B}(\mathbb{R}) \times \mathcal{F}$ -mesurable.
2. $f(t)$ est \mathcal{F}_t adaptée.
3. $P\left(\omega, \int_a^b f^2(t, \omega) dt < \infty\right) = 1$.

Le coefficient μ est la dérive, σ est le coefficient de diffusion, ils vérifient les conditions suivantes :

- C_1 . $\mu(t)$ est \mathcal{F}_t adaptée et $P\left(\cap_{t>0} \left\{ \int_0^t |\mu(s, \omega)| ds < \infty \right\}\right) = 1$.
- C_2 . $\sigma \in \mathbf{A}_d$ et $P\left(\cap_{t>0} \left\{ \int_0^t \sigma^2(s, \omega) ds < \infty \right\}\right) = 1$.

THÉORÈME 12 *Soit X le processus d'Itô donné par (3.4). Soit $f(t, x)$ une fonction de $C^2(]0, +\infty[\times \mathbb{R})$. dans \mathbb{R} . Alors le processus $Y_t = f(t, X_t)$ est aussi un processus d'Itô avec*

$$dY_t = \frac{\partial f}{\partial t}(t, X_t) dt + \frac{\partial f}{\partial x}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X_t) (dX_t) \cdot (dX_t) \quad (3.5)$$

avec la règle de multiplication

$$dt \cdot dt = dt \cdot dB_t = 0, dB_t \cdot dB_t = dt.$$

Exemple. Si on prends dans (3.4) $\mu_t = 0$ et $\sigma_t = 1$ c'est à dire que $X_t = B_t$, en appliquant la formule (3.5) avec $f(t, x) = x^2$ on obtient

$$d(B_t^2) = 2B_t dB_t + dt,$$

donc

$$B_t^2 = 2 \int_0^t B_s dB_s + t,$$

et obtient facilement la formule (3.3).

3.2.1 Formule d'Itô multidimensionnel

Soit $B_t = (B_t^{(1)}, \dots, B_t^{(d)})$ un mouvement Brownien d -dimensionnel.

DÉFINITION 28 *Un processus $X_t = (X_t^{(1)}, \dots, X_t^{(p)})$ est un processus d'Itô de dimension p si*

$$X_t = X_0 + \int_0^t U_s ds + \int_0^t \Sigma_s dB_s \quad (3.6)$$

Avec $U_t = (\mu_t^{(1)}, \dots, \mu_t^{(p)})$ un vecteur tel que chaque terme $\mu_t^{(j)}$ vérifie la condition C_1 ci-dessus.

$\Sigma_t = (\sigma_t(i, j))_{1 \leq i \leq p, 1 \leq j \leq d}$ est une matrice telle que chaque terme $\sigma_t(i, j)$ vérifie la condition C_2 ci-dessus.

THÉORÈME 13 *Soit X le processus d'Itô donné par (3.6). Soit $f(t, x) = (f_1(t, x), \dots, f_q(t, x))$ une fonction de $C^2([0, +\infty[\times \mathbb{R}^p)$ dans \mathbb{R}^q . Alors le processus $Y_t = f(t, X_t)$ est aussi un processus d'Itô avec pour tout $1 \leq k \leq q$,*

$$dY_t^{(k)} = \frac{\partial f_k}{\partial t}(t, X_t)dt + \sum_{i=1}^p \frac{\partial f_k}{\partial x_i}(t, X_t)dX_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^p \frac{\partial^2 f_k}{\partial x_i \partial x_j}(t, X_t)(dX_t^{(i)}).(dX_t^{(j)}) \quad (3.7)$$

avec la règle de multiplication

$$dt.dt = dt.dB_t^{(i)} = 0, dB_t^{(i)}.dB_t^{(j)} = \delta_i^j dt,$$

δ_i^j est le symbole de Kronecker $\delta_i^j = 1$ si $i = j$ et 0 sinon.

3.3 Intégration par partie

DÉFINITION 29 *Le crochet : Le crochet d'une martingale de carré intégrable continue M est l'unique processus croissant $\langle M \rangle$ tel que $M_t^2 - \langle M \rangle_t$ soit une martingale.*

Si X et Y sont deux martingales continues leur crochet $\langle X, Y \rangle$ est l'unique processus à variation bornée tel que $X_t Y_t - \langle X, Y \rangle_t$ soit une martingale.

Soient (X_t) et (Y_t) deux processus d'Itô. Alors pour tout $t \geq 0$, on a

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t. \quad (3.8)$$

qu'on écrit encore sous forme différentielle

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d \langle X, Y \rangle_t.$$

3.4 Exercices

Exercice 1 Posons

$$X_t = \int_0^t (\cos s) dB_s.$$

1. Montrer que X_t existe.
2. Montrer que $E(X_s X_t) = \frac{\min(s,t)}{2} + \frac{1}{4} \sin(2 \min(s, t))$.
3. Calculer $E[X_t | \mathcal{F}_s]$ pour $s \leq t$.
4. Montrer que $X_t = (\cos t) B_t + \int_0^t (\sin s) B_s ds$.

Exercice 2 Montrer que si f est une fonction déterministe de carré intégrable

$$E \left(B_t \int_0^t f(s) dB_s \right) = \int_0^t f(s) ds.$$

Exercice 3 Soit $t_1 < t_2$. Calculer $E \left(\int_{t_1}^{t_2} (B_t - B_{t_1}) dt \right)$. Montrer que

$$\int_{t_1}^{t_2} (B_t - B_{t_1}) dt = \int_{t_1}^{t_2} (t_2 - t) dB_t.$$

Calculer $Var \left(\int_{t_1}^{t_2} (B_t - B_{t_1}) dt \right)$.

Exercice 4 Montrer que les processus suivants sont des processus d'Itô

1. $X_t = B_t^2$,
2. $X_t = t e^{B_t}$,
3. $X_t = B_t^4 - 4t B_t$.

Exercice 5 On considère le processus

$$dX_t = \frac{X_t}{t-1} dt + dB_t; 0 \leq t < 1, X_0 = 0.$$

1. Montrer que

$$X_t = (1-t) \int_0^t \frac{dB_s}{1-s}, 0 \leq t < 1.$$

2. Montrer que $(X_t, t \geq 0)$ est un processus gaussien. Calculer son espérance et sa covariance.

Exercice 6 Soit $Y_t = \int_0^t e^s dB_s$ et $Z_t = \int_0^t Y_s dB_s$. Calculer $dZ_t, E(Z_t)$ et $var(Z_t)$

Exercice 7 Soient $f \in \mathbf{A}(0, t)$ et

$$X_t = \int_0^t f(s) dB_s - \frac{1}{2} \int_0^t f^2(s) ds.$$

Posons $Y_t = e^{X_t}$ et $Z_t = e^{-X_t}$.

1. Calculer dY_t et dZ_t .

2. En supposant que $f(t) = f$ une constante déterministe montrer que $E(Y_t) = 1$.

Exercice 8 Soit

$$Z_t = e^{(\alpha - \gamma^2/2)t + \gamma B_t} \left(\delta + \beta \int_0^t e^{-(\alpha - \gamma^2/2)s - \gamma B_s} ds \right),$$

α, β, γ et δ sont des constantes. Calculer dZ_t .

Exercice 9 Soit $Y_t = tX_t^{(1)}X_t^{(2)}$ avec

$$\begin{aligned} dX_t^{(1)} &= f_1(t)dt + \sigma_1(t)dB_t \\ dX_t^{(2)} &= \sigma_2(t)dB_t \end{aligned}$$

Calculer dY_t .

Exercice 10 Soient $B_t^{(1)}$ et $B_t^{(2)}$ deux mouvements Browniens standards et indépendants. Posons $V_t = \left(B_t^{(1)}\right)^2 + \left(B_t^{(2)}\right)^2$, calculer dV_t .

Montrer que l'on peut écrire $W_t = \sqrt{V_t}$ sous la forme

$$dW_t = \frac{B_t^{(1)}dB_t^{(1)} + B_t^{(2)}dB_t^{(2)}}{2W_t}.$$

Exercice 11 Soient les deux processus $X^{(1)}$ et $X^{(2)}$ définis par

$$dX_t^{(i)} = X_t^{(i)} \left(\mu dt + \sigma_i dB_t^{(i)} \right), i = 1, 2,$$

où $B_t^{(1)}$ et $B_t^{(2)}$ deux mouvements Browniens standards et indépendants, μ et σ_i sont supposées déterministes.

Posons $Y_t = (X_t^{(1)} + X_t^{(2)})/2$ et $Z_t = \sqrt{X_t^{(1)} X_t^{(2)}}$. Calculer dY_t .
Montrer que

$$dZ_t = Z_t \left\{ \left(\mu - \frac{\sigma_1^2 + \sigma_2^2}{2} \right) dt + \frac{\sigma_1}{2} dB_t^{(1)} + \frac{\sigma_2}{2} dB_t^{(2)} \right\}.$$

Chapitre 4

Equations différentielles stochastiques

DÉFINITION 30 Une équation différentielle stochastique est une équation de la forme

$$X_t = x + \int_0^t \mu(s, X_s) ds + \int_0^t \Sigma(s, X_s) dB_s, \quad (4.1)$$

ou sous forme simplifiée

$$dX_t = \mu(t, X_t) dt + \Sigma(t, X_t) dB_t, X_0 = x; \quad (4.2)$$

μ et B sont des vecteurs de dimension n et m respectivement, Σ une matrice.

THÉORÈME 14 On suppose qu'il existe deux constantes $C_1, C_2 > 0$ telles que
1. Pour tout $t \in [0, T]$, on a

$$\|\mu(t, x)\|^2 + \|\Sigma(t, x)\|^2 \leq C_1 (1 + \|x\|^2).$$

2. les deux fonctions $\mu(t, \cdot)$ et $\Sigma(t, \cdot)$ sont lipschitziennes

$$\|\mu(t, x) - \mu(t, y)\| + \|\Sigma(t, x) - \Sigma(t, y)\| \leq C_2 \|x - y\|.$$

Soit Z un v.a indépendant de B_t et de carré intégrable, alors l'équation différentielle

$$dX_t = \mu(t, X_t) dt + \Sigma(t, X_t) dB_t, X_0 = Z$$

admet une solution unique X_t qui est $\sigma(Z) \vee \mathcal{F}_t$ -mesurable, ayant des trajectoires continues et telle que

$$E \left(\int_0^T \|X_s\|^2 ds \right) < \infty.$$

4.0.1 Equations différentielles homogènes

On suppose que $\dim(X_t) = m = 1, \mu(t, x) = \mu(x), \sigma(t, x) = \sigma(x)$.

THÉORÈME 15 Soient $(B_t, t \in \mathbb{R}^+)$ un mouvement Brownien standard de filtration naturelle $(\mathcal{F}_t, t \in \mathbb{R}^+)$, $x \in \mathbb{R}$ et μ et σ sont lipschitziennes. Alors il existe un unique processus $(X_t, t \in \mathbb{R}^+)$ continu et adapté à $(\mathcal{F}_t, t \in \mathbb{R}^+)$ tel que

$$X_t = x + \int_0^t \mu(X_s) ds + \int_0^t \sigma(X_s) dB_s \text{ p.s} \quad (4.3)$$

ou encore

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t, X_0 = x. \quad (4.4)$$

De plus,

$$E(\sup_{0 \leq t \leq T} X_t^2) < \infty \text{ pour tout } T > 0.$$

Remarque - La solution (X_t) donnée par (4.3) ci-dessus est appelée une solution forte de l'équation (4.4) car le mouvement Brownien est fixé à l'avance et on cherche la solution dans l'espace filtré $L^2(\Omega, \mathcal{F}_t, P)$.

4.0.2 Solution faible

Soient $x_0 \in \mathbb{R}$ et $\mu : \mathbb{R} \rightarrow \mathbb{R}, \sigma : \mathbb{R} \rightarrow \mathbb{R}^+$. On considère l'équation différentielle stochastique

$$dX_t = \mu(X_t) dt + \sigma(X_t) dB_t, X_0 = x_0. \quad (4.5)$$

DÉFINITION 31 Une solution faible de l'équation (4.4) est un processus continu (X_t) tel que les processus (M_t) et (N_t) définis par

$$M_t = X_t - X_0 - \int_0^t \mu(X_s) ds \text{ et } N_t = M_t^2 - \int_0^t \sigma(X_s)^2 ds$$

soient des martingales.

Le mouvement brownien standard (B_t) ne figure plus dans la solution faible. Ainsi, une solution faible d'une équation différentielle stochastique est une solution en loi.

THÉORÈME 16 *Supposons μ, σ continues, g bornée et supposons que l'équation (4.5) admette une solution forte (X_t) . Alors (X_t) est une solution faible de (4.5).*

Remarque. Une équation différentielle stochastique peut admettre une solution faible, mais pas de solution forte; il existe donc plus souvent une solution faible qu'une solution forte.

Exemple. La solution faible de l'équation différentielle stochastique

$$dX_t = \text{sgn}(X_t)dB_t, \quad X_0 = 0 \quad (4.6)$$

(où $\text{sgn}(x)$ est la fonction signe c'est à dire que $\text{sgn}(x) = 1$ si $x \geq 0$ et -1 sinon) est un processus continu (X_t) tel que X_t et $(X_t^2 - \int_0^t 1ds = X_t^2 - t)$ sont des martingales. Ainsi la solution faible de l'équation ci-dessus est donc un mouvement Brownien standard B'_t différent du mouvement Brownien standard (B_t) (car B_t ne peut pas être une solution de (4.6)).

4.0.3 Propriété de Markov

On note $(X_t^{s,x}, t \geq s)$ la solution de (4.5) partant de x à l'instant s , soit

$$X_t^{s,x} = x + \int_s^t \mu(u, X_u^{s,x})du + \int_s^t \Sigma(u, X_u^{s,x})dB_u.$$

Sous les conditions du théorème 14 on peut montrer que

$$X_t^{0,x} = X_t^{s, X_s^{0,x}},$$

ce qui montre que la solution de (4.5) est un processus de Markov par rapport à la filtration \mathcal{F}_t :

$$E(f(X_t) | \mathcal{F}_s) = E(f(X_t) | X_s) = g(t, s, X_s),$$

où $g(t, s, x) = E(f(X_t^{s,x}))$, $t \geq s$.

En particulier si

$$X_t^{s,x} = x + \int_s^t \mu(X_u^{s,x})du + \int_s^t \Sigma(X_u^{s,x})dB_u$$

on obtient un processus de Markov homogène et on a

$$E(f(X_t) | \mathcal{F}_s) = E(f(X_t) | X_s) = g(t, s, X_s) = h(t - s, X_s),$$

avec $g(t, s, x) = E(f(X_t^{s,x})) = E(f(X_{t-s}^{0,x}))$ et $h(u, x) = E(f(X_u^{0,x}))$, $t \geq s$.

4.0.4 Théorème de Girsanov

THÉORÈME 17 Soit $(B_t; t \geq 0)$ un mouvement Brownien sur un espace $(\Omega, \mathcal{F}; P)$ et (\mathcal{F}_t) sa filtration naturelle. Soit

$$L_t = \exp \left(\int_0^t \mu_s dB_s - \frac{1}{2} \int_0^t \mu_s^2 ds \right), \quad t \leq T.$$

On suppose que L_t est une martingale par rapport à la filtration \mathcal{F}_t sous la probabilité P .

Soit la mesure Q sur la tribu \mathcal{F}_T

$$dQ(\omega) = L_T(\omega) dP(\omega).$$

Alors Q est une probabilité sur \mathcal{F}_T et $B_t = \tilde{B}_t + \int_0^t \mu_s ds$ où \tilde{B}_t est un mouvement Brownien sous la probabilité Q .

Q est la transformée de Girsanov de P .

Remarques.1. Sous la condition de Novikov $E_P \left(\exp \frac{1}{2} \int_0^T \mu_s^2 ds \right) < 1$, L_T est une variable positive d'espérance 1 sous P et L est une P -martingale.

Si L n'est pas d'espérance 1, L est une surmartingale d'espérance strictement plus petite que 1.

2. Pour une application du théorème de Girsanov en finance voir A. Conze (2012).

4.1 Quelques équations du monde de la finance

4.1.1 Modèle de Black & Scholes

DÉFINITION 32 Un mouvement Brownien géométrique est donné par

$$X_t = \exp \left((\mu - \sigma^2/2)t + \sigma B_t \right), \quad (4.7)$$

μ et σ des constantes, B_t est un mouvement Brownien standard.

En appliquant la formule d'Itô à la fonction $f(t, x) = \exp \left((\mu - \sigma^2/2)t + \sigma x \right)$, on obtient

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \quad X_0 = 1; \quad (4.8)$$

c'est l'équation de Black Scholes où X_t est le prix de l'action, μ le taux de rendement et σ la volatilité de l'action.

4.1.2 Modèle de Vasicek

Le taux d'intérêt court instantané r_t (voir Vasicek (1977)) suit un processus d'Ornstein-Uhlenbeck à coefficients constants :

$$dr_t = a[b - r_t]dt + \sigma dB_t, r_0 = x$$

avec a, b, σ et x des constantes positives ; la solution est donnée par

$$r_t = r_0 e^{-at} + b(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dB_s.$$

Démonstration. On applique la formule d'Itô à $f(t, x) = x e^{at}$.

4.1.3 Modèle CIR

Le modèle (CIR) a été proposé par Cox, Ingersoll et Ross (1985) pour modéliser le taux court instantané r_t , ils supposent que celui-ci est solution de l'équation différentielle stochastique :

$$dr_t = a[b - r_t]dt + \sigma \sqrt{r_t} dB_t, r_0 = x,$$

avec la condition

$$\sigma^2 < 2ab.$$

Remarque. Aucune solution explicite n'est connue pour cette équation, mais on peut montrer que r_t suit une loi de Khi-deux décentrée.

4.2 Exercices

Exercice 1 Soit B_t un mouvement Brownien. Posons $Y_t = B_t^k$.

1. Montrer que Y_t vérifie l'équation différentielle suivante

$$dY_t = k B_t^{k-1} dB_t + \frac{1}{2} k(k-1) B_t^{k-2} dt.$$

2. Posons $m_k(t) = E(B_t^k)$, montrer que

$$m_k(t) = \frac{1}{2} k(k-1) \int_0^t m_{k-2}(s) ds,$$

calculer $m_k(t)$ pour $k = 2, 4, 6$.

Exercice 2 Soit l'équation différentielle

$$dX_t = \mu X_t dt + dB_t, X_0 = x.$$

1. Posons $Y_t = e^{-\mu t} X_t$, écrire l'équation différentielle associée à Y_t .
2. Montrer que la solution de cette dernière est donnée par

$$Y_t = y + \int_0^t e^{-\mu s} dB_s, y = x.$$

3. Calculer $E(Y_t), E(Y_t^2)$.

Exercice 3 (Mouvement Brownien géométrique)

1. Montrer que la solution Y_t de l'équation

$$dX_t = \mu X_t dt + \sigma X_t dB_t, Y_0 = 1,$$

est donnée par

$$X_t = \exp((\mu - \sigma^2/2)t + \sigma B_t).$$

2. Montrer que si $\mu \geq 0$ alors X_t est une sous-martingale par rapport à la filtration (\mathcal{F}_t) .

Exercice 4 Soit l'équation différentielle stochastique

$$dX_t = (\mu_1 + \mu_2 X_t)dt + (\sigma_1 + \sigma_2 X_t)dB_t, X_0 = x, t \in [0, T]. \quad (4.9)$$

1. Montrer que (4.9) admet une unique solution.
2. Soit $m_k(t) = E(X_t^k)$ le moment d'ordre k de X_t .
Montrer que $m_1(t)$ est solution de l'équation différentielle déterministe

$$\frac{dy(t)}{dt} - \mu_2 y(t) = \mu_1, y(0) = x. \quad (4.10)$$

3. Ecrire l'équation différentielle stochastique associée à $Y_t = X_t^2$; en déduire que $m_2(t)$ est solution de l'équation différentielle

$$\frac{dy(t)}{dt} - (2\mu_1 + \sigma_2^2)y(t) = 2(\mu_1 + \sigma_1\sigma_2)m_1(t) + \sigma_1^2, y(0) = x^2. \quad (4.11)$$

2. Résoudre les deux équations (4.10) et (4.11).

Exercice 5 (Modèle de Black et Scholes avec effet hétéroscédastique). Soit l'équation différentielle stochastique

$$dS_t = \mu S_t dt + \sigma(t) S_t dB_t, S_0 = x,$$

On suppose que la volatilité σ est une fonction déterministe.

1. Montrer que

$$S_t = x \exp \left(\mu t + \int_0^t \sigma(s) dB_s - \frac{1}{2} \int_0^t \sigma^2(s) ds \right).$$

2. Calculer $E(S_t)$.

Exercice 6 (Modèle Cox, Ingersoll et Ross (1985) pour le taux court instantané r_t). On suppose que r_t est solution de l'équation différentielle stochastique :

$$dr_t = a[b - r_t]dt + \sigma\sqrt{r_t}dB_t, r_0 = x,$$

avec la condition

$$\sigma^2 < 2ab.$$

Calculer $E(r_t)$ et $var(r_t)$.

Exercice 7 On suppose que le processus (X_t) est à valeurs dans $]0, 1[$ et solution de l'équation différentielle stochastique

$$dX_t = -\sigma^2 X_t^2(1 - X_t)dt + \sigma X_t(1 - X_t)dB_t, X_0 = x.$$

Montrer que

$$X_t = \frac{x \exp(\sigma B_t - \sigma^2 t/2)}{x \exp(\sigma B_t - \sigma^2 t/2) + 1 - x}.$$

Exercice 8 (Girsanov) Soit μ une fonction déterministe et L_t le processus solution de l'équation

$$dL_t = -L_t \mu_t dB_t, L_0 = 1.$$

1. Montrer que

$$L_t = \exp \left(- \int_0^t \mu_s dB_s - \frac{1}{2} \int_0^t \mu_s^2 ds \right), t \leq T.$$

38 CHAPITRE 4. EQUATIONS DIFFÉRENTIELLES STOCHASTIQUES

2. Montrer que L_t est une martingale par rapport à la filtration \mathcal{F}_t sous la probabilité P .

Soit la mesure Q sur la tribu \mathcal{F}_T

$$dQ(\omega) = L_T(\omega)dP(\omega).$$

3. En déduire que Q est une probabilité sur \mathcal{F}_T et $\tilde{B}_t = B_t + \int_0^t \mu_s ds$ est un mouvement Brownien sous la probabilité Q .

4. Montrer que

$$E_P(L_T \log L_T) = \frac{1}{2} \int_0^T \mu_s^2 ds.$$

Deuxième partie

Partie II : Séries temporelles

Chapitre 5

Discrétisation des équations différentielles

On considère l'équation différentielle stochastique

$$dX_t = \mu(t, X_t)dt + \Sigma(t, X_t)dB_t, X_0 = x; \quad (5.1)$$

μ et B sont des vecteurs dimension n et m respectivement, Σ une matrice. On notera Δ le pas de discrétisation, c'est à dire le temps qui s'écoule entre deux instants où on observe le processus. $[0, T]$ est l'intervalle dans lequel les trajectoires du processus vont être observées.

Posons

$$t_k = \frac{k\Delta}{T}, k = 1, \dots, N = \left\lceil \frac{T}{\Delta} \right\rceil$$

les instants où le processus est observés.

Si on pose

$$X_k = X(t_k), k = 1, \dots, N.$$

La suite des valeurs X_1, \dots, X_N est une série **temporelle**.

5.1 Discrétisation exacte

DÉFINITION 33 *Un processus \hat{X} est une discrétisation exacte du processus X si*

$$X(t_k) \stackrel{\text{loi}}{=} \hat{X}(t_k)$$

pour tout $k = 1, \dots, N = \left\lceil \frac{T}{\Delta} \right\rceil$.

PROPOSITION 3 *Si l'équation différentielle stochastique a une solution explicite alors le processus X admet une discrétisation exacte.*

Les deux modèles de Black and Scholes et Vasicek possèdent une discrétisation exacte :

1. Modèle de Black Scholes

$$X(t_k) = \exp((\mu - \sigma^2/2)t_k + \sigma B_{t_k}).$$

On obtient

$$\begin{aligned} X(t + \Delta) &= \exp((\mu - \sigma^2/2)(t + \Delta) + \sigma B_{t+\Delta}) \\ &= X(t) \exp((\mu - \sigma^2/2)\Delta + \sigma(B_{t+\Delta} - B_t)) \\ &\stackrel{\text{loi}}{=} X(t) \exp((\mu - \sigma^2/2)\Delta + \sigma(B_\Delta)) \\ &\stackrel{\text{loi}}{=} X(t) \exp((\mu - \sigma^2/2)\Delta + \sigma\sqrt{\Delta}Z_t) \end{aligned}$$

avec $Z_t \rightsquigarrow N(0, 1)$ variable Gaussienne centrée réduite.

Remarque Z_t dépend de t vient du fait qu'on applique l'approximation de Taylor sur l'intervalle $[t, t + \Delta]$.

Donc une discétisation exacte du mouvement Brownien géométrique est donnée par

$$\hat{X}(t + \Delta) = \hat{X}(t) \exp((\mu - \sigma^2/2)\Delta + \sigma\sqrt{\Delta}Z_t), \hat{X}(0) = 1. \quad (5.2)$$

2. Modèle de Vasicek

Le taux d'intérêt court instantané r_t (voir Vasicek (1977)) suit un processus d'Orstein-Uhlenbeck à coefficients constants :

$$dr_t = a[b - r_t]dt + \sigma dB_t, r_0 = x$$

dont la solution est donnée par

$$r_t = r_0 e^{-at} + b(1 - e^{-at}) + \sigma e^{-at} \int_0^t e^{as} dB_s$$

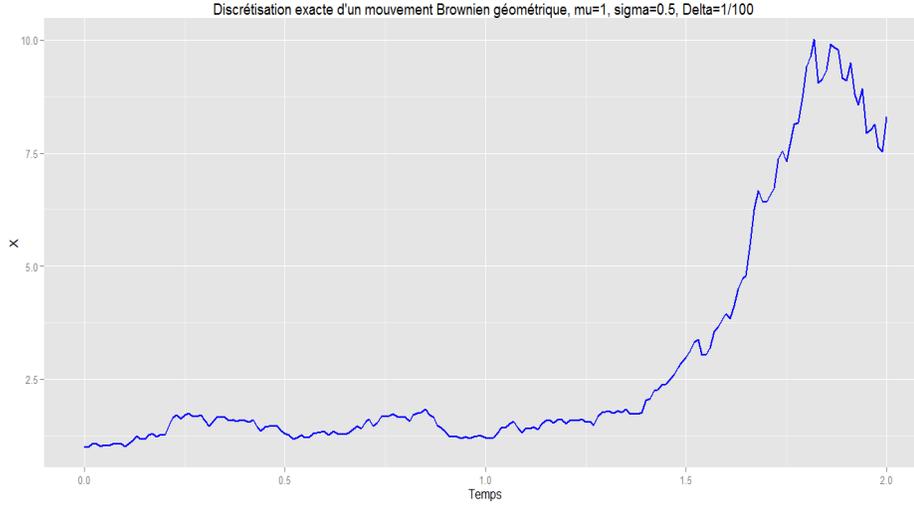


FIGURE 5.1 –

$$\begin{aligned}
 r_{t+\Delta} &= r_0 e^{-a(t+\Delta)} + b(1 - e^{-a(t+\Delta)}) + \sigma e^{-a(t+\Delta)} \int_0^{t+\Delta} e^{as} dB_s \\
 &= r_0 e^{-a(t+\Delta)} + b(1 - e^{-a(t+\Delta)}) + \sigma e^{-a(t+\Delta)} \left(\int_0^t e^{as} dB_s + \int_t^{t+\Delta} e^{as} dB_s \right) \\
 &= r_t e^{-a\Delta} + b(1 - e^{-a\Delta}) + \sigma e^{-a(t+\Delta)} \int_t^{t+\Delta} e^{as} dB_s \\
 &\stackrel{\text{loi}}{=} r_t e^{-a\Delta} + b(1 - e^{-a\Delta}) + \sigma \sqrt{\frac{1 - e^{-2a\Delta}}{2a}} Z_t,
 \end{aligned}$$

avec $Z_t \rightsquigarrow N(0, 1)$ car

$$\int_t^{t+\Delta} e^{as} dB_s \rightsquigarrow N(0, v), v = \int_t^{t+\Delta} e^{2as} ds.$$

Donc une discrétisation exacte du processus d'Orstein-Uhlenbeck est donnée par

$$\hat{r}_{t+\Delta} = \hat{r}_t e^{-a\Delta} + b(1 - e^{-a\Delta}) + \sigma \sqrt{\frac{1 - e^{-2a\Delta}}{2a}} Z_t. \quad (5.3)$$

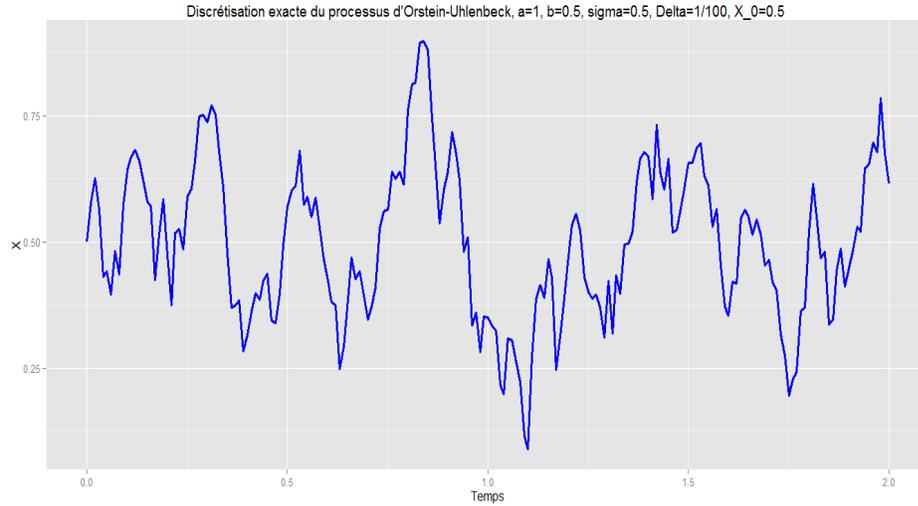


FIGURE 5.2 –

5.2 Discrétisation approchée

Lorsque la solution de l'équation différentielle n'est pas explicite, on recourt souvent à des approximations discrètes du processus continu. On utilise alors des schémas basés sur des développements d'Itô-Taylor de l'équation (5.1) à des ordres différents.

5.2.1 Schéma d'Euler

On approxime la solution aux points de la subdivision par

$$\hat{X}(0) = x,$$

$$\hat{X}(t_k) = \hat{X}(t_{k-1}) + \Delta \mu(t_{k-1}, \hat{X}(t_{k-1})) + \Sigma(t_{k-1}, \hat{X}(t_{k-1})) (B_{t_k} - B_{t_{k-1}}).$$

Ceci vient du fait que sur l'intervalle $[t_{k-1}, t_k]$ on utilise les approximations de Taylor au premier ordre suivantes

$$\mu(t, \hat{X}(t)) = \mu(t_{k-1}, \hat{X}(t_{k-1})) + R_1(t_{k-1}, t_k, \hat{X}(t_{k-1})),$$

$$\Sigma(t, \hat{X}(t)) = \Sigma(t_{k-1}, \hat{X}(t_{k-1})) + R_2(t_{k-1}, t_k, \hat{X}(t_{k-1})),$$

et on néglige les restes R_1 et R_2 .

Dans le cas scalaire ($m = n = 1$) le schéma devient :

$$\hat{X}(t + \Delta) = \hat{X}(t) + \Delta\mu \left(t, \hat{X}(t) \right) + \sigma \left(t, \hat{X}(t) \right) \sqrt{\Delta} Z_t.$$

$Z_t \rightsquigarrow N(0, 1)$.

Si on applique le schéma d'Euler au mouvement Brownien géométrique solution de (4.8) on obtient

$$\tilde{X}(t + \Delta) = \tilde{X}(t) \left(1 + \mu\Delta + \sigma\sqrt{\Delta}Z_t \right), \quad \tilde{X}(0) = 1. \quad (5.4)$$

Ce processus est à comparer à \hat{X} donné par (5.2) obtenu par la discrétisation exacte.

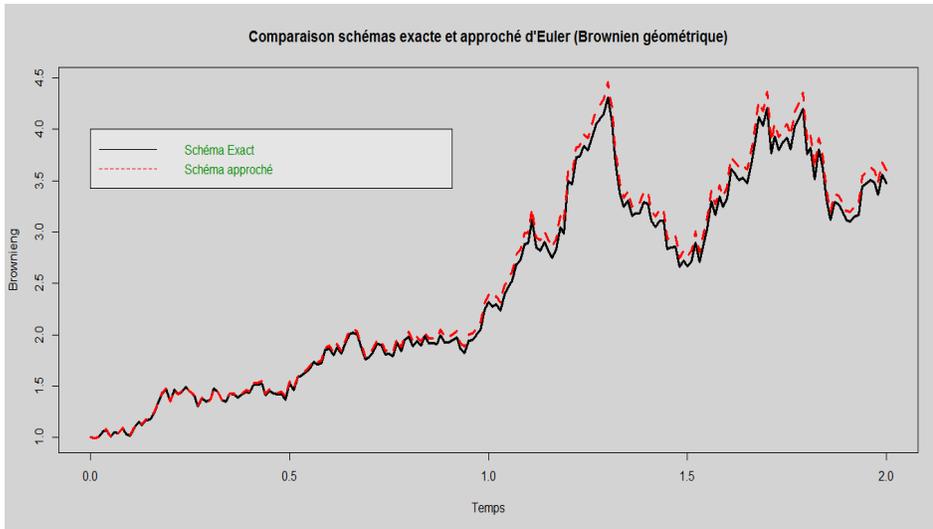


FIGURE 5.3 –

Dans le cas du processus d'Orstein-Uhlenbeck on obtient la discrétisation

$$\tilde{r}_{t+\Delta} = ab\Delta + (1 - a\Delta)\tilde{r}_t + \sigma\sqrt{\Delta}Z_t, r_0 = x$$

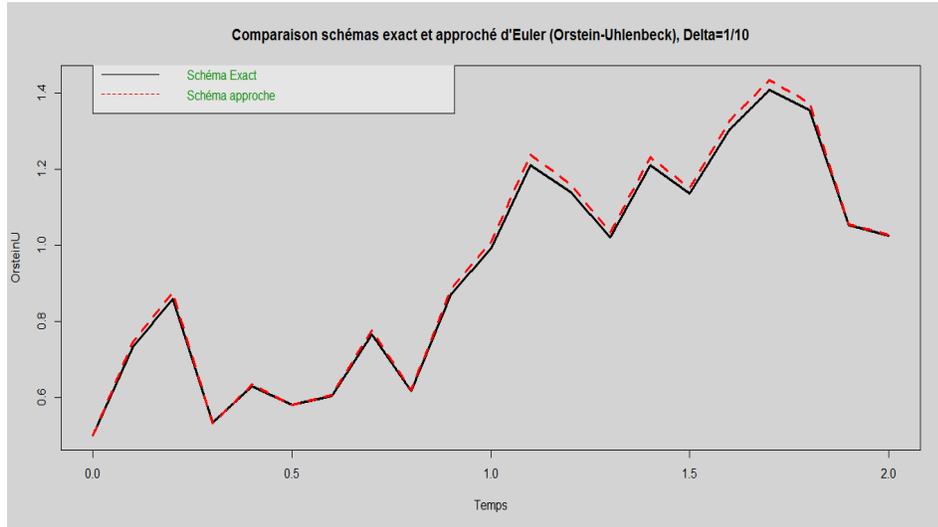


FIGURE 5.4 –

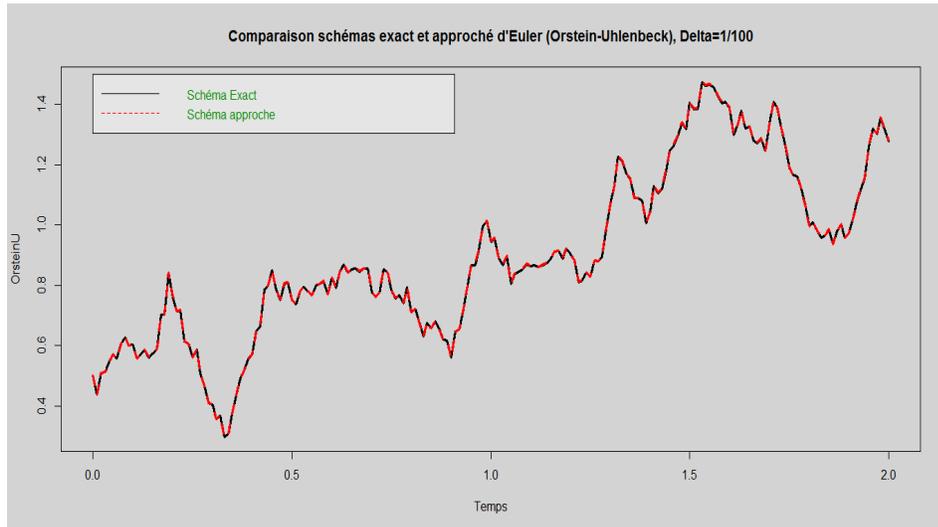


FIGURE 5.5 –

5.2.2 Schéma de Milstein

On applique la formule d'Itô aux deux fonctions $\mu(X_t, t)$ et $\sigma(X_t, t)$ on obtient l'approximation de Milstein

$$\hat{X}(t+\Delta) = \hat{X}(t) + \Delta \mu(t, \hat{X}(t)) + \sigma(t, \hat{X}(t)) \sqrt{\Delta} Z_t + \frac{1}{2} \sigma(t, \hat{X}(t)) \frac{\partial \sigma(t, \hat{X}(t))}{\partial x} \Delta (Z_t^2 - 1).$$

$Z_t \rightsquigarrow N(0, 1)$.

Si on applique le schéma de Milstein au mouvement Brownien géométrique solution de (4.8) on obtient

$$\tilde{X}(t + \Delta) = \tilde{X}(t) \left(1 + \mu\Delta + \sigma\sqrt{\Delta}Z_t + \frac{1}{2}\sigma^2\Delta(Z_t^2 - 1) \right), \quad \tilde{X}(0) = 1. \quad (5.5)$$



FIGURE 5.6 –

La figure 5.7 montre que lorsque le pas de discrétisation est égale à $\Delta = 1/10$ le schéma de Milstein est plus proche au schéma exact. La différence entre les performances des schémas d'Euler et Milstein semble se réduire quand le pas s'approche de zéro (voir la figure 5.8)

Dans le cas du processus d'Orstein-Uhlenbeck les deux discrétisations Euler et Milstein sont identiques car $\sigma(X_t, t) = \sigma$.

Si on considère le modèle CIR le schéma d'Euler nous donne

$$\tilde{r}_{t+\Delta} = ab\Delta + (1 - a\Delta)\tilde{r}_t + \sigma\sqrt{\Delta\tilde{r}_t}Z_t, r_0 = x$$

Alors que le schéma de Milstein nous donne

$$\tilde{r}_{t+\Delta} = ab\Delta + (1 - a\Delta)\tilde{r}_t + \sigma\sqrt{\Delta\tilde{r}_t}Z_t + \frac{1}{4}\sigma^2\Delta(Z_t^2 - 1), r_0 = x$$

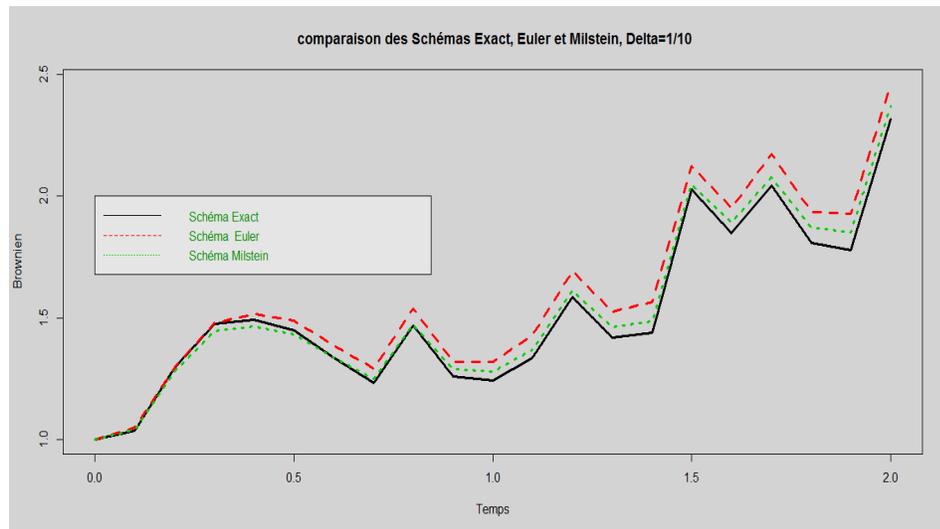


FIGURE 5.7 – Euler-Milstein, $\Delta = 1/10$.

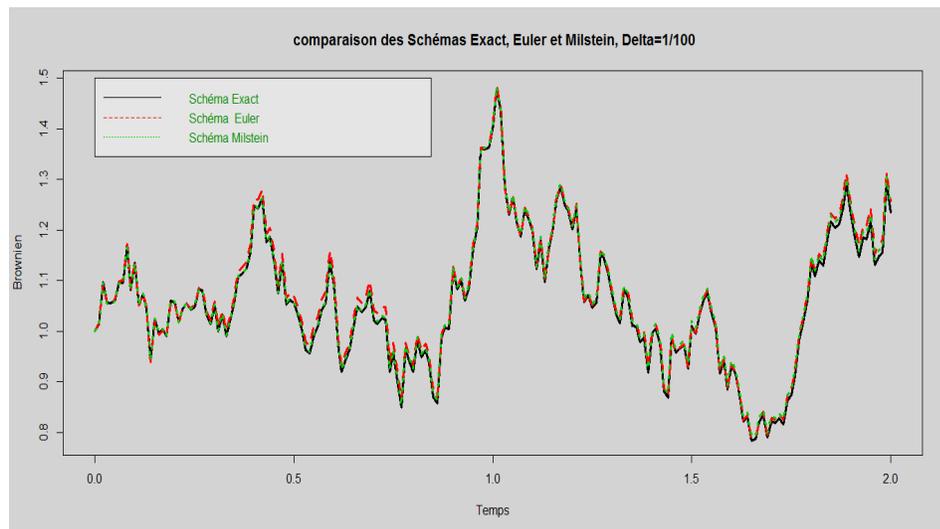
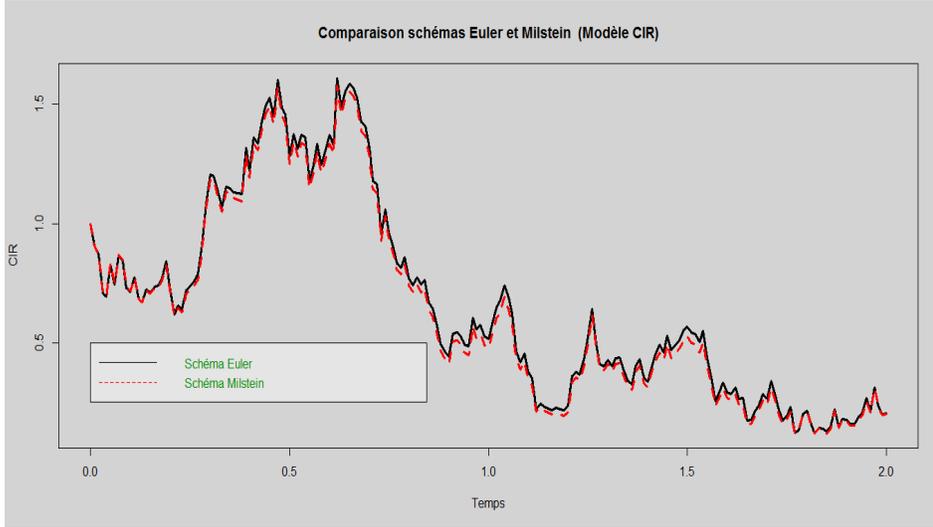


FIGURE 5.8 – Euler-Milstein, $\Delta = 1/100$.

La figure 5.9 montre que les deux schémas Euler et Milstein donnent à peu près les mêmes approximations lorsque le pas de discrétisation est très petit

FIGURE 5.9 – Euler-Milstein pour le modèle de CIR, $\Delta = 1/100$.

(ici $\Delta = 1/100$).

5.3 Convergence des Schémas de discrétisation

DÉFINITION 34 Soit $(\hat{X}_{t,\Delta})$ une discrétisation du processus (X_t) sur l'intervalle $[0, T]$. On dit que $(\hat{X}_{t,\Delta})$ converge fortement à l'ordre r vers (X_t) si $\exists C > 0$, telle que

$$E \left(\left| \hat{X}_{T,\Delta} - X_T \right| \right) \leq C \Delta^r. \quad (5.6)$$

On dit que $(\hat{X}_{t,\Delta})$ converge fortement vers (X_t) si

$$\lim_{\Delta \rightarrow 0} E \left(\left| \hat{X}_{T,\Delta} - X_T \right| \right) = 0. \quad (5.7)$$

Remarque. La convergence forte implique que $\hat{X}_{T,\Delta} - X_T$ converge en probabilité vers 0 lorsque le pas de discrétisation tend vers 0.

Chapitre 6

Estimation des équations différentielles

Estimation

Pour estimer les paramètres de l'équation différentielle, on se base souvent sur l'échantillon obtenu par discrétisation exacte (si la solution de l'équation est explicite) ou approchée selon le schéma d'Euler ou Milstein (si la solution de l'équation n'est pas connue).

6.0.1 Estimation du Mouvement Brownien géométrique

Considérons le mouvement géométrique (ou modèle de Black Scholes)

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \quad X_0 = 1. \quad (6.1)$$

avec μ et σ deux paramètres inconnus. Posons

$$Y_t = \log(X_t), Y_k = Y_{t_k}, k = 1, \dots, N = \left\lceil \frac{T}{\Delta} \right\rceil.$$

Comme

$$dY_t = (\mu - \sigma^2/2)dt + \sigma dB_t, \quad Y_0 = 0. \quad (6.2)$$

Posons $m = \mu - \sigma^2/2$, on déduit que

$$Y_{t_k} = Y_{t_{k-1}} + m\Delta + \sigma (B_{t_k} - B_{t_{k-1}}), k = 1, \dots, N, \quad (6.3)$$

ou encore

$$Y_k = Y_{k-1} + m\Delta + \sigma u_k, k = 1, \dots, N, \quad (6.4)$$

où $u_k = B_{t_k} - B_{t_{k-1}}$. Puisque les accroissements d'un mouvement Brownien sont indépendants, on déduit que $(u_k, k = 1, \dots, N)$ est une suite i.i.d. Gaussienne centrée et de variance Δ . Par conséquent le processus (Y_k) défini par (6.4) est un ARIMA(0,1,0), c'est un processus ayant une racine unitaire et donc non stationnaire. Pour la stationnariser on le différencie c'est à dire on considère

$$Z_k = Y_k - Y_{k-1}, k = 2, \dots, N, Z_1 = Y_1.$$

On en déduit que $(Z_k, k = 1, \dots, N)$ est une suite i.i.d. Gaussienne de moyenne $\mu\Delta$ et de variance $\sigma\Delta$. L'estimation de $\theta = (\mu, \sigma)$ peut se baser sur l'échantillon (Z_1, \dots, Z_N) .

Ainsi les estimateurs du maximum de vraisemblance (et aussi empiriques) sont donnés par

$$\hat{m} = \frac{1}{N\Delta} \sum_{k=1}^N Z_k = \frac{1}{N\Delta} Y_N, \quad (6.5)$$

et

$$\hat{\sigma}^2 = \frac{1}{N\Delta} \sum_{k=1}^N (Z_k - \hat{\mu}\Delta)^2. \quad (6.6)$$

6.0.2 Estimation du processus d'Orstein-Uhlenbeck

On considère le processus d'Orstein-Uhlenbeck :

$$dr_t = a[b - r_t]dt + \sigma dB_t, r_0 = x$$

a, b et σ sont inconnus. On rappelle que la discrétisation exacte du processus d'Orstein-Uhlenbeck est donnée par

$$\hat{r}_{t_k} = \hat{r}_{t_{k-1}} e^{-a\Delta} + b(1 - e^{-a\Delta}) + \sigma e^{-at_k} \int_{t_{k-1}}^{t_k} e^{as} dB_s. \quad (6.7)$$

Posons

$$Z_k = \int_{t_{k-1}}^{t_k} e^{as} dB_s,$$

La suite Z_k suit une loi normale de moyenne zéro et de variance $\sigma_k^2 = \frac{1}{2a} (e^{2at_k} - e^{2at_{k-1}})$, de plus d'après le théorème on a

$$\text{cov}(Z_k, Z_{k+j}) = 0, \quad \forall j \neq 0,$$

par conséquent $u_k = \sigma e^{-at_k} Z_k$ est une suite i.i.d. Gaussienne de moyenne 0 et de variance

$$\sigma_u^2 = \sigma^2 \frac{1 - e^{-2a\Delta}}{2a}.$$

Donc en posant $\hat{r}_k = \hat{r}_{t_k}$ on obtient

$$\hat{r}_k = \hat{r}_{k-1} e^{-a\Delta} + b(1 - e^{-a\Delta}) + u_k, k = 1, \dots, N, \quad (6.8)$$

c'est à dire que le processus (\hat{r}_k) est un AR(1).

En calculant l'estimateur des moindres carrés des paramètres $\theta = (\mu, a_1, \sigma_u^2)$ (voir chapitre 7) dans le processus

$$\hat{r}_k = \mu + a_1 \hat{r}_{k-1} + u_k, k = 1, \dots, N, \quad (6.9)$$

on déduit les estimations de (a, b, σ^2) par

$$\hat{a} = -\frac{\log(\hat{a}_1)}{\Delta}, \hat{b} = \frac{\hat{\mu}}{1 - e^{-\hat{a}\Delta}}, \hat{\sigma}^2 = \frac{2\hat{a}\hat{\sigma}_u^2}{1 - e^{-2\hat{a}\Delta}}.$$

Chapitre 7

Processus autorégressif-moyenne mobile (ARMA)

— Processus du second ordre

DÉFINITION 35 On appelle **processus aléatoire du second ordre à temps discret**, toute suite réelle $X_t : \Omega \rightarrow \mathbb{R}$ définie sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, indexée par $t \in \mathbb{N}$ ou \mathbb{Z} telle que :

$$E(|X_t|^2) < \infty \text{ pour tout } t.$$

i) Moyenne : $\mu : \mathbb{Z} \rightarrow \mathbb{R}$, $\mu(t) = E(X_t)$, $\forall t \in \mathbb{Z}$.

ii) Auto-covariance : $K : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$, $K(t, s) = E((X_t - \mu(t))(X_s - \mu(s)))$, $\forall t, s \in \mathbb{Z}$. Ces fonctions s'expriment encore à l'aide du produit scalaire :

$$\langle X, Y \rangle = E(XY)$$

défini sur l'espace de probabilité $L^2(\Omega, \mathcal{F}, P)$

$$\mu(t) = \langle X_t, 1 \rangle, \quad K(t, s) = \langle X_t, X_s \rangle - \mu(t)\mu(s).$$

Le processus X_t est dit centré si $\mu = 0$.

THÉORÈME 18 La fonction d'auto-covariance est symétrique de type positif, i.e.

$$K(t, s) = K(s, t); \quad \sum_t \sum_s a(t)K(t, s)a(s) \geq 0$$

pour toute fonction $a : \mathbb{Z} \rightarrow \mathbb{R}$ à support fini (quels que soient les entiers $(t_j)_{1 \leq j \leq k}$, la matrice $[K(t_i, t_j)]_{1 \leq i, j \leq k}$ est symétrique positive).

— Processus stationnaire au second ordre

DÉFINITION 36 Soit X un processus ; on dit que X est **strictement (ou fortement) stationnaire** si pour toute suite finie d'instants t_1, \dots, t_k élément de \mathbb{Z} et tout entier $r \in \mathbb{Z}$, les lois jointes de $(X_{t_1}, \dots, X_{t_k})$ et de $(X_{r+t_1}, \dots, X_{r+t_k})$ sont les mêmes (lois jointes invariantes par translation dans le temps).

DÉFINITION 37 Un processus est **stationnaire au second ordre (on dit aussi faiblement stationnaire)** si moyenne et auto-covariance sont invariantes par translation dans le temps, i.e. :

$$\begin{cases} E(X_t) = \mu, \forall t \in \mathbb{Z}, \\ \text{et} \\ E(X_t - \mu)(X_s - \mu) = E(X_{t+r} - \mu)(X_{s+r} - \mu), \quad \forall (t, s, r) \in \mathbb{Z}^3. \end{cases}$$

où μ est une constante et $K(t, s)$ ne dépend plus que de $t - s$, donc il existe une fonction $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$, telle que : $K(t, s) = \gamma(t - s)$.

Alors pour un processus faiblement stationnaire, on a

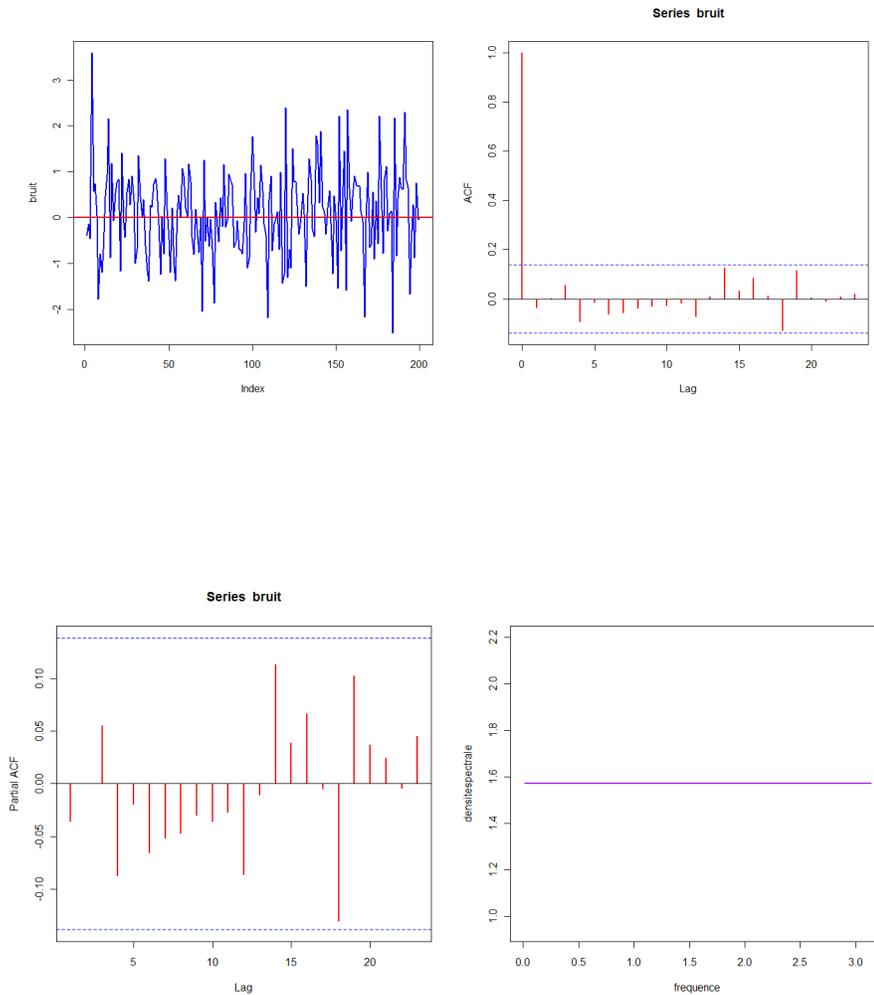
$$\left\{ \begin{array}{l} \text{var}(X_t) = \gamma(0), \forall t \in \mathbb{Z}, \\ |\gamma(k)| \leq \gamma(0), \forall k \in \mathbb{Z}, \quad (\text{ par l'inégalité de Shwartz}) \\ \gamma(k) = \gamma(-k), \forall k \in \mathbb{Z}, \\ \rho(j) = \frac{\gamma(j)}{\gamma(0)}, \forall j \in \mathbb{Z}, \quad \text{fonction d'auto-corrélation ou mémoire de } X_t \\ r(j) = \frac{\text{cov}(X_1 - \mathbf{P}_{[X_2, \dots, X_j]}(X_1), X_{j+1} - \mathbf{P}_{[X_2, \dots, X_j]}(X_{j+1}))}{\text{var}(X_1 - \mathbf{P}_{[X_2, \dots, X_j]}(X_1))} \\ \text{fonction d'auto-corrélation partielle .} \end{array} \right.$$

Analyse spectrale d'un processus stationnaire au second ordre

THÉORÈME 19 (Herglotz-Fourier) Soit $X = (X_t)_{t \in \mathbb{Z}}$ un processus stationnaire au second ordre, de fonction d'auto-covariance $\gamma(k)$ et de densité spectrale $f(\lambda)$. Alors

$$\begin{aligned} \gamma(k) &= \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda \quad \text{pour tout } k \in \mathbb{Z} \\ f(\lambda) &= \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{-ik\lambda} \gamma(k) \end{aligned}$$

Exemples. 1) Bruit blanc : $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$ un processus centré tel que $E(\varepsilon_t \varepsilon_s) = \sigma^2 \delta_t^s$. Alors $f(\lambda) = \sigma^2/2\pi$.



ii) Moyenne mobile : Soit $(\varepsilon_t)_{t \in \mathbb{Z}}$ un bruit blanc, un processus $MA(q)$ est défini par

$$X_t = \sum_{k=0}^q c_k \varepsilon_{t-k}, \quad c_0 = 1.$$

$$\begin{aligned}
\gamma(k) &= E(X_{t+k}X_t) \\
&= \sum_{l,m} c_l c_m E(\varepsilon_{t+k-l}\varepsilon_{t-m}), \quad |k| \leq q, \\
&= \sigma^2 \sum_{l=0}^{q-k} c_{k+l}c_l, \quad |k| \leq q,
\end{aligned}$$

$$\gamma(k) = 0, \quad \text{si } |k| > q;$$

donc un tel processus admet une densité spectrale (continue)

$$\begin{aligned}
f(\lambda) &= \frac{1}{2\pi} \sum_k \gamma(k)e^{-ik\lambda}, \\
&= \frac{\sigma^2}{2\pi} \sum_{|k| \leq q} \sum_{l=0}^{q-k} c_{k+l}c_l e^{-ik\lambda}
\end{aligned}$$

posons $k + l = m$,

$$\begin{aligned}
f(\lambda) &= \frac{\sigma^2}{2\pi} \sum_{m=0}^q \sum_{l=0}^q c_m c_l e^{-i(m-l)\lambda} \\
&= \frac{\sigma^2}{2\pi} \left| \sum_{l=0}^q c_l e^{-il\lambda} \right|^2
\end{aligned}$$

Programme R

```
MA1 <- arima.sim(list(order=c(0,0,1), ma=0.7),200)
```

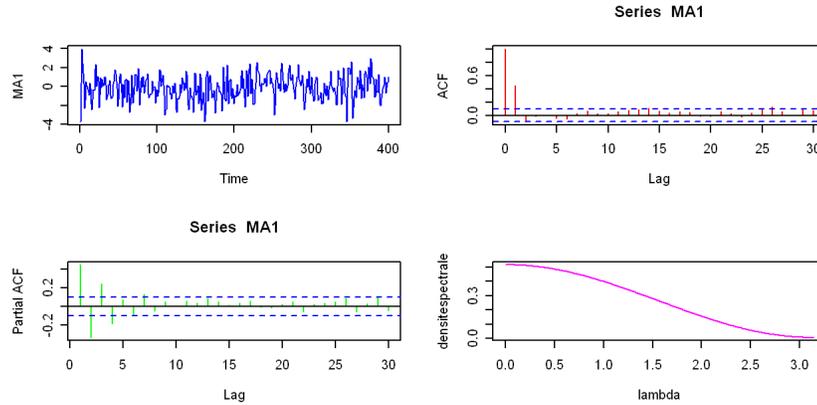
```
acf(MA1)
```

```
pacf(MA1)
```

— **Processus ARMA**

DÉFINITION 38 *soit X un processus stationnaire au second ordre et centré. Nous dirons que X est un processus ARMA(p, q), s'il existe un bruit blanc ε défini sur le même (Ω, \mathcal{F}, P) et des nombres réels $a_1, \dots, a_p, c_1, \dots, c_q$ tels que*

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=1}^q c_j \varepsilon_{t-j} + \varepsilon_t. \quad (7.1)$$



En particulier : X est dit autorégressif d'ordre p , $\text{AR}(p)$, si

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \varepsilon_t \quad (7.2)$$

— **Processus $\text{AR}(p)$**

- Ces processus jouent un rôle important dans les applications.
- X_n est une extrapolation linéaire des p valeurs précédentes X_{n-1}, \dots, X_{n-p} à un bruit blanc près.

Si A désigne le polynôme unitaire ($A(0) = 1$) de degré p

$$A(x) = 1 - \sum_{k=1}^p a_k x^k, \quad (7.3)$$

l'équation (7.2) peut alors s'écrire, à l'aide de l'opérateur retard B , ($BX_t = X_{t-1}$) :

$$A(B)X_t = \varepsilon_t, \quad \forall t \in \mathbb{Z}. \quad (7.4)$$

THÉORÈME 20 *Un processus autorégressif d'ordre p , possède une densité spectrale définie par*

$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2} \quad (7.5)$$

(σ^2 est la variance de ε), et le polynôme A ne peut donc posséder de zéros de module 1.

Réciproquement si un processus centré stationnaire au second ordre X possède une densité spectrale de la forme (7.5) pour un nombre réel $\sigma > 0$ et un polynôme unitaire A de degré p (nécessairement sans zéro de module 1) ce processus est autorégressif d'ordre p les $(a_j)_{1 \leq j \leq p}$ étant donnés par (7.3).

THÉORÈME 21 La fonction d'auto-corrélation partielle d'un processus $AR(p)$ vérifie $r(j) = 0 \forall j > p$.

THÉORÈME 22 La densité spectrale d'un processus $AR(p)$ s'écrit d'une seule manière sous la forme

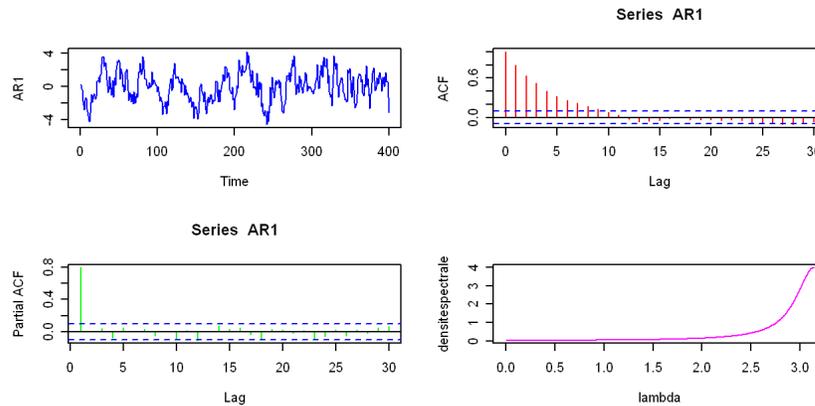
$$f(\lambda) = \frac{\sigma^2}{2\pi |A(e^{-i\lambda})|^2},$$

pour un réel $\sigma > 0$, et un polynôme unitaire A de degré p dont toutes les racines sont de module > 1 .

La représentation (7.2) du processus X est alors appelée représentation canonique de X .

Programme R

```
AR1 <- arima.sim(list(order=c(1,0,0), ar=0.7), 400).
acf(AR1)
pacf(AR1)
```



DÉFINITION 39 Soit (X_t) le processus ARMA défini par l'équation (7.1). On dit que (X_t) est *causal* si les racines de $A(z)$ sont toutes strictement à l'extérieur du disque unité.

DÉFINITION 40 On dit que (X_t) est *inversible* si les racines de $C(z)$ sont toutes strictement à l'extérieur du disque unité.

THÉORÈME 23 (représentation canonique) Soit (X_t) un processus ARMA stationnaire au second ordre.

i) Sa densité spectrale est donnée par :

$$f(\lambda) = \frac{\sigma^2 |C(e^{-i\lambda})|^2}{2\pi |A(e^{-i\lambda})|^2} \quad (7.6)$$

ii) Il existe deux polynômes $\tilde{A}(z)$ et $\tilde{C}(z)$ n'ayant que des racines strictement à l'extérieur du disque unité et un bruit blanc (η_t) tel que :

$$\tilde{A}(B)X_t = \tilde{C}(B)\eta_t,$$

cette représentation est unique et dite canonique.

$$A(z) = \prod_{j=1}^p (1 - \alpha_j z); \quad C(z) = \prod_{j=1}^q (1 - \beta_j z);$$

$$|\alpha_j| < 1, \text{ si } 1 \leq j \leq r; \quad |\alpha_j| > 1, \text{ si } r+1 \leq j \leq p;$$

$$|\beta_j| < 1, \text{ si } 1 \leq j \leq s; \quad |\beta_j| > 1, \text{ si } s+1 \leq j \leq q;$$

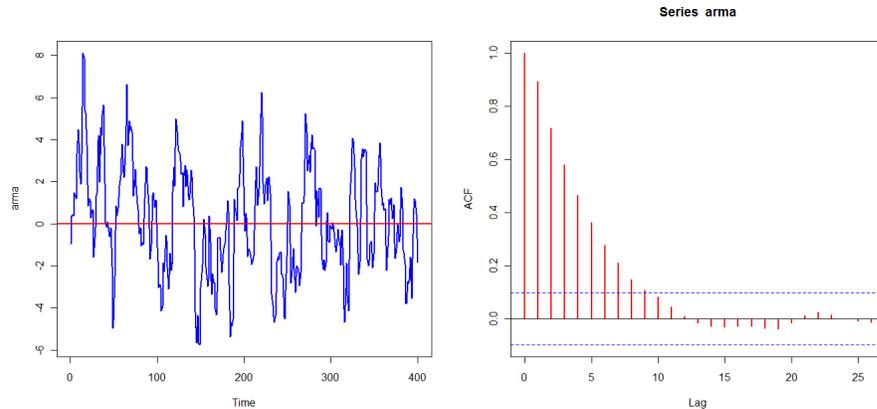
$$\tilde{A}(z) = \prod_{j=1}^r (1 - \alpha_j z) \prod_{j=r+1}^p \left(1 - \frac{1}{\alpha_j} z\right);$$

$$\tilde{C}(z) = \prod_{j=1}^s (1 - \beta_j z) \prod_{j=s+1}^q \left(1 - \frac{1}{\beta_j} z\right);$$

$$\sigma_\eta^2 = \sigma_\varepsilon^2 \frac{\prod_{j=s+1}^q |\beta_j|^2}{\prod_{j=r+1}^p |\alpha_j|^2}.$$

Programme R

```
arma <- arima.sim(list(order = c(1,0,1), ar = 0.8,ma=0.6), 400)
acf(arma)
```



— Estimation

- L'estimation des paramètres d'un processus ARMA(p, q) peut se faire par plusieurs méthodes.
- Critères d'information pour un ARMA(p, q), avec $\hat{\sigma}^2$ un estimateur de $\sigma^2 = \text{var}(u_t)$:

$$\mathbf{Akaike} \text{ (AIC)} : AIC(p, q) = \ln(\hat{\sigma}^2) + 2(p + q)/T,$$

$$\mathbf{Schwarz} \text{ (BIC)} : BIC(p, q) = \ln(\hat{\sigma}^2) + 2(p + q) \ln(T)/T.$$

- **Estimation d'un AR(p)** : Sur R il existe quatre méthodes d'estimation utilisant le critère d'information AIC pour déterminer un choix optimal du degré p . Usage R
`ar(x, aic= TRUE, order.max = NULL, method=c("yule-walker", "burg", "ols", "mle", "yw"))`

Programme R

```
fit4 <- ar(AR1, aic=TRUE, "burg")
Call : ar(x = AR1, aic = TRUE, method = "burg")
Coefficients :
1
0.6872
```

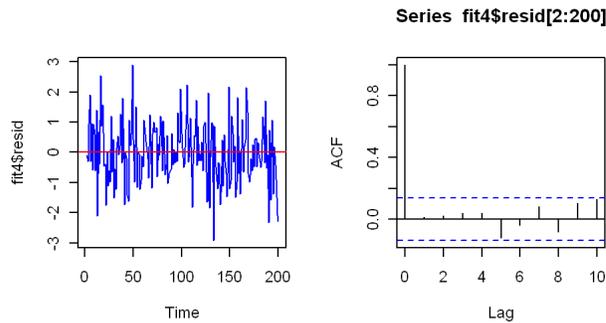
Order selected 1 sigma^2 estimated as 1.030

Interprétation : Le modèle optimal est un AR(1), $\hat{a}_1 = 0.6872$.

— Analyse des résidus

Programme R

```
ts.plot(fit4$resid,type="l",col="blue")
abline(0,0,col="red")
acf(fit4$resid[2 :200],10)
```



Estimation d'un ARMA(p,q)

- Plusieurs bibliothèques dans R : stats, tseries, forecast

Exemples : 1. Simulation et estimation d'un ARMA(2,1)

Programme R

```
library(forecast),
x = arima.sim(model = list(ar = c(0.5,-0.5), ma = 0.9), n = 100)
fit5= Arima(x, order = c(1, 0, 1), include.mean = FALSE, method = "CSS")
fit5
```

Series : x

ARIMA(2,0,1) with zero mean

Coefficient(s) :

ar1	ar2	ma1
0.5433	-0.4902	0.9212

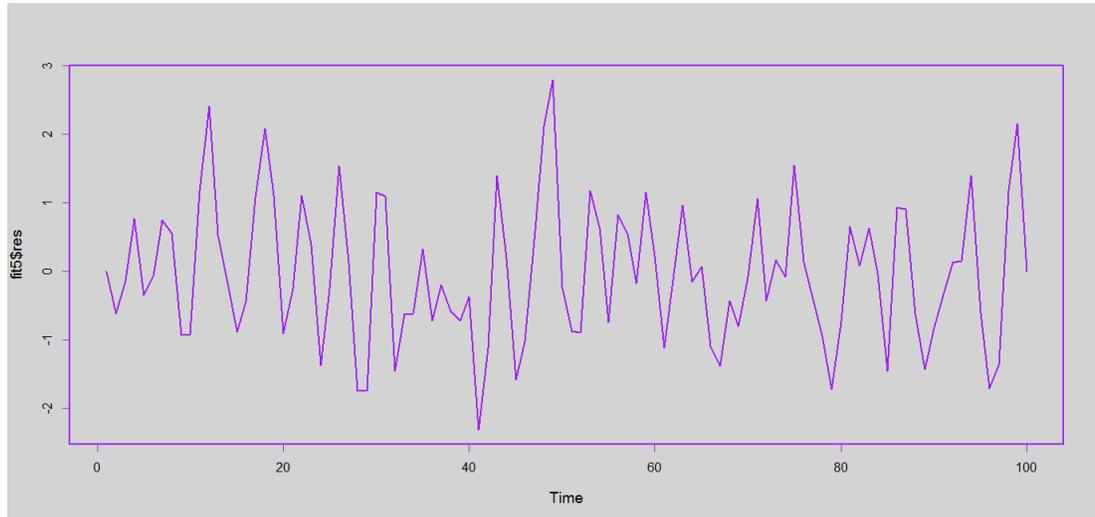
Prévision d'un ARMA(p,q)

Programme R

```
prev=forecast(fit5, 10)
```

```
prev
```

```
Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
```



101 1.35190230 0.1554438 2.548361 -0.477923 3.181728
 102 -0.33485669 -2.3585393 1.688826 -3.429812 2.760099
 103 -0.77675402 -2.8085627 1.255055 -3.884137 2.330629
 104 -0.19498763 -2.3356032 1.945628 -3.468777 3.078801
 105 0.27173701 -1.9036965 2.447171 -3.055301 3.598775
 106 0.21271623 -1.9670427 2.392475 -3.120937 3.546370
 107 -0.02994989 -2.2230254 2.163126 -3.383969 3.324070
 108 -0.11197295 -2.3055141 2.081568 -3.466705 3.242759
 109 -0.03659035 -2.2320328 2.158852 -3.394230 3.321049
 110 0.03535024 -2.1609699 2.231670 -3.323632 3.394332

Mesure de la qualité de prévision.

Dans l'échantillon initial, (X_1, \dots, X_T) on considère seulement $T_1 = [(1-\varepsilon)T]$ observations avec $\varepsilon > 0$. Les $L = T - [(1-\varepsilon)T]$ seront à prévoir par le modèle.

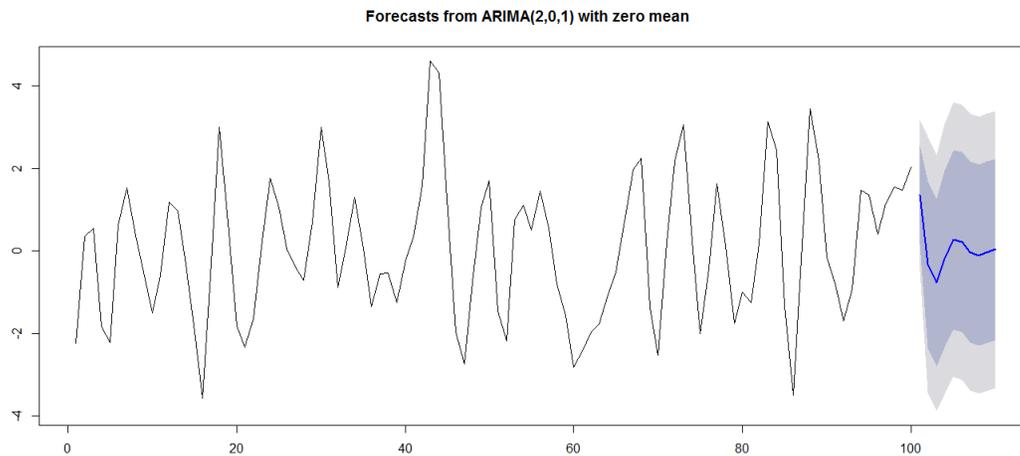
On peut alors considérer plusieurs critères :

1. Mean Absolute Percentage Error

$$MAPE = \frac{100}{L} \sum_{r=1}^L \left| \frac{X_{T_1+r} - \hat{X}_{T_1+r/T_1}}{X_{T_1+r}} \right|$$

2. Root of Mean Square Error

$$RMSE = \left(\sum_{r=1}^L \frac{(X_{T_1+r} - \hat{X}_{T_1+r/T_1})^2}{L} \right)^{1/2}.$$



Exemple : Simulation et estimation d'un ARMA(1,1)

Programme R :

```
library(forecast)
```

```
x = arma.sim(model = list(ar = 0.5, ma = 0.7), n = 100)
```

```
Eaprenti <- x[1 :90]
```

```
Etemoin <- x[91 :100]
```

```
fit6 = Arima(Eaprenti, order=c(1,0,1), include.mean = FALSE, method = "CSS")
```

```
summary(fit6)
```

on obtient

Series : Eaprenti

ARIMA(1,0,1) with zero mean

Coefficients :

```
ar1 ma1
```

```
0.4763 0.6721
```

```
s.e. 0.1041 0.0911
```

σ^2 estimated as 0.9299 : part log likelihood=-124.44

Training set error measures :

ME RMSE MAE MPE MAPE MASE

66 CHAPITRE 7. PROCESSUS AUTORÉGRESSIF-MOYENNE MOBILE (ARMA)

Training set 0.04666327 0.9589592 0.7288203 332.362 516.5434 0.7560344

ACF1

Training set -0.05117232

Programme R :

```
prev = forecast(fit6,10)
```

```
accuracy(prev, Etemoin)
```

on obtient

ME RMSE MAE MPE MAPE MASE

Training set 0.04666327 0.9589592 0.7288203 332.3620 516.5434 0.7560344

Test set -0.65031702 1.2874606 1.0271002 56.7831 118.0325 1.0654522

ACF1

Training set -0.05117232

Test set NA.

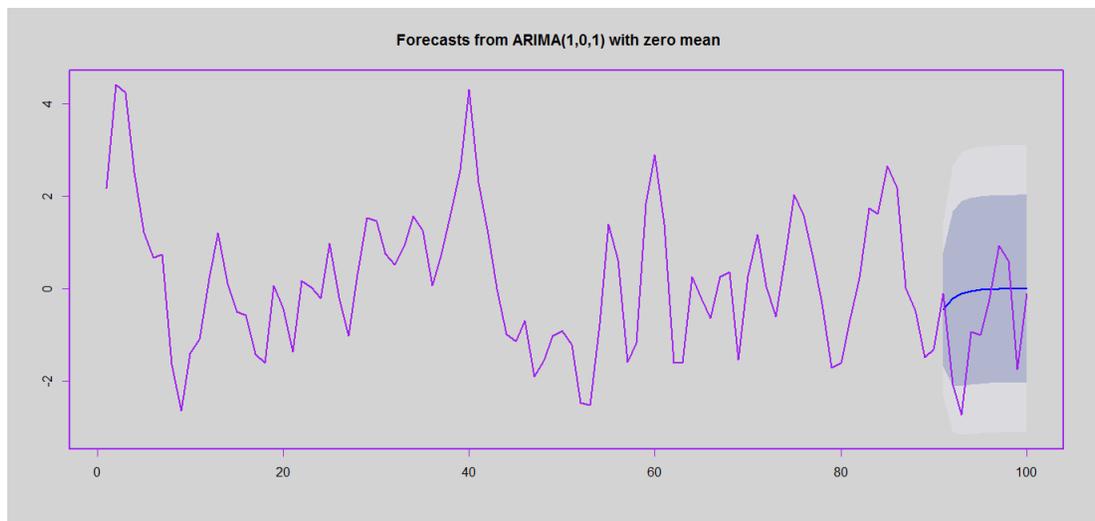


FIGURE 7.1 – Prévisions et intervalle de confiance

7.1 Un exemple réel

On considère la série temporelle **Vent** qui représente la vitesse du vent quotidienne mesurée à New York sur la période Mai à Septembre 1973 (sé-

rie constituée de 153 observations) et qui se trouve dans le jeu de données `airquality` du logiciel .

```
Programme R
data(airquality)
Vent <- airquality$Wind
ts.plot(Vent, main = "Trajectoire de la série Vent")
```

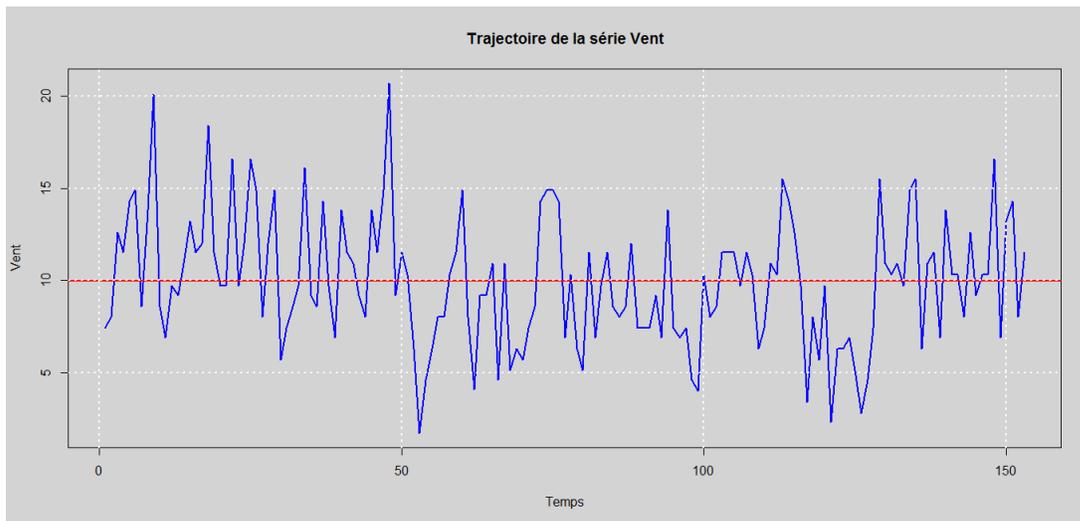


FIGURE 7.2 – Trajectoire de la série Vent

Étude de la stationnarité

La figure 7.2 montre bien que la série fluctue autour de sa moyenne $m=9.96$, donc à priori on peut supposer qu'elle est stationnaire. On teste alors sa stationnarité par le test KPSS.

```
Programme R
library(tseries)
kpss.test(Vent)
KPSS Test for Level Stationarity
data : Vent
KPSS Level = 0.5103, Truncation lag parameter = 2,
p-value = 0.03934.
Le test KPSS valide la stationnarité à 1% de risque.
Rappelons les étapes de la modélisation :
```



```

3 x x o o o o o o o o o o o o
4 x x x x o o o o o o o o o o
5 x o o x x o o o o o o o o o
6 x o o x o o o o o o o o o o
7 x x x x o o o o o o o o o o

```

Les autocorrélations étendues montrent bien que la série chronologique du vent peut aussi être modélisée par un MA(3).

Nous considérons plutôt le modèle AR(1) qui possède moins de paramètres à estimer.

2. Estimation des paramètres a_i, b_j .

```

Programme R
modele1 <- Arima(Vent, order = c(1, 0, 0))
modele1

```

```

Series : Vent
ARIMA(1,0,0) with non-zero mean
Coefficients :
ar1 intercept
0.310 9.955
s.e. 0.077 0.390

```

```

sigma2 estimated as 11.1 : log likelihood=-401.54
AIC=809.08 AICc=809.24 BIC=818.17

```

3. Réduction du modèle

```

Programme R
t_stat(modele1)
ar1 intercept
t.stat 4.038835 25.5419
p.val 0.000054 0.0000

```

Le logiciel indique que les deux paramètres $a_i, 0 \leq i \leq 1$ sont significatifs. On ne peut donc pas simplifier le modèle.

4. Validation du modèle : Tests de diagnostic des résidus.

Programme R

```
Box.test(modele1$residuals, lag = n/4, type = "Box-Pierce")
```

Box-Pierce test

```
data : modele1$residuals
```

```
X-squared = 27.9364, df = 38.25, p-value = 0.8902
```

Une p-value égale à 0.8902 indique que l'hypothèse (H_0) : *"les résidus forment un bruit blanc"* est confortablement non rejetée. On confirme avec le diagnostic graphique.

Programme R

```
detach(package :TSA)
```

```
library(forecast)
```

```
tsdiag(modele1)
```

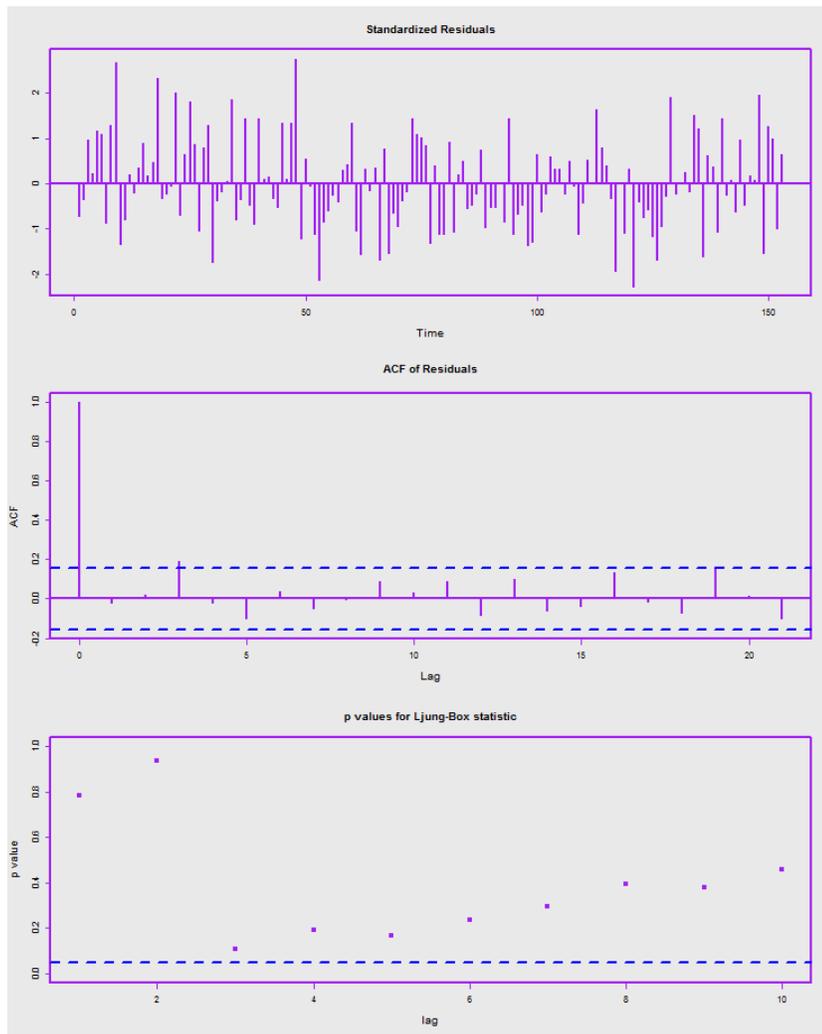


FIGURE 7.4 – Diagnostique graphique de validation des résidus

Le modèle estimé est valide. Par conséquent le modèle AR(1) estimé pour le vent est convenable (adéquat).

Ainsi l'évolution du $Vent_t$ peut être décrite par le modèle AR(1) suivant :

$$Vent_t - 9.96 = 0.31 \times (Vent_{t-1} - 9.96) + \epsilon_t, \quad (7.7)$$

avec $(\epsilon_t)_t$ un bruit blanc de variance $\sigma_\epsilon^2 = 11.1$.

II. Modélisation automatique

La modélisation automatique fait appel à un système expert qui choisira un modèle optimal basé sur une stratégie fixée par l'utilisateur. Ainsi dans le logiciel, les fonctions `ar()` et `auto.arima()` sont basées sur les critères d'information.

a) La fonction `ar`

```
Programme R
modele2 <- ar(Vent)
modele2
Call :
ar(x = Vent)
Coefficients :
1 2 3
0.276 0.029 0.157
```

Order selected 3 σ^2 estimated as 11.1

La fonction `ar` sélectionne le modèle AR(p) qui minimise le critère d'information AIC. Avec cette fonction on obtient un AR(3) :

$$Vent_{t-m} = 0.28 \times (Vent_{t-1-m}) + 0.03 \times (Vent_{t-2-m}) + 0.16 \times (Vent_{t-3-m}) + \epsilon_t \quad (7.8)$$

avec $m = 9.9575$ et (ϵ_t) un bruit blanc de variance $\sigma_\epsilon^2 = 11.1$.

b) La fonction `auto.arima`

```
Programme R
modele3 <- auto.arima(Vent, d = 0, max.p = 5, max.q = 5, ic = "aic")
modele3
Series : Vent
ARIMA(1,0,0) with non-zero mean
Coefficients :
```

```
arl intercept
0.310 9.955
s.e. 0.077 0.390
sigma2 estimated as 11.1 : log likelihood=-401.54
AIC=809.08 AICc=809.24 BIC=818.17
```

La procédure `auto.arima()` propose le même modèle $AR(1)$ pour la série `Vent` que celui sélectionné par la procédure manuelle.

5. Sélection du modèle

a) Critère d'information.

Si on prend le critère AIC comme stratégie de sélection, le tableau suivant résume les critères d'information obtenus pour les trois modèles

Méthode	Modèle	critère AIC
Manuelle (fonction <code>Arima</code>)	AR(1)	809.08
Automatique (fonction <code>ar</code>)	AR(3)	808.42
Automatique (fonction <code>auto.arima</code>)	AR(1)	809.08

Ainsi le modèle AR(3) donné par (7.8) est légèrement préférable aux deux autres.

b) Pouvoir prédictif

On peut aussi sélectionner parmi le modèle AR(3) et le modèle AR(1) celui qui a le bon ajustement ou celui qui a un pouvoir prédictif meilleur sur un échantillon témoin.

```
Programme R
summary(modele1)
Series : Vent
ARIMA(1,0,0) with non-zero mean
Coefficients :
arl intercept
0.310 9.955
s.e. 0.077 0.390
sigma2 estimated as 11.1 : log likelihood=-401.54
AIC=809.08 AICc=809.24 BIC=818.17
In-sample error measures :
```

74 CHAPITRE 7. PROCESSUS AUTORÉGRESSIF-MOYENNE MOBILE (ARMA)

```
ME RMSE MAE MPE
0.00599268 3.33730610 2.69345777 -15.92485613
MAPE
35.07431223
Programme R
summary(modele2)
Series : Vent
ARIMA(3,0,0) with non-zero mean
Coefficients :
```

```
ar1 ar2 ar3 intercept
0.275 0.027 0.158 9.956
s.e. 0.080 0.083 0.080 0.487
sigma2 estimated as 10.8 : log likelihood=-399.21
AIC=808.42 AICc=808.83 BIC=823.58
In-sample error measures :
```

```
ME RMSE MAE MPE
0.00968248 3.28603815 2.65043659 -15.54793827
MAPE
34.53050260
```

Le critère d'ajustement basé sur les mesures "RMSE", "MAE" et "MAPE" indique que le modèle AR(3) a un bon ajustement aux données, que le modèle AR(1), résultat attendu car théoriquement plus on augmente le nombre de paramètres plus les résidus diminuent.

On considère l'échantillon d'apprentissage $E_1 = (Vent_1, \dots, Vent_{145})$ et l'échantillon témoin $E_2 = (Vent_{146}, \dots, Vent_{153})$, E_2 contient 5% des observations.

```
Programme R
modele1 <- Arima(Vent[1 :145], order = c(1, 0,0))
modele2 <- Arima(Vent[1 :145], order = c(3, 0,0))
fit1 = forecast(modele1, h = 8)
fit2 = forecast(modele2, h = 8)
accuracy(fit1, Vent[146 :153])
```

```
ME RMSE MAE MPE MAPE MASE
1.562672 3.392180 2.769166 7.036918 23.592118 0.990887
Programme R
accuracy(fit2, Vent[146 :153])
```

```

ME RMSE MAE MPE MAPE MASE
1.560761 3.385856 2.747195 7.064515 23.335940 0.983025
Programme R
plot(fit1)
lines(Vent[1 :153])
plot(fit2)
lines(Vent[1 :153])

```

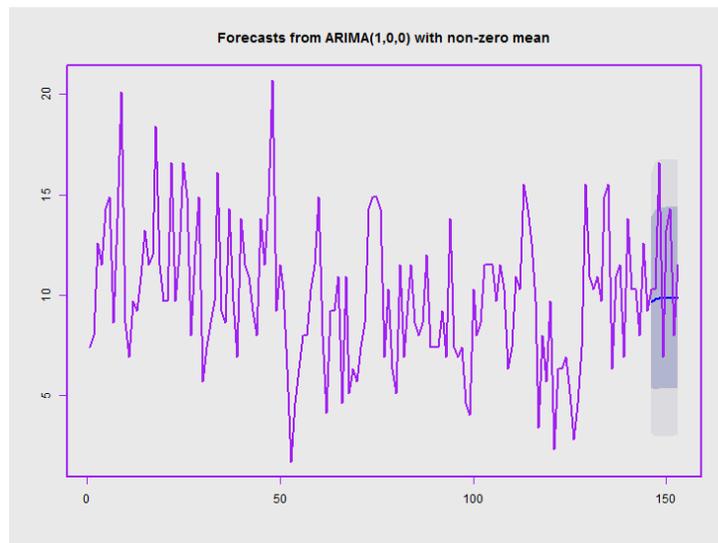


FIGURE 7.5 – Prévisions basées sur le modèle 1

Les critères "RMSE" et "MAPE" calculés sur l'échantillon témoin E_2 sont en faveur du modèle 2. C'est donc ce dernier qui selectionné comme modèle final à la série temporelle Vent.

6. Prévisions

```

Programme R
prev <- forecast(modele2, h = 20)
shapiro.test(modele2$res)
Shapiro-Wilk normality test

```

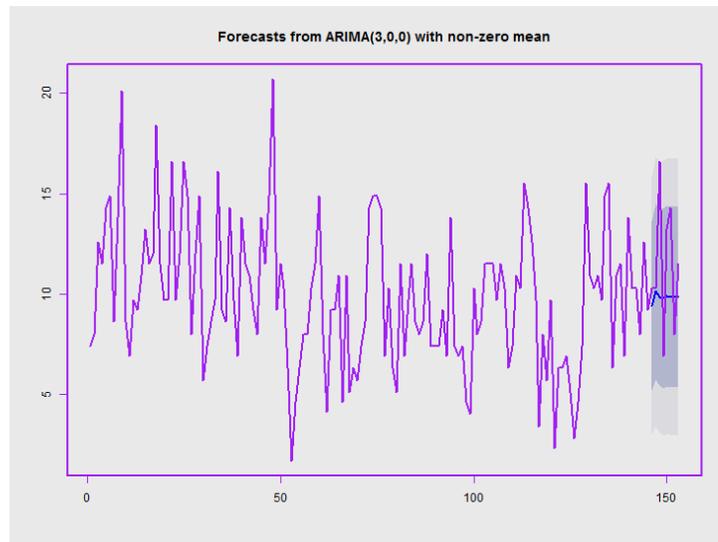


FIGURE 7.6 – Prévisions basées sur le modèle 2

```

data : modele2$res
W = 0.991, p-value = 0.4879
jarque.bera.test(modele2$res)
JarqueBeraTest
data : modele2$res
X - squared = 1.8343, df = 2, p - value = 0.3997
lillie.test(modele2$res)
Lilliefors (Kolmogorov-Smirnov) normality test
data : modele2$res
D = 0.0519, p-value = 0.4417
Les résidus sont gaussiens. On peut donc ajouter les intervalles de prévision.

```

```

Programme R
plot(prev)

```

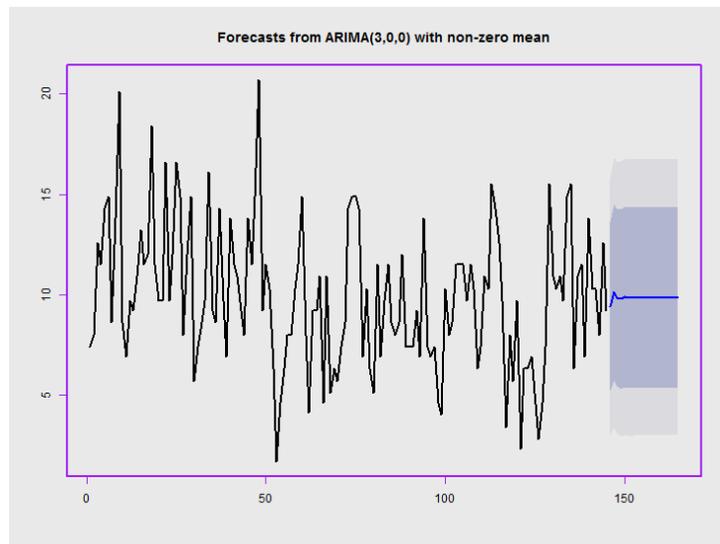


FIGURE 7.7 – Prévission pour la série Vent

Chapitre 8

Processus non stationnaires

8.1 les modèles ARIMA

Il existe plusieurs processus non stationnaires au second ordre, nous nous contentons de présenter deux structures de modèles qui sont plus proche des modèles ARMA.

Modèles ARIMA

DÉFINITION 41 *Nous dirons que $X = (X_t)$ est un processus ARIMA(p, d, q), s'il existe un entier naturel d tel que le processus $Y_t = (1 - B)^d X_t$ est un ARMA(p, q). Autrement dit il existe deux polynômes $A(z)$ et $C(z)$ de degré p et q respectivement et un bruit blanc (ε_n) défini sur le même (Ω, F, P) tels que*

$$(1 - B)^d A(B)X_t = C(B)\varepsilon_t \quad (8.1)$$

Programme R

```
e <- rnorm(200)
```

```
Marchealeatoire <- cumsum(e)
```

```
ts.plot(Marchealeatoire,type="l",col="blue")
```

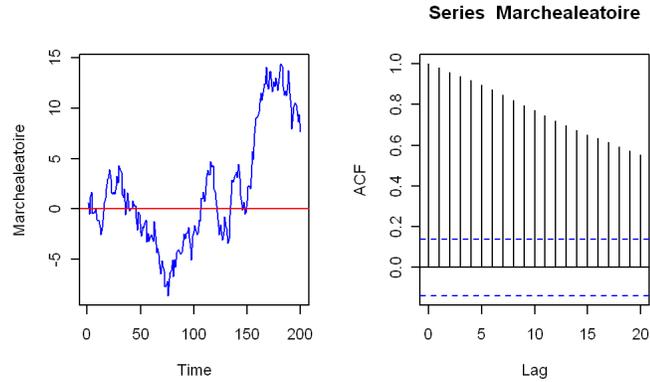
```
abline(0,0,col="red") acf(Marchealeatoire,20)
```

MODELISATION ARIMA

Etapes de la modélisation

— Stationnarisation par différentiation.

— L'opérateur de différentiation $\Delta = 1 - B$, (ou Δ^d) élimine les tendances, d est estimé en effectuant des tests de stationnarité sur



la série brute puis sur les séries résiduelles. Un estimateur de d est le nombre total de fois où on rejette la stationnarité.

— La série résiduelle stationnaire sera modélisée par un ARMA.

- Estimation : R offre la possibilité d'utiliser deux méthodes : 1) la méthode de maximum de vraisemblance "ML" et la méthode des moindres carrés conditionnels "CSS"
- Validation : Le modèle retenu est-il compatible ? : Tests d'adéquation sur les résidus : absence d'autocorrélation, normalité, homoscédasticité,...

Test de non-stationnarité

Test de Dickey et Fuller augmenté

Robuste à l'autocorrélation par rapport au test de Dickey-Fuller

$$\begin{cases} H_0 : y_t \sim I(1) \\ H_1 : y_t \text{ n'est pas } I(1) \end{cases}$$

On exécute la régression

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + \sum_{j=1}^p \psi_j \Delta y_{t_j} + u_t,$$

$$ADF = T\hat{\pi} / (1 - \hat{\psi}_1 - \dots - \hat{\psi}_p).$$

Test de Phillips-Perron Robuste à l'hétéroscédasticité

On effectue la régression :

$$\Delta y_t = \beta' D_t + \pi y_{t-1} + u_t,$$

1

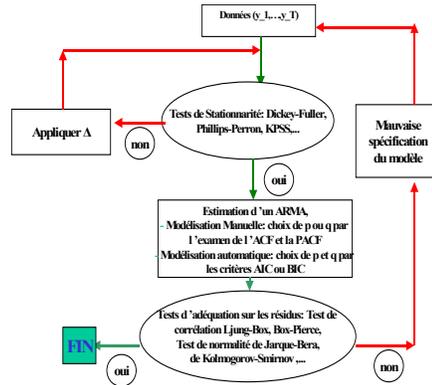


Schéma général de la modélisation d'une série temporelle par un modèle ARIMA.

$$PP = T\hat{\pi} - \frac{1}{2}T^2 \frac{SE(\hat{\pi})}{\hat{\sigma}^2} (\hat{\lambda}^2 - \hat{\sigma}^2),$$

$$\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T \hat{u}_t^2 \hat{u}_t : \text{résidus de la régression,}$$

$$SE(\hat{\pi}) = \text{écart-type de } \hat{\pi},$$

$$\hat{\lambda}^2 = \hat{\sigma}^2 + 2 \sum_{j=1}^q (1 - \frac{j}{q+1}) \hat{\gamma}(j), (Newey - West),$$

$$\hat{\gamma}(j) = \frac{1}{T} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}.$$

Tests de stationnarité : Test de KPSS

$$y_t = \beta' D_t + \mu_t + u_t, \mu_t = \mu_{t-1} + \varepsilon_t, var(\varepsilon_t) = \sigma_\varepsilon^2.$$

TABLE 8.1 – Valeurs critiques de ADF et PP : $P((ADF < c) = \alpha)$. Source : Hamilton (1994), page 763

• $d_t = 0$ (processus centré)

α	n	25	50	100	250	500	∞
0.025		-2.26	-2.25	-2.24	-2.23	-2.23	-2.23
0.05		-1.95	-1.95	-1.95	-1.95	-1.95	-1.95
0.95		1.33	1.31	1.29	1.29	1.28	1.28
0.975		1.70	1.66	1.64	1.63	1.62	1.62

• $d_t = 1$ (processus avec constante)

α	n	25	50	100	250	500	∞
0.025		-3.33	-3.22	-3.17	-3.14	-3.13	-3.12
0.05		-3	-2.93	-2.89	-2.88	-2.87	-2.88
0.95		0	-0.03	-0.05	-0.06	-0.07	-0.07
0.975		0.34	0.29	0.26	0.24	0.24	0.23

• $d_t = (1, t)'$ (processus avec tendance)

α	n	25	50	100	250	500	∞
0.025		-3.95	-3.80	-3.73	-3.69	-3.68	-3.68
0.05		-3.60	-3.50	-3.45	-3.43	-3.42	-3.41
0.95		-0.80	-0.87	-0.90	-0.92	-0.93	-0.94
0.975		-0.50	-0.58	-0.62	-0.64	-0.65	-0.66

$$\begin{cases} H_0 : \sigma_\varepsilon^2 = 0 \\ H_1 : \sigma_\varepsilon^2 \neq 0 \end{cases}$$

$$KPSS = \left(T^{-2} \sum_{t=1}^T \widehat{S}_t^2 \right) / \widehat{\lambda}^2.$$

$\widehat{S}_t = \sum_{j=1}^t \widehat{u}_j, \widehat{u}_t$: résidus de la régression de y_t sur D_t .

$D_t = 1$ veut dire qu'on effectue la régression de y_t sur 1 ce qui veut dire que soustrait la moyenne empirique \bar{y} des y_t , ou encore on centre les données. $D_t = (1, t)'$ veut dire qu'on soustrait une tendance linéaire des y_t .

Exemple : Simulation et estimation d'un ARIMA :

Programme R

```
ari <- arima.sim(list(order = c(1,1,0), ar = 0.7), n = 400)
```

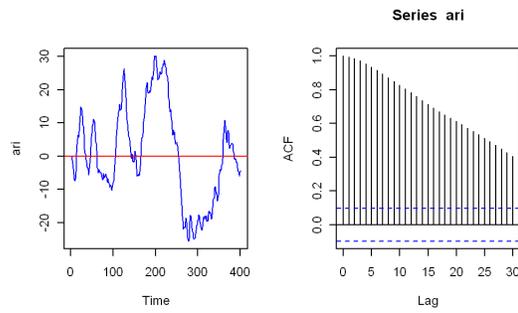
```
plot(ari)
```

```
acf(ari)
```

TABLE 8.2 – Valeurs critiques de KPSS $P(KPSS > c) = \alpha$. Source : Kwiatkowski(1991), page 166.

D_t	α	0.10	0.05	0.025	0.01
$D_t = 1$		0.347	0.436	0.574	0.739
$D_t = (1, t)'$		0.119	0.146	0.176	0.216

Tests de Stationnarité



- ADF :

Programme R

```
library(tseries)
```

```
adf.test(ari)
```

Augmented Dickey-Fuller Test : data : ari

Dickey-Fuller = -2.4035, Lag order = 7, p-value = 0.4068

alternative hypothesis : stationary

On ne rejette pas l'hypothèse nulle d'une racine unitaire.

- PP

Programme R

```
library(tseries)
```

```
pp.test(ari)
```

Phillips-Perron Unit Root Test :data : ari

Dickey-Fuller Z(alpha) = -6.1984, Truncation lag parameter = 5

p-value = 0.7634

- KPSS

Programme R

```
library(tseries)
kpss.test(ari)
KPSS Test for Level Stationarity : data : ari
KPSS Level = 1.3669, Truncation lag parameter = 4, p-value = 0.01
Warning message :
In kpss.test(ari) : p-value smaller than printed p-value
On rejette l'hypothèse nulle de stationnarité.
Conclusion : Tous les tests confirment l'existence d'une racine unitaire.
```

Stationnarisation

```
Programme R
dari <- diff(ari)
Tests de Stationnarité de la série différentiée :
```

- ADF

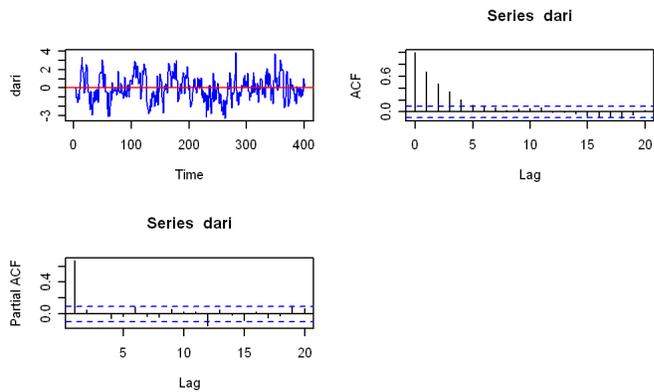
```
Programme R
library(tseries)
adf.test(dari)
Augmented Dickey-Fuller Test :data : dari
Dickey-Fuller = -5.8272, Lag order = 7, p-value = 0.01
alternative hypothesis : stationary
Warning message :In adf.test(dari) : p-value smaller than printed p-value
donc la série différentié est stationnaire.
```

- KPSS

```
Programme R
library(tseries)
kpss.test(ari)
KPSS Test for Level
Stationarity :data : dari
KPSS Level = 0.0976, Truncation lag parameter = 4, p-value = 0.1
on ne rejette pas l'hypothèse nulle de stationnarité.
La série différentiée est stationnaire, on l'a modélise alors par un ARMA.
L'examen de la corrélation partielle nous indique bien que la série peut être
modélisable par un AR(1).
```

Estimation

```
Programme R
fit7 <- Arima(ari, order = c(1, 1, 0),method="ML")
fit7
Series : ari
ARIMA(1,1,0)
```



Coefficients :

ar1

0.6684

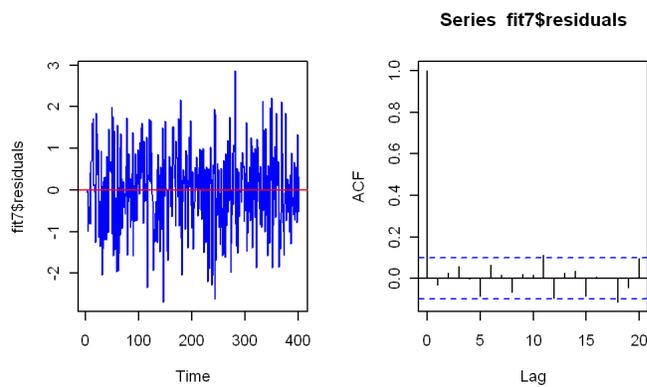
s.e. 0.0370

σ^2 estimated as 1.038 : log likelihood=-575.

AIC=1154.53 AICc=1154.56 BIC=1162.51

Tests d'adéquations

Programme R



```
Box.test(fit7$residuals, lag = 10, type = "Box-Pierce")
```

Box-Pierce test : data : fit7\$residuals

X-squared = 8.728, df = 10, p-value = 0.5581

La p-value est grande donc les résidus ne sont pas corrélés.

Programme R

```
library(nortest)
```

```
lillie.test(fit7$residuals)
```

Lilliefors (Kolmogorov-Smirnov) normality test :data : fit7\$residuals

D = 0.0407, p-value = 0.1114

Donc les données suivent une loi de Gauss.

8.1.1 Un exemple réel, modélisation des indices boursiers (bourse de Paris).

On considère les prix de clôture de l'indice boursier CAC40 sur la période 1991-1998 (1860 observations).

Programme R

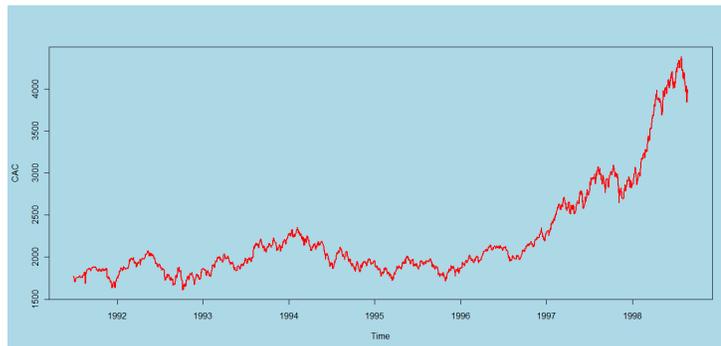
```
data(EuStockMarkets)
```

```
CAC <- EuStockMarkets[,3]
```

```
ts.plot(CAC, type="l", col="blue")
```

```
acf(CAC,20)
```

Tests de Stationnarité



- ADF

Programme R

```
library(tseries)
```

```
adf.test(CAC)
```

Augmented Dickey-Fuller Test : data : CAC

Dickey-Fuller = -0.249, Lag order = 12, p-value = 0.99

alternative hypothesis : stationary

On ne rejette pas l'hypothèse nulle d'une racine unitaire.

Stationnarisation

Programme R

```
dCAC <- diff(log(CAC))
```

Tests de Stationnarité de la série différenciée

•ADF

Programme R

```
library(tseries)
```

```
adf.test(dCAC)
```

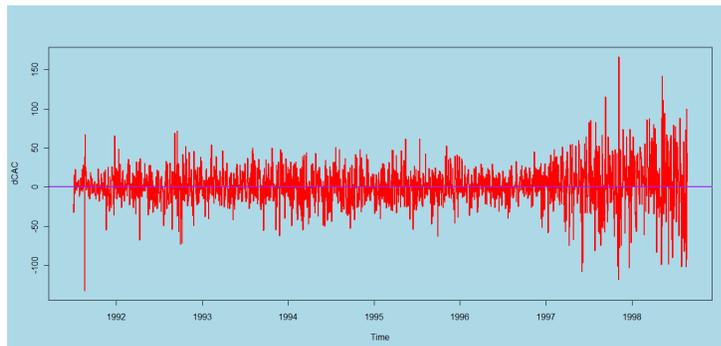
Augmented Dickey-Fuller Test :data : dCAC

Dickey-Fuller = -11.4471, Lag order = 12, p-value = 0.01

alternative hypothesis : stationary

La série différenciée est donc stationnaire.

La fonction d'autocorrélation nous indique que la série différenciée est un



bruit blanc.

Test d'autocorrélation

Programme R

```
Box.test(dCAC, lag = 10, type = "Box-Pierce")
```

Box-Pierce test : data : dCAC

X-squared = 19.9932, df = 10, p-value = 0.02932

Avec 1% de risque on ne rejette pas l'hypothèse que les accroissements du CAC40 est un bruit blanc.

Conclusion : Le CAC40 peut alors être décrit par le mouvement Brownien géométrique (voir le modèle de Black-Scholes (4.7) et son estimation (6.5) et

(6.6)) suivant

$$CAC_t = \exp((\mu - \sigma^2/2)/t + \sigma B_t), \mu = 0.0004978941, \sigma = 0.0001216802, \quad (8.2)$$

avec B_t un mouvement Brownien standard.

8.2 Les modèles SARIMA

Les modèles SARIMA "Seasonal-Integrated-AutoRegressive-Moving-Average" sont utilisés pour modéliser les données saisonnières (mensuelles, trimestrielles) ou cycliques, les portions de trajectoires obtenues par la translation $Y_n \rightarrow Y_{n+s}$ ont de fortes similarités.

Exemple :

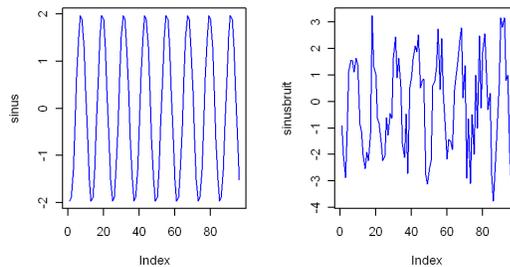
1) Soit le processus (X_t) définit par :

$$X_t = a \sin\left(\frac{\pi t}{6} + 4\right) + \varepsilon_t,$$

où (ε_t) est un bruit blanc. On a

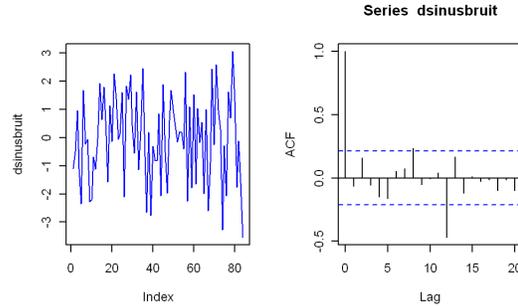
$$E(X_t) = a \sin\left(\frac{\pi t}{6} + 4\right)$$

Le processus (Y_t) définit par $Y_t = (1 - B^{12})X_t = X_t - X_{t-12}$ vérifie : $Y_t =$



$\varepsilon_t - \varepsilon_{t-12}$ c'est un MA(12), donc stationnaire au second ordre.

Programme R



```
dsinusbruit <- diff(sinusbruit, lag=12)
plot(dsinusbruit)
acf(dsinusbruit,20)
```

DÉFINITION 42 *Nous dirons que $X = (X_n)$ est un processus SARIMA(p, d, s, q), s'il vérifie l'équation :*

$$(1 - B)^d (1 - B^s) A(B)X_n = C(B)\varepsilon_n \quad (8.3)$$

Modélisation par les modèles SARIMA

Souvent on utilise une structure plus générale, on suppose que la série (ou le processus la modélisant) satisfait l'équation :

$$(1 - B)^d (1 - B^s)^D \phi(B)\Phi(B^s)X_t = \theta(B)\Theta(B^s)\varepsilon_t \quad (8.4)$$

p, P, d, s, D, q, Q sont des entiers naturels inconnus ; les polynômes $\phi(z), \Phi(z), \theta(z)$ et $\Theta(z)$ sont à coefficients inconnus.

$$\Phi(B^s) = 1 + \Phi_1 B^s + \dots + \Phi_P B^{sP}, \Theta(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{sQ}$$

Dans la structure (8.4), on appelle :

d : Différence, s : saison, D : différence saisonnière, $\phi(z)$ la partie autorégressive, $\Phi(z)$ la partie autorégressive saisonnière, $\theta(z)$ la partie moyenne-mobile., $\theta(z)$ la partie moyenne-mobile saisonnière.

Usage sur R :

`Arima(x, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = s), xreg = NULL, include.mean = TRUE, delta = 0.01, transform.pars =`

TRUE, fixed = NULL, init = NULL, method = c("ML", "CSS"), n.cond,
optim.control = list())

On utilise souvent $\Theta(B^s)$ si on applique aux données le filtre $(1 - B^s)^D$.

Modélisation Identique à celle de la modélisation ARIMA sauf la première étape de stationnarisation qui est remplacée par la suivante :

Stationnarisation par filtrage

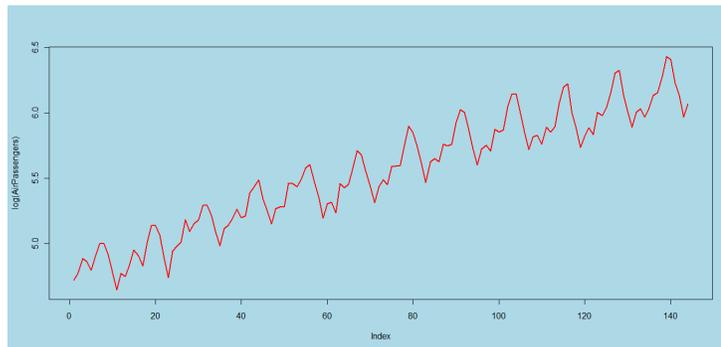
- L'opérateur $(1 - T^s)^D$, élimine la périodicité (désaisonnalisation).
- L'opérateur de différentiation $\Delta = 1 - B$, (ou Δ^d) élimine les tendances, d est estimé en effectuant des tests de stationnarité sur la série brute puis sur les séries résiduelles. Un estimateur de d est le nombre total de fois où on rejette la stationnarité. Ses deux opérations rendent la série stationnaire et donne une estimation des entiers d, s, D .

8.2.1 Deux exemples réels

Exemple 1 : Modélisation de la série des voyageurs

On considère la série "AirPassengers" qui représente le nombre mensuel de voyageur de la compagnie "Airline" pendant la période 1949-1960.

Programme R

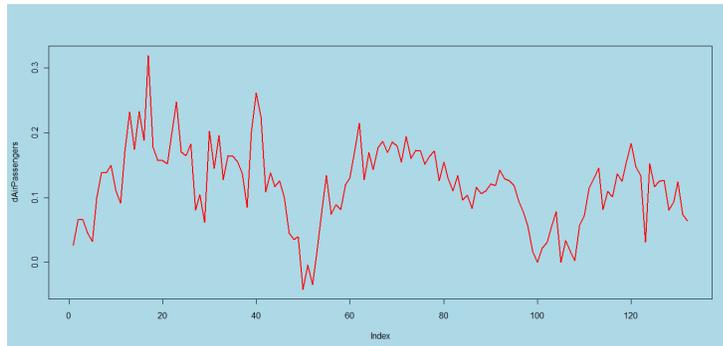


```
dAirPassengers <- diff(AirPassengers, lag=12)
```

L'effet saisonnier semble disparaître.

Programme R

```
library(tseries)
adf.test(dCAC)
```



Augmented Dickey-Fuller Test :data : dAirPassengers

Dickey-Fuller = -3.1519, Lag order = 5, p-value = 0.09899
 alternative hypothesis : stationary
 Donc il y a une racine unitaire.

Programme R

```
fit8 <- Arima(log(AirPassengers),c(0,1,1), seasonal = list(order=c(0, 1 ,1),
period=12))
```

```
fit8
```

```
Series : log(AirPassengers)
```

```
ARIMA(0,1,1)(0,1,1)[12]
```

```
Coefficients :
```

```
ma1 sma1
```

```
-0.4018 -0.5569
```

```
s.e. 0.0896 0.0731
```

```
 $\sigma^2$  estimated as 0.001348 : log likelihood=244.7
```

```
AIC=-483.4 AICc=-483.21 BIC=-474.77
```

C'est à dire que le modèle proposé pour la série est donné par

$$(1 - B)(1 - B^{12}) \log(AirPassengers_t) = (1 - 0.4013B)(1 - 0.65544B^{12})\varepsilon_t. \quad (8.5)$$

Les résidus sont représentés dans la figure 8.1.

Tests d'adéquation

1. Test d'autocorrélation

Programme R

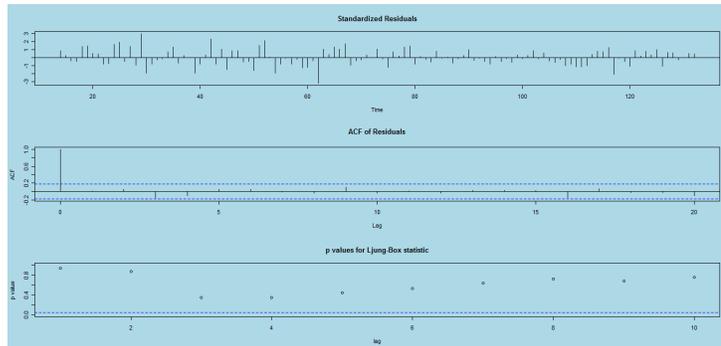


FIGURE 8.1 – Résidus du modèle (8.5)

```
Box.test(fit8$residuals, lag = 10, type = "Box-Pierce")
```

```
Box-Pierce test :data : fit8$residuals
```

```
X-squared = 7.8203, df = 10, p-value = 0.6464
```

Les résidus sont non corrélés.

2. Test de normalité

Programme R

```
library(nortest)
```

```
sf.test(fit8$residuals)
```

Shapiro-Francia normality test

```
data : fit8$residuals
```

```
W = 0.9886, p-value = 0.30036
```

La normalité n'est pas rejetée.

Pouvoir Prédictif

Programme R

```
library(forecast)
```

```
fit81 = Arima(log(AirPassengers)[1 :120], c(0,1,1), seasonal= list(order=c(0,
1 ,1), period=12))
```

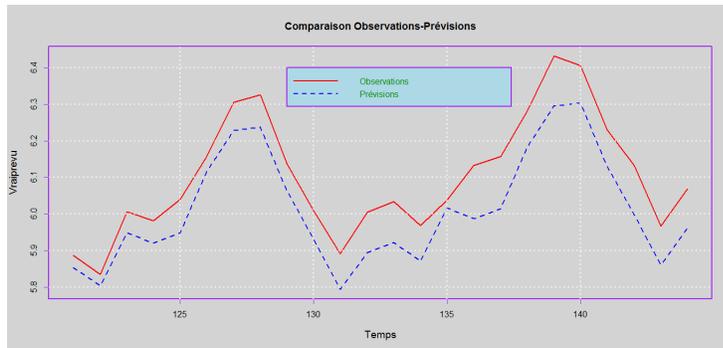
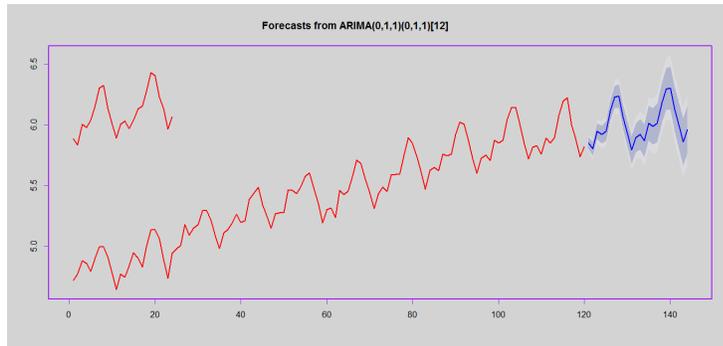
```
fit= forecast(fit81, h=24)
```

```
plot(fit, col="red", lwd=2)
```

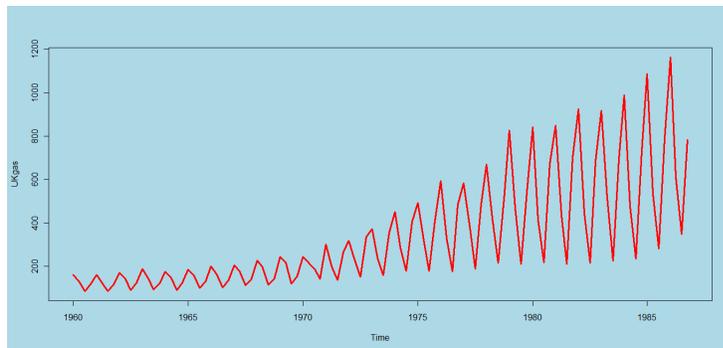
```
lines(log(AirPassengers)[121 :144], lwd=2, col="red" )
```

Exemple 2 : Modélisation de la série du gaz

On considère la série temporelle "UKgas" qui représente la consommation trimestrielle du gaz en angleterre du premier trimestre de l'année 1960 au quatrième trimestre de l'année 1986.



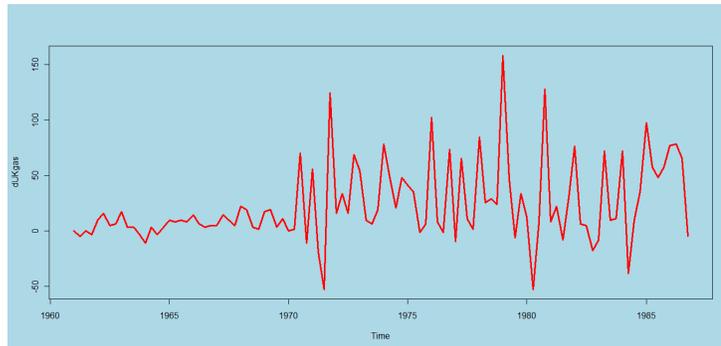
Programme R



$dUKgas < -diff(UKgas, lag=4)$

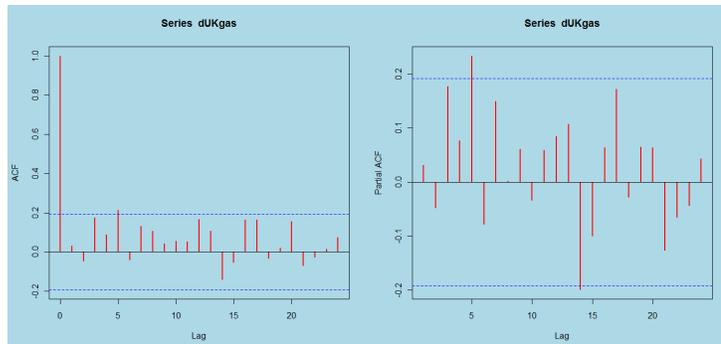
L'effet saisonnier semble disparaître.

Programme R



```
library(tseries)
adf.test(dUKgas)
Augmented Dickey-Fuller Test :data : dUKgas
Dickey-Fuller = -3.7419, Lag order = 4, p-value = 0.02438
alternative hypothesis : stationary
Avec un risque de 5% on rejette la présence de racine unitaire.
```

1. Test d'autocorrélation



```
Programme R
Box.test(dUKgas, lag = 24, type = "Box-Pierce")
Box-Pierce test :data : dUKgas
X-squared = 28.7167, df = 24, p-value = 0.2310
Les observations de dUKgas ne sont pas corrélés
```

2. Test de normalité

```
Programme R
library(nortest)
```

`lillie.test(dUKgas)`

Lilliefors (Kolmogorov-Smirnov) normality test

data : dUKgas

D = 0.1886, p-value = 9.11e-10

La normalité est fortement rejetée. Donc le modèle pour la consommation du gaz est le suivant

$$(1 - B^4) UKgas_t = 23.19712 + \varepsilon_t$$

Pouvoir prédictif

