

N° d'ORDRE : 13/2013- M/MT

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DES SCIENCES ET TECHNOLOGIE DE HOUARI BOUMEDIENNE
FACULTÉ DES MATHÉMATIQUES



MEMOIRE
Présenté pour l'obtention du Diplôme de Magister.

EN RECHERCHE OPÉRATIONNELLE
SPÉCIALITÉ : MATHÉMATIQUE DISCRÈTE & OPTIMISATION

Par : Meriem BENSALLOUA

Sujet

**Estimation des Paramètres de Dérive dépendants
du temps d'une Equation différentielle Stochastique**

Soutenu publiquement le 07/02/2013, devant le jury composé de :

Mr- M.MOULAI	Professeur ,	à l'U.S.T.H.B.	Président
Mr- K.BOUKHETALA	Professeur ,	à l'U.S.T.H.B.	Directeur de thèse
Mr- R.OUAFI	Professeur ,	à l'U.S.T.H.B.	Examineur
Mr- A.TATACHAK	Maître de Conférences A ,	à l'U.S.T.H.B.	Examineur

Table des matières

Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Remerciements.	vii
Résumé	viii
Introduction Générale	1
1 Généralités, Le Mouvement Brownien, et L'intégrale Stochastique	4
1.1 Introduction	5
1.2 Généralités	5
1.2.1 Tribus	5
1.2.2 Filtration	5
1.2.3 Processus aléatoires, Fonctions mesurables et progressivement mesurables	6
1.2.4 Processus Adapté	6
1.2.5 Martingale à temps continu	7
1.2.6 Processus de Markov	7
1.3 Mouvement Brownien	7
1.3.1 Promenade aléatoire	8
1.3.2 Continuité des trajectoires	9
1.3.3 Le mouvement brownien comme processus gaussien	10
1.3.4 Mouvement Brownien et bruit blanc	11
1.3.5 Approximation formelle par un bruit blanc gaussien	12
1.3.6 Martingales browniennes	12
1.4 Intégrale stochastique et calcul d'Itô	14
1.4.1 Construction de l'intégrale stochastique	14
1.4.2 Différentielle Stochastique	17
1.4.3 Formule d'Itô	17
1.4.4 Différentielle Stochastique Multidimensionnelle	18
1.5 processus de diffusion	20
1.5.1 Description générale	20
1.5.2 Propriétés d'un processus de diffusion	21

1.5.3	Transformation élémentaire d'un processus	21
1.6	Conclusion	23
2	Equations Différentielles Stochastiques	24
2.1	Introduction	25
2.2	Définition	25
2.3	Solution faible et forte d'une Equation Différentielle Stochastique	26
2.4	Existence et unicité de la solution	27
2.5	Equations différentielles stochastiques de type linéaire	31
2.5.1	Résolution des EDS de type linéaire	31
2.5.2	Exemples d'applications d'une EDS linéaire	35
2.6	Propriété de Markov des solutions des EDS	37
2.7	Simulation des Equations Différentielles Stochastiques	38
2.7.1	Discrétisation Exacte	39
2.7.2	Discrétisation Approximative	40
2.8	Conclusion	42
3	Estimation des paramètres de dérive d'une Equation Différentielle Stochastique	43
3.1	Introduction	44
3.2	Modèle Mathématique	44
3.2.1	Existence et unicité de la solution de l'équation (3.1)	44
3.3	Structure statistique du problème d'estimation	46
3.3.1	Théorème de Girsanov	47
3.4	La fonction de log-vraisemblance pour l'estimation des paramètres de dérive dépendants du temps	48
3.5	Existence de l'estimateur du maximum de vraisemblance	49
3.6	Application Numérique "Simulation exacte de la trajectoire $(\pi_t)_{t \geq 0}$ "	50
3.7	Conclusion	51
4	Métaheuristiques et Inférences Statistiques	52
4.1	Introduction	52
4.2	Algorithmes génétiques	53
4.3	Présentation des Algorithmes Génétiques (GA)	53
4.4	Codage et population initiale	54
4.4.1	Codage Binaire	54
4.4.2	Codage Réel	55
4.4.3	Codage en base n	55
4.5	Algorithmes Génétiques à Codage Binaire (BCGA)	56
4.5.1	Evaluation	56
4.5.2	Opérateur de Sélection	56
4.5.3	Opérateur de Reproduction avec Croisement et Mutation	59
4.6	Algorithmes Génétiques à Codage Réel (RCGA)	64
4.6.1	Le Croisement	64

4.6.2	La mutation	65
4.7	RCGA pour l'estimation des paramètres de dérive dépendant du temps d'une EDS	66
4.7.1	Formulation du problème d'optimisation	66
4.7.2	Présentation de l'Algorithmes Génétiques à Codage Réel	67
4.7.3	Structure de l'Algorithmes Génétiques à Codage Réel (RCGA)	68
4.7.4	Implémentation de l'Algorithmes Génétiques à Codage Réel	69
4.7.5	Application Numérique	75
4.7.6	Comparaisons Expérimentales	78
4.8	Conclusion	85
<i>Annexe</i>		86
4.9	Annexe A : Scaling	86
4.10	Annexe B : Codage	87
4.11	Annexe C : Code Source des opérateurs de La RCGA	88
<i>CONCLUSION GÉNÉRALE</i>		90
<i>Bibliographie</i>		91

Table des figures

1.1	Promenade aléatoire	10
2.1	Discrétisation Exacte (Modèle de Black et Scholes)	40
2.2	Discrétisation Approximative (Schéma d'Euler)	41
2.3	Discrétisation Approximative (Schéma de Milstein)	42
3.1	Discrétisation par le Schéma d'Euler sur le système (3.1)	51
4.1		58
4.2		60
4.3	One-point crossover	60
4.4	Two-point crossover	60
4.5		61
4.6	N-points crossover	61
4.7	Uniform crossover	61
4.8		62
4.9	Principe du croisement discret	64
4.10	Structure de la RCGA	68
4.11	La solution π_{τ_n} simulée pour 100 individus à 1000 discrétisation	71
4.12	La solution π_{τ_n} simulée pour 100 individus à 10000 discrétisation	71
4.13	Evaluation de $(\phi_1^{(\tau_n)}, \phi_2^{(\tau_n)})$ par rapport à L'_{τ_n}	72
4.14		75
4.15		76
4.16		76
4.17		77
4.18		77
4.19		78
4.20	Test 1 : $\theta = 1, \gamma = 50$	79
4.21	Test 2 : $\theta = 1, \gamma = 100$	79
4.22	Test 3 : $\theta = 1, \gamma = 500$	80
4.23	Test 4 : $\theta = 10, \gamma = 50$	80
4.24	Test 5 : $\theta = 10, \gamma = 100$	81
4.25	Test 6 : $\theta = 10, \gamma = 500$	81
4.26	Test 7 : $\theta = 100, \gamma = 50$	82

4.27	Test 8 : $\theta = 100, \gamma = 100$	82
4.28	Test 9 : $\theta = 100, \gamma = 500$	83

Liste des tableaux

Remerciements.

En premier lieu, je tiens à exprimer toute ma reconnaissance à mon directeur de mémoire, Kamel BOUKHETALA, avec qui j'ai eu un énorme plaisir à travailler. Outre sa grande culture générale et mathématique, j'ai particulièrement apprécié son souci de pédagogie et son goût pour les raisonnements clairs et subtils qu'il aimait partager.

Je remercie vivement les Professeurs Mustapha MOULAI, Rachid OUAFI et Abdelkader TAT-CHAK d'avoir accepté de composer mon jury et de l'intérêt qu'ils ont témoigné pour ce travail.

Des remerciements plus particuliers sont adressés à Abderahim KESSIRA, Isma DAHMANI et Scherazade ZERROUKHAT pour leurs soutiens et leurs encouragements constants.

Je remercie ma famille et plus particulièrement mes parents pour le soutien moral et financier qu'ils m'ont apporté durant mes études. Ce travail leur est entièrement dédié.

RÉSUMÉ

Dans ce travail, on étudie le problème de l'estimation des paramètres de dérive dépendants du temps d'un système dynamique régi par une équation différentielle stochastique non linéaire unidimensionnelle excitée par des bruits brownien.

D'une part, on montre que si les paramètres des bruits sont connus on peut, sous certaines conditions, construire une structure statistique. Puis on utilise le lemme de Randon-Nikodym pour approcher le problème d'estimation par la méthode du maximum de vraisemblance. On donne la forme générale de la fonction log-vraisemblance, et de sa dérivée première et seconde. D'autres part, on formule le problème d'estimation des paramètres de dérive d'une EDS analytiquement par un problème d'optimisation (maximisation de log-vraisemblance), où les paramètres de dérive sont les paramètres inconnus qu'on cherche à déterminer.

Dans la pratique, ce problème est un problème d'optimisation stochastique : sa fonction objectif (log vraisemblance) évolue au cours du temps. L'approche principalement adoptée dans cette étude consiste à adapter des algorithmes génétiques à codage réel RCGA pour l'estimation des paramètres qui sont en fonction du temps.

Les RCGA, offrent généralement l'avantage d'être mieux adaptées aux problèmes d'optimisation numérique continu. ils sont plus pratique, plus rapide, plus précis (surtout pour les larges domaines qui requièrent une représentation trop longue avec les algorithmes génétiques à codage binaire BCGA) ; ils sont plus consistants de générations en générations, car les opérateurs agissent directement sur le phénotype et non pas sur le génotype.

Mots clés : Equation Différentielle Stochastique, Estimation des paramètres, log-vraisemblance, l'algorithme génétique à codage réel.

INTRODUCTION GÉNÉRALE

Les équations différentielles stochastiques (EDS) sont un choix naturel pour modéliser l'évolution temporelle des systèmes dynamiques qui sont soumis à des influences aléatoires. Par exemple, en physique la dynamique des ions dans les conducteurs super-ioniques sont modélisés par l'équation de Langevin (Dieterichet et al (1980)). On retrouve d'autres applications en mécanique (Gelb (1974)), en biologie (Jennrich et Bright (1976)), en médecine (Jones (1984)), en économétrie (Bergstom (1976,1988)), en finance (Black et Scholes (1973)), en géophysique (Arato (1982)) et en océanographie (Alder et Al (1996)).

Le travail que nous présentons dans le cadre de cette étude, concerne une classe particulière de modèles stochastiques du système dynamique évoluant en temps continu. Cette classe est décrite par des équations différentielles stochastiques non linéaire unidimensionnelle excitées par des bruits brownien, en présence de fonction de commande (ou contrôle ou stratégie).

Soit

$$\begin{cases} dX_t = (\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t))dt + \sigma dW_t \\ X_0 = 0, \quad t \geq 0 \end{cases} \quad (1)$$

Où X_t est l'état du système à l'instant t , $g(X_t)$ est le contrôle appliqué au système à l'instant t , W_t est un mouvement brownien standard. $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ sont des paramètres de dérive dépendants du temps qui sont inconnus, σ est le paramètre de diffusion que nous supposons constant dans notre étude.

En pratique, on rencontre une variété d'exemples concrets qui peuvent être modélisés par l'équation (1). Par exemple en électronique, dans un système de communication le signal d'entrée est une équation linéaire contrôlée multidimensionnelle et le signal de sortie est une équation non linéaire [6].

Nous nous intéressons principalement au problème d'estimation de paramètres de dérive dépendants du temps au vu de l'observation d'une trajectoire de l'état du système pour un contrôle choisi dans une classe qu'on définira. Diverses méthodes d'estimations des paramètres inconnus peuvent être utilisées, Breton (1977) et Boukhatala (1984) ont utilisé la méthode de maximum de vraisemblance pour l'estimation des paramètres de dérive constants.

L'estimation des paramètres des modèles en temps continu n'est pas immédiate. De nombreux processus, tels que le modèle (1) n'admet pas de discrétisation exacte [6], et comme sa fonction de vraisemblance déterminée est trop complexe, à calculer, on est alors conduit à utiliser des méthodes heuristiques, comme l'algorithme génétique.

Les algorithmes génétiques offrent généralement l'avantage d'être mieux adaptés aux problèmes d'optimisation numérique continus. Ils sont très utilisés pour la résolution de problème d'optimisation ils sont reconnus comme des outils très efficace, plus particulièrement lorsque les paramètres à résoudre sont compliqués et que la fonction à optimiser n'est pas régulière ou que ses dérivées soient inaccessibles, mal conditionnées ou complexe à calculer, et que les méthodes d'optimisation classique soient totalement inadaptées.

Dans notre étude, le problème d'estimation de modèle (1) se formule mathématiquement par un problème d'optimisation (maximisation de log vraisemblance), où les paramètres de dérives sont les paramètres inconnus qu'on cherche à déterminer.

Le premier pas dans l'implémentation des algorithmes génétiques est de créer une population d'individus initiaux. En effet, les algorithmes génétiques agissent sur une population d'individus, et non pas sur un individu isolé. Dans notre problème la population initiale est composé de N individus, chaque individu a deux paramètres qui sont les paramètres de dérive dépendant du temps, les valeurs des paramètres sont créés aléatoirement.

Par analogie avec la biologie, chaque individu de la population est codé par un chromosome ou génotype (Hollond 1975) en ce qui concerne la représentation binaire d'un individu, et phénotype pour ce qui est de sa valeur réelle correspondante dans l'espace de recherche.

Dans notre problème nous nous basons plus sur un codage binaire des paramètres à optimiser, mais nous travaillons directement sur les paramètres eux mêmes (Jean-Michel Renders (1995)). Ces versions des Algorithmes Génétiques, que nous appellerons Algorithmes Génétiques à Codage Réel (RCGA), offrent généralement des outils d'optimisation adaptés à des fonctions, des problèmes difficiles, complexe, irrégulier par rapport aux algorithmes génétiques à codage binaire (BCGA) qui ont cependant un coût calculatoire très important (Michalewicz 1992).

Les algorithmes génétiques à codage réel (RCGA), sont souvent utilisés pour estimer les paramètres dans divers applications. par exemple : (Jamie Alcock et Kevin Burrage (2003)), en estimation des paramètres d'une équation différentielle ordinaire, (A.Khalik) en 2007 pour le problème d'identification des paramètres.

Dans le présent mémoire, Nous résolvons les deux problèmes suivants :

- a. Déterminer la fonction log-vraisemblance de modèle (1).
- b. Estimer les paramètres de dérive dépendants du temps par une adaptation de la méthode d'estimation par l'algorithme génétique à codage réel (RCGA).

Dans un premier temps, nous présentons au prochain chapitre les principaux aspects et définition des processus stochastique, plus particulièrement le mouvement brownien. Ce dernier

constitue la base de la théorie classique des EDS et comme sa trajectoire est non différentiable alors il est indispensable de se tourner vers une nouvelle description mathématique qui est l'intégration stochastique. Notons que cette intégrale n'obéit pas une règle de calcul différentiel classique étant donné que le processus de Wiener est à variation non bornée ceci explique le passage vers un nouveau calcul, dit "calcul d'Itô", basé sur la formule d'Itô. Nous terminons ce chapitre en donnant un petit aperçu sur le processus de diffusion qui est solution d'une EDS.

Dans le deuxième chapitre, nous présentons les caractéristiques des équations différentielles stochastiques de type Itô, telles que l'existence d'une solution forte et faible. Par la suite nous parlerons de la méthode de résolution des EDS basée essentiellement sur le calcul d'Itô. Les équations de type linéaire seront traitées en premier lieu en donnant les solutions exactes de divers EDS. Nous concluons ce chapitre par un aperçu sur l'approximation par discrétisation du temps sur laquelle se base l'approche numérique de résolution des EDS.

Dans le troisième chapitre, nous traitons un exemple d'identification de paramètres de dérive d'une équation différentielle stochastique linéaire dans un cas unidimensionnel, dans le but de construire une structure statistique pour le problème d'estimation des paramètres de dérive qui sont en fonctions du temps. Par la suite nous donnons la structure statistique dans la quelle sera résolu le problème d'estimation. En vertu du théorème de Girsanov cette structure est dominée par une mesure qui ne dépend pas des paramètres inconnus. Puis nous utilisons le lemme de Randon-Nikodym pour approcher le problème d'estimation par la méthode du maximum de vraisemblance. Nous donnons la forme générale de la fonction log vraisemblance, sa dérivée première et seconde. et une simulation numérique du modèle par discrétisation du temps.

Le quatrième chapitre est consacré aux algorithmes génétiques. Nous présentons une description détaillée de l'algorithme génétique simple dans laquelle nous rappelons les définitions relatives à leur fonctionnement. Par la suite nous introduisons les algorithmes génétiques à codage réel, version amélioré des algorithmes génétiques simple. Nous concluons par une application numérique de l'algorithme génétiques à codage réel pour estimer les paramètres de dérive dépendants du temps de modèle étudié.

1

Généralités, Le Mouvement Brownien, et L'intégrale Stochastique

1.1 Introduction

Le mouvement brownien est un phénomène aléatoire jouant un rôle fondamental dans la théorie des processus stochastiques. Son évolution au cours du temps est extrêmement désordonnée. Il en résulte que la construction d'une intégrale par rapport à ce processus ne rentre pas dans le cadre de l'intégration classique. Dans ce chapitre, nous introduirons la théorie du calcul stochastique, ou calcul d'Itô, du nom d'un célèbre mathématicien japonais du 20^{ème} siècle, nous permettant de donner une définition probabiliste à cette intégrale, mais aussi d'étudier des équations différentielles perturbées aléatoirement par ce type d'intégrales. Les solutions de ces équations, que l'on appelle équations différentielles stochastiques, définissent de nouveaux processus, les processus de diffusion, qui sont à la base du calcul probabiliste moderne.

1.2 Généralités

1.2.1 Tribus

Définition 1.1 Soit Ω est un espace fondamentale dont les éléments sont notés ω et $\mathfrak{F}$ une tribu sur Ω et $(\Omega, \mathfrak{F}, P)$ un espace probabilisé (qu'on suppose complet). Une tribu ou $(\sigma - \text{algre})$ sur Ω est une famille de parties de Ω telle que

- a. $\emptyset \in \mathfrak{F}$
- b. $\forall A \in \mathfrak{F}$ alors $C_A \in \mathfrak{F}$ (stable par passage complémentaire)
- c. stable par union et intersection dénombrable.

Notation

Pour tout n dans $\mathbb{N}^*$, on note par $B(\mathbb{R}^n)$ (ou $B_{\mathbb{R}^n}$), la tribu borélienne sur $\mathbb{R}^n$ et par $\sigma(X)$ la plus petite tribu qui rend une variable aléatoire X mesurable et on a $\sigma(X) = \{X^{-1}(E); E \in B(\mathbb{R}^n)\}$. On appelle processus stochastique, une famille de variables aléatoire $(X_t)_{t \in \mathbb{R}_+}$ définies sur cette espace.

1.2.2 Filtration

Définition 1.2 Soit $(\Omega, \mathfrak{F}, P)$ un espace probabilisé, une filtration $(\mathfrak{F}_t)_{t \geq 0}$ est une famille croissante de sous tribus de $\mathfrak{F}$.

Le tribu $\mathfrak{F}_t$ représente l'information dont on dispose à l'instant t . On dit qu'un processus $(X_t)_{t \geq 0}$ est adapté à $(\mathfrak{F}_t)_{t \geq 0}$, si pour chaque t , X_t est $\mathfrak{F}_t$ mesurable.

Remarque

Dans la suite, les filtrations que l'on considèrera, possèdent la propriété suivante :

$$\text{Si } a \in A \text{ et si } P(a) = 0, \text{ alors pour tout } t, a \in \mathfrak{F}_t.$$

Ceci exprime que $\mathfrak{F}_t$ contient tous les ensembles de mesure nulle de $\mathfrak{F}$. Le but de cette hypothèse technique est de permettre d'affirmer que si $X = Y$ P p.s et que Y est $\mathfrak{F}_t$ -mesurable alors X est aussi $\mathfrak{F}_t$ -mesurable.

1.2.3 Processus aléatoires, Fonctions mesurables et progressivement mesurables

Définition 1.3 Soit S un ensemble. On appelle processus stochastique indexé par S et à valeur dans $\mathbb{R}^d$ une famille $(X_t)_{t \in S}$ d'applications mesurables de $(\Omega, \mathfrak{F})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, pour tout $t \in S$, $(X_t)_{t \in S}$ est une variable aléatoire.

Dans la suite $S = \mathbb{N}$, ce qui correspond au cas discret, ou $S = \mathbb{R}$, ce qui correspond au cas continu. La caractéristique de base d'un processus stochastique est le fait que la loi de la variable X_t soit fonction de t . Une réalisation d'un processus est appelée trajectoire. C'est donc la suite des réalisations des variables aléatoires X_t . Les réalisations d'une même variable aléatoire pouvant être différentes, les réalisations d'un même processus peuvent donner des trajectoires différentes.

Définition 1.4 Soit $(\Omega, \mathfrak{F})$ et $(\Omega', \mathfrak{F}')$ deux espaces mesurables. Une application $f : \Omega \rightarrow \Omega'$ est dite mesurable par rapport à $(\mathfrak{F}, \mathfrak{F}')$ si $f^{-1}(A) \in \mathfrak{F}, \forall A \in \mathfrak{F}'$ où $f^{-1}(A) = \{w \in \Omega / f(w) \in A\}$ si X est une variable aléatoire $\mathfrak{F}$ mesurable et f une fonction Borélienne alors $f(X)$ est $\mathfrak{F}'$ mesurable.

Définition 1.5 Soient X et Y deux processus. X est une modification de Y si, pour tout $t \geq 0$, les v.a X_t et Y_t sont égales P.ps :

$$\forall t \geq 0, p(X_t = Y_t) = 1$$

On dit que X et Y sont indistingables si, P-p.s, les trajectoires de X et de Y sont les mêmes c'est à dire $p(X_t = Y_t, \forall t \geq 0) = 1$

Définition 1.6 Un processus $(X_t)_{t \in \mathbb{R}_+}$ est dit mesurable si la fonction $X_t : (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+)) \times (\Omega, \mathfrak{F}) \rightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ qui à $(t, w) \in (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+)) \times (\Omega, \mathfrak{F})$ associe $X_t(w) \in (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ est mesurable.

Définition 1.7 Un processus $(X_t)_{t \in \mathbb{R}_+}$ est dit progressivement mesurable si $\forall T > 0$, la fonction $X_t : ([0, T], \mathcal{B}_{[0, T]}) \times (\Omega, \mathfrak{F}_T) \rightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ qui à $(t, w) \in ([0, T], \mathcal{B}_{[0, T]}) \times (\Omega, \mathfrak{F}_T)$ associe $X_t(w) \in (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ est mesurable.

1.2.4 Processus Adapté

Définition 1.8 Un processus $(X_t)_{t \in \mathbb{R}_+}$ est dit adapté à la filtration $(\mathfrak{F}_t)_{t \in \mathbb{R}_+}$ si X_t est $\mathfrak{F}_t$ mesurable.

remarque

Un processus progressivement mesurable est mesurable et adapté

1.2.5 Martingale à temps continu

Comme dans le cas des modèles à temps discret, la notion de martingale à temps continu est un outil essentiel pour expliciter **La notion d'arbitrage**¹.

Définition 1.9 Soit $(\Omega, \mathfrak{F}, P)$ un espace probabilisé et $(\mathfrak{F}_t)_{t \geq 0}$ une filtration de cet espace. Une famille adaptée $(M_t)_{t \geq 0}$ de variables aléatoires intégrables, (c'est-à-dire vérifiant $E(|M_t|) < +\infty$ pour tout t) est :

- Une martingale si, pour tout $s \leq t$, $E(M_t/\mathfrak{F}_s) = M_s$.
- Une surmartingale si, pour tout $s \leq t$, $E(M_t/\mathfrak{F}_s) \leq M_s$.
- Une sousmartingale si, pour tout $s \leq t$, $E(M_t/\mathfrak{F}_s) \geq M_s$.

Remarque

On déduit de cette définition que, si $(M_t)_{t \geq 0}$ est une martingale, alors $E(M_t) = E(M_0)$, pour tout t .

1.2.6 Processus de Markov

Définition 1.10 On dit qu'un processus est de Markov, si son comportement dans le futur ne dépend du passé qu'à travers le présent.

mathématiquement cela se traduit par : soit X un processus et $(\mathfrak{F}_t)$ sa filtration canonique. On dit que le processus est de Markov si, pour tout t , pour toute variable bornée $Y \in \mathfrak{F}_\infty$ on a $E(Yo\theta/\mathfrak{F}_t) = E(Yo\theta/X_t)$ où o est l'opérateur de translation défini sur les applications coordonnées par $X_uo\theta_s = X_{u+s}$

Remarque

On peut définir aussi un processus de Markov de cette façon :

Pour tout n pour tout fonction bornée g définie sur Δ^n pour $t_1 < t_2 < \dots < t_n$

$E(g(X_{s+t_1}, X_{s+t_2}, \dots, X_{s+t_n})/\mathfrak{F}_s) = E(g(X_{s+t_1}, X_{s+t_2}, \dots, X_{s+t_n})/X_s)$ ceci implique en particulier que pour toute fonction f borélienne bornée

$$E(f(X_t)/\mathfrak{F}_s) = E(f(X_t)/X_s), \forall t > s$$

Le processus est dit de Markov fort si la propriété précédente est vraie pour tout couple de temps d'arrêt finis T, S avec $T > S$

1.3 Mouvement Brownien

Le botaniste Robert Brown observe en 1928 le mouvement irrégulier des particules de pollen en suspension dans l'eau. En 1877, Delsaux explique les changements incessants de direction de

1. L'arbitrage est une combinaison de plusieurs opérations permettant de réaliser un Bénéfice sans Risque (en théorie au moins !) en tirant partie des seules imperfections susceptibles d'apparaître entre différents marchés. L'arbitrage permet d'assurer l'égalité des Prix à un même moment. Il assure la fluidité entre les différents marchés et contribue à leur liquidité. C'est l'opération de base qui permet de garantir l'efficacité des marchés

trajectoire par les chocs entre les particules de pollen et les molécules d'eau. Un mouvement de ce type est qualifié de "mouvement au hasard".

En 1900, Bachelier, en vue d'étudier les cours de la bourse met en évidence le caractère "markovien" du mouvement Brownien : la position d'une particule à l'instant $t + s$ dépend de sa position en t , et ne dépend pas de sa position avant t . Il convient d'insister sur le caractère précurseur de Bachelier et le fait que la théorie du mouvement Brownien a été développée pour la Bourse, avant de l'être pour la physique.

En 1905, Einstein détermine la densité de transition du mouvement Brownien par l'intermédiaire de l'équation de la chaleur et relie ainsi le mouvement Brownien et les équations aux dérivées partielles de type parabolique. Dans la même année, Smoluchowski décrit le mouvement Brownien comme une limite de promenade aléatoire.

La première étude mathématique rigoureuse est faite par N. Wiener (1923) qui exhibe également une démonstration de l'existence de Brownien. P. Lévy (1948) s'intéresse aux propriétés fines des trajectoires du Brownien. Depuis, le mouvement Brownien continue de passionner les probabilistes, aussi bien pour l'étude de ses trajectoires que pour la théorie de l'intégration stochastique (Wiener, Itô, Watanabe, Meyer, Yor, LeGall, Salminen, Durrett, Chung, Williams, Knight, Pitman,...).

Nous donnons quelques éléments mathématiques nécessaires à la compréhension des modèles à temps continu. En particulier, nous introduirons le mouvement brownien, qui est l'outil majeur du modèle de Black et Scholes et sert à construire la plupart des modèles d'actifs en finance. Puis nous étendrons la notion de l'intégrale stochastique d'Itô et nous introduirons le calcul différentiel qui lui est associé : le calcul d'Itô.

Définition 1.11 Soit le processus stochastique $\{X_t, t \geq 0\}$. X_t est un processus de Wiener s'il vérifie les conditions suivantes :

- a. $X_t \rightsquigarrow N(0, c^2 t), t \geq 0$ et $X(0) = 0$
- b. $\{X_t, t \geq 0\}$ a des accroissements indépendants : $X_t - X_s \perp X_s, \forall t > s \geq 0$
- c. $\{X_t, t \geq 0\}$ a des accroissements stationnaires : $(X_t - X_s) \rightsquigarrow N(0, c^2(t - s))$

Définition 1.12 Soit le processus stochastique $\{X_t, t \geq 0\}$. X_t est un mouvement Brownien standard s'il vérifie les conditions suivantes avec $c = 1$:

- a. $X_t \rightsquigarrow N(0, t), t \geq 0$ et $X(0) = 0$
- b. $\{X_t, t \geq 0\}$ a des accroissements indépendants : $X_t - X_s \perp X_s, \forall t > s \geq 0$
- c. $\{X_t, t \geq 0\}$ a des accroissements stationnaires : $(X_t - X_s) \rightsquigarrow N(0, (t - s))$

1.3.1 Promenade aléatoire

On peut montrer que le mouvement Brownien s'obtient comme limite de promenades aléatoires renormalisées. Cette propriété est exploitée pour des simulations.

Soit, sur un espace de probabilité $(\Omega, \mathfrak{F}, P)$ une famille de variables aléatoires de Bernoulli indépendantes équadistribuées

$$P(X_i = 1) = P(X_i = -1) = 1/2, i \in N^*$$

On associe à cette famille la suite $(S_n, n \geq 0)$ définie par

$$S_0 = 0, S_n = \sum_{i=1}^n X_i$$

S_n est appelée promenade aléatoire. (Jeu de pile ou face).

On a $E(S_n) = 0, \text{Var}(S_n) = n$

Remarquons que la suite $(S_m - S_n, m \geq n)$ est indépendante de $(S_0, S_1, \dots, S_n)$ et que $S_m - S_n$ possède la même loi que S_{m-n} .

On procède alors à une double renormalisation. Soit N fixé

- on ramène l'intervalle de temps $[0, N]$ à $[0, 1]$
- on change l'échelle des valeurs prises par S_n

Plus précisément, on définit une famille de variables aléatoires indexées par les réels de la forme $\frac{k}{N}, k \in N$, par

$$U_{k/N} = \frac{1}{\sqrt{N}} S_k$$

On a $E(U_{k/N}) = 0$ et $\text{Var}(U_{k/N}) = \frac{k}{N}$

Les propriétés d'indépendance et de stationarité de la promenade aléatoire restent vérifiées, soit

- si $k \geq k', U_{k/N} - U_{k'/N}$ est indépendante de $(U_{p/N}; p \leq k')$
- si $k \geq k', U_{k/N} - U_{k'/N}$ a même loi que $U_{(k-k')/N}$

On définit un processus à temps continu $(U_t, t \geq 0)$ à partir de $U_{k/N}$ en imposant à la fonction $t \rightarrow U_t$ d'être affine entre k/N et $(k+1)/N$. pour cela, N étant fixé, on remarque que pour tout $t \in \mathbb{R}_+$ il existe $k(t) \in N$ unique tel que $\frac{k(t)}{N} \leq t \leq \frac{k(t)+1}{N}$ et on pose

$$U_t^N = U_{k/N} + N(t - \frac{k}{N})(U_{\frac{k+1}{N}} - U_{\frac{k}{N}})$$

où $k = k(t)$

Pour $t = 1$ on a $U_1^N = \frac{1}{\sqrt{N}} S_N$. Le théorème central-limite implique que U_1^N converge en loi vers une variable aléatoire gaussienne centrée réduite.

On montre alors que le processus U^N converge (au sens de la convergence en loi) vers un mouvement Brownien X .

En particulier $U_t^N \xrightarrow{L} X_t$ et $(U_{t_1}^N, \dots, U_{t_k}^N) \xrightarrow{L} (X_{t_1}, \dots, X_{t_k})$ pour tout k -uplet $(t_1, \dots, t_k)$

1.3.2 Continuité des trajectoires

Un processus aléatoire $\{X_t, t \geq 0\}$ est continu lorsque

$$\lim_{h \rightarrow 0} X_{t+h} - X_t = 0$$

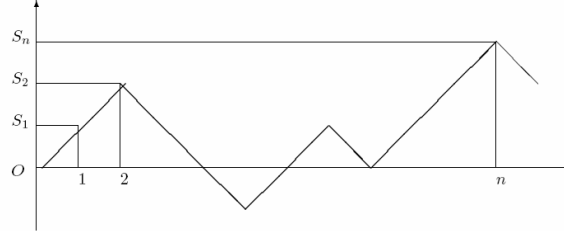


FIGURE 1.1 – Promenade aléatoire

Selon le type de convergence de cette variable aléatoire, on obtient une continuité plus ou moins forte. La plus fiable des notions de continuité est liée à la convergence en loi. Elle est évidemment vérifiée. Nous allons démontrer une continuité en probabilité pour le mouvement brownien standard.

Proposition 1.13 Soit $\varepsilon > 0$ et $\{X_t, t \geq 0\}$ un mouvement brownien standard (dimension 1). On a

$$\lim_{h \rightarrow 0} \frac{1}{h} P(|X_{t+h} - X_t| > \varepsilon) = 0$$

Démonstration. Soit $h > 0$. Par définition, l'accroissement brownien $X_{t+h} - X_t$ admet pour loi $N(0, 1)$. Donc

$$\begin{aligned} \frac{1}{h} P(|X_{t+h} - X_t| > \varepsilon) &= \frac{2}{h} \int_{\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi h}} e^{-\frac{x^2}{2h}} dx \\ &< \frac{2}{h} \int_{\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{h^{3/2}} e^{-\frac{\varepsilon x}{2h}} dx \\ &= \frac{\sqrt{2}}{\sqrt{\pi}} \frac{1}{h^{3/2}} \frac{2h}{\varepsilon} e^{-\varepsilon^2/2h} \end{aligned}$$

Le dernier terme converge vers 0 lorsque $h \rightarrow 0$

1.3.3 Le mouvement brownien comme processus gaussien

Proposition 1.14 Le mouvement brownien standard est un processus aléatoire gaussien de moyenne nulle et de fonction de covariance $\forall s, t \in \mathfrak{R}_+, K(s, t) = \min(s, t)$

Démonstration. Soit $X_t, t \geq 0$ un mouvement brownien standard. Soient $0 < t_1 < t_2 < \dots < t_n$. Les variables réelles $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes et gaussiennes. Le vecteur aléatoire

$$Z = (X_{t_1}, \dots, X_{t_n})^T$$

se déduit par une transformation linéaire : il s'agit donc d'un vecteur gaussien. De plus, pour tout $s \leq t$, nous avons, par indépendance des accroissements,

$$E[X_s X_t] = E[X_s(X_t - X_s)] + E[X_s^2] = 0 + s = \min(s, t)$$

1.3.4 Mouvement Brownien et bruit blanc

Soit un mouvement brownien $(X_t)_{t \geq 0}$ défini sur un espace filtré $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \geq 0}, P)$. On a les deux résultats suivants :

- a. $\lim_{t \rightarrow \infty} \frac{X_t}{t} = 0$ P p.s.
- b. $\lim_{t \rightarrow \infty} \sup \frac{X_t}{\sqrt{2t \log \log(t)}} = +1$ et $\lim_{t \rightarrow \infty} \inf \frac{X_t}{\sqrt{2t \log \log(t)}} = -1$ P p.s.

Il existe t_n tel que si $n \rightarrow \infty$ alors $t_n \rightarrow \infty$ et il existe t'_n tel que si $n \rightarrow \infty$ alors $t'_n \rightarrow \infty$. D'après la loi du logarithme itérée, on aura presque sûrement les résultats suivants :

$$\lim_{t_n \rightarrow \infty} \sup \frac{X_{t_n}}{\sqrt{2t_n \log \log(t_n)}} = +1$$

$$\lim_{t'_n \rightarrow \infty} \inf \frac{X_{t'_n}}{\sqrt{2t'_n \log \log(t'_n)}} = -1$$

proposition 1.15 Si $(X_t)_{t \geq 0}$ est un processus de mouvement brownien, alors les processus $tX(\frac{1}{t})$ et $X(t+h) - X(h)$, $h \geq 0$ sont des mouvements brownien

On utilise les résultats précédents et on obtient, presque sûrement :

$$\lim_{t \rightarrow 0} \sup \frac{X_t}{\sqrt{(2/t) \log \log(1/t)}} = +1 \quad (1.1)$$

$$\lim_{t \rightarrow 0} \inf \frac{X_t}{\sqrt{(2/t) \log \log(1/t)}} = -1 \quad (1.2)$$

pour $h = s$ il existe t_n tel que d'après (1.1) et (1.2) nous avons :

$$\frac{X(t+s) - X(s)}{t_n} \geq (1 - \varepsilon) \sqrt{\frac{2 \log \log(1/t_n)}{t_n}} \quad (1.3)$$

$$\frac{X(t+s) - X(s)}{t_n} \leq (-1 + \varepsilon) \sqrt{\frac{2 \log \log(1/t_n)}{t_n}} \quad (1.4)$$

Si $n \rightarrow \infty$ alors :

$$(1 - \varepsilon) \sqrt{\frac{2 \log \log(1/t_n)}{t_n}} \rightarrow +\infty$$

$$(-1 + \varepsilon) \sqrt{\frac{2 \log \log(1/t_n)}{t_n}} \rightarrow -\infty$$

Donc, d'après ces résultats, le m.b. est non différentiable au point $t = s$, alors c'est une fonction aléatoire P p.s. non différentiable.

proposition 1.16 Le M.B.S. est une fonction aléatoire P p.s. non différentiable et à variations non bornées.

Pour cela, on essaye d'approximer formellement la dérivation du mouvement brownien en utilisant la notion du bruit blanc puisque il a des caractéristiques connues.

1.3.5 Approximation formelle par un bruit blanc gaussien

Définition 1.17 Un processus $\{Z_t\}_{t \geq 0}$ est dit un bruit blanc (White noise) d'espérance nulle et de variance σ^2 , et on note $\{Z_t\}_{t \geq 0} \rightarrow WN(0, \sigma^2)$ si seulement si :

- $E(Z_t) = 0 \forall t \in T \subset \mathbb{R}$
- $Y_Z(h) = cov(Z_t, Z_{t+h}) = \sigma^2$ si $h = 0$

Si les variables Z_t sont indépendantes et identiquement distribuées. Par analogie avec le bruit blanc en temps discret, on cherche à définir un processus stochastique $B_t (t \geq 0)$ vérifiant : $\forall t > 0$ et $\forall h > 0$

- $E(B_t) = 0$
- $E(B_t B_{t+h}) = \delta_0(h)$

Où δ_0 est la mesure de Dirac en 0. Un tel processus n'existe pas son idéalisation est la dérivée d'un mouvement brownien : Pour $\Delta t > 0$, considérons le processus : $C_t = \frac{X_{t+\Delta t} - X_t}{\Delta t}$ Il est facile de montrer grâce aux propriétés du mouvement brownien de dimension 1 que :

- a. $E(C_t) = 0$
- b. $E(C_t C_{t+h}) = \frac{1}{\Delta t} (1 - \frac{|h|}{\Delta t}) \mathbb{I}_{[0,1]}(\frac{|h|}{\Delta t})$

Lorsque $\Delta t \rightarrow 0 : E(C_t C_{t+h}) \rightarrow \delta_0(h)$. Il est donc clair que la dérivée formelle $\frac{dX_t}{dt}$ a les propriétés d'un bruit blanc gaussien, on peut donc considérer l'approximation du bruit blanc en terme de covariance $B_t = \frac{dX_t}{dt}$

1.3.6 Martingales browniennes

proposition 1.18 Si $(X_t)_{t \geq 0}$ est un $\mathfrak{F}_t$ -mouvement brownien standard :

- a. X_t est une $\mathfrak{F}_t$ -martingale.
- b. $X_t^2 - t$ est une $\mathfrak{F}_t$ -martingale.
- c. $(\exp(\sigma X_t - (\sigma^2/2)t))$ est une $\mathfrak{F}_t$ -martingale.

Dans les finances, ces trois processus peuvent être l'objet d'un projet d'investissement ou d'un risque.

Démonstration :

1- Soit $(X_t)_{t \geq 0}$ un $\mathfrak{F}_t$ -mouvement brownien standard et $0 \leq s \leq t$ alors $E(X_t - X_s / \mathfrak{F}_s) = E(X_t - X_s) = 0$ et $E(X_t / \mathfrak{F}_s) - E(X_s / \mathfrak{F}_s) = X_s - 0 = X_s$

Et donc, M_t est $\mathfrak{F}_t$ -martingale.

2- Sous les mêmes hypothèses qui précèdent, on a alors :

$$\begin{aligned} E(X_t^2 - X_s^2 / \mathfrak{F}_s) &= E((X_t - X_s)^2 + 2X_s(X_t - X_s) / \mathfrak{F}_s) \\ &= E((X_t - X_s)^2 / \mathfrak{F}_s) + 2X_s E(X_t - X_s / \mathfrak{F}_s) \\ &= E((X_t - X_s)^2 / \mathfrak{F}_s) + 2X_s E(X_t - X_s) \\ &= E(X_{t-s}^2 / \mathfrak{F}_s) \\ &= t - s \end{aligned}$$

Par conséquent

$$E(X_t^2 - t / \mathfrak{F}_s) = E(X_s^2 / \mathfrak{F}_s) - s = X_s^2 - s$$

Et donc, $X_t^2 - t$ est $\mathfrak{F}_s$ - martingale.

3- Soit G une variable aléatoire centrée réduite $\mathbb{N}(0, 1)$, alors pour tout $\lambda \geq 0$, on a :

$$E(e^{\lambda g}) = \int_{-\infty}^{+\infty} e^{\lambda x} e^{-x^2/2} dx / \sqrt{2\pi} = e^{\lambda^2/2} \text{ Ainsi, on a}$$

$$\begin{aligned} E(e^{\{\sigma X_t - \frac{\sigma^2}{2}t\}} / \mathfrak{F}_s) &= E(e^{\{\sigma X_t - \frac{\sigma^2}{2}t + \sigma X_s - \sigma X_s\}} / \mathfrak{F}_s) \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}t\}} E(e^{\sigma(X_t - X_s)} / \mathfrak{F}_s) \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}t\}} E(e^{\sigma(X_t - X_s)}) \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}t\}} E(e^{\sigma X_{t-s}}) \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}t\}} E(e^{\sigma G \sqrt{t-s}}) \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}t\}} e^{\frac{\sigma^2}{2}(t-s)} \\ &= e^{\{\sigma X_s - \frac{\sigma^2}{2}s\}} \end{aligned}$$

Et donc $e^{\{\sigma X_t - \frac{\sigma^2}{2}t\}}$ est une $\mathfrak{F}_t$ -martingale

Si $(M_t)_{t \geq 0}$ est une martingale, la relation $E(M_t / \mathfrak{F}_s) = M_s$ peut être étendue à des temps aléatoires si ces temps sont des temps d'arrêt bornés. Ce résultat porte le nom de théorème d'arrêt.

théorème 1.19 Théorème d'arrêt

Si $(M_t)_{t \geq 0}$ est une martingale continue par rapport à une filtration $(\mathfrak{F}_t)_{t \geq 0}$, et si τ_1 et τ_2 sont deux temps d'arrêt tels que $\tau_1 \leq \tau_2 \leq K$, K étant une constante réelle finie, alors M_{τ_2} est intégrable et :

$$E(M_{\tau_2} / \mathfrak{F}_{\tau_1}) = M_{\tau_1} \quad \text{P.ps}$$

Remarque

Ce résultat entraîne que, si τ est un temps d'arrêt borné, alors $E(M_\tau) = E(M_0)$ (il suffit d'appliquer le théorème en remplaçant l'égalité précédente par :

$$E(M_{\tau_2} / \mathfrak{F}_{\tau_1}) \geq M_{\tau_1} \quad \text{P.ps}$$

Nous allons donner un exemple d'application de ce résultat au calcul des temps d'atteinte d'un point par le mouvement brownien.

proposition 1.20 Soit $(X_t)_{t \geq 0}$ un $\mathfrak{F}_t$ -mouvement brownien, on pose :

$$T_x = \inf\{s \geq 0, X_s = x\}; (x \in \mathbb{R})$$

Alors T_x est un temps d'arrêt P.p.s fini.

Et de plus on a :

$$\mathcal{L}T_x = E(e^{-\lambda T_x}) = e^{-\sqrt{2\lambda}|x|}$$

théorème 1.21 Inégalités de Doob Si $(M_t)_{0 \leq t \leq T}$ est une martingale continue, on a :

$$E(\sup_{0 \leq t \leq T} |M_t|^2) \leq 4E(|M_T|^2).$$

1.4 Intégrale stochastique et calcul d'Itô

Le calcul différentiel donne un cadre à la notion d'équation différentielle ordinaire. Il sert de modèle pour des phénomènes variables dans le temps. Lorsqu'on a voulu ajouter à ces équations des perturbations aléatoires, on a été gêné par la non différentiabilité du MB. Du coup on a commencé par construire une intégrale par rapport au MB, pour ensuite définir la notion d'équation différentielle stochastique. Et il a fallu donner un sens à $\int_0^t f(s, \omega) dW_s$

1.4.1 Construction de l'intégrale stochastique

Soit $(W_t)_{t \geq 0}$ un $\mathfrak{F}_t$ -mouvement brownien standard sur espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \geq 0}, P)$. Nous allons donner un sens à $\int_0^t f(s, \omega) dW_s$ pour une classe de processus $f(s, \omega)$ adaptés à la filtration $(\mathfrak{F}_t)_{t \geq 0}$. On va commencer par construire l'intégrale stochastique sur un ensemble de processus dits élémentaires. Dans toute la suite, on fixe T un réel strictement positif et fini.

Définition 1.22 On appelle processus élémentaire $(H_t)_{0 \leq t \leq T}$ un processus de la forme :

$$H_t(\omega) = \sum_{i=1}^p \phi_i(\omega) \mathbf{1}_{]t_{i-1}, t_i]}(t)$$

où $0 = t_0 < t_1 < \dots < t_p = T$ et ϕ_i est $F_{t_{i-1}}$ -mesurable et bornée.

On définit l'intégrale stochastique d'un processus élémentaire H, le processus continu $(I(H)_t)_{0 \leq t \leq T}$, si $t \in]t_k, t_{k+1}]$ on a :

$$I(H)_t = \sum_{1 \leq i \leq k} \phi_i(W_{t_i} - W_{t_{i-1}}) + \phi_{k+1}(W_t - W_{t_k})$$

Notons que $I(H)_t$ peut s'écrire :

$$I(H)_t = \sum_{1 \leq i \leq p} \phi_i(W_{t_i \wedge t} - W_{t_{i-1} \wedge t}),$$

ce qui prouve la continuité de la fonction $t \mapsto I(H)_t$. on notera $\int_0^t H_s dW_s$ pour $I(H)_t$. On a alors le résultat essentiel suivant :

Proposition 1.23 Si $(H_t)_{0 \leq t \leq T}$ est un processus élémentaire :

- $(\int_0^t H_s dW_s)_{0 \leq t \leq T}$ est une F_t -martingale continue,
- $E((\int_0^t H_s dW_s)^2) = E(\int_0^t H_s^2 ds)$,
- $E(\sup_{t \leq T} |\int_0^t H_s dW_s|^2) \leq 4E(\int_0^T H_0^2 ds)$

Démonstration : Pour démontrer cette proposition nous allons utiliser des processus à temps discret. En effet, pour établir que $(\int_0^t H_s dW_s)$ est une martingale, il suffit de prouver que, pour tout $t > s$:

$$E(\int_0^t H_u dW_u / \mathfrak{F}_s) = \int_0^s H_u dW_u$$

Si l'on ajoute s et t à la subdivision $t_0 = 0 < t_1 < \dots < t_p = T$, et si on pose $M_n = \int_0^{t_n} H_s dW_s$ et $g_n = F_{t_n}$ pour $0 \leq n \leq p$, il suffit de vérifier que M_n est une g_n -martingale. Pour démontrer ceci, remarquons que :

$$M_n = \int_0^{t_n} H_s dW_s = \sum_{i=1}^n \phi_i (W_{t_i} - W_{t_{i-1}})$$

avec ϕ_i qui est g_{i-1} -mesurable. D'autre part $X_n = W_{t_n}$ est une g_n -martingale (en effet, $(W_t)_{t \geq 0}$ est un mouvement brownien). $(M_n)_{n \in [0, p]}$ apparaît donc comme une transformée de la martingale $(X_n)_{n \in [0, p]}$. Il est facile de démontrer que $(M_n)_{n \in [0, p]}$ une martingale.

Le deuxième point s'obtient, en remarquant que :

$$\begin{aligned} E(M_n^2) &= E((\sum_{i=1}^n \phi_i (X_i - X_{i-1}))^2) \\ &= \sum_{i=1}^n \sum_{j=1}^n E(\phi_i \phi_j (X_i - X_{i-1})(X_j - X_{j-1})) \end{aligned}$$

De plus, si $i < j$, on a :

$$\begin{aligned} E(\phi_i \phi_j (X_i - X_{i-1})(X_j - X_{j-1})) &= E(E(\phi_i \phi_j (X_i - X_{i-1})(X_j - X_{j-1}) / g_{j-1})) \\ &= E(E(\phi_i \phi_j (X_i - X_{i-1})(X_j - X_{j-1}) / g_{j-1})) \end{aligned}$$

Comme X_j est une martingale, on a $E(X_j - X_{j-1} / g_{j-1}) = 0$. On en déduit que, si $i < j$:

$$E((\phi_i \phi_j (X_i - X_{i-1})(X_j - X_{j-1}))) = 0$$

Si $j > i$ on obtient le même résultat. Enfin si $i = j$, on a :

$$\begin{aligned} E(\phi_i^2 (X_i - X_{i-1})^2) &= E(E(\phi_i^2 (X_i - X_{i-1})^2 / g_{i-1})) \\ &= E(\phi_i^2 E((X_i - X_{i-1})^2 / g_{i-1})) \end{aligned}$$

Finalement :

$$E((X_i - X_{i-1})^2 / g_{i-1}) = E((W_{t_i} - W_{t_{i-1}})^2) = t_i - t_{i-1}$$

En regroupant ces résultats on obtient :

$$E((\sum_{i=1}^n \phi_i^2 (X_i - X_{i-1})^2) = E(\sum_{i=1}^n \phi_i^2 (t_i - t_{i-1})).$$

La continuité de $t \rightarrow \int_0^t H_s dW_s$ est clair par définition. Le troisième point est une conséquence de l'inégalité de Doob(1.3.7) appliquée à la martingale continue $(\int_0^t H_s dW_s)_{t \geq 0}$

Remarque On pose par définition :

$$\int_t^T H_s dW_s = \int_0^T H_s dW_s - \int_0^t H_s dW_s$$

Si $t \leq T$, et si $A \in \mathfrak{F}_t$, alors $s \rightarrow 1_A 1_{\{t < s\}}, H_s$ reste un processus élémentaire et il est facile de vérifier, que :

$$\int_0^T 1_A H_s 1_{t < s} dW_s = 1_A \int_t^T H_s dW_s \quad (1.5)$$

On vient de définir et donner des propriétés de l'intégrale stochastique pour les processus élémentaires, nous allons maintenant étendre cette intégrale à une classe de processus adaptés :

$$\mathfrak{H} = \{(H_t)_{0 \leq t \leq T}, \text{processus adapté à } (F_t)_{t \geq 0}, E(\int_0^T H_s^2 ds) < +\infty\}$$

Proposition 1.24 Soit $(W_t)_{t \geq 0}$ un F_t -brownien. Alors il existe une unique application linéaire J de $\mathfrak{H}$ dans l'espace des F_t -martingales continues définies sur $[0, T]$, telle que :

- a. Si $(H_t)_{t \leq T}$ est un processus élémentaire, P p.s pour tout $0 \leq t \leq T$ $J(H)_t = I(H)_t$
- b. Si $t \leq T$, $E(J(H)_t^2) = E(\int_0^t H_s^2 ds)$

Cette application linéaire est unique au sens suivant, si J et J' sont deux prolongements linéaires vérifiant les propriétés précédentes alors : P p.s $\forall 0 \leq t \leq T$, $J(H)_t = J'(H)_t$

On note, si $H \in \mathfrak{H}$, $\int_0^t H_s dW_s = J(H)_t$. De plus cette intégrale stochastique vérifie les propriétés suivantes :

Proposition 1.25 si $(H_t)_{0 \leq t \leq T}$ un processus de $\mathfrak{H}$ alors :

- a. On a

$$E(\sup_{t \leq T} |\int_0^t H_s dW_s|^2) \leq 4E(\int_0^T H_s^2 ds)$$

- b. si τ est un $\mathfrak{F}_t$ -temps d'arrêt :

$$P \text{ p.s } \int_0^\tau H_s dW_s = \int_0^T 1_{s \leq \tau} H_s dW_s$$

1.4.2 Différentielle Stochastique

Parallèlement à l'intégrale stochastique, on peut introduire un processus stochastique qui donner la définition formelle de différentielle stochastique avec laquelle on peut modéliser les variations instantanées aléatoires.

Définition 1.26 Soit $(X_t)_{0 \leq t \leq T}$ un processus tel que pour chaque $(t_1, t_2) \in [0, T] \times [0, T]$, $t_1 < t_2$:

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} a(t)dt + \int_{t_1}^{t_2} b(t)dW_t \quad (1.6)$$

Où $a \in C_1([0, T])$ et $b \in C_1([0, T])$. Alors X_t admet une différentielle stochastique

$$dX_t = a(t)dt + b(t)dW_t \quad \text{sur } [0, T] \quad (1.7)$$

Remarque Si X_t admet une différentielle stochastique de la forme (1.7), alors pour tout $t > 0$, on a

$$X_t = X_0 + \int_0^t a(s)ds + \int_0^t b(s)dW_s \quad (1.8)$$

Alors

- a. Les trajectoires de $(X_t)_{0 \leq t \leq T}$ sont continues presque partout ;
- b. Pour $t \in [0, T]$, X_t est $\mathfrak{F}_t = \sigma(W_s, 0 \leq s \leq t)$ -mesurable, ainsi $X_t \in C_1([0, T])$

1.4.3 Formule d'Itô

Le lemme d'Itô, ou encore formule d'Itô est l'un des principaux résultats de la théorie du calcul stochastique. Ce lemme offre un moyen de manipuler le mouvement brownien ou les solutions d'équations différentielles stochastiques (EDS).

La formule d'Itô a été démontrée pour la première fois par le mathématicien japonais Kiyoshi Itô dans les années 1940.

Le mathématicien Wolfgang Doeblin avait de son côté ébauché une théorie similaire avant de se suicider à la défaite de son bataillon en juin 1940. Ses travaux furent envoyés dans un pli cacheté à l'Académie des sciences qui ne fut ouvert qu'en 2000.

Soient $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \geq 0}, P)$ un espace probabilisé filtré et $(W_t)_{t \geq 0}$ un $\mathfrak{F}_t$ -mouvement brownien. on appelle processus d'Itô, un processus $(X_t)_{0 \leq t \leq T}$ à valeur dans $\mathbb{R}$ tel que :

$$X_t = X_0 + \int_0^t a_s \cdot ds + \int_0^t b_s \cdot dW_s \text{ P.ps.} \quad (1.9)$$

Avec :

- X_0 est $\mathfrak{F}_0$ -mesurable
- $(K_t)_{0 \leq t \leq T}$ et $(H_t)_{0 \leq t \leq T}$ des processus adaptés à $\mathfrak{F}_t$
- $\int_0^t |K_s| \cdot ds < +\infty$ P.ps
- $\int_0^t |H_s|^2 \cdot ds < +\infty$ P.ps

Théorème 1.27 (Formule d'Itô)

Si $dX_t = a(t)dt + b(t)dW_t$

et si $f(t, x) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ est une fonction deux fois continûment dérivable et soient f_x, f_{xx} , et f_t ses dérivées partielles, alors la différentielle stochastique de processus $f(t, X_t)$ est donné par

$$df(t, X_t) = \{f_t(t, X_t) + 1/2 f_{xx}(t, X_t)b^2(t) + f_x(t, X_t)a(t)\}dt + f_x(t, X_t)b(t)dW_t \quad (1.10)$$

1.4.4 Différentielle Stochastique Multidimensionnelle

Définition 1.28 Soit $(X_t)_{0 \leq t \leq T}$ un processus m-dimensionnelle et

$$a : [0, T] \times \Omega \rightarrow \mathbb{R}^m, a \in C_{1W}([0, T]), \quad (1.11)$$

$$b : [0, T] \times \Omega \rightarrow \mathbb{R}^{mn}, b \in C_{1W}([0, T]) \quad (1.12)$$

La différentielle stochastique dX_t de X_t est donnée par :

$$dX_t = a(t)dt + b(t)dW_t \quad (1.13)$$

Si, pour tous $0 \leq t_1 < t_2 \leq T$

$$X_{t_2} - X_{t_1} = \int_{t_1}^{t_2} a(t)dt + \int_{t_1}^{t_2} b(t)dW_t$$

Théorème 1.29 (Formule de Itô Multidimensionnelle) Soit $f(t, x) : \mathbb{R}_+ \times \mathbb{R}^m \rightarrow \mathbb{R}$ est une fonction deux fois continûment différentiable sur $\mathbb{R}^m$ et soient $f_x, f_{x_i x_j}$ et f_t ses dérivés partielles. Soit X_t est un processus m-dimensionnel, affecté par une différentielle stochastique

$$dX_t = a(t)dt + b(t)dW_t$$

Où $a = (a_1, \dots, a_m)' \in C_W([0, T])$ et $b = (b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in C_W([0, T])$ alors $f(t, X_t)$ a une différentielle stochastique

$$\begin{aligned} df(t, X_t) &= \left(\{f_t(t, X_t) + \sum_{i=1}^m f_{x_i}(t, X_t)a_i(t) + 1/2 \sum_{l=1}^n \sum_{i,j=1}^m f_{x_i x_j}(t, X_t)b_{il}(t)b_{jl}(t)\} \right) dt \\ &+ \sum_{l=1}^n \sum_{i=1}^m f_{x_i}(t, X_t)b_{il}(t)dW_{lt} \quad (*) \end{aligned}$$

Si on met $a_{ij} = (bb')_{ij}, i, j = 1, \dots, m :$

$$L = 1/2 \sum_{i,j=1}^m a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^m a_i \frac{\partial}{\partial x_i} + \frac{\partial}{\partial t}$$

et d'introduire l'opérateur gradient

$$\nabla_x = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_m} \right)^t$$

Alors, en notation vectorielle, l'équation (*) peut être mise sous la forme

$$df(t, X_t) = Lf(t, X_t)dt + \nabla_x f(t, X_t) \cdot b(t) dW_t \quad (1.14)$$

où $\nabla_x f(t, X_t) \cdot b(t) dW(t)$ est le produit scalaire de deux vecteurs m-dimensionnel

Lemme 1.30 Si $W_1, \dots, W_n$ sont des processus de Wiener qui sont indépendants et

$$dX_{it} = a_i(t)dt + \sum_{j=1}^n (b_{ij}(t)dW_{jt}), \quad i = 1, 2$$

Alors

$$d(X_1 X_2)_t = X_{1t} dX_{2t} + X_{2t} dX_{1t} + \sum_{j=1}^n b_{1j} b_{2j} dt \quad (1.15)$$

Preuve Soient X_{1t} et X_{2t} deux processus qui admettent les différentielles stochastiques

$$dX_{it} = a_i(t)dt + b_i(t)dW_t \quad (1.16)$$

Soit $f(t, X_1, X_2) = X_{1t} X_{2t}$. La différentielle stochastique du produit $Y(t) = X_{1t} X_{2t}$ se repose sur l'indépendance et la dépendance des deux processus de Wiener W_{1t} et W_{2t} .

Dans le premier cas la différentielle (1.16) peut être mise sous la forme vectorielle :

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix} dt + \begin{bmatrix} b_1(t) & 0 \\ 0 & b_2(t) \end{bmatrix} d \begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

et grâce à la formule d'Itô la différentielle du produit sera :

$$dY_t = [a_1(t)X_{2t} + a_2(t)X_{1t}]dt + b_1(t)X_{2t}dW_{1t} + b_2(t)X_{1t}dW_{2t}$$

Ou plus explicitement

$$dY_t = X_{1t}dX_{2t} + X_{2t}dX_{1t} \quad (1.17)$$

Lorsque $W_{1t} = W_{2t} = W_t$, la différentielle du vecteur $X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}$ est

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix} dt + \begin{bmatrix} b_1(t) \\ b_2(t) \end{bmatrix} dW_t$$

Par conséquent, la formule d'Itô donne :

$$\begin{aligned} dY_t &= [a_1(t)X_{2t} + a_2(t)X_{1t} + b_1(t)b_2(t)]dt + [b_1(t)X_{2t} + b_2(t)X_{1t}]dW_t \\ &= X_{1t}(a_2(t)dt + b_2(t)dW_t) + X_{2t}(b_1(t)dt + b_1(t)dW_t) + b_1(t)b_2(t)dt \end{aligned}$$

Ce qui est équivalent à

$$dY_t = X_{1t}dX_{2t} + X_{2t}dX_{1t} + b_1(t)b_2(t)dt \quad (1.18)$$

Le terme $b_1(t)b_2(t)dt$ est le terme correcteur qui fait la différence par rapport au calcul différentiel ordinaire.

ce résultat est obtenu par application directe de (1.15).

Cas particulier

Si $W_{1t} = W_{2t} = W_t$, (1.18) implique $dW_t^2 = 1dt + 2W_t dW_t$

1.5 processus de diffusion

1.5.1 Description générale

Un processus stochastique à temps continu $(X_t)_{t \geq 0}$ est dit de diffusion s'il est markovien et à trajectoires presque sûrement continues .

Plusieurs phénomènes physiques, biologiques, économiques et sociaux ont été modélisés par un processus de diffusion. Citons par exemple : les mouvements des molécules dans les particules soumises à des interactions, Systèmes de communication avec bruit et la croissance de la population. Probabilités de premier passage, temps moyen d'absorption, distributions de temps d'occupation et les distributions stationnaires sont des fonctions qui peuvent être calculées par un processus de diffusion unidimensionnel, tandis que le processus de diffusion multidimensionnel est utilisé dans la formulation. Le calcul et la résolution des équations différentielles de deuxième degré.

Considérons un processus de diffusion $\{X(t), t \geq 0\}$ où l'espace des états est un intervalle I (I peut être fermé, ouvert, fini ou infini). Pour tout point z dans I , on définit T_z comme la variable aléatoire représentant l'instant de premier passage de processus par la valeur z . Dans le cas où z n'est jamais atteinte, on met par convention $T_z = \infty$. Le processus est dit régulier si :

$$Pr\{T_z < \infty / X(0) = x\} > 0, \forall x \in I$$

Dans ce qui suit, on ne considère que les processus de diffusions réguliers.

Tous les processus de diffusion satisfont la condition suivante :

$$\begin{aligned} \lim_{t \rightarrow s} \frac{1}{(t-s)} Pr\{|X(t) - x| > \varepsilon / X(s) = x\} &= 0, \quad \forall x \in I, \forall \varepsilon > 0 \\ \lim_{t \rightarrow s} \frac{1}{(t-s)} \int_{|x-y| > \varepsilon} f(s, x; t, y) dy &= 0, \quad \forall x \in I, \forall \varepsilon > 0 \end{aligned} \quad (1.19)$$

Cette relation signifie que la probabilité d'avoir des grandes transitions (supérieur à ε) pour un intervalle de temps suffisamment petit est nulle. Ce fait est du à la continuité des trajectoires de processus de diffusion.

Tous les processus de diffusion existants possèdent deux conditions de base décrivant la moyenne et la variance des transitions infinitésimales respectivement :

$$\begin{aligned} \lim_{t \rightarrow s} \frac{1}{(t-s)} \int_{|x-y| \leq \varepsilon} (x-y) f(s, x; t, y) dy &= a(s, x), \quad \forall x \in I, \forall \varepsilon > 0 \\ \lim_{t \rightarrow s} \frac{1}{(t-s)} E(X(t) - X(s) | X(s) = x) &= a(s, x), \quad \forall x \in I \end{aligned} \quad (1.20)$$

et

$$\begin{aligned} \lim_{t \rightarrow s} \frac{1}{(t-s)} \int_{|x-y| \leq \varepsilon} (x-y)^2 f(s, x; t, y) dy &= b^2(s, x), \quad \forall x \in I, \forall \varepsilon > 0 \\ \lim_{t \rightarrow s} \frac{1}{(t-s)} E\left(\left(X(t) - X(s)\right)^2 | X(s) = x\right) &= b^2(s, x), \quad \forall x \in I \end{aligned} \quad (1.21)$$

La fonction $a(s, x)$ est appelée coefficient de dérive ou la moyenne infinitésimales, et $b^2(s, x)$ est appelé coefficient de diffusion ou variance infinitésimale.

On peut définir aussi un processus de diffusion par un processus markovien $(X_t)_{t \geq 0}$ vérifiant les relations (1.19), (1.20) et (1.21).

1.5.2 Propriétés d'un processus de diffusion

Les coefficients infinitésimaux Généralement, les fonctions $a(s, x)$ et $b^2(s, x)$ sont continues pour tout s et x , et pour un processus régulier la fonction $b^2(s, x)$ est positive pour x dans I . Si le processus est homogène les fonctions $a(s, x) = a(x)$ et $b^2(s, x) = b^2(x)$ sont indépendantes de s . La classification de toutes les diffusions possibles par la prescription des coefficients infinitésimaux $a(s, x)$ et $b^2(s, x)$ dépend de la définition de mouvement qu'ils représentent. En plus des relations infinitésimales (2) et (3), la relation des moments infinitésimaux d'ordre supérieur est vraie :

$$\lim_{t \rightarrow s} \frac{1}{(t-s)} E\left(\left|X(t) - X(s)\right|^r | X(s) = x\right) = 0, \quad r = 3, 4, \dots \quad (1.22)$$

Proposition 1.31 : le processus du mouvement brownien est un processus de diffusion en milieu homogène de coefficient de dérive $a(s, x) = 0$ et coefficient de diffusion $b^2(s, x) = 1$

1.5.3 Transformation élémentaire d'un processus

Soit $\{X(t), t \geq 0\}$ un processus de diffusion régulier où l'espace des états est un intervalle I dont les extrémités sont les points l et r , on suppose que $a(x)$ et $b^2(x)$ sont les coefficients

infinitésimaux de $\{X(t)\}$. Soit g une fonction strictement monotone sur I avec une dérivée seconde continue $g''(x)$ pour $l < x < r$. D'où le processus $Y(t) = g[X(t)]$ est un processus de diffusion régulier sur l'intervalle dont les extrémités sont $g(l)$ et $g(r)$, et $\{Y(t)\}$ a comme coefficients infinitésimaux

$$\begin{aligned} a_Y(y) &= \frac{1}{2}b^2(x)g''(x) + a(x)g'(x) \\ b_Y^2(y) &= b^2(x)[g'(x)]^2 \end{aligned}$$

où $y = g(x)$

Preuve. On considère le cas où la fonction g est strictement croissante car la preuve est similaire dans le où elle est strictement décroissante.

Comme g est deux fois continûment différentiable, la formule de Taylor donne :

$$g(x + \Delta x) = g(x) + \Delta x g'(x) + \frac{1}{2}(\Delta x)^2 g''(x) + \frac{1}{2}(\Delta x)^2 [g''(\xi) - g''(x)] \quad \text{où } x \leq \xi \leq x + \Delta x$$

En remplaçant x par $X(t)$ et ΔX par $X(t+h) - X(t)$, on aura :

$$g(X(t+h)) = g(X(t)) + \Delta X g'(X(t)) + \frac{1}{2}(\Delta X)^2 g''(X(t)) + \frac{1}{2}(\Delta X)^2 [g''(\xi(\omega)) - g''(X(t))]$$

où $\xi(\omega)$ lie entre $X(t)$ et $X(t+h)$ et ω signifie la réalisation particulière de processus qui rend $\xi(\omega)$ aléatoire. Comme $X(t) = x$ donc $g(x) = g(X(t)) = T(y) = y$ et l'équation précédente devient :

$$Y(t+h) - Y(t) = \Delta X g'(X(t)) + \frac{1}{2}(\Delta X)^2 g''(X(t)) + \frac{1}{2}(\Delta X)^2 [g''(\xi(\omega)) - g''(X(t))] \quad (*)$$

En divisant sur h et en passant à la moyenne puis à la limite limite, on trouve :

$$\lim_{h \rightarrow 0} \frac{1}{h} E[Y(t+h) - Y(t) | Y(t) = y] = a(x)g'(x) + \frac{1}{2}b^2(x)g''(x) + \frac{1}{2} \lim_{h \rightarrow 0} \frac{1}{h} E[(\Delta X)^2 \{g''(\xi(\omega)) - g''(X(t))\}]$$

La continuité de g'' assure la convergence de $g''(\xi(\omega))$ à $g''(X(t))$, donc :

$$\lim_{h \rightarrow 0} \frac{1}{h} E[(\Delta X)^2 \{g''(\xi(\omega)) - g''(X(t))\}] = 0$$

ça signifie que la moyenne infinitésimale de $Y(t)$ vaut :

$$\begin{aligned} a_Y(y) &= \lim_{h \rightarrow 0} \frac{1}{h} E[Y(t+h) - Y(t) | Y(t) = y] \\ &= a(x)g'(x) + \frac{1}{2}b^2(x)g''(x) \end{aligned}$$

En calculant le carré de $(*)$, on trouve :

$$[Y(t+h) - Y(t)]^2 = (\Delta X)^2 [g'(X(t))]^2 + R - h$$

où R_h est le reste et qui contient $(\Delta X)^3$ comme terme d'ordre maximum.

D'après (1.22) $\lim_{h \rightarrow 0} \frac{1}{h} E[|\Delta X|^r | X(t) = x] = 0$ pour tout $r \geq 3$. Par la suite la variance infinitésimale de $Y(t)$ est :

$$\begin{aligned} b_Y^2(y) &= \lim_{h \rightarrow 0} \frac{1}{h} E[\{Y(t+h) - Y(t)\}^2 | Y(t) = y] \\ &= b^2(x)[g'(x)]^2 \end{aligned}$$

La condition (1) de $Y(t) = g[X(t)]$ est déjà vérifiée, donc le processus $Y(t)$ est un processus de diffusion.

1.6 Conclusion

En général la formule d'Itô ne suffit pas pour résoudre tous les problèmes réels. La résolution des équations différentielles stochastiques est bien souvent impossible. Cependant, on est amené à utiliser des méthodes numériques afin de déterminer des solutions approchées. Pour cela, il faut à priori savoir si l'EDS possède une solution (propriété d'existence de la solution) et vérifiant de préférence la propriété d'unicité, sachant bien sûr sa valeur initiale.

2

Equations Différentielles Stochastiques

2.1 Introduction

Le concept d'équation différentielle stochastique (EDS) généralise celui d'équation différentielle ordinaire aux processus stochastiques. La formulation théorique de ce problème à elle seule a posé problème aux mathématiciens, il a fallu attendre les années 1940 et les travaux du mathématicien japonais Itô Kiyoshi pour la définition de l'intégrale stochastique.

Dans ce chapitre, nous introduirons la notion d'existence d'une solution forte et faible d'une EDS. Après nous parlerons de la méthode de résolution des EDS basée essentiellement sur le calcul d'Itô. Enfin, nous introduirons l'approximation par discrétisation du temps sur laquelle se base l'approche numérique de résolution des EDS.

2.2 Définition

Soient $\begin{cases} a : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d \\ b : \mathbb{R}^+ \times \mathbb{R}^d \rightarrow \mathfrak{M}_{d,m}(\mathbb{R}) \end{cases}$, deux fonctions mesurables bornées. Soit $x \in \mathbb{R}^d$ une condition initiale.

Une solution de l'EDS

$$E_x(a, b) = \begin{cases} X_0 = x \\ dX_t = a(t, X_t)dt + b(t, X_t)dW_t \end{cases}$$

est constituée par :

- Un espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \geq 0}, P)$,
 $\mathfrak{F} = \sigma(W_s, 0 \leq s \leq t)$.
- Un $(\mathfrak{F}_t)$ -mouvement brownien $W = (W_1, \dots, W_m)$ à valeurs dans $\mathbb{R}^m$.
- Un processus $X = \{X_t, t \geq 0\}$ continu $(\mathfrak{F}_t)$ -adapté tel que

Les intégrales $\int_0^t a(s, X_s)ds$ et $\int_0^t b(s, X_s)dW_s$ aient un sens et tel que l'égalité

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s$$

Soient satisfaites pour tout t P.p.s.

Autrement dit, pour tout $i = 1, \dots, d$ on a

$$X_t^i = X_0^i + \int_0^t a^i(s, X_s)ds + \sum_{j=1}^m \int_0^t b_{ij}(s, X_s)dW_s^j$$

Le but de ce chapitre est d'étudier les équations différentielles stochastiques dans le cas unidimensionnel : Soit $d = m = 1$

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s \quad (2.1)$$

Où $(W_t)_{t \in \mathbb{R}_+}$ est un processus de Wiener sur espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t \geq 0}, P)$, $\mathfrak{F} = \sigma(W_s, 0 \leq s \leq t)$.

Le coefficient $a(t, X_t)$ de dt est appelé dérive et le coefficient $b(t, x_t)$ de dW_t est appelé terme de diffusion, sont deux fonctions mesurables sur $[0, T] \times \mathbb{R}$.

Nous appellerons solution sur $[0, T]$ tout processus d-dimensionnel $\{X_t, t \geq 0\}$ qui satisfait

- a. X_0 est $\mathfrak{F}_0$ -mesurable ;
- b. $|a(t, X_t)|^{1/2}, b(t, X_t) \in C_1([0, T])$;
- c. X_t est différentiable , alors

$$\begin{cases} dX_t = a(t, X_t)dt + b(t, X_t)dW_t & t \in [0, T] \\ X_0 = 0 \end{cases} \quad (2.2)$$

Lemme 2.1 (Lemme de Gronwall) Soit f une application localement intégrable de $\mathbb{R}_+$ dans $\mathbb{R}$, α et β deux applications croissantes et non négatives tel que pour tout $t \in [0, T]$

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

Alors on a :

$$f(t) \leq \alpha(t) + \beta(t) \int_0^t e^{\beta(t)(t-s)} \alpha(s)ds$$

(Démonstration : voir [9])

2.3 Solution faible et forte d'une Equation Différentielle Stochastique

Définition 2.2 On appelle solution faible de l'équation différentielle stochastique (2.2) sur $[0, T]$

$$X_t = X_0 + \int_0^t a(t, X_s)ds + \int_0^t b(t, X_s)dW_s \quad \text{P.ps} \quad (2.3)$$

Où $(W_t)_{t \in \mathbb{R}_+}$ est un processus de Wiener sur un espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t \geq 0}, P)$, $\mathfrak{F} = \sigma(W_s, 0 \leq s \leq t)$, pour tout $t \in [0, T]$ et P est une loi de probabilité sur $(\Omega, \mathfrak{F})$.

Remarque 2.3 On dit que la solution faible de (2.2) est unique si la loi obtenue P est unique.

Définition 2.4 Etant donné W un mouvement brownien standard défini sur une base $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t \geq 0}, P)$, $\mathfrak{F} = \sigma(W_s, 0 \leq s \leq t)$, on appelle solution forte de (2.2) un processus π_t défini sur cette même base tel que l'on ait :

$$\pi_t = X_0 + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s) dW_s \quad (2.4)$$

On dit que la solution forte est unique si pour deux solutions π_t^1 et π_t^2 , processus sur $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t \geq 0}, P)$, $\mathfrak{F} = \sigma(W_s, 0 \leq s \leq t)$

$$P\{\sup_{t \in [0, T]} \|\pi_t^1 - \pi_t^2\| > 0\} = 0$$

Remarque 2.5 Dans le cas d'équations différentielles stochastiques de Itô, l'unicité de la solution forte (unicité en trajectoires) implique l'unicité de la solution faible (unicité en lois)

2.4 Existence et unicité de la solution

Nous étudions dans cette partie l'existence et l'unicité des solutions des équations différentielles stochastiques (EDS). Lorsque la solution de l'EDS ne change pas pour une autre valeur du processus de Wiener, alors on parle de solution forte. Dans le cas contraire, c'est-à-dire que si on trouve une solution correspondante à un processus de Wiener bien particulier, alors elle est dite solution faible. Certaines équations différentielles stochastiques admettent des solutions faibles et non fortes.

Théorème 2.6 [9] soient $a(t, X_t)$ et $b(t, X_t)$ deux fonctions mesurables sur $[0, T] \times \mathbb{R}$ auxquelles on impose les hypothèses ci-dessous :

- a. (Condition de Lipschitz) : Pour tout $t \in [0, T]$ et pour tous $(x, y) \in \mathbb{R} \times \mathbb{R}$
 $|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leq K^* |x - y|$
- b. (Borne à accroissement linéaire) : Pour tout $t \in [0, T]$ et tout $x \in \mathbb{R}$
 $|a(t, x)| \leq K(1 + |x|), |b(t, x)| \leq K(1 + |x|), (K^*, K \text{ constants})$
- c. (Condition sur la valeur initiale) : X_{t_0} est $\mathfrak{F}_0$ -mesurable avec $E(|X_{t_0}|^2) \leq \infty$

Alors :

L'équation (2.2) admet une solution forte unique presque sûrement à trajectoires continues

L'hypothèse de mesurabilité et la condition de Lipchitz sont un moyen de prouver l'unicité et l'existence de la solution par la méthode des approximations successives.

Preuve : Unicité. Soit $(X_1(t))_{t \in [0, T]}$ et $(X_2(t))_{t \in [0, T]} \in C([0, T])^2$ deux solutions de l'équation (2.6).

Alors d'après la troisième hypothèse sur la condition initiale de théorème (2.6) on a :

$$\begin{aligned} X_1(t) - X_2(t) &= \\ &= \int_0^t [a(s, X_1(s)) - a(s, X_2(s))] ds + \int_0^t [b(s, X_1(s)) - b(s, X_2(s))] dW_s \\ &= \int_0^t \tilde{a}(s) ds + \int_0^t \tilde{b}(s) dW_s, \quad t \in [0, T] \end{aligned}$$

Où $\tilde{a}(s) = a(s, X_1(s)) - a(s, X_2(s))$ et $\tilde{b}(s) = b(s, X_1(s)) - b(s, X_2(s))$

On a $(a + b)^2 \leq 2(a^2 + b^2)$, alors on obtient

$$|X_1(t) - X_2(t)|^2 \leq 2 \left(\int_0^t \tilde{a}(s) ds \right)^2 + 2 \left(\int_0^t \tilde{b}(s) dW_s \right)^2$$

D'après l'inégalité de Cauchy-Schwarz

$$\left(\int_0^t \tilde{a}(s) ds \right)^2 \leq t \left(\int_0^t |\tilde{a}(s)|^2 ds \right)$$

En prenant l'espérance, on obtient

$$E \left[\left(\int_0^t \tilde{a}(s) ds \right)^2 \right] \leq t E \left[\int_0^t |\tilde{a}(s)|^2 ds \right]$$

de plus, d'après la deuxième hypothèse de théorème (2.6)

$$\begin{aligned} E \left[\int_0^T (b(s, X_i(s)))^2 ds \right] &\leq E \left[\int_0^T (K + |X_i(s)|)^2 ds \right] \\ &\leq 2K^2 E \left[\int_0^T (1 + |X_i(s)|^2) ds \right] \\ &< +\infty \end{aligned}$$

Pour $i = 1, 2$ et $X_i(s) \in C([0, T])$ cela montre que $b(s, X_i(s)) \in C([0, T])$ alors pour $i = 1, 2$ on a $\tilde{b}(s) \in C([0, T])$, puis d'après les propriétés d'espérance de mouvement brownien on a

$$E \left[\left(\int_0^t \tilde{b}(s) dW_s \right)^2 \right] = E \left[\int_0^t (\tilde{b}(s))^2 ds \right]$$

D'où

$$E[(X_1(t) - X_2(t))^2] \leq 2t E \left[\int_0^t (\tilde{a}(s))^2 ds \right] + 2E \left[\int_0^t (\tilde{b}(s))^2 ds \right]$$

D'après la 1ère hypothèse (Condition de Lipchitz) du théorème (2.6), on a :

$$\begin{aligned} |\tilde{a}(s)|^2 &\leq (K^*)^2 |X_1(s) - X_2(s)|^2, \\ |\tilde{b}(s)|^2 &\leq (K^*)^2 |X_1(s) - X_2(s)|^2, \end{aligned}$$

par conséquent

$$\begin{aligned} E[|X_1(t) - X_2(t)|^2] &\leq 2t(K^*)^2 \int_0^t E[|X_1(s) - X_2(s)|^2]ds + 2(K^*)^2 \int_0^t E[|X_1(s) - X_2(s)|^2]ds \\ &\leq 2T(K^*)^2 \int_0^t E[|X_1(s) - X_2(s)|^2]ds + 2(K^*)^2 \int_0^t E[|X_1(s) - X_2(s)|^2]ds \\ &= 2(K^*)^2(T+1) \int_0^t E[|X_1(s) - X_2(s)|^2]ds \end{aligned}$$

En utilisant le lemme de Gronwall,

$$E[|X_1(t) - X_2(t)|^2] = 0, \quad \forall t \in [0, T]$$

on obtient :

$$X_1(t) - X_2(t) = 0, P.p.s \quad \forall t \in [0, T]$$

Existence. Prouvons l'existence de la solution par la méthode des approximations successives.

Soit

$$\begin{cases} X_0(t) = X^0, \\ X_n(t) = X^0 + \int_0^t a(s, X_{n-1}(s))ds + \int_0^t b(s, X_{n-1}(s))dW_s \quad \forall t \in [0, T], n \in \mathbb{N}^* \end{cases}$$

On suppose que $X^0 - \mathfrak{F}_0$ -mesurable et par la condition sur la valeur initiale du théoème (2.6), il est évident que $X^0 \in ([0, T])$. Par induction, on va montrer que :

$$\forall n \in \mathbb{N} : E[|X_{n+1}(t) - X_n(t)|^2] \leq \frac{(ct)^{n+1}}{(n+1)!} \quad (2.5)$$

où

$$c = \max\{4K^2(T+1)(1 + E[|X^0|^2]), 2(K^*)^2(T+1)\}, \quad (2.6)$$

Et $\forall n \in \mathbb{N} : X_{n+1} \in C([0, T])$

Sous les hypothèses (1) et (2) du théorème (2.6), on obtient

$$E[|b(s, X^0)|^2] \leq E[K^2(1 + |X^0|)^2] \leq 2K^2(1 + E[|X^0|^2]) < +\infty$$

D'après l'inégalité :

$$(|x| + |y|)^2 \leq 2|x|^2 + 2|y|^2 \quad (2.7)$$

On a $b(s, X^0) \text{ et } a(s, X^0) \in C([0, T])^2$

Utilisons l'inégalité (2.7), on obtient

$$\begin{aligned} |X_1(t) - X^0|^2 &= \left| \int_0^t a(s, X^0)ds + \int_0^t b(s, X^0)dW_s \right|^2 \\ &\leq 2 \left| \int_0^t a(s, X^0)ds \right|^2 + 2 \left| \int_0^t b(s, X^0)dW_s \right|^2 \end{aligned}$$

Et par l'inégalité de Schwarz

$$\begin{aligned} \left| \int_0^t a(s, X^0) ds \right|^2 &\leq t \int_0^t |a(s, X^0)|^2 ds \\ &\leq T \int_0^t |a(s, X^0)|^2 ds \end{aligned}$$

D'après les propriétés d'espérance de mouvement brownien on a :

$$E\left[\left|\int_0^t b(s, X^0) dW_s\right|^2\right] = E\left[\int_0^t |b(s, X^0)|^2 ds\right]$$

D'après l'hypothèse de Borne à accroissement linéaire du théorème (2.6)

$$\begin{aligned} E[|X_1(t) - X^0|^2] &= \\ &\leq 2TE\left[\int_0^t |a(s, X^0)|^2 ds\right] + 2E\left[\int_0^t |b(s, X^0)|^2 ds\right] \\ &\leq 2TE\left[\int_0^t K^2(1 + |X^0|)^2 ds\right] + 2E\left[\int_0^t K^2(1 + |X^0|)^2 ds\right] \\ &\leq (2TK^2 + 2K^2)E\left[\int_0^t (1 + |X^0|)^2 ds\right] \\ &= (2TK^2 + 2K^2)E\left[\int_0^t (1 + |X^0|)^2 ds\right] \\ &= 2K^2(T + 1)tE[(1 + |X^0|)^2] \\ &\leq 4K^2(T + 1)t(1 + E[|X^0|^2]) = cst \end{aligned}$$

La dernière inégalité est une conséquence directe de (2.7).

On suppose que (2.5) est vrai pour $n = 1$ et on va monter par induction qu'elle est valable pour $(n + 1)$.

Soit $a(s, X_n(s)) \in C([0, T])$ et $b(s, X_n(s)) \in C([0, T])$

$$\begin{aligned} |X_{n+1}(t) - X_n(t)|^2 &\leq \left(\int_0^t |a(s, X_n(s)) - a(s, X_{n-1}(s))| ds\right)^2 + \left|\int_0^t (b(s, X_n(s)) - b(s, X_{n-1}(s))) dW_s\right|^2 \\ &\leq 2\left(\int_0^t |a(s, X_n(s)) - a(s, X_{n-1}(s))| ds\right)^2 + 2\left(\int_0^t (b(s, X_n(s)) - b(s, X_{n-1}(s))) dW_s\right)^2 \end{aligned}$$

D'après l'inégalité de Schwarz

$$\begin{aligned} \left(\int_0^t |a(s, X_n(s)) - a(s, X_{n-1}(s))| ds\right)^2 &\leq t \int_0^t |a(s, X_n(s)) - a(s, X_{n-1}(s))|^2 ds \\ &\leq T(K^*)^2 \int_0^t |X_n(s) - X_{n-1}(s)|^2 ds \end{aligned}$$

où la dernière inégalité est due au 1ère hypothèse du théorème(2.6). En outre, d'après les propriétés d'espérance de mouvement brownien on a

$$\begin{aligned} E\left[\left(\int_0^t (b(s, X_n(s)) - b(s, X_{n-1}(s))) dW_s\right)^2\right] &= E\left[\int_0^t |b(s, X_n(s)) - b(s, X_{n-1}(s))|^2 ds\right] \\ &\leq (K^*)^2 E\left[\int_0^t |X_n(s) - X_{n-1}(s)|^2 ds\right] \end{aligned}$$

Finalement, D'après la 1ère hypothèse du théorème (2.6) on obtient

$$\begin{aligned} E[|X_{n+1}(t) - X_n(t)|^2] &\leq 2T(K^*)^2 E\left[\int_0^t |X_n(s) - X_{n-1}(s)|^2 ds\right] + 2(K^*)^2 E\left[\int_0^t |X_n(s) - X_{n-1}(s)|^{2ds}\right] \\ &\leq CE\left[\int_0^t |X_n(s) - X_{n-1}(s)|^2 ds\right] \\ &\leq \int_0^t \frac{(Cs)^n}{n!} ds = \frac{(Ct)^{n+1}}{(n+1)!} \end{aligned}$$

(La suite du preuve voir [9]) $\square$.

2.5 Equations différentielles stochastiques de type linéaire

2.5.1 Résolution des EDS de type linéaire

Définition 2.7 Soit $\mathcal{L}(\mathbb{R}^n, M_{n,m}(\mathbb{R}))$ un espace des fonctions linéaires continues de $\mathbb{R}^n$ dans $M_{n,m}$, l'équation différentielle stochastique

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t \quad (2.8)$$

est linéaire si $a(t, X_t)$ et $b(t, X_t)$ de la forme :

$$\begin{aligned} a(t, X_t) &= [a_1(t)X_t + a_2(t)] \text{ avec } \begin{cases} a_1 : [0, T] \rightarrow \mathbb{R}^n \\ a_2 : [0, T] \rightarrow M_{n,n}(\mathbb{R}) \end{cases} \text{ et} \\ b(t, X_t) &= [b_1 X_t + b_2(t)] \text{ avec } \begin{cases} b_1 : [0, T] \rightarrow M_{n,m}(\mathbb{R}) \\ b_2 : [0, T] \rightarrow \mathcal{L}(\mathbb{R}^n, M_{n,m}(\mathbb{R})) \end{cases} \end{aligned}$$

Alors l'équation (2.8) est de type scalaire dont la forme générale est

$$dX_t = [a_1(t)X_t + a_2(t)]dt + [b_1(t)X_t + b_2(t)]dW_t \quad (2.9)$$

Définition 2.8 Une équation différentielle stochastique linéaire est dite homogène si $a_2(t) \equiv b_2(t) \equiv 0$. Elle est dite linéaire au sens faible si $b_1(t) \equiv 0$.

Examinon tout d'abord le cas des équations différentielles stochastiques linéaire au sens faible ; quand $b_1(t) \equiv 0$, le bruit apparaît comme terme additif et l'équation (2.9) devient :

$$dX_t = [a_1(t)X_t + a_2(t)]dt + b_2(t)dW_t \quad (2.10)$$

L'équation homogène obtenu à partir de (2.10) est une équation différentielle ordinaire

$$\frac{dX_t}{dt} = a_1(t)X_t \quad (2.11)$$

où la solution fondamentale est donnée par

$$\phi_{t,t_0} = \exp\left[\int_{t_0}^t a_1(s)ds\right]$$

En appliquant la formule d'Itô (Théorème 1.27) à la transformation $f(t, x) = \phi_{t,t_0}^{-1}x$ on obtient :

$$d(\phi_{t,t_0}^{-1}X_t) = \left[\frac{d\phi_{t,t_0}^{-1}}{dt}X_t + (a_1(t)X_t + a_2(t))\phi_{t,t_0}^{-1}\right] + b_2(t)\phi_{t,t_0}^{-1}dW_t$$

Or

$$\frac{d\phi_{t,t_0}^{-1}}{dt} = d\phi_{t,t_0}^{-1}a_1(t) \Rightarrow d\phi_{t,t_0}^{-1}X_t = a_2(t)\phi_{t,t_0}^{-1}dt + b_2(t)\phi_{t,t_0}^{-1}dW_t$$

Cette dernière expression peut se mettre sous la forme intégrale :

$$\phi_{t,t_0}^{-1}X_t = \phi_{t_0,t_0}^{-1}X_{t_0} + \int_{t_0}^t a_2(s)\phi_{s,t_0}^{-1}ds + \int_{t_0}^t b_2(s)\phi_{s,t_0}^{-1}dW_s$$

Sous la condition $\phi_{t_0,t_0} = 1$, la solution de (2.10) est

$$X_t = \phi_{t,t_0}[X_{t_0} + \int_{t_0}^t (a_2(s)\phi_{s,t_0}^{-1})ds + \int_{t_0}^t b_2(s)\phi_{s,t_0}^{-1}dW_s] \quad (2.12)$$

$$\phi_{t,t_0} = \exp\left[\int_{t_0}^t a_1(s)ds\right] \quad (2.13)$$

Le processus X_t (2.12) est une combinaison linéaire de variables aléatoires gaussiennes car $\int_{t_0}^t f(s)dW_s$ est gaussienne pour toute fonction déterministe f ; en plus, X_{t_0} est supposé déterministe (constante) ou une variable gaussienne. dès lors, X_t est un processus gaussien où la moyenne et le moment d'ordre deux satisfont les équations différentielles ordinaires. toutes fois dans le cas général la solution de (2.8) n'est pas toujours gaussienne.

Lorsque $a_2(t) \equiv 0$ et $b_2(t) \equiv 0$, l'expression (2.9) se réduit à l'équation différentielle stochastique homogène et linéaire :

$$dX_t = a_1(t)X_t dt + b_1(t)X_t dW_t \quad (2.14)$$

La solution fondamentale de cette dernière inégalité est donné par

$$\phi_{t,t_0} = \phi_{t,t_0} \exp\left[\int_{t_0}^t (a_1(s) - 1/2b_1^2(s))ds + \int_{t_0}^t b_1(s)dW_s\right] \quad (2.15)$$

Où ϕ_{t,t_0} est généralement égale 1.

Preuve La solution fondamentale (2.13) d'une équation linéaire au sens limité satisfait l'équation différentielle ordinaire $dLn\phi_{t,t_0} = a_1(t)dt$. En utilisant cette définition, la formule d'Itô (Théorème 1.27) appliquée à la transformée de la solution fondamentale $Ln\phi_{t,t_0}$ de (2.14) sera :

$$\begin{aligned} d(Ln\phi_{t,t_0}) &= [a_1(t)\phi_{t,t_0}\phi_{t,t_0}^{-1} - 1/2b_1^2(t)\phi_{t,t_0}^2\phi_{t,t_0}^{-2}]dt + b_1(t)\phi_{t,t_0}\phi_{t,t_0}^{-1}dW_t \\ &= [a_1(t) - 1/2b_1^2(t)]dt + b_1(t)dW_t \end{aligned}$$

ce qui implique

$$Ln\phi_{t,t_0} = \int_{t_0}^t [a_1(s) - 1/2b_1^2(s)]ds + \int_{t_0}^t b_1(s)dW_s$$

Proposition 2.9 Le processus $\phi_{t,t_0}^{-1}X_t$ qui est fonction de la solution X_t de l'équation générale (2.10) possède une forme explicite de l'intégrale stochastique de la forme :

$$\phi_{t,t_0}^{-1}X_t = X_{t_0} + \int_{t_0}^t [a_2(s) - b_1(s)b_2(s)]\phi_{s,t_0}^{-1}ds + \int_{t_0}^t [b_2(s)\phi_{s,t_0}^{-1}]dW_s$$

Similairement, on applique la formule d'Itô à ϕ_{t,t_0}^{-1} et on obtient :

$$d\phi_{t,t_0}^{-1} = [-a_1(t) + b_1^2(t)]\phi_{t,t_0}^{-1}dt - b_1(t)\phi_{t,t_0}^{-1}dW_t \quad (2.16)$$

Ce qui implique que la solution de l'équation générale (2.9) est donné par

$$X_t = \phi_{t,t_0}[X_{t_0} + \int_{t_0}^t [a_2(s) - b_1(s)b_2(s)]\phi_{s,t_0}^{-1}ds + \int_{t_0}^t [b_2(s)\phi_{s,t_0}^{-1}]dW_s] \quad (2.17)$$

Où ϕ_{t,t_0} est donné par l'expression (2.16)

Preuve Les deux termes ϕ_{t,t_0}^{-1} et X_t possèdent des différentielles stochastiques incluant le même processus de Wiener W_t , ce qui implique que la différentielle de leur produit (voir 1.18) :

$U(X_1(t), X_2(t)) = X_1(t)X_2(t)$ avec $X_1(t) = \phi_{t,t_0}^{-1}$ et $X_2(t) = X(t)$ donne

$$d\phi_{t,t_0}^{-1}X(t) = \phi_{t,t_0}^{-1}([a_1(t)X(t) + a_2(t)]dt + [b_1(t)X(t) + b_2(t)]dW_t) + X(t) \\ (-[a_1(t) + b_1^2(t)]\phi_{t,t_0}^{-1}dt - b_1(t)\phi_{t,t_0}^{-1}dW_t) - b_1(t)\phi_{t,t_0}^{-1}[b_1(t)X_t + b_2(t)]dt$$

Ce qui équivalent à

$$d\phi_{t,t_0}^{-1}X_t = [a_2(t) - b_1(t)b_2(t)]\phi_{t,t_0}^{-1}dt + b_2(t)\phi_{t,t_0}^{-1}dW_t$$

En intégrant sur l'intervalle de temps $[t_0, t]$ et en supposant que $\phi_{t_0,t_0} = 1$, on obtient :

$$\phi_{t,t_0}^{-1}X_t = X_{t_0} + \int_{t_0}^t [a_2(s) - b_1(s)b_2(s)]\phi_{s,t_0}^{-1}ds + \int_{t_0}^t [b_2(s)\phi_{s,t_0}^{-1}]dW_s$$

Corollaire 2.10 Dans le cas d'une équation différentielle stochastique de type

$dX_t = aX_tdt + bX_t dW_t$, on applique l'équation (2.17) avec

$$\begin{cases} a_1(t) \equiv a \\ b_1(t) \equiv b \end{cases}$$

et $a_2(t) = b_2(t) = 0$ et la solution donné dans ce cas par

$$X_t = X_0 \exp[(a - 1/2b^2)t + bW_t]$$

Solution générales de quelques EDS de type linéaire

Nous exposons dans cette partie la solution générale de diverses équations différentielles stochastiques linéaires où le bruit, présenté par le mouvement brownien, soit additif ou multiplicatif. Nous rappelons que la solution générale d'une équation linéaire de la forme

$$dX(t) = [a_1(t)X_t + a_2(t)dt] + [b_1(t)X_t + b_2(t)]dW_t$$

est donné par l'expression

$$X_t = \phi_{t,t_0}[X_{t_0} + \int_{t_0}^t [a_2(s) - b_1(s)b_2(s)]\phi_{s,t_0}^{-1}ds + \int_{t_0}^t [b_2(s)\phi_{s,t_0}^{-1}]dW_s]$$

Où la solution fondamentale est $\phi_{t,t_0} = \exp[\int_{t_0}^t (a_1(s) - 1/2b_1^2(s))ds + \int_{t_0}^t b_1(s)dW_s]$

$$\text{a. Si } \begin{cases} a_1(t) \equiv -\alpha \\ b_2(t) \equiv \sigma \\ a_2(t) \equiv b_1(t) \equiv 0 \end{cases} \quad \text{Alors}$$

$$X_t = e^{-\alpha t}(X_0 + \sigma \int_0^t e^{\alpha s} dW_s) \quad (2.18)$$

$$\text{b. Si } \begin{cases} a_1(t) \equiv a \\ a_2(t) \equiv b \\ b_1(t) \equiv 0 \\ b_2(t) \equiv c \end{cases} \quad \text{Alors}$$

$$X_t = e^{at}(X_0 + b/a(1 - e^{-at}) + c \int_0^t e^{-as} dW_s) \quad (2.19)$$

$$\text{c. Si } \begin{cases} a_1(t) \equiv a(t) \\ a_2(t) \equiv b(t) \\ b_1(t) \equiv c(t) \\ b_2(t) \equiv 0 \end{cases} \quad \text{Alors}$$

$$X_t = \phi_{t,t_0}[X_{t_0} + \int_{t_0}^t \phi_{s,t_0}^{-1}b(s)ds + \int_{t_0}^t c(s)\phi_{s,t_0}^{-1}dW_s] \quad (2.20)$$

Où $\phi_{t,t_0} = \exp(\int_{t_0}^t a(s)ds)$

$$\text{d. Si } \begin{cases} a_1(t) \equiv a \\ b_1(t) \equiv b \\ a_2(t) \equiv b_2(t) \equiv 0 \end{cases} \quad \text{Alors}$$

$$X_t = X_0 \exp[(a - 1/2b^2)t + bW_t] \quad (2.21)$$

$$\text{e. Si } \begin{cases} a_1(t) & \equiv a \\ a_2(t) & \equiv c \\ b_1(t) & \equiv b \\ b_2(t) & \equiv d \end{cases} \text{ Alors}$$

$$X_t = \phi_t [X_0 + (c - bd) \int_0^t \phi_s^{-1} ds + d \int_0^t \phi_s^{-1} dW_s] \quad (2.22)$$

$$\text{Où } \phi_t = \exp[(a - 1/2b^2)t + bW_t]$$

$$\text{f. Si } \begin{cases} a_1(t) & \equiv a(t) \\ b_1(t) & \equiv b(t) \\ a_2(t) & \equiv b_2(t) \equiv 0 \end{cases}$$

$$X_t = X_0 \exp\left[\int_0^t (a(s) - 1/2b^2(s))ds + \int_0^t b(s)dW_s\right] \quad (2.23)$$

2.5.2 Exemples d'applications d'une EDS linéaire

Exemple 1 : Si $m = n = 1$, et si f, g sont deux fonctions continues, l'unique solution de $\begin{cases} dX_t = fX_t dt + gX_t dW_t \\ X_0 = 1 \end{cases}$ est : $X_t = X_0 \exp[(f - 1/2g^2)t + gW_t]$, pour tout $t \in [0, T]$.

Exemple 2 : Prix d'action

Si P_t est le prix d'une action à l'instant t , il est possible de modéliser son évolution en supposant que $\frac{dP_t}{P_t}$, la variation relative du prix, évolue selon l'équation différentielle stochastique

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t$$

où μ et σ sont deux constantes appelées dérive et volatilité de l'action. Alors

$$dP_t = \mu P_t dt + \sigma P_t dW_t$$

et par la formule d'Itô, la solution est :

$$P_t = P_0 \exp[(\mu - \frac{\sigma^2}{2})t + \sigma W_t]$$

Exemple 3 : Pont brownien

$$\text{La solution de } \begin{cases} dX_t = (X_t/1-t)dt + dW_t \\ X_0 = 0 \end{cases}$$

est

$$(\forall 0 \leq t < 1) X_t = (1-t) \int_0^1 \frac{1}{1-s} dW_s$$

et est appelée le pont brownien entre l'origine au temps 0 et au temps 1.

Exemple 4 : Equation de Langevin

Une amélioration possible du mouvement brownien consiste à prendre en compte dans le mouvement des particules une force de frottement. Dans le cas monodimensionnel :

$$dX_t = -bX_t dt + \sigma dW_t$$

où $b > 0$ est le coefficient de frottement, et σ est le coefficient de diffusion. X_t est alors la vitesse de la particule brownienne. La solution, pour X_0 indépendant du mouvement, est :

$$(\forall t \geq 0) X_t = X_0 e^{-bt} + \sigma \int_0^t e^{-b(t-s)} dW_s$$

On remarque alors que

$$E(X_t) = e^{-bt} E(X_0)$$

et que

$$\begin{aligned} E(X_t^2) &= E(e^{-2bt} X_0^2 + 2\sigma e^{-bt} X_0 \int_0^t e^{-b(t-s)} dW_s + \sigma^2 (\int_0^t e^{-b(t-s)} dW_s)^2) \\ &= e^{-2bt} E(X_0^2) + 2\sigma e^{-bt} E(X_0) E(\int_0^t e^{-b(t-s)} dW_s) + \sigma^2 \int_0^t e^{-2b(t-s)} ds \\ &= e^{-2bt} E(X_0^2) + \frac{\sigma^2}{2b} (1 - e^{-2bt}) \end{aligned}$$

et la variance $V(X_t)$ est donnée par :

$$V(X_t) = e^{-2bt} V(X_0) + \frac{\sigma^2}{2b} (1 - e^{-2bt})$$

En supposant $V(X_0) < \infty$, il s'en suit que lorsque $t \rightarrow \infty$

$$\begin{cases} E(X_t) \rightarrow 0 \\ V(X_t) \rightarrow \frac{\sigma^2}{2b} \end{cases}$$

La distribution de X_t approche $\mathbb{N}(0, \frac{\sigma^2}{2b})$ lorsque $t \rightarrow \infty$. Ainsi, quelque soit la distribution initiale, la solution de l'équation différentielle stochastique pour des temps très grands suit approximativement une loi gaussienne centrée et de variance $\frac{\sigma^2}{2b}$, qui représente un équilibre entre la force aléatoire de perturbation dW_t (que nous avons appelé un bruit blanc) et la force de frottement $-bX_t$

Exemple 5 : Processus d'Ornstein-Uhlenbeck

Le mouvement brownien a été construit pour modéliser le déplacement d'une particule microscopique en suspension dans un liquide, soumise à l'agitation thermique. Une critique de cette modélisation est que les accroissements sont indépendants et ne dépendent pas de la vitesse de la particule au début de chaque période. Un modèle plus approprié est alors donné par l'équation d'Ornstein-Uhlenbeck :

$$\begin{cases} d^2 Y_t = -b dY_t + \sigma dW_t \\ Y_0 = y_0, dY_0 = y_1 \end{cases}$$

où Y_t est la position de la particule à l'instant t , y_0 et y_1 sont des variables aléatoires gaussiennes données. Si $X_t = dY_t$, le processus de vitesse satisfait l'équation de Langevin :

$$\begin{cases} dX_t = -bX_t dt + \sigma dW_t \\ X_0 = y_0 \end{cases}$$

et la solution est alors

$$(\forall t \geq 0) X_t = e^{-bt} X_0 + \sigma \int_0^t e^{-b(t-s)} dW_s$$

La position Y_t satisfait alors

$$(\forall t \geq 0) Y_t = Y_0 + \int_0^t X_s ds$$

et on montre qu'alors : $E(Y_t) = E(Y_0) + (\frac{1-e^{-bt}}{b})E(Y_1)$ et

$$V(Y_t) = V(Y_0) + \frac{\sigma^2 t}{b^2} + \frac{\sigma^2}{2b^3}(-3 + 4e^{-bt} - e^{-2bt})$$

2.6 Propriété de Markov des solutions des EDS

supposons que $E_x(a, b)$ ait une unique solution forte $\{X_t(x), t \geq 0\}$. Par linéarité de l'intégrale, on voit que pour tout $s, t \geq 0$,

$$X_{s+t}(x) = X_s(x) + \int_s^{s+t} a(u, X_u(x)) du + \int_s^{s+t} b(u, X_u(x)) dW_u \quad (2.24)$$

On peut écrire

$$X_{s+t}(x) = X_s(x) + \int_0^t a(s+u, X_{s+u}(x)) du + \int_0^t b(s+u, X_{s+u}(x)) dW'_u \quad (2.25)$$

et par unicité, ceci entraîne que $X_{s+t}(x) = X_t^s(X_s(x))$ pour tout $t \geq 0$, où $X_t^s(x)$ est l'unique solution de $E_x(a(s+.,.), b(s+.,.))$ portée par le Brownien W' .

Pour tout $s, t \geq 0$ et toute fonction borélienne f , on déduit alors que

$$E[f(X_{t+s})/\mathfrak{F}_s^B] = E[f(X_{t+s})/X_s] = \Phi(t; s, X_s)$$

Où $\Phi(t; s, x) = E[f(X_t^s(x))]$. Ceci signifie que X vérifie la **propriété de Markov inhomogène**. Dans le cas où les coefficients a et b ne dépendent pas du temps, X vérifie la **propriété de Markov homogène**,

$$E[f(X_{t+s})/\mathfrak{F}_s^B] = E[f(X_{t+s})/X_s] = \Phi(t; X_s)$$

avec $\Phi(t; x) = E[f(X_t(x))]$.

Dans le cas homogène, on peut étendre la propriété aux temps d'arrêt finis :

Théorème 2.11 [Propriété de Markov forte] Soit X l'unique solution forte de $E_x(a, b)$ avec a et b ne dépendant pas du temps. Soit T un temps d'arrêt fini p.s. Pour la filtration naturelle du Brownien porteur. Alors pour tout $t \geq 0$ et toute fonction f mesurable bornée,

$$E[f(X_{T+t})/\mathfrak{F}_T^B] = E[f(X_{T+t})/X_T] = \Phi(t; X_T) \quad (2.26)$$

où l'on a noté $\Phi(t; x) = E[f(X_t(x))]$

Dans le cas inhomogène, on montre que le processus espace-temps $t \mapsto (t, X_t)$ vérifie la propriété de Markov forte, au sens où pour tout temps d'arrêt fini T et toute fonction f deux variables,

$$E[f(T+t, X_{T+t})/\mathfrak{F}_T^B] = E[f(T+t, X_{T+t}/T, X_T)] = \Phi(t; T, X_T)$$

avec la notation $\Phi(t; s, x) = E[f(s+t, X_t^s(x))]$.

L'opérateur $\mathfrak{L}$ qui à $f \in C^2(\mathbb{R}, \mathbb{R})$ fait correspondre

$$\mathfrak{L}f : x \mapsto a(x)f'(x) + 1/2b^2(x)f''(x)$$

s'appelle le **générateur infinitésimal** du processus X . Il vérifie

$$\mathfrak{L}f(x) = \lim_{t \rightarrow 0} \frac{E_x[f(X_T)] - f(x)}{t}$$

pour tout $x \in \mathbb{R}^d$.

Le générateur infinitésimal a aussi un sens dans le cas inhomogène : si l'on considère l'EDS

$$X_t = X_0 + \int_0^t a(s, X_s)ds + \int_0^t b(s, X_s)dW_s,$$

alors on peut définir son générateur infinitésimal

$$\mathfrak{L}(f)(t, x) = a(t, x)f'_x(t, x) + \frac{b^2(t, x)}{2}f''_{xx}(t, x).$$

Si f est fonction de $\mathbb{R}_+ \times \mathbb{R}$ dans $\mathbb{R}$ telle que

$$\mathfrak{L}(f)(t, x) + f'_t(t, x) = 0$$

alors le processus $\{f(t, X_t), t \geq 0\}$ est une martingale locale.

2.7 Simulation des Equations Différentielles Stochastiques

Le traitement de la solution d'une équation différentielle stochastique par simulation est effectuée en élaborant des algorithmes numériques en temps discret. La simulation des trajectoires par la discrétisation du temps est l'approche la plus efficace pour les résoudre.

Néanmoins, ce n'est pas toutes les approximations par discrétisation du temps qui convergent vers la solution exacte quand la taille maximale des pas Δt tend vers zéro.

Ce problème a été déjà rencontré, en particulier dans la méthode de Runge-Kutta qui est une méthode numérique déterministe pour l'approximation des équations différentielles ordinaires (EDO).

Prenons le cas d'un processus défini par l'EDS suivante :

$$\begin{cases} dX_t = a(t, X_t)dt + b(t, X_t)dW_t & t \in [t_0, T] \\ X_0 = x_0 \end{cases} \quad (2.27)$$

Où $(W_t)_{t \in \mathbb{R}_+}$ est un processus de Wiener sur espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F})_{t \geq 0}, P)$,

$$\mathfrak{F} = \sigma(W_s, t_0 \leq s \leq t).$$

La mise en oeuvre pratique de ce processus va nécessiter la discrétisation de celui-ci. Pour cela on met l'équation (2.27) sous sa forme intégrale :

$$X_t = X_0 + \int_0^t a(s, X_s) ds + \int_0^t b(s, X_s) dW_s \quad (2.28)$$

Si le processus considéré ne dispose pas d'une discrétisation exacte, un développement d'Itô-Taylor de l'équation (2.28) nous permet de disposer d'une version discrétisée approximative. Cette approximation est d'autant plus précise que le développement intervient à un ordre élevé.

2.7.1 Discrétisation Exacte

La simulation d'un processus d'Itô pourra être effectuée directement (sans erreur de discrétisation) dès lors que celui-ci admet une discrétisation exacte. Rappelons la définition d'une discrétisation exacte.

Définition 2.12 un processus X admet une discrétisation exacte dès lors que l'on peut résoudre explicitement l'EDS qui lui associée.

Exemple : Cas d'une EDS de type linéaire (modèle de Black et Scholes)

$$\begin{cases} dX_t = \alpha X_t dt + B X_t dW_t & t \in [t_0, T] \\ X_0 \in \mathbb{R} \end{cases} \quad (2.29)$$

La solution explicite de cette équation est obtenue en appliquant la forme d'Itô (1.27) avec $a(t, X_t) = \alpha X_t$ et $b(t, X_t) = B X_t$.

La solution est donc :

$$X_t = \exp\{(\alpha - 1/2B^2)t + BW_t\} \quad (2.30)$$

$\forall t \in [t_0, T]$ et $\forall W = \{W_t, t \geq 0\}$

La trajectoire de la solution exacte est obtenue en évaluant le processus d'Itô aux instants de discrétisation

$$X_{\tau_n} = X_0 \exp\{(\alpha - 1/2B^2)\tau_n + B \sum_{i=1}^n \Delta W_{i-1}\}$$

où

$\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ est l'incrément du processus de Wiener d'ordre n , et

$\Delta_n = \tau_{n+1} - \tau_n$ est l'incrément du temps d'ordre n .

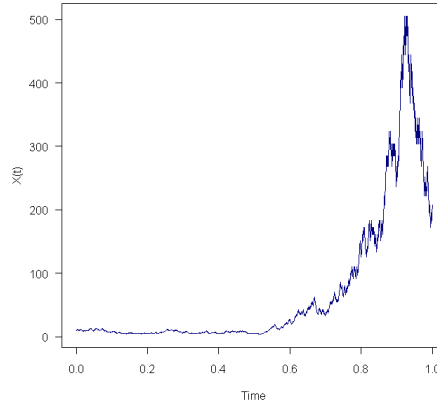


FIGURE 2.1 – Discrétisation Exacte (Modèle de Black et Scholes)

2.7.2 Discrétisation Approximative

Lorsque la discrétisation exacte n'existe pas, il convient de se tourner vers des approximations discrètes du processus continu sous-jacent. Les schémas d'Euler et de Milstein sont les procédés de discrétisation les plus répandus. Tous deux sont des développements d'Itô-Taylor de l'équation (2.28) à des ordres différents (ordre 1 pour Euler, ordre 2 pour Milstein).

Dans la suite, nous ferons référence au critère de convergence forte pour classer les procédés de discrétisation. Une discrétisation approximative $\tilde{X}$ converge fortement vers le processus continu X lorsque l'erreur commise sur la valeur finale (à la date T) de la trajectoire obtenue par le processus discrétisé est en moyenne asymptotiquement négligeable, i.e lorsque $\forall T > 0$,

$$\lim_{\Delta \rightarrow 0} E[|\tilde{X}_T^\Delta - X_T|] = 0 \quad (2.31)$$

La vitesse de convergence de l'équation (2.31) nous permet d'introduire un ordre entre les procédés de discrétisation. Ainsi le processus discrétisé $\tilde{X}$ converge fortement à l'ordre γ vers le processus X si : $\exists K > 0$, $\exists \delta_0 > 0$, $\forall \Delta \in]0, \delta_0]$

$$E[|\tilde{X}_T^\Delta - X_T|] \leq K\Delta^\gamma$$

Où Δ est la taille du pas équidistante.

Schéma d'Euler

Considérons le processus d'Itô $X = \{X_t, t_0 \leq t \leq T\}$ ayant la forme différentielle stochastique (2.28) l'approximation d'Euler consiste à discrétiser l'intervalle de temps $[t_0, T]$ en N pas $t_0 = \tau_0 < \tau_1 < \dots < \tau_n = T$, en générant itérativement des valeurs approchées aux trajectoires qui sont des valeurs exactes du processus d'Itô. Cette méthode appliquée à l'équation différentielle (2.27) implique l'approximation d'Euler qui est un processus stochastique à temps continu défini

par :

$$\begin{cases} Y_{n+1} = Y_n + a(\tau_n, Y_n)\Delta_n + b(\tau_n, Y_n)\Delta W_n \\ Y_0 = X_0 \quad n = 0, 1, 2, \dots, N-1 \end{cases}$$

tel que :

- * $Y_n = Y(\tau_n)$ est la valeur de l'approximation à l'instant de discrétisation τ_n .
- * $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ est l'incrément du processus de Wiener d'ordre n .
- * $\Delta_n = \tau_{n+1} - \tau_n$ est l'incrément du temps d'ordre n .

Pour obtenir la suite $\{Y_n\}_{n=0,1,\dots,N}$ aux instants $\{\tau_n, n = 0, 1, \dots, N\}$, on a besoin de générer les incréments aléatoires $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ pour $n = 0, 1, \dots, N-1$, du processus de Wiener $W = W_t, t \geq 0$. Ces variables aléatoires sont gaussiennes de moyenne nulle et de variance Δ_n : $\Delta W_n \rightarrow N(0, \Delta_n)$

Remarque :

Si les pas sont équidistants $\Delta_n = \Delta = \frac{T-t_0}{N}$ alors $\tau_n = t_0 + n\Delta$ et tel que pour N grand $\Delta \in [0, 1]$. Dans certaines applications, le temps de discrétisation n'est pas équidistant comme on a supposé précédemment, mais une variable aléatoire où l'on doit spécifier la taille maximale du pas $\delta = \max \Delta_n$.

Soit le temps de discrétisation $(\tau)_\delta = \{\tau_n, n = 0, 1, \dots\}$, qui est une variable aléatoire vérifiant $0 \leq \tau_0 < \tau_1 < \dots \leq \infty$ avec $\sup_{n_t} (\tau_{n+1} - \tau_n) \leq \delta$ et $n_t < \infty$, P.s $\forall t \in \mathbb{R}^+$, où τ_{n+1} est A_{τ_n} mesurable $\forall t \in \mathbb{R}^+$.

L'ensemble $\{A_t, t \geq 0\}$ est la famille croissante des σ -algèbres, associé généralement au processus d'Itô ou de Wiener.

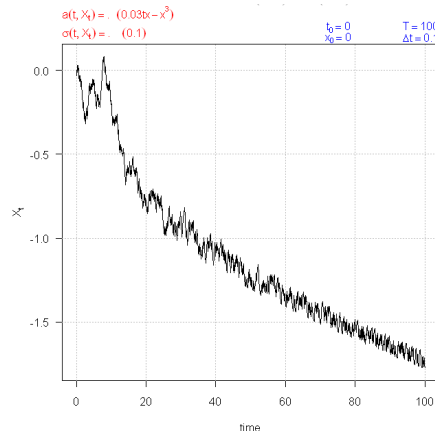


FIGURE 2.2 – Discrétisation Approximative (Schéma d'Euler)

Schéma de Milstein

Ce schéma d'approximation est proposé par Milstein, son ordre de convergence est supérieur à celui d'Euler, il est égal à 2. Le processus discrétisé est alors défini par

$$Y_{n+1} = Y_n + a(t, Y_n)\Delta_n + b(t, Y_n)\Delta W_n + 1/2b(t, Y_n)\frac{db}{dY_n}(t, Y_n)\{(\Delta W_n)^2 - \Delta_n\}$$

Remarquons que si la volatilité b est une constante, les procédés de discrétisation d'Euler et de Milstein conduiront à la même discrétisation.

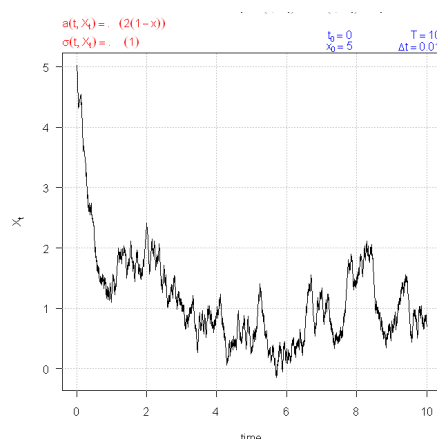


FIGURE 2.3 – Discrétisation Approximative (Schéma de Milstein)

2.8 Conclusion

Les équations différentielles stochastiques sont des outils largement employés en finance et en assurance, notamment pour modéliser les taux d'intérêts et les cours d'actions.

L'estimation des paramètres est une étape délicate dans la simulation de trajectoires d'un processus continu car elle peut être l'origine d'un biais, nous abordons dans le chapitre suivant l'estimation des paramètres de dérive dépendant du temps d'une équation différentielle stochastique linéaire, contrôlée, où la fonction de contrôle est donnée.

3

Estimation des paramètres de dérive d'une Equation Différentielle Stochastique

3.1 Introduction

Les modèles mathématiques basés sur un système d'équations différentielles stochastiques (EDS) sont largement utilisés dans de nombreuses applications en finance. Les solutions de ces modèles ne peuvent pas être exprimées par des fonctions élémentaires dans la plupart des cas. Ces solutions contiennent des paramètres inconnus qui sont généralement estimés par les données expérimentales obtenus à partir des conditions initiales bien définies. Le type de problème est généralement appelé l'estimation des paramètres. Dans notre étude Le coefficient de dérive contient des paramètres inconnus. La méthode de maximum de vraisemblance est utilisée pour montrer l'existence d'une famille d'estimateurs consistants, caractérisant le système.

3.2 Modèle Mathématique

On considère un système dynamique stochastique à temps continu régi par une équation de la forme :

$$\begin{cases} dX_t = (\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t))dt + \sigma dW_t \\ X_0 = 0, \quad t \geq 0 \end{cases} \quad (3.1)$$

X_t est l'état du système à l'instant t , $g(X_t)$ est le contrôle à l'instant t , donné par :

$$g(X_t) = \begin{cases} \theta X_t, & \text{si } |X_t| < 1/\theta \\ 1, & \text{si } X_t \geq 1/\theta, \quad \theta > 0 \\ -1, & \text{si } X_t \leq -1/\theta \end{cases}$$

W_t est un mouvement brownien standard, $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ sont des paramètres de dérive dépendant du temps qui sont inconnus, σ est le paramètre de diffusion que nous supposons constant dans notre étude.

Le problème est donc d'estimer $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ au vu de l'observation d'une trajectoire de l'état du système pour une réalisation de l'expérience dans l'intervalle de temps $[0, T]$.

D'abord, on définit une structure statistique et on montre qu'elle est dominée. On utilise la méthode de **maximum de vraisemblance** pour estimer les paramètres inconnus $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$.

3.2.1 Existence et unicité de la solution de l'équation (3.1)

Dans le but de construire une structure statistique pour le problème d'estimation des paramètres $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$, on montre l'existence et l'unicité d'une solution de l'équation (3.1). Donc l'existence d'une famille de lois de probabilité sur l'espace des trajectoires.

Lemme 3.1 L'équation différentielle stochastique (3.1) admet une solution forte, unique, presque sûrement à trajectoires continues.

Preuve :

On définit l'application $g, g : \mathbb{R} \rightarrow \mathbb{R}$ tel que :

$$g(X_t) = \begin{cases} \theta X_t, & \text{si } |X_t| < 1/\theta \\ 1, & \text{si } X_t \geq 1/\theta, \quad \theta > 0 \\ -1, & \text{si } X_t \leq -1/\theta \end{cases}$$

On pose $a(t, X) = \varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X)$, g étant impaire et continue, il suffit donc, pour assurer l'existence et l'unicité de la solution, de vérifier les hypothèses (1) et (2) du théorème(2.6), sur l'intervalle $[0, +\infty[$.

Pour l'hypothèse (1), on a :

Premier cas : si $X_t, Y_t \in]0, 1/\theta[$

on a

$$\begin{aligned} |a(t, X_t) - a(t, Y_t)| &= |\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t) - (\varphi_1^{(t)} Y_t + \varphi_2^{(t)} g(Y_t))| \\ &\leq |\varphi_1^{(t)} (X_t - Y_t) + \theta \varphi_2^{(t)} (X_t - Y_t)| \\ &\leq (|\varphi_1^{(t)}| + \theta |\varphi_2^{(t)}|) |X_t - Y_t| \end{aligned}$$

D'où la condition de Lipchitz est vérifiée.

Pour vérifier le deuxième hypothèse de théorème (2.6) on a :

$$\begin{aligned} |a(t, X_t)| &= |\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t)| \\ &= |\varphi_1^{(t)} X_t + \theta \varphi_2^{(t)} X_t| \\ &\leq (|\varphi_1^{(t)}| + \theta |\varphi_2^{(t)}|) |X_t| \\ &\leq (|\varphi_1^{(t)}| + \theta |\varphi_2^{(t)}|) (1 + |X_t|) \end{aligned}$$

D'où l'hypothèse de Borne à accroissement linéaire est vérifiée.

2ème cas : si $X_t, Y_t \in [1/\theta, +\infty[$

$$\begin{aligned} |a(t, X_t) - a(t, Y_t)| &= |\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t) - \varphi_1^{(t)} Y_t - \varphi_2^{(t)} g(Y_t)| \\ &= |\varphi_1^{(t)} X_t + \varphi_2^{(t)} - \varphi_1^{(t)} Y_t - \varphi_2^{(t)}| \\ &\leq |\varphi_1^{(t)}| |X_t - Y_t| \end{aligned}$$

Alors :

D'où le premier hypothèse est vérifiée.

Pour le deuxième hypothèse on a :

$$\begin{aligned} |a(t, X_t)| &= |\varphi_1^{(t)} X_t + \varphi_2^{(t)} g(X_t)| \\ &\leq |(\varphi_1^{(t)} X_t + \varphi_2^{(t)})| \\ &\leq K |1 + X_t| \end{aligned}$$

D'où l'hypothèse de Borne à accroissement linéaire est vérifiée.

3ème cas : si $X_t \in [1/\theta, +\infty[$, $Y_t \in [0, 1/\theta]$

$$|a(X_t) - a(Y_t)| = |a(X_t - a(1/\theta) + a(1/\theta) - a(Y_t))|$$

Le membre de droite de l'égalité précédente est majoré par :

$$|\varphi_1^{(t)}|(X_t - 1/\theta) + (|\varphi_1^{(t)}| + |\varphi_2^{(t)}\theta|)(1/\theta - Y_t) \leq (|\varphi_1^{(t)}| + |\varphi_2^{(t)}\theta|)|X_t - Y_t|$$

Pour le deuxième hypothèse on a : Pour tout $X_t \in \mathbb{R}^n$, $|a(X_t)| = |\varphi_1^{(t)}X_t + \varphi_2^{(t)}g(X_t)|$ Le membre de droite de l'égalité précédente est majoré par :

$$\max(|\varphi_1^{(t)}|, |\varphi_2^{(t)}|)(1 + |X_t|).$$

Par conséquent :

Pour tout X_t et Y_t dans $\mathbb{R}^n$ et

$$K = \max(|\varphi_1^{(t)}| + |\varphi_2^{(t)}|\theta, \max(|\varphi_1^{(t)}|, |\varphi_2^{(t)}|))$$

On a :

$$|a(X_t) - a(Y_t)| \leq K|X_t - Y_t|$$

et

$$|a(x_t)| \leq K(1 + |X_t|)$$

pour tout X_t et Y_t dans $\mathbb{R}^n$. Le lemme est démontré.

Conclusion :

Alors l'équation (3.1) admet une solution $(\pi_t)_{t \geq 0}$ forte unique presque sûrement à trajectoires continues.

3.3 Structure statistique du problème d'estimation

Etant donné des entiers p et q , $p \geq 1$, $q \geq 1$, et T un nombre réel $T > 0$ on désigne par :

- * C (resp C_T) l'espace des applications continues de $\mathbb{R}^+$ (resp $[0, T]$) dans $\mathbb{R}^p$,
- * S (resp S_T) la tribu sur C engendrée par les applications coordonnées $\pi_t : X \rightarrow X_t$, pour tout $t \geq 0$ (resp $0 \leq t \leq T$),
- * Q (resp Q^T) la loi du mouvement brownien canonique sur (C, S) (resp (C_T, S_T)).

On observe sur un intervalle de temps $[0, T]$, $T > 0$ une fonction continue $X(\cdot) = (X_t; t \geq 0)$ élément de l'espace C_T , trajectoire de l'état d'un système dynamique stochastique.

Pour $\varphi \in (H, \mathbb{R})$ et $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P_{\varphi}^T, (W_t)_{t \geq 0})$ donnée, on a d'après le lemme (3.1). L'existence et l'unicité de la solution forte de (3.1), et donc l'unicité en loi, au sens que toutes les solutions induisent la même mesure sur l'espace des trajectoires C_T .

On est donc conduit à étudier le problème d'estimation du paramètre φ , au vu de l'observation de $X(\cdot)$ sur l'intervalle $[0, T]$, dans la structure statistique $(C_T, S_T, P_{\varphi}^T, \varphi \in (H, \mathbb{R}), \varphi_i^{(t)} < 0)$.
Où P_{φ}^T désigne la restriction à S_T de la loi $P_{\varphi^{(t)}}$ du processus $(X_t; t \geq 0)$ solution de l'équation (3.1).

3.3.1 Théorème de Girsanov

Le théorème de Girsanov (Voir [26]) est le résultat fondamental utilisé pour l'équivalence et l'absolue continuité des lois de processus de type diffusion.

Théorème 3.2 Soient $\tilde{W}$ et W deux processus $\mathbb{R}^P$ -mesurables définis sur l'espace probabilisé filtré $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t \geq 0}, P)$, satisfaisant les propriétés suivantes :

- a. W est un mouvement brownien, P -dimensionnel,
- b. $\tilde{W} = \int_0^t B(\cdot) ds + W_t$, P.ps, où $B(\cdot)$ est une application de Ω dans $\mathbb{R}^P$, mesurable et telle que :
 $P[\int_0^T \|B(\cdot)\|^2 ds < \infty] = 1$
- c. $E\{\exp[\int_0^T \langle B(\cdot), dW_s \rangle - 1/2 \int_0^T \|B(\cdot)\|^2 ds]\} = 1$

Alors :

Le processus $\tilde{W}$ est un mouvement brownien relativement à la mesure de probabilité $\tilde{P}$ définie par :

$$\frac{d\tilde{P}}{dP} = \exp\left[\int_0^T \langle B(\cdot), dW_s \rangle - \frac{1}{2} \int_0^T \|B(\cdot)\|^2 ds\right]$$

les hypothèses a et c sont satisfaites pour notre problème, et nous avons le lemme suivant :

Lemme 3.3 Pour tout $\varphi \in (H, \mathbb{R})$ et tout $T > 0$, les lois P_{φ}^T et Q^T sont équivalentes, la dérivée de **Randon-Nikodym** est :

$$\frac{dP_{\varphi}^T}{dQ^T} = \exp\left\{1/\sigma^2 \left(\int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\pi_s - \frac{1}{2} \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s))^2 ds \right)\right\} \quad (3.2)$$

De plus relativement à P_{φ}^T , π_t est solution de l'équation (3.1) pour le processus de Wiener $(N_t)_{t \geq 0}$ définit par :

$$N_t = \pi_t - \int_0^t (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) ds \quad (3.3)$$

On a d'après le théorème de Girsanov :

$$E_Q^T \left\{ 1/\sigma^2 \left(\exp \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\pi_s - \frac{1}{2} \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s))^2 ds \right) \right\} = 1 \quad (3.4)$$

et de plus, pour tout $T > 0$ et tout $\varphi \in (H, \mathbb{R})$, les lois P_{φ}^T, Q^T sont équivalentes.

Remarque :

Le lemme précédent est la clé de notre problème de l'identification du paramètre φ^t dans l'équation (3.1). En effet, il montre que la structure statistique est dominée et de plus, il fournit explicitement la fonction de log-vraisemblance pour l'estimation de φ^t par la méthode du maximum de vraisemblance.

3.4 La fonction de log-vraisemblance pour l'estimation des paramètres de dérive dépendants du temps

D'après (3.2), une fonction log-vraisemblance $L_T(\varphi^{(t)}, \pi_t)$ est donnée par :

$$L_T(\varphi^{(t)}, \pi_t) = 1/\sigma^2 \left[\int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\pi_s - \frac{1}{2} \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s))^2 ds \right] \quad (3.5)$$

$$\begin{aligned} L_T(\varphi^{(t)}, \pi_t) &= 1/\sigma^2 \left[\int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\pi_s - \frac{1}{2} \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s))^2 ds \right] \\ &= 1/\sigma^2 \left[\int_0^T \varphi_1^{(s)} \pi_s d\pi_s + \int_0^T \varphi_2^{(s)} g(\pi_s) d\pi_s - 1/2 \left(\int_0^t (\varphi_1^{(s)})^2 \pi_s^2 ds \right. \right. \\ &\quad \left. \left. + 2 \int_0^T \varphi_1^{(s)} \varphi_2^{(s)} \pi_s g(\pi_s) ds + \int_0^T (\varphi_2^{(s)})^2 (g(\pi_s))^2 ds \right) \right] \end{aligned}$$

On désigne par $L_T^1(\varphi^{(t)}, \pi_t)$ et $L_T^2(\varphi^{(t)}, \pi_t)$ les dérivées première et seconde de $L_T(\varphi^{(t)}, \pi_t)$ obtenues par dérivation formelle de (3.5)

Lemme 3.4

$$\begin{aligned} \text{a. } L_T^1(\varphi^{(t)}, \pi_t) &= 1/\sigma^2 \int_0^T \begin{bmatrix} d\varphi_1^{(s)} \pi_s \\ d\varphi_2^{(s)} g(\pi_s) \end{bmatrix} dN_s \\ \text{b. } L_T^2(\varphi^{(t)}, \pi_t) &= 1/\sigma^2 \left[- \int_0^T \begin{bmatrix} (d\varphi_1^{(s)})^2 \pi_s^2 \\ (d\varphi_2^{(s)} g(\pi_s))^2 \end{bmatrix} ds + \int_0^T \begin{bmatrix} d^2 \varphi_1^{(s)} \pi_s \\ d^2 \varphi_2^{(s)} g(\pi_s) \end{bmatrix} dN_s \right] \end{aligned}$$

Preuve (a) :

Par dérivation formelle de (3.5) on obtient :

$$\begin{aligned}
 L_T^1(\varphi^{(t)}, \pi_t) &= \begin{bmatrix} \frac{\partial L_T(\varphi^{(t)}, \pi_t)}{\partial \varphi_1^{(t)}} \\ \frac{\partial L_T(\varphi^{(t)}, \pi_t)}{\partial \varphi_2^{(t)}} \end{bmatrix} \\
 &= 1/\sigma^2 \left[\begin{aligned} &\int_0^T d\varphi_1^{(s)} \pi_s d\pi_s - \int_0^T \varphi_1^{(s)} d\varphi_1^{(s)} \pi_s^2 ds - \int_0^T d\varphi_1^{(s)} \varphi_2^{(s)} \pi_s g(\pi_s) ds \\ &\int_0^T d\varphi_2^{(s)} g(\pi_s) d\pi_s - \int_0^T \varphi_1^{(s)} d\varphi_2^{(s)} \pi_s g(\pi_s) ds - \int_0^T \varphi_2^{(s)} d\varphi_2^{(s)} (g(\pi_s))^2 ds \end{aligned} \right] \\
 &= 1/\sigma^2 \left[\begin{aligned} &\int_0^T d\varphi_1^{(s)} \pi_s d\pi_s - \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\varphi_1^{(s)} \pi_s ds \\ &\int_0^T d\varphi_2^{(s)} g(\pi_s) d\pi_s - \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\varphi_2^{(s)} g(\pi_s) ds \end{aligned} \right] \\
 &= 1/\sigma^2 \left[\begin{aligned} &\int_0^T d\varphi_1^{(s)} \pi_s [d\pi_s - (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) ds] \\ &\int_0^T d\varphi_2^{(s)} g(\pi_s) [d\pi_s - (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) ds] \end{aligned} \right] \\
 &= 1/\sigma^2 \int_0^T \begin{bmatrix} d\varphi_1^{(s)} \pi_s dN_s \\ d\varphi_2^{(s)} g(\pi_s) dN_s \end{bmatrix}
 \end{aligned}$$

Preuve (b) :

Par dérivation seconde de (3.5) on obtient :

$$\begin{aligned}
 L_T^2(\varphi^{(t)}, \pi_t) &= \begin{bmatrix} \frac{\partial^2 L_T(\varphi^{(t)}, \pi_t)}{\partial \varphi_1^{(t)}} \\ \frac{\partial^2 L_T(\varphi^{(t)}, \pi_t)}{\partial \varphi_2^{(t)}} \end{bmatrix} \\
 &= 1/\sigma^2 \left[\begin{aligned} &\int_0^T d^2\varphi_1^{(s)} \pi_s^2 d\pi_s - \int_0^T (d\varphi_1^{(s)})^2 \pi_s^2 ds - \int_0^T \varphi_1^{(s)} d^2\varphi_1^{(s)} \pi_s^2 ds \\ &\quad - \int_0^T d^2\varphi_1^{(s)} \varphi_2^{(s)} \pi_s g(\pi_s) ds \\ &\int_0^T d^2\varphi_2^{(s)} g(\pi_s) d\pi_s - \int_0^T \varphi_1^{(s)} d^2\varphi_2^{(s)} \pi_s g(\pi_s) ds - \\ &\int_0^T (d\varphi_2^{(s)})^2 (g(\pi_s))^2 ds - \int_0^T \varphi_2^{(s)} d^2\varphi_2^{(s)} (g(\pi_s))^2 ds \end{aligned} \right] \\
 &= 1/\sigma^2 \left[\begin{aligned} &-\int_0^T (d\varphi_1^{(s)})^2 \pi_s^2 ds + \int_0^T d^2\varphi_1^{(s)} \pi_s [d\pi_s - (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) ds] \\ &-\int_0^T (d\varphi_2^{(s)})^2 (g(\pi_s))^2 ds + \int_0^T d^2\varphi_2^{(s)} g(\pi_s) [d\pi_s - (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) ds] \end{aligned} \right] \\
 &= 1/\sigma^2 \left[\begin{aligned} &-\int_0^T (d\varphi_1^{(s)})^2 \pi_s^2 ds + \int_0^T d^2\varphi_1^{(s)} \pi_s dN_s \\ &-\int_0^T (d\varphi_2^{(s)})^2 (g(\pi_s))^2 ds + \int_0^T d^2\varphi_2^{(s)} g(\pi_s) dN_s \end{aligned} \right]
 \end{aligned}$$

3.5 Existence de l'estimateur du maximum de vraisemblance

D'après la preuve (a), le système d'équations, $L_T^1(\varphi^{(t)}, \pi_t) = 0$ peut s'écrire sous la forme suivante :

$$\begin{aligned}
 L_T^1(\varphi^{(t)}, \pi_t) &= 1/\sigma^2 \left[\int_0^T \begin{bmatrix} d\varphi_1^{(s)} & 0 \\ 0 & d\varphi_2^{(s)} \end{bmatrix} \begin{pmatrix} \pi_s \\ g(\pi_s) \end{pmatrix} d\pi_s - \int_0^T \begin{bmatrix} d\varphi_1^{(s)} & 0 \\ 0 & d\varphi_2^{(s)} \end{bmatrix} \begin{bmatrix} \pi_s^2 & \pi_s g(\pi_s) \\ \pi_s g(\pi_s) & g(\pi_s)^2 \end{bmatrix} \begin{pmatrix} \varphi_1^{(s)} \\ \varphi_2^{(s)} \end{pmatrix} ds \right] \\
 &= 0
 \end{aligned}$$

L'estimateur du maximum de vraisemblance s'obtient donc par résolution du système (3.6), Ce dernier est très compliqué à résoudre à cause des paramètres qui dépendent du temps, nous proposons de le résoudre dans le cas où les paramètres sont constants (Voir [6]) alors, le système (3.6) s'écrit sous la forme suivante :

$$1/\sigma^2 \left[\int_0^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \pi_s \\ g(\pi_s) \end{pmatrix} d\pi_s - \int_0^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_s^2 & \pi_s g(\pi_s) \\ \pi_s g(\pi_s) & g(\pi_s)^2 \end{bmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} ds \right] = 0 \quad (3.6)$$

Ce qui revient à écrire :

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \int_0^T \begin{bmatrix} \pi_s^2 & \pi_s g(\pi_s) \\ \pi_s g(\pi_s) & g(\pi_s)^2 \end{bmatrix} ds = \int_0^T \begin{pmatrix} \pi_s \\ g(\pi_s) \end{pmatrix} d\pi_s$$

Alors, si P_φ^T -presque sûrement, la matrice

$$\int_0^T \begin{bmatrix} \pi_s^2 & \pi_s g(\pi_s) \\ \pi_s g(\pi_s) & g(\pi_s)^2 \end{bmatrix} ds$$

est non singulière (Voir [6]), et la dérivée seconde de (3.7) est négative alors on définit l'estimateur $\hat{\varphi}$ comme l'unique solution de (3.7), soit, P_φ^T -presque sûrement,

$$\begin{aligned} \hat{\varphi} &= \left[\int_0^T \begin{bmatrix} \pi_s^2 & \pi_s g(\pi_s) \\ \pi_s g(\pi_s) & g(\pi_s)^2 \end{bmatrix} ds \right]^{-1} \left[\int_0^T \begin{pmatrix} \pi_s \\ g(\pi_s) \end{pmatrix} d\pi_s \right] \\ &= \left[\int_0^T \begin{bmatrix} \pi_s \\ g(\pi_s) \end{bmatrix}^2 ds \right]^{-1} \left[\int_0^T \begin{pmatrix} \pi_s \\ g(\pi_s) \end{pmatrix} d\pi_s \right] \end{aligned}$$

3.6 Application Numérique "Simulation exacte de la trajectoire $(\pi_t)_{t \geq 0}$ "

On suppose les valeurs actuelles des paramètres $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ comme suit :

$$\begin{aligned} \varphi_1^{(t)} &= \sin(t) \\ \varphi_2^{(t)} &= \cos(t) \end{aligned}$$

La discrétisation par le schéma d'Euler sur le système (3.1) donne :

Pour tout $t \in [0, T]$, on a :

$$\begin{cases} Y_{n+1} = Y_n + (\varphi_1^{(\tau_n)} Y_n + \varphi_2^{(\tau_n)} g(Y_n)) \Delta_n + \sigma \Delta W_n \\ Y_0 = X_0 = 0 \quad n = 0, 1, 2, \dots, N-1 \end{cases}$$

tel que :

* $Y_n = Y(\tau_n)$ est la valeur de l'approximation à l'instant de discrétisation τ_n .

* $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ est l'incrément du processus de Wiener d'ordre n.

* $\Delta_n = \tau_{n+1} - \tau_n$ est l'incrément du temps d'ordre n .

Pour obtenir la suite $\{Y_n\}_{n=0,1,\dots,N}$ aux instants $\{\tau_n, n = 0, 1, \dots, N\}$, on a besoin de générer les incréments aléatoires $\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}$ pour $n = 0, 1, \dots, N-1$, du processus de Wiener $W = W_t, t \geq 0$. Ces variables aléatoires sont gaussiennes de moyenne nulle et de variance $\Delta_n = T - 0/N$.

Ce qui revient à écrire :

$$Y_{n+1} = \begin{cases} Y_n + ((\sin(\tau_n)Y_n + \cos(\tau_n)))\frac{T}{N} + \sigma(W_{\tau_{n+1}} - W_{\tau_n}) & \text{si } |Y_n| < 1/\theta \\ Y_n + (\sin(\tau_n)Y_n + \cos(\tau_n))\frac{T}{N} + \sigma(W_{\tau_{n+1}} - W_{\tau_n}) & \text{si } Y_n \geq 1/\theta \\ Y_n + (\sin(\tau_n)Y_n - \cos(\tau_n))\frac{T}{N} + \sigma(W_{\tau_{n+1}} - W_{\tau_n}) & \text{si } Y_n \leq -1/\theta \end{cases}$$

avec $Y_0 = X_0 = 0$ et $n = 0, 1, 2, \dots, N-1$

On suppose $\theta = \sigma = 1$ et pour tout $t \in [0, 1]$ on a la simulation suivante :

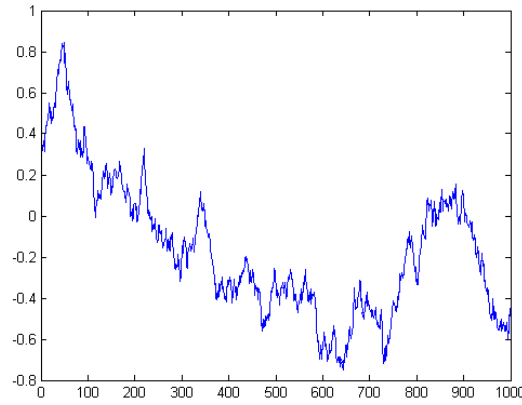


FIGURE 3.1 – Discrétisation par le Schéma d'Euler sur le système (3.1)

3.7 Conclusion

L'estimation des paramètres des modèles en temps continu n'est pas immédiate. De nombreux processus, tels que le modèle (3.1) n'admet pas de discrétisation exacte, et comme sa fonction de vraisemblance déterminée est trop complexe, à calculer, on est alors conduit à utiliser des méthodes heuristiques, comme l'algorithme génétique. Le problème d'estimation de modèle (3.1) se formule mathématiquement par un problème d'optimisation (maximisation de log vraisemblance), où les paramètres de dérive sont les paramètres inconnus qu'on cherche à déterminer.

4

Métaheuristiques et Inférences Statistiques

4.1 Introduction

Les heuristiques (du verbe grec *heuriskein*, qui signifie "trouver") sont un outil essentiel de l'optimisation appliquée, et pour de nombreux problèmes pratiques de grande taille le seul applicable.

Feignebaum et Feldman (1963) définissent une heuristique comme une règle d'estimation, une stratégie, une astuce, une simplification, ou tout autre sorte de système qui limite drastiquement la recherche des solutions dans l'espace des configurations possibles. Newell, Shaw et Simon (1957) précisent qu'un processus heuristique peut résoudre un problème donné, mais n'offre pas la garantie de le faire.

La métaheuristique, elle, se place à un niveau plus général encore, et intervient dans toutes les situations où l'ingénieur ne connaît pas d'heuristique efficace pour résoudre un problème donné, ou lorsqu'il estime qu'il ne dispose pas du temps nécessaire pour en déterminer une.

En 1996, I.H. Osman et G. Laporte ont défini la métaheuristique comme "un processus itératif qui subordonne et qui guide une heuristique, en combinant intelligemment plusieurs concepts pour explorer et exploiter tout l'espace de recherche. Des stratégies d'apprentissage sont utilisées pour structurer l'information afin de trouver efficacement des solutions optimales, ou presque-optimales".

Les métaheuristiques qui comprennent notamment la méthode du recuit simulé (Kirkpatrick et al., 1983), la méthode de recherche Tabou (Glover, 1989), les algorithmes de colonies de fourmis, les algorithmes génétiques etc . . . , présentent des inconvénients :

- Raisonnement par analogie avec les systèmes réels (physique, biologie, éthologie, . . .).
 - Stochastiques
- et partagent aussi les mêmes inconvénients :

- Difficulté de réglage des paramètres de la méthode
- Temps de calcul élevé.

Dans ce chapitre, nous nous intéressons uniquement aux algorithmes génétiques qui ont connu ces quinze dernières années un développement considérable grâce à l'augmentation vertigineuse de la puissance des calculateurs et notamment suite à l'apparition des architectures massivement parallèles qui exploitent leur parallélisme intrinsèque.

Le chapitre est organisé en trois parties. La première est consacrée aux algorithmes génétiques comme étant une méthode d'optimisation mono-objectif. Après une présentation générale du principe de fonctionnement d'un algorithme génétique, nous ferons une description détaillée de ses différentes composantes. La deuxième partie est consacrée aux algorithmes génétiques à codage réels (RCGA), une version améliorée des AG. Enfin, une application des algorithmes génétiques à codage réel en inférence statistique.

4.2 Algorithmes génétiques

Les algorithmes génétiques (AG) développés par J. Holland (Holland, 1992) présentent des qualités intéressantes pour la résolution des problèmes d'optimisation complexes. Leurs fondements théoriques furent exposés par Goldberg (Goldberg, 1994). Ils tentent de simuler le processus d'évolution des espèces dans leur milieu naturel : soit une transposition artificielle de concepts basiques de la génétique et des lois de survie énoncés par Darwin. Rappelons que la génétique représente un individu par un code, c'est-à-dire un ensemble de données (appelées chromosomes), identifiant complètement l'individu. La reproduction est, dans ce domaine, un mixage aléatoire de chromosomes de deux individus, donnant naissance à des individus enfants ayant une empreinte génétique nouvelle, héritée des parents. La mutation génétique est caractérisée dans le code génétique de l'enfant par l'apparition d'un chromosome nouveau, inexistant chez les individus parents.

Ce phénomène génétique d'apparition de " mutants " est rare mais permet d'expliquer les changements dans la morphologie des espèces, toujours dans le sens d'une meilleure adaptation au milieu naturel. La disparition de certaines espèces est expliquée par " les lois de survie " selon lesquelles seuls les individus les mieux adaptés auront une longévité suffisante pour générer une descendance. Les individus peu adaptés auront une tendance à disparaître. C'est une sélection naturelle qui conduit de génération en génération à une population composée d'individus de plus en plus adaptés.

4.3 Présentation des Algorithmes Génétiques (GA)

Selon Lerman et Ngounet (1995) un Algorithme Génétique est défini par :

Définition 4.1 (Individu : solution potentielle)

$\forall i \in [1, l]$, une séquence A de longueur $\mathcal{L}(A)$ est une suite $A = \{a_1, a_2, \dots, a_l\}$, où $a_i \in \{0, 1\}$.

Définition 4.2 (Population)

C'est un ensemble de solutions potentielles.

Définition 4.3 (Chromosome)

Un chromosome est une représentation d'une solution potentielle, est donc une suite de bits en codage binaire, appelée aussi chaîne binaire. Dans le codage non binaire, tel que le codage réel, la suite A ne contient qu'un point a avec $a \in \mathbb{R}$.

Définition 4.4 (fonction Fitness)

C'est une fonction définie par l'utilisateur (la fonction de coût ou la fonction de profit), appelée fonction d'adéquation (traduction de mot anglais fitness).

4.4 Codage et population initiale

Le premier pas dans l'implémentation des Algorithmes Génétiques est de créer une population d'individus initiaux. En effet, les Algorithmes Génétiques agissent sur une population d'individus, et non pas sur un individu isolé.

Par analogie avec la biologie, chaque individu de la population est codé par un chromosome ou génotype (Hollond 1975) en ce qui concerne la représentation binaire d'un individu, et phénotype pour ce qui est de sa valeur réelle correspondante dans l'espace de recherche (Voir [22]). Il existe principalement trois types de codage : le codage binaire, le codage réel et le codage en base n .

4.4.1 Codage Binaire

Ce codage a été le premier à être utilisé dans le domaine des AG. Il présente plusieurs avantages : alphabet minimum 0, 1, facilité de mise en point d'opérateurs génétiques et existence de fondements théoriques (théorie sur les schémas), voir [22]. Néanmoins ce type de codage présente quelques inconvénients :

- a. Les performances de l'algorithme sont dégradées devant les problèmes d'optimisation de grande dimension à haute précision numérique. Pour de tels problèmes, les AG basés sur les chaînes binaires ont de faibles performances comme le montre michalewicz (Michalewicz, 1992), voir [15].

- b. La distance de Hamming entre deux nombres voisins (nombre de bits différents) peut être assez grande dans le codage binaire : l'entier 7 correspond à la chaîne 0111 et la chaîne 1000 correspond à l'entier 8. Or la distance de hamming entre ces deux chaînes est de 4, ce qui crée bien souvent une convergence, et non pas l'obtention de la valeur optimale.

4.4.2 Codage Réel

Il a le mérite d'être simple. Chaque chromosome est en fait un vecteur dont les composantes sont les paramètres du processus d'optimisation. Par exemple, si on recherche l'optimum d'une fonction de n variables $f(x_1, x_2, \dots, x_{n-1}, x_n)$, on peut utiliser tout simplement un chromosome contenant les n variables : Avec ce type de codage, la procédure d'évaluation des chromosomes est plus rapide vu l'absence de l'étape de transcodage (du binaire vers le réel). Les résultats donnés par Michalewicz (Michalewicz, 1992) voir [15], montrent que la représentation réelle aboutit souvent à une meilleure précision et un gain important en termes de temps d'exécution.

4.4.3 Codage en base n

Dans ce type de codage¹, les gènes constituant un chromosome sont des chiffres exprimés dans une base de numération n , ce qui permet de représenter n valeurs discrètes. L'AG démarre avec une population composée de N individus dans le codage retenu. Le choix des individus conditionne fortement la rapidité de l'algorithme. Si la position de l'optimum dans l'espace de recherche est totalement inconnue, il est intéressant que la population soit répartie sur tout l'espace de recherche. Si par contre des informations à priori sur le problème sont disponibles, il paraît évident de générer les individus dans un espace particulier afin d'accélérer la convergence. Disposant d'une population initiale souvent non homogène, la diversité de la population doit être entretenue aux cours des générations afin d'explorer le plus largement possible l'espace de recherche. C'est le rôle des opérateurs de croisement et de mutation.

1. Le lecteur pourra consulter l'Annexe pour une étude complète du codage

4.5 Algorithmes Génétiques à Codage Binaire (BCGA)

L'Algorithme Génétique à Codage Binaire, est un algorithme itératif de recherche globale dont le but d'optimiser une fonction définie par l'utilisateur (fitness) ; pour atteindre cet objectif, l'algorithme travaille en parallèle sur une population de l'espace de recherche.

La fonction d'adéquation d'un individu est normalement indépendante des autres individus de la population (elle est explicitement donnée par l'utilisateur). Chaque individu (Chromosome) est constitué d'un ensemble d'éléments appelés caractéristiques où gènes, pouvant prendre plusieurs valeurs (allèles) appartenant à un alphabet non nécessairement numérique. Dans l'algorithme de base, les allèles possibles sont 0 et 1, et un chromosome est donc une chaîne binaire. Le but est donc de chercher la combinaison optimale de ces éléments, qui donne lieu au maximum d'adéquation. A chaque itération, appelée génération, est créée une nouvelle population avec le même nombre d'individus. Cette nouvelle génération consiste généralement en des individus mieux "adaptés" à l'environnement tel qu'il est représenté par la fonction d'adaptation. Au fur et à mesure des générations, les individus vont tendre en général vers l'optimum de la fonction d'adaptation.

La génération d'une nouvelle population à partir de la précédente s'effectue en cinq étapes

- a. **Initialisation** : une population initiale de taille N chromosomes est tirée aléatoirement.
- b. **Evaluation** : chaque chromosome est décodé puis évalué.
- c. **Sélection** : création d'une nouvelle population de N chromosomes par l'utilisation d'une méthode de sélection appropriée.
- d. **Reproduction** : croisement et mutation de certains chromosomes au sein de la nouvelle population.
- e. **Retour** à la phase 2 tant que la condition d'arrêt du problème n'est pas satisfaite.

4.5.1 Evaluation

L'Algorithme Génétique évalue la fonction d'adéquation f_i de chaque individu i de l'ancienne population.

4.5.2 Opérateur de Sélection

Cet opérateur est peut-être le plus important puisqu'il permet aux individus d'une population de survivre de se reproduire ou de mourir. En règle générale, la probabilité de survie d'un individu sera directement reliée à son efficacité relative au sein de la population.

Il existe plusieurs types de sélection, citons par exemple :

Sélection déterministe

C'est la méthode de sélection la plus simple à mettre en oeuvre, puisqu'elle consiste à choisir les n meilleurs individus parmi la population, n étant un paramètre fixé (égal à la

moitié de la population par exemple).

Elle présente l'inconvénient de faire se conserver les éléments ayant une mauvaise performance, si la performance de la population est globalement mauvaise.

Sélection par tournois

Lors de la sélection par tournoi (Tournament selection), k individus de la population sont choisis aléatoirement et celui, dont la performance est la plus élevée, est retenu pour participer à la reproduction. L'opération répétée autant de fois qu'il ya d'individus à sélectionner.

La probabilité qu'un individu de rang i soit sélectionné après compétition est donnée ci-après :

$$P_s(a_i^t) = \begin{cases} \frac{C_{N-i}^{k-1}}{C_N^k} & 1 \leq i \leq N-k+1 \\ 0 & \text{si } N-k+2 \leq i \leq N \end{cases} \quad (4.1)$$

Dans le cas particulier du tournoi de deux individus ($k = 2$), qualifié de tournoi binaire probabiliste (Binary Tournament Selection), (4.1) se réduit à (4.2).

$$P_s(a_i^t) = 2/N \frac{N-i}{N-1} \quad (4.2)$$

Sélection par roulette (roulette wheel)

Cette méthode est la plus connue et la plus utilisée. Avec cette méthode, chaque individu a une chance d'être sélectionné proportionnelle à sa performance, donc plus les individus sont adaptés au problème, plus ils ont de chances d'être sélectionnés ; plus précisément, l'opérateur de sélection sélectionne chaque individu i avec une probabilité $f_i / \sum f_j$ (La somme étant prise sur les n individus constituant la population).

Pour utiliser l'image de la " Roue du Forain ", chaque individu se voit attribué un secteur dont l'angle est proportionnel à son adaptation, sa " fitness ".

On fait tourner la roue et quand elle cesse de tourner on sélectionne l'individu correspondant au secteur désigné par une sorte de " curseur ", curseur qui pointe sur un secteur particulier de celle-ci après qu'elle se soit arrête de tourner. Par exemple : avec une population de quatre individus, dont les valeurs de performance seraient de 169, 576, 64 et 361, on obtiendrait la roulette suivante :

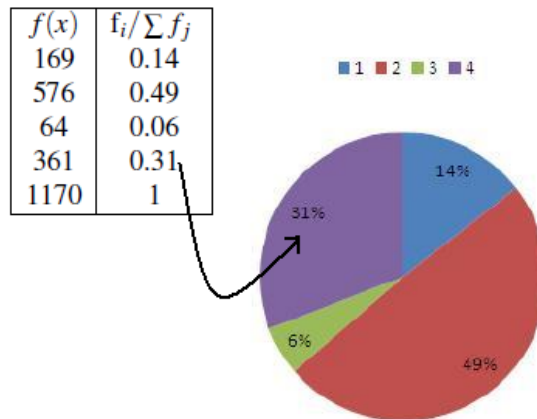


FIGURE 4.1 –

Avec cette technique, il faut s'arranger pour que f_i prenne des valeurs positives, en bornant inférieurement la fonction à 0, ou à une très petite valeur, pour que tout individu, même mauvais, ait quelque chance d'être sélectionné.

Cette méthode, bien que largement répandue, a pas mal d'inconvénients.

En effet, elle a une forte variance. Il n'est pas impossible que sur n sélections successives destinées à désigner les parents de la nouvelle génération P' , la quasi-totalité, voire pire la totalité des n individus sélectionnés soient des individus ayant une fitness vraiment mauvaise et donc que pratiquement aucun individu voire aucun individu a forte fitness ne fasse partie des parents de nouvelle génération. Ce phénomène est bien sûr très dommageable car cela va complètement à l'encontre du principe des Algorithmes Génétiques qui veulent que les meilleurs individus soient sélectionnés de manière à converger vers une solution la plus optimale possible.

A l'inverse, on peut arriver à une domination écrasante d'un individu " localement supérieur " ceci entraîne une grave perte de diversité. Imaginons par exemple qu'on ait un individu ayant une fitness très élevée par rapport au reste de la population, disons dix fois supérieurs, il n'est pas impossible qu'après quelques générations successive on se retrouve avec une population ne contenant que des copies de cet individu. Le problème est que cet individu avait une fitness très élevée, mais seulement en comparaison des autres individus. On se retrouve donc face à problème connu sous le nom de " convergence prématurée " ; l'évolution se met donc à stagner et on atteindra alors jamais l'optimum, on restera bloqué sur un optimum local.

Il existe certaines techniques pour essayer de limiter ce phénomène, comme par exemple le " Scaling " ², qui consiste à effectuer un changement d'échelle de manière à augmenter ou diminuer de manière forcée la fitness d'un individu par rapport à un autre selon leur écart de fitness, (Voir [25]).

2. Le lecteur pourra consulter l'Annexe pour une étude complète du Scaling

Sélection par rang

La sélection par rang trie d'abord la population par fitness. Ensuite, chaque chromosome se voit associé un rang en fonction de sa position. Ainsi le plus mauvais chromosome aura le rang 1, le suivant le rang 2, et ainsi de suite jusqu'au meilleur chromosome qui aura le rang m (pour une population de m chromosomes). La sélection par rang d'un chromosome est la même que par roulette, mais les proportions sont en relation avec le rang plutôt qu'avec la valeur de l'évaluation. Le tableau suivant fournit un exemple de sélection par rang. Avec cette méthode de sélection, tous les chromosomes ont une chance d'être sélectionnés. Cependant, elle conduit à une convergence plus lente vers la bonne solution. Ceci est dû au fait que les meilleurs chromosomes ne diffèrent pas énormément des plus mauvais. Par exemple une sélection par rang pour 6 chromosomes :

Chromosome	1	2	3	4	5	6	Totale
Probabilité Initial	0.89	0.05	0.01	0.04	0.03	0.02	1
Rang	6	5	1	4	3	2	21
Probabilité Final	0.29	0.24	0.05	0.19	0.14	0.09	1

4.5.3 Opérateur de Reproduction avec Croisement et Mutation

On distingue les opérateurs de croisement, qui génèrent de nouveaux individus à partir de plusieurs (un couple, le plus souvent), et les opérateurs de mutation, qui transforment un seul individu.

Pour pouvoir croiser, il faut déjà représenter les solutions.

C'est dans les années 60 que John Holland, le père des Algorithmes Génétiques, a mis au point un système de représentation binaire, inspiré du modèle phénotype-génotype du monde vivant. Ce modèle est devenu classique, quoique controversé.

Dans ce système, les solutions du problème sont comparables à des chromosomes, et sont constituées d'une série de gènes, chaque gène étant associé à une séquence de bits, à une chaîne binaire codée en fonction d'une règle de conversion particulière. Croiser des individus consiste alors à " croiser " les séquences binaires de deux parents, pour former les " génomes " des enfants.

Il existe plusieurs méthodes :

Croisement à un-point (One-point crossover)

La technique de croisement d'un point est la technique de croisement la plus simple, mais elle est très efficace.

La marche à suivre du croisement d'un-point est de choisir un point de croisement au hasard.

Les gènes allant jusqu'au point de croisement sont copiés à l'enfant respectif, les gènes restants sont copiés à l'enfant alternatif. Supposer que les chromosomes de parent 1 et de parent 2 sont comme suit : Si la position 3 est aléatoirement choisie comme position

Parent 1 :	3	6	7	1	8	4	2	5
Parent 2 :	1	8	4	7	3	2	5	6

FIGURE 4.2 –

de croisement, alors les deux enfants sont comme suit : Où le premier enfant a les trois

Enfant 1 :	3	6	7	7	3	2	5	6
Enfant 2 :	1	8	4	1	8	4	2	5

FIGURE 4.3 – One-point crossover

premiers gènes du parent1 et les autres gènes du parent 2, le deuxième enfant a les trois premiers gènes du parent 2 et les autres gènes du parent1.

Croisement à deux-points (Two-point crossover)

La marche à suivre du croisement de deux points est assimilée à celle du croisement d'un point sauf que deux positions sont choisies et seulement les bits entre les deux positions sont permutés. La première partie et la dernière partie de chromosomes sont préservées. Avec les mêmes parents ci-dessus, les positions 2 et 5 sont produites comme positions de croisement, alors les deux enfants sont comme suit :

Enfant 1 :	3	6	4	7	3	4	2	5
Enfant 2 :	1	8	7	1	8	2	5	6

FIGURE 4.4 – Two-point crossover

Croisement à N-points (N-points crossover)

La marche à suivre du croisement N-point est également-assimilée au croisement à un point sauf que des positions de N sont choisies et seulement les bits entre une position de croisement impair et la suivante position sont permuté et vice versa.

Les éléments entre deux positions consécutives sont alternativement échangés et maintenus

Parent1 :	3	6	2	1	8	4	2	5
Parent2 :	1	8	4	7	3	2	3	6

FIGURE 4.5 –

Enfant1 :	3	8	2	1	3	4	2	6
Enfant2 :	1	6	4	7	8	2	3	5

FIGURE 4.6 – N-points crossover

Croisement Uniforme (Uniform crossover)

On définit de manière aléatoire un " masque " qui est une chaîne de nombre compris entre 0 et 1 de même longueur que les chromosomes parents.

Ce masque est destiné à savoir, pour chaque gènes, de quel parent les fils devront hériter du gène s'y trouvant, si le nombre aléatoire produit pour une position donnée est moins de 0.5, on copie le gène du parent 1 sur le gène de l'enfant 1, et le gène du parent2 sur le gène de l'enfant 2 et vice versa L'opérateur de croisement permet la création de deux nouveaux

Parent1 :	3	6	2	1	8	4	2	5
Parent2 :	1	8	4	7	3	2	3	6
Masque :	0.5	0.2	0.4	0.6	0.8	0.4	0.2	0.8
Enfant1 :	1	6	2	7	3	4	2	6
Enfant2 :	3	8	4	1	8	2	3	5

FIGURE 4.7 – Uniform crossover

individus. Toutefois, un individu sélectionné lors de la reproduction ne subit pas nécessairement l'action d'un croisement. Ce dernier ne s'effectue qu'avec une certaine probabilité. Plus cette probabilité est élevée et plus la population subira de changement.

Quoi qu'il en soit, il se peut que l'action conjointe de reproduction et du croisement soit insuffisante pour assurer la réussite de l'AG.

Ainsi, dans le cas du codage binaire, certaines informations (c.à.d. caractères de l'alphabet) peuvent disparaître de la population. Ainsi aucun individu de la population initiale ne contient de 1 en dernière position de la chaîne, et que ce 1 fasse partie de la chaîne optimale pour trouver tous les croisements possibles ne permettront pas de faire apparaître ce 1 initialement inconnue.

En effet, imaginons par exemple que nous avons une population d'individus possédant un seul chromosome. Considérons un gène particulier de ce chromosome, on l'appellera g , gène ayant 2 allèles possible 0 et 1 ; si aucun individu de la population initiale ne possède l'allèle 1 pour ce gène, aucun croisement possible ne permettra d'introduire cet allèle pour notre gène g . Si la solution optimale au problème est telle que notre gène l'allèle 1, il nous sera impossible d'atteindre cette solution optimale simplement par sélection et croisement. C'est pour remédier entre autre à ce problème que l'opérateur de mutation est utilisé.

Opérateur de Mutation

Le rôle de cet opérateur est de modifier aléatoirement, avec une certaine probabilité, la valeur d'un composant de l'individu dans le cas du codage binaire, chaque bit $a_i \in \{0, 1\}$ est remplacé selon une probabilité P_M par son inverse $a_i = 1 - a_i$. C'est ce qu'illustre la figure suivante. Tout comme plusieurs lieux de croisement peuvent être possibles, nous pouvons très bien admettre qu'une même chaîne puisse subir plusieurs mutations.

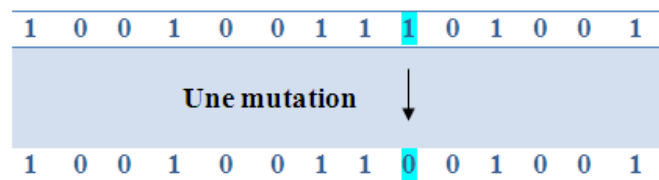


FIGURE 4.8 –

L'opérateur de mutation modifie donc de manière complètement aléatoire les caractéristiques d'une solution, ce qui permet d'introduire et de maintenir la diversité au sein de notre population de solution. Cet opérateur joue le rôle d'un " élément perturbateur ", il introduit de " bruit " au sein de la population. Ainsi il dispose de 3 grands avantages :

- Il garantit la diversité de la population, ce qui est primordial pour les algorithmes génétiques.
- Il permet d'éviter un phénomène connu sous le nom de **dérive génétique**, on parle de dérive génétique quand certains gènes favorisés par le hasard se répandent au détriment des autres et sont ainsi présents au même endroit sur tous les chromosomes.

Le fait que l'opérateur de mutation puisse entraîner de manière aléatoire des changements au niveau de n'importe quel locus permet d'éviter l'installation de cette situation défavorable.

La mutation alors entraînant des inversions de bits de manière aléatoire permet de réintroduire des différences entre les individus et donc de nous extirper de cette situation

- La mutation permet d’atteindre **la propriété d’ergodicité**, L’ergodicité est une propriété garantissant que chaque point de l’espace de recherche puisse être atteint.

En effet, une mutation pouvant intervenir de manière aléatoire au niveau de n’importe quel locus, on a la certitude mathématique que n’importe quelle permutation de notre chaîne de bits peut apparaître au sein de la population et donc que tout point de l’espace de recherche peut être atteint.

Grâce à cette propriété on est donc sûr de pouvoir atteindre l’optimum global. (Voir [28])

Elitisme

La stratégie élitiste consiste à conserver le meilleur individu de chaque génération. Ainsi l’élitisme empêche l’individu le plus performant de disparaître au cours de la sélection ou que ses bonnes combinaisons soient affectées par les opérateurs de croisement et de mutation. Après chaque évaluation de la performance des individus à une génération G donnée, le meilleur individu de la génération précédente ($G - 1$) est réintroduit dans la population si aucun des individus de la génération G n’est meilleur que lui. Par cette approche, la performance du meilleur individu de la population courante est monotone de génération en génération. Il apparaît que l’élitisme améliore considérablement les performances de l’algorithme génétique pour certaines classes de problèmes, mais peut les dégrader à d’autres classes, en augmentant le taux de convergences prématurées.

Paramètres de base d’un Algorithme Génétique

Les opérateurs de l’Algorithme Génétique sont guidés par un certain nombre de paramètres fixés à l’avance.

La valeur de ces paramètres influence la réussite ou non d’un Algorithme Génétique. Ces paramètres sont les suivants :

- La taille de la population, N , et la longueur du codage de chaque individu I (dans le cas du codage binaire). Si N est trop grand le temps de calcul de l’Algorithme peut s’avérer très important, et si N est trop petit, il peut converger trop rapidement vers un mauvais chromosome.

Cette importance de la taille est essentiellement due à la notion de parallélisme implicite qui implique que le nombre d’individus traité par l’algorithme est au moins proportionnelle au cube de nombre d’individus.

- La probabilité de croisement P_C . Elle dépend de la forme de la fonction fitness. Son choix est en général heuristique. Plus elle est élevée, plus la population subit de changements importants. Les valeurs généralement admises sont comprises entre 0.5 et 0.9.
- La probabilité de mutation P_M est généralement faible puisque une probabilité élevée risque de conduire à une solution sous optimale, tandis que une valeur de bonne pratique est 0.05. (Voir [22])

Remarque Le choix de la fonction de fitness retenue est important et dépend du problème à résoudre et de l'espace de recherche qui en découle.

4.6 Algorithmes Génétiques à Codage Réel (RCGA)

De nombreuses versions modifiées des Algorithmes Génétiques originaux ne se basent plus sur un codage binaire des paramètres à optimiser, mais travaillent directement sur les paramètres eux mêmes (Voir [28] et [22]). Ces versions des Algorithmes Génétiques, que nous appellerons **Algorithmes Génétiques à Codage Réel (RCGA)**, offrent généralement l'avantage d'être mieux adaptées aux problèmes d'optimisation numérique continus, d'accélérer la recherche et de rendre plus aisé le développement de méthode hybrides avec des méthodes plus classiques.

Ce faisant, ces versions sont sans doute plus proches des besoins et des habitudes des praticiens industriels pour la résolution de problèmes réels ; c'est pourquoi elles sont plus répandues dans le milieu industriels que les Algorithmes Génétiques originaux (Voir [10]). Les algorithmes génétiques à codage réel sont similaires à ceux présentés précédemment, tant en ce qui concerne leur structure que leurs mécanismes de sélection. Par contre, ils sont caractérisés par des opérateurs de recombinaison génétiques qui utilisent directement les valeurs des paramètres réels pour évoluer vers l'optimum. Le chromosome des individus est donc représenté par les n variables objets dont l'intervalle de variation est défini par les contraintes de domaine du problème.

4.6.1 Le Croisement

Deux types de croisement différents sont employés par les algorithmes génétiques à codage réel :

- **Le croisement discret** est assez similaire aux méthodes de croisement utilisées dans le cas du codage binaire. Il consiste à permuter les variables de chaque parent entre des sites du chromosome choisis aléatoirement. Le croisement discret peut être à un site, multi-sites ou uniforme comme le montre la figure (4.9). cette technique a un taux d'exploration relativement limité car elle ne fait apparaître aucune valeur nouvelle pour les paramètres.

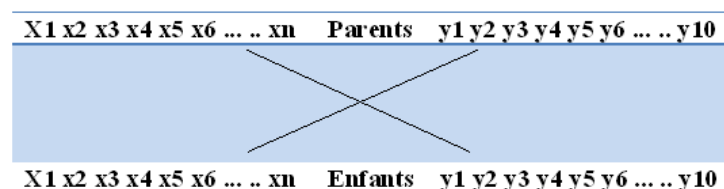


FIGURE 4.9 – Principe du croisement discret

- **Le croisement continu** assure l'exploration de l'espace en testant des nouvelles valeurs de variables objets. Il est appliqué paramètre par paramètre en utilisant le mécanisme suivant :

$$\begin{cases} x'_i = x_i + a_1(y_i - x_i) \\ y'_i = y_i + a_2(x_i - y_i) \end{cases} \quad \text{Où } x_i \text{ et } y_i \text{ désignent les paramètres (de même rang) des deux parents, } x'_i \text{ et } y'_i \text{ étant les paramètres (de rang correspondant) des deux enfants obtenus. } a_1 \text{ et } a_2 \text{ sont deux variables prises dans l'intervalle } [0, 1].$$
Lorsque a_1 et a_2 sont deux variables aléatoires distinctes, le croisement est continu asymétrique. Lorsque a_1 et a_2 désignent une même valeur aléatoire, le croisement est continu symétrique. Enfin, si $a_1 = a_2 = 1/2$, le croisement est continu arithmétique. Il conduit alors à un unique enfant représentant une moyenne arithmétique des parents. Dans, ce cas il est nécessaire d'effectuer un nombre de croisement deux fois plus important si nous voulons obtenir autant d'enfants que de parents à chaque génération.

4.6.2 La mutation

Dans le cas d'un codage réel, on utilise principalement deux opérateurs de mutation : la mutation uniforme et la mutation non uniforme (Michalewicz, 1992). En supposant fixée la probabilité de mutation P_M , un tirage au sort pour chaque gène x_k d'un chromosome I permet de décider si ce gène doit être ou non modifié. Nous supposons que le gène prend ses valeurs dans un intervalle $[x_k^{min}, x_k^{max}]$.

Pour la mutation uniforme, qui est une simple extension de la mutation binaire, on remplace le gène x_k sélectionné par une valeur quelconque x'_k tirée aléatoirement dans l'intervalle $[x_k^{min}, x_k^{max}]$.

Pour la mutation non uniforme, le calcul de la nouvelle valeur d'un gène est un peu plus complexe. Le gène x_k subit des modifications importantes durant les premières générations, puis graduellement décroissantes au fur et à mesure que l'on progresse dans le processus d'optimisation. Pour une génération G , on tire au sort une valeur binaire qui décidera si le changement doit être positif ou négatif. La nouvelle valeur x'_k du gène x_k est donnée par :

$$x'_k = \begin{cases} x_k + \Delta(G, x_k^{max} - x_k) & \text{si rand}=0 \\ x_k - \Delta(G, x_k - x_k^{min}) & \text{si rand}=1 \end{cases}$$

Où $\Delta(G, y)$ est une fonction qui définit l'écart entre la nouvelle valeur et la valeur initiale à la génération G et rand est nombre aléatoire qui prend les valeurs 0 ou 1.

Dans (Cordon et al., 2001), les auteurs proposent d'utiliser une fonction $\Delta(G, y)$ correspondante à une décroissance exponentielle de l'écart à travers les générations. Cette fonction est définie par :

$$\Delta(G, y) = y \times (1 - r^{(1-G/\gamma)^\beta}) \quad (4.3)$$

Où γ est l'indice de génération pour la quelle l'amplitude de la mutation s'annule, β est un paramètre de l'opérateur de mutation (souvent $\beta = 5$), r est un nombre aléatoirement dans l'intervalle $[0, 1]$ et G le numéro de la génération.

Si par contre, l'intervalle de variations du gène x_k n'est pas connu, une mutation gaussienne est souvent utilisée. Le principe de base de ce type de mutation est d'ajouter un bruit gaussien centré $N(0, \sigma)$ au gène que l'on désire faire muter :

$x'_k = x_k + N(0, \sigma)$, où σ représente la variance.

La mutation est traditionnellement considérée comme un opérateur intervenant à la marge, dans la mesure où sa probabilité est en général assez faible (de l'ordre de 0.01). Mais elle confère aux algorithmes génétiques une propriété très importante : l'ergodicité (tous les points de l'espace de recherche peuvent être atteints). Cet opérateur est donc d'une grande importance et il est loin d'être marginal, d'ailleurs un algorithme génétique peut converger sans l'opérateur de croisement et certaines applications fonctionnent de cette manière (Hwang, 1999).

4.7 RCGA pour l'estimation des paramètres de dérive dépendant du temps d'une EDS

4.7.1 Formulation du problème d'optimisation

L'estimation des paramètres des modèles en temps continu n'est pas immédiate. De nombreux processus, tels que le modèle (3.1) étudié dans le chapitre (3), n'admet pas de discrétisation exact. Et sa fonction de vraisemblance calculée s'avère, trop complexe, à maximiser, on est alors conduit à utiliser des méthodes heuristiques. Comme l'algorithme génétique à codage réel, pour l'estimation des paramètres du modèle considéré (Voir [24]). En premier lieu, comme dans tous les problèmes d'optimisation il faut bien définir le problème à optimiser, la quantité à optimiser, et aussi savoir quels paramètres on veut déterminer.

La première chose à faire est de définir notre problème d'optimisation.

Le problème d'estimation des paramètres de dérive d'une EDS se formule mathématiquement par un problème d'optimisation (maximisation de log vraisemblance), sous la forme suivante :

$$(P) \begin{cases} \text{Max} & L_T(\varphi^{(t)}; \pi_t) \\ \varphi^{(t)} \in \mathcal{L} = \{p_0, p_1, \dots, p_p\} \subseteq \mathbb{R}^p \end{cases}$$

- $L_T(\varphi^{(t)}; \pi_t)$: fonction log vraisemblance donnée par l'égalité (3.5).
- π_t : solution forte unique presque sûrement à trajectoires continues du modèle (3.1).
- $\varphi^{(t)} = (\varphi_1^{(t)}, \varphi_2^{(t)})$: les paramètres de dérive du modèle (3.1) qu'on les cherche à déterminer.
- $\mathcal{L}$: L'ensemble de développement limité de l'ensemble des paramètres $\phi^{(t)} = (\varphi_1^{(t)}, \varphi_2^{(t)})$ à l'ordre (p).

4.7.2 Présentation de l'Algorithme Génétique à Codage Réel

Individu Un individu est une solution potentielle tel que :

$$I = (\varphi_1^{(t)}, \varphi_2^{(t)}).$$

Chromosome Un chromosome est un vecteur dont les composantes sont les paramètres du problème à optimiser.

$$I = Ch = (\varphi_1^{(t)}, \varphi_2^{(t)}).$$

Population C'est un ensemble de solutions potentielles.

Lorsqu'on crée une population initiale, les valeurs des paramètres sont créées aléatoirement avec une distribution uniforme sur l'intervalle spécifié. Comme exemple, on considère une population composée de N individus. On veut estimer m variables $(x_1, x_2, \dots, x_m)$, donc chaque chromosome aura m gènes. Donc la population est :

$$\text{population} = \begin{cases} I_1 = (x_1^1, x_1^2, \dots, x_1^m) \\ I_2 = (x_2^1, x_2^2, \dots, x_2^m) \\ I_3 = (x_3^1, x_3^2, \dots, x_3^m) \\ \vdots \\ I_N = (x_n^1, x_n^2, \dots, x_n^m) \end{cases}$$

Chaque individu I_i est composé d'un chromosome à m gènes dont les valeurs sont choisies aléatoirement.

fonction Fitness Dans notre étude la fonction fitness est la fonction de log-vraisemblance calculée dans le chapitre (3), elle est donnée par :

$$L_T(\varphi^{(t)}, \pi_t) = 1/\sigma^2 \left[\int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s)) d\pi_s - \frac{1}{2} \int_0^T (\varphi_1^{(s)} \pi_s + \varphi_2^{(s)} g(\pi_s))^2 ds \right] \quad (4.4)$$

4.7.3 Structure de l'Algorithmes Génétiques à Codage Réel (RCGA)

L'organigramme fonctionnel de la Figure suivante illustre la structure de l'algorithme génétique à codage réel (RCGA). Nous détaillerons les diverses phases qui le constituent et présenterons les mécanismes associés à chacune d'entre elles dans les sections suivantes.

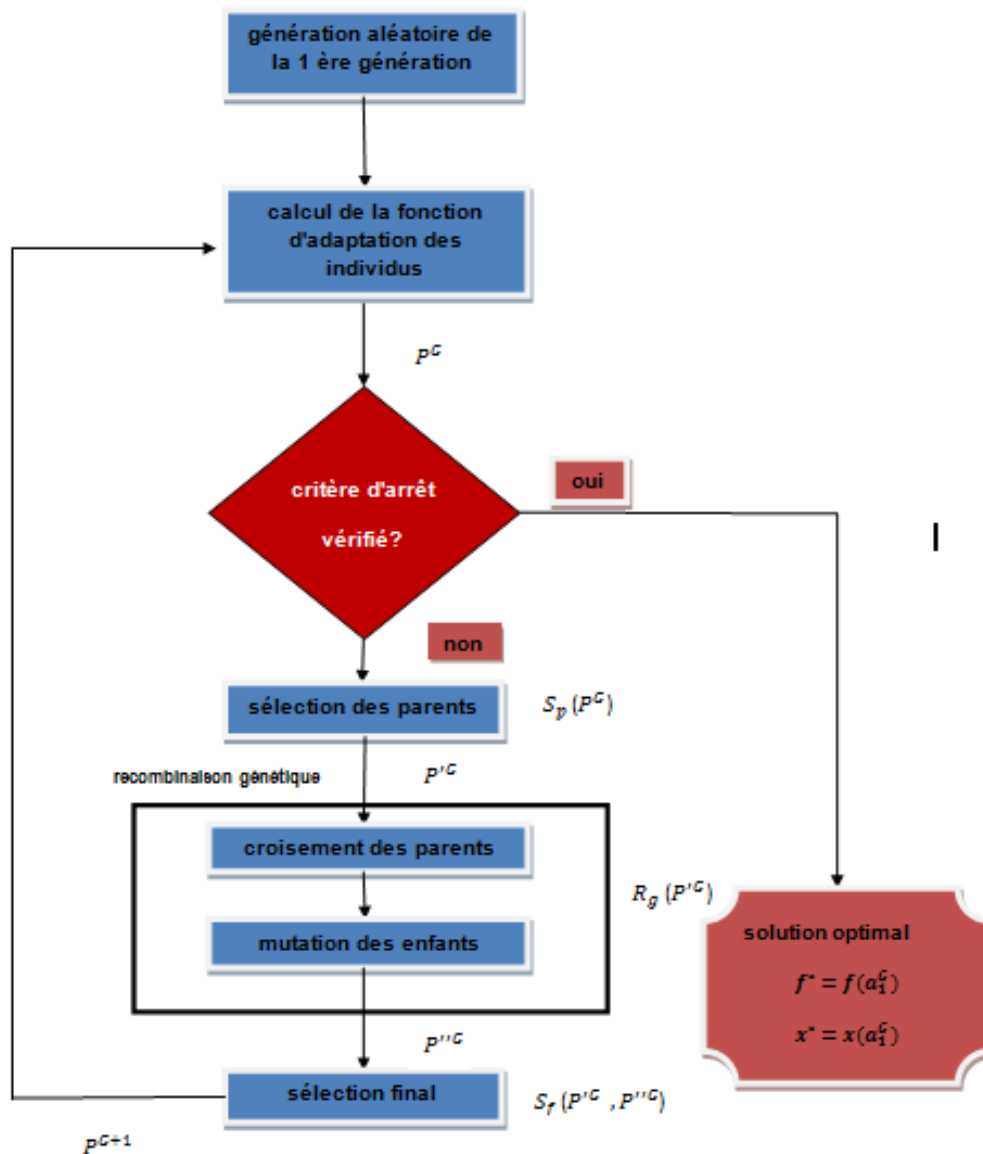


FIGURE 4.10 – Structure de la RCGA

4.7.4 Implémentation de l'Algorithmes Génétiques à Codage Réel

Les Algorithmes Génétiques à codage réel (RCGA) sont similaires à ceux présentés précédemment, tant en ce qui concerne leur structure que leurs mécanismes de sélection. Par contre, ils sont caractérisés par des opérateurs de recombinaison génétiques qui utilisent directement les valeurs des paramètres réels pour évoluer vers l'optimum. Nous détaillerons dans la suite les diverses phases qui constituent cet algorithme.

(a) Population initiale

Dans notre étude on a deux paramètres à estimer $(\varphi_1^{(t)}, \varphi_2^{(t)})$ qui sont en fonction de temps dont les valeurs sont choisies de la façon suivante :

Pour tout $t \in [0, T]$, on suppose les valeurs actuelles des paramètres $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ les DL_p au voisinage 0 pour tout entier p des fonction usuelles $\sin(t)$ respectivement $\cos(t)$.

$$\begin{aligned}\varphi_1^{(t)} &= \sum_{i=0}^p \frac{(-1)^i}{(2i+1)!} t^{2i+1} + o(t^{2p+1}) \\ \varphi_2^{(t)} &= \sum_{i=0}^p \frac{(-1)^i}{(2i)!} t^{2i} + o(t^{2p})\end{aligned}$$

Pour p donnée

On considère une population composé de p^2 individus, chaque individu I_{ij} composé de deux paramètres $(\varphi_1^{(t)}(i), \varphi_2^{(t)}(j))$, tel que :

Pour $i, j = 1, \dots, p$

$\varphi_1^{(t)}(i)$ = Le DL_i de la fonction $\sin(t)$ au voisinage 0 pour tout entier i .

$\varphi_2^{(t)}(j)$ = Le DL_j de la fonction $\cos(t)$ au voisinage 0 pour tout entier j .

Donc la population est donné sous la forme suivante :

$$\text{population} = \begin{cases} I_{11} = (\varphi_1^{(t)}(1), \varphi_2^{(t)}(1)) \\ I_{12} = (\varphi_1^{(t)}(1), \varphi_2^{(t)}(2)) \\ \vdots \\ I_{kl} = (\varphi_1^{(t)}(k), \varphi_2^{(t)}(l)) \\ \vdots \\ I_{pp} = (\varphi_1^{(t)}(p), \varphi_2^{(t)}(p)) \end{cases}$$

Une fois la population initiale créée, il faut ensuite procéder à la première étape du calcul, l'évaluation

(b) **Evaluation**

On évalue la performance de chaque individu à l'aide de la fonction objective. La performance est une mesure de l'exactitude de la solution obtenue avec les paramètres d'un individu. Cette fonction objective varie d'un problème à l'autre. elle varie aussi si on essaie de maximiser ou de minimiser un problème. Elle est calculée à partir des valeurs des paramètres de chaque individu. On obtient donc une valeur de fonction objective pour chaque individu, donc N valeurs au total pour l'ensemble de la population.

La fonction objective dans le travail effectué ici est la fonction log vraisemblance discrétisée $L'_T(\varphi^{(t)}; \pi_t)$ de $L_T(\varphi^{(t)}; \pi_t)$.

La fonction log-vraisemblance discrétisée $L'_T(\varphi^{(t)}; \pi_t)$

Soit D le nombre de discrétisation de l'intervalle $[0, T]$, tel que :

$$0 = \tau_0 < \tau_1 < \dots < \tau_D \leq T.$$

Pour tout $n=0, \dots, D$, on a :

$$\begin{aligned} L'_{\tau_n}(\varphi^{\tau_n}; \pi_{\tau_n}) &= \sum_{n=1}^D (\varphi_1^{(\tau_n)} \pi_{\tau_n} + \varphi_2^{(\tau_n)} g(\pi_{\tau_n})) \Delta \pi_n - \frac{1}{2} \sum_{n=1}^D [(\varphi_1^{(\tau_n)} \pi_{\tau_n})^2 + 2\varphi_1^{(\tau_n)} \pi_{\tau_n} \varphi_2^{(\tau_n)} g(\pi_{\tau_n}) \\ &\quad + (\varphi_2^{(\tau_n)} g(\pi_{\tau_n}))^2] \Delta_n \quad (5.1) \end{aligned}$$

* $\Delta \pi_n = (\pi_{\tau_{n+1}} - \pi_{\tau_n})$ est l'incrément du solution $(\pi_t)_{t \geq 0}$ simulée à l'instant de discrétisation τ_n .

* $\Delta_n = \tau_{n+1} - \tau_n$ est l'incrément du temps d'ordre n .

$$* g(\pi_{\tau_n}) = \begin{cases} \theta \pi_{\tau_n}, & \text{si } |\pi_{\tau_n}| < 1/\theta \\ 1, & \text{si } \pi_{\tau_n} \geq 1/\theta, \\ -1, & \text{si } \pi_{\tau_n} \leq -1/\theta \end{cases} \quad \theta > 0$$

Pour obtenir les valeur de $L'_{\tau_n}(\varphi^{\tau_n}; \pi_{\tau_n})$ aux instants $\{\tau_n, n = 1, \dots, D\}$, on a besoin de calculer les incréments $(\pi_{\tau_{n+1}} - \pi_{\tau_n})$ pour $n = 1, \dots, D$, du solution simulée $(\pi_t)_{t \geq 0}$.

La solution $(\pi_t)_{t \geq 0}$ simulée pour p^2 individus

Pour tout $t \in [0, 1]$, $p = 10$ et $\theta = 1$

La solution π_{τ_n} simulée pour chaque individu est donnée par les graphes suivants :

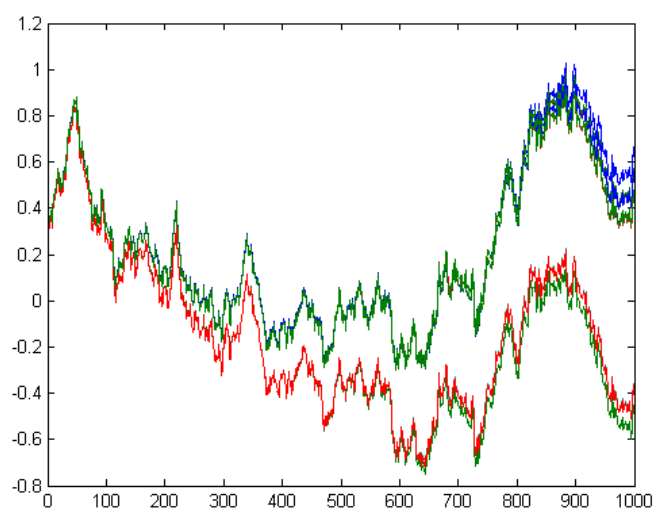


FIGURE 4.11 – La solution π_{τ_n} simulée pour 100 individus à 1000 discrétisation

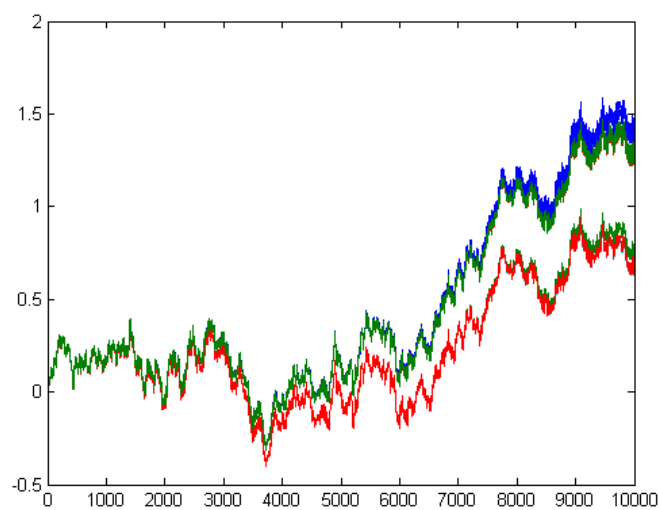


FIGURE 4.12 – La solution π_{τ_n} simulée pour 100 individus à 10000 discrétisation

Une fois la solution π_{τ_n} est simulée, ensuite on évalue la performance de chaque individu à l'aide de la fonction objective L'_{τ_n} donnée par la formule (5.1). On obtient une valeur de fonction objective pour chaque individu, au total p^2 valeurs pour l'ensemble de la population.

Pour tout $t \in [0, 1], D = 1000, \theta = 1$, et $p = 10$, on a :

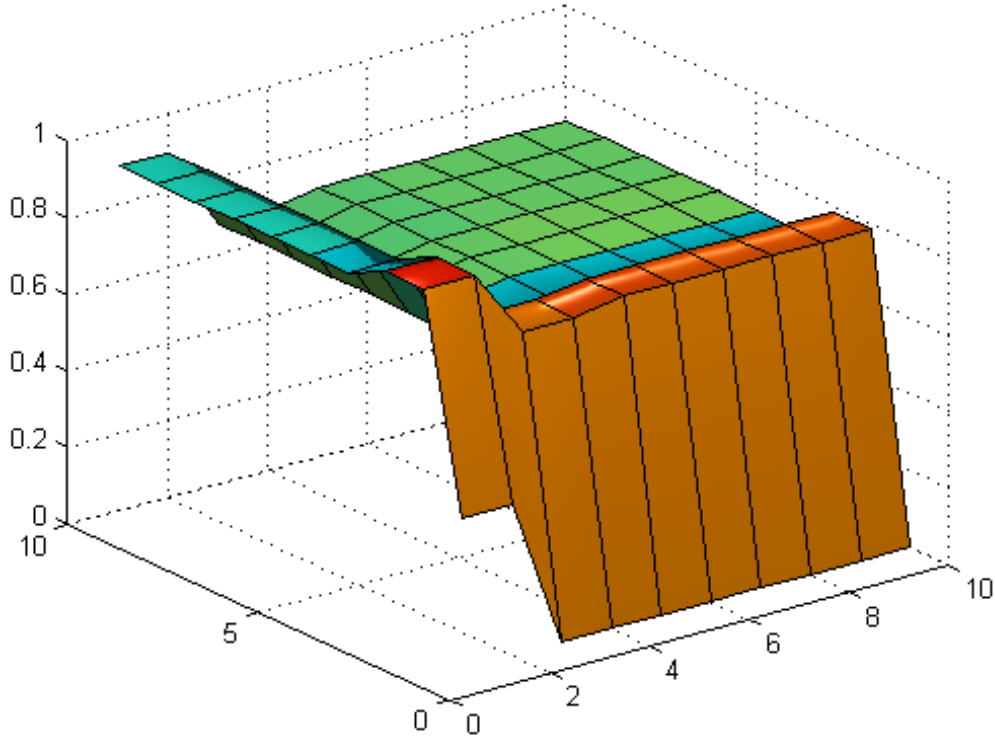


FIGURE 4.13 – Evaluation de $(\varphi_1^{(\tau_n)}, \varphi_2^{(\tau_n)})$ par rapport à L'_{τ_n}

(c) Sélection

Une fois la performance de chacun des individus évaluée, il est recommandé de réaliser un réajustement de la fonction d'adaptation avant la sélection (voir l'annexe), il faut ensuite classer les individus et leur associer une probabilité de reproduction. la méthode utilisée ici est la sélection proportionnelle RWS (Voir [28]), elle est implémentée de la façon suivante

L'algorithme

Algorithme 1 Pseudo code de RWS

- i. Evaluer la fitness L_i , correspondante pour chaque individu de la population P
 - ii. Calculer la probabilité p_i de i^{me} individu sélectionner de la population tels que :

$$p_i = \frac{L_i}{\sum_{j=1}^n L_j}, N : \text{la taille de la population}$$
 - iii. Calculer la probabilité cumulative, q_i , pour chaque individu i tels que :

$$q_i = \sum_{j=1}^i p_j$$
 - iv. Générer un nombre aléatoire $r \in [0, 1]$
 - v.
$$\begin{cases} \text{Si } r < q_1 & \text{Alors sélectionner le 1er individu } x_1 \\ \text{Si } q_{i-1} < r \leq q_i & \text{Alors sélectionner le } i\text{eme individu } x_i \end{cases}$$
 - vi. Répéter les étapes (iv-v) N fois pour la création de N individus de la nouvelle population P^G
-

On a maintenant un ensemble d'individus de taille p^2 qui ont survécu à la sélection et qui représentent les meilleures solutions obtenues.

(d) Croisement

Il faut ensuite choisir les paires d'individus qui vont croiser pour former des rejetons. Chaque paire d'individus est choisie aléatoirement parmi l'ensemble de sélection. On prend une paire d'individus (I_{ij} et I_{kl}) et on vérifie s'il ya croisement, selon la probabilité de croisement P_C (qui est typiquement entre 0.5 et 1). Dans notre étude on utilise le croisement heuristique (Voir [1]) pour créer des rejetons, ce type de croisement assure l'exploration de l'espace en testant des nouvelles valeurs de variables objets. Ce type de croisement est implémenté de la façon suivante :

L'algorithme

Algorithme 2 Pseudo code de Croisement Heuristique

- i. On tire de façon aléatoire deux individus (I_{ij}, I_{kl}) de la population P^G
 - ii. Générer un nombre aléatoire r
 - Si $r > P_C$ copier les deux individus (parents) (I_{ij}, I_{kl}) dans la nouvelle population P'^G
 - Sinon
$$\begin{cases} I'_{ij} = I_{ij} + a_1(I_{kl} - I_{ij}) \\ I'_{kl} = I_{kl} + a_2(I_{ij} - I_{kl}) \end{cases} \quad (a_1, a_2) \text{ deux nombres aléatoires dans } [0, 1]$$
 - iii. Répéter les étapes (i-ii) N fois pour la création de N individus pour la nouvelle population P'^G
-

Une fois les rejetons (enfants) créés, on obtient à nouveau une population P'^G de p^2 individus qui, en théorie, ont une meilleure performance en moyenne que les individus de la génération précédente.

(e) **Mutation**

La prochaine étape est la mutation. On utilise une mutation non-uniforme, de Michalewicz (Voir [15]). On vérifie d'abord s'il y a mutation, selon la probabilité de mutation P_M (qui se situe habituellement entre 0.00001 et 0.1).

Pour un paramètre $\phi^{(t)}$, s'il y a mutation, le changement s'effectue de la façon suivante :

L'algorithme

Algorithme 3 Pseudo code Mutation non uniforme

- i. On tire de façon aléatoire un individu $I = (\phi_1^{(t)}, \phi_2^{(t)})$
 - ii. Générer un nombre aléatoire K
 - Si $K \geq P_M$ copier I dans la nouvelle génération P''^G
 - Sinon pour $i = 1, 2$

$$\phi_i^{(t)}(G+1) = \begin{cases} \phi_i^{(t)}(G) + [\text{Max}\phi_i^{(t)}(G) - \phi_i^{(t)}(G)](1 - r^{(1-G/\gamma)^b}) & (\text{pile}) \\ \phi_i^{(t)}(G) - [\phi_i^{(t)}(G) - \text{Min}\phi_i^{(t)}(G)](1 - r^{(1-G/\gamma)^b}) & (\text{face}) \end{cases}$$
 - iii. Répéter les étapes (i-ii) N fois pour la création de N individus pour la nouvelle population P''^G
-

Remarque :

- r = nombre aléatoire entre 0 et 1 (distribution uniforme).
- G = génération actuelle.
- γ = génération maximale.
- b = degré d'extinction d'amplitude.

Le processus de sélection, reproduction et mutation est alors répété. Chaque fois qu'on effectue ces trois opérations, cela construit une génération.

(f) **Elitisme**

Une technique très souvent utilisée avec l'algorithme génétique est l'élitisme. L'élitisme est la conservation du meilleur individu dans la génération suivante.

Après le croisement, on compare le meilleur individu de la nouvelle génération avec le meilleur individu de la génération précédente. Le meilleur individu est conservé et l'autre est détruit. Cette technique permet de s'assurer que la fonction objective augmente (diminue) toujours si on maximise (minimise) une fonction.

4.7.5 Application Numérique

Pour $P = 10, D = 1000, T = 1, \theta = 1, N = 100, P_C = 0.9, P_M = 0.1, \gamma = 100, b = 5$

On déroule l'RCGA sous Matlab 7.0 à une itérations on trouve les résultats suivants :

La figure (4.14) montre un graphique de la fonction log-vraisemblance discrétisée après ajustement.

La fonction log-vraisemblance discrétisée possède de nombreux maxima locaux. Cependant, la fonction a des maxima globaux, comme l'indique la figure suivante.

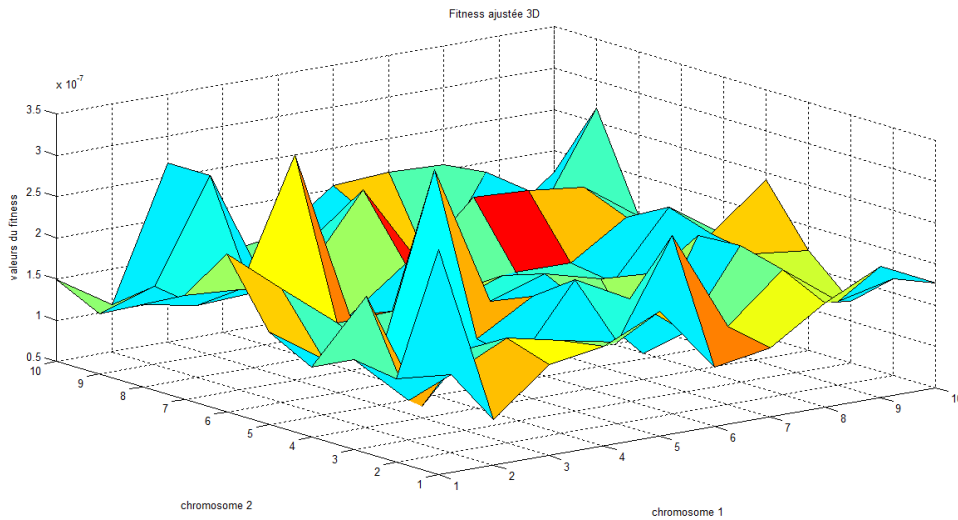


FIGURE 4.14 –

La figure (4.15) présente les contours où se localise les maxima globaux et maxima locaux.

La figure (4.16) montre la fonction log-vraisemblance discrétisée après utilisation des opérations de l'algorithme génétique à codage réel (croisement mutation et elitisme).

Le graphique (4.17) montre la valeur maximale du fitness pour chaque génération, on remarque que l'algorithme converge vers la valeur maximale de la fonction fitness qui correspond à la meilleur valeur des paramètres à estimer.

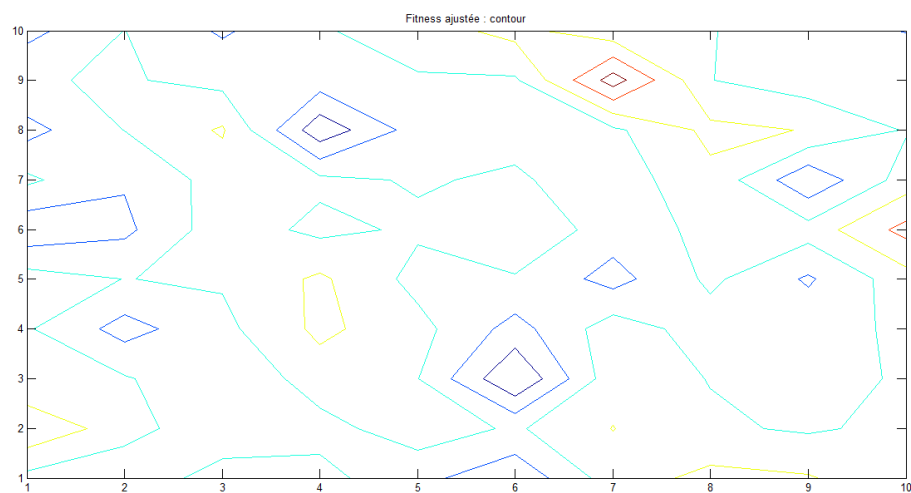


FIGURE 4.15 –

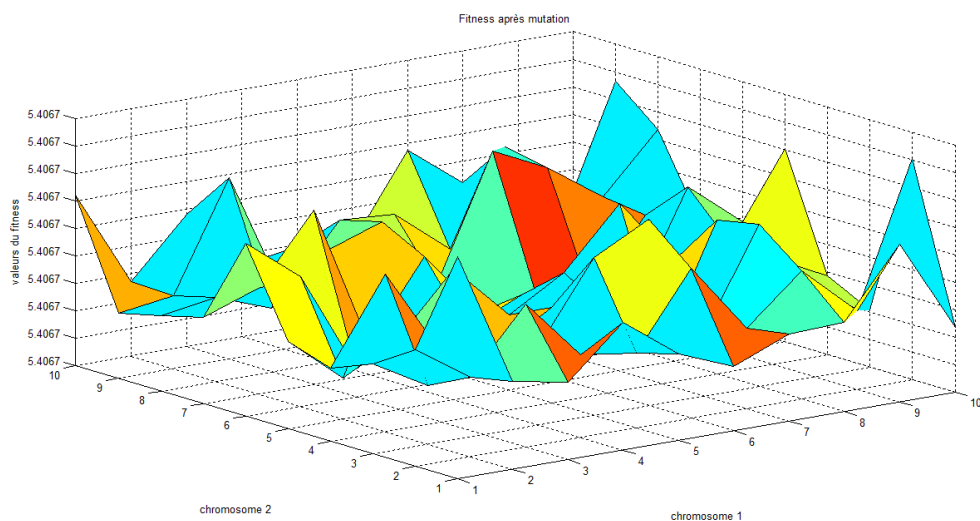


FIGURE 4.16 –

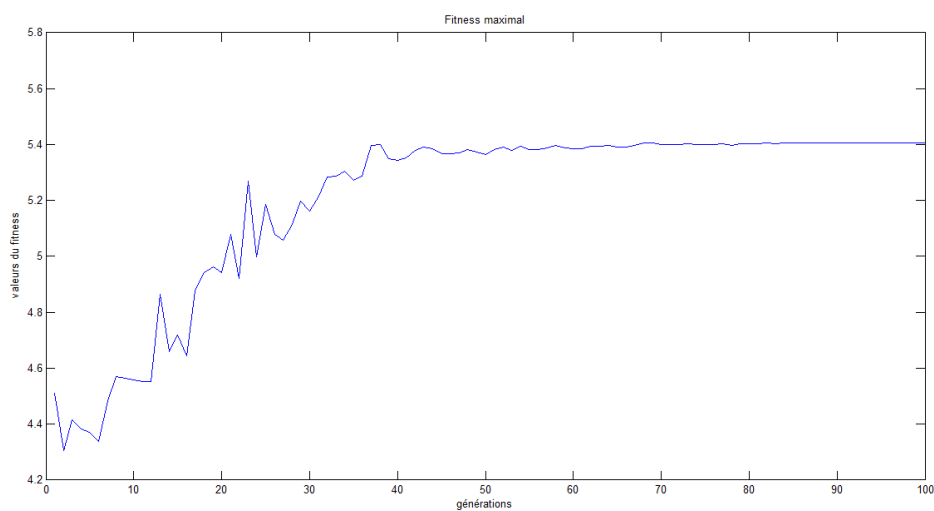


FIGURE 4.17 –

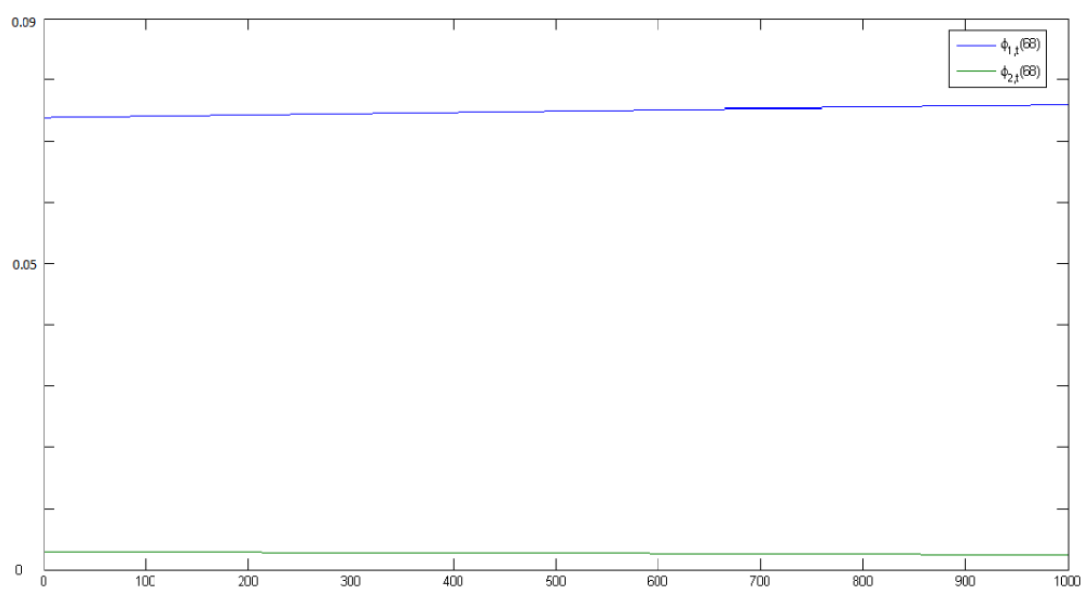


FIGURE 4.18 –

Commentaire

La méthode de RCGA nous donne une bonne estimation des solutions $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ à la 68^{ème} génération comme l'indique la figure (4.18). Les estimateurs $\varphi_1^{(t)}$ et $\varphi_2^{(t)}$ nous fournit une solution π_t -ajustée qui ajuste mieux la solution initiale π_t .

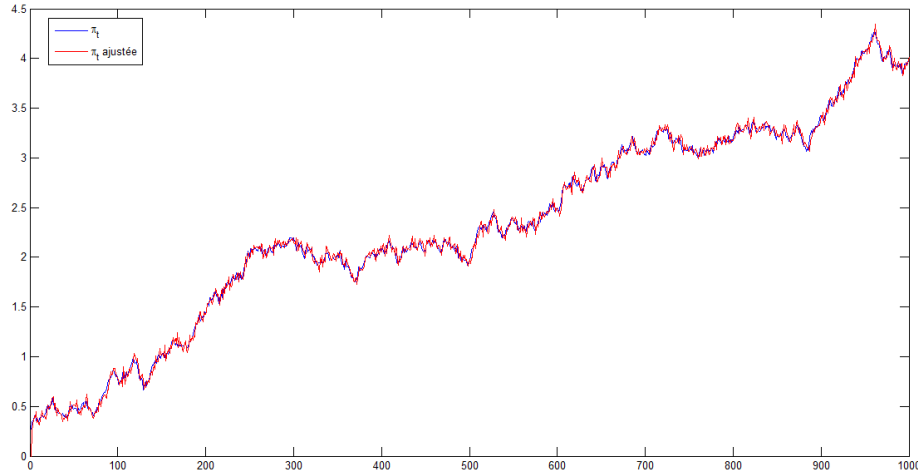


FIGURE 4.19 –

4.7.6 Comparaisons Expérimentales

Pour analyser et comparer les performances de la méthode RCGA présentée précédemment, nous avons choisi d'optimiser la fonction $L_T(\varphi^{(t)}; \pi_t)$ pour trois valeurs du paramètre $\theta = 1, 10$ et 100 , ce qui correspond à des caractéristiques fonctionnelles très différentes. les critères de comparaison sont les suivants :

- (a) **Fiabilité** : elle est déterminée par la valeur du paramètre (compris entre $[-1, 1]$).
- (b) **Temps de calcul** : il est évalué par le nombre d'évaluation de la fonction.

Pour une taille du population est égale à $N = 100$, nous avons effectué 100 expériences avec des points initiaux choisis aléatoirement. Les paramètres utilisés pour chacune sont donnés comme suit :

Données techniques relatives au méthode de RCGA utilisée

- Taille de la population : 100.
- Probabilité de croisement $P_c = [0.5, 1]$ avec un pas de discrétisation $\Delta = 0.1$.
- Probabilité de mutation $P_m = [0.01, 0.1]$ avec un pas de discrétisation $\Delta = 0.01$.
- Sélection : RWS.
- Critère d'arrêt : le nombre maximum de génération.

Les valeurs reportées ici sont des moyennes effectuées sur 100 expériences (voir les résultats sur les tableaux).

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,13	0,16	0,13	0,17	0,11	0,17	0,15	0,12	0,13	0,24
0,6	0,16	0,18	0,18	0,1	0,11	0,12	0,12	0,07	0,25	0,13
0,7	0,08	0,13	0,11	0,11	0,16	0,18	0,14	0,11	0,09	0,14
0,8	0,17	0,2	0,14	0,14	0,11	0,15	0,13	0,13	0,15	0,24
0,9	0,17	0,17	0,21	0,24	0,13	0,18	0,12	0,16	0,15	0,14
1	0,14	0,15	0,18	0,13	0,19	0,09	0,12	0,13	0,2	0,08
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,15273	0,16629	0,16795	0,17384	0,18111	0,18306	0,18593	0,20349	0,18216	0,1786
0,6	0,17349	0,17517	0,17431	0,18356	0,18942	0,18276	0,18141	0,18327	0,18395	0,20429
0,7	0,19246	0,19389	0,19806	0,1939	0,19424	0,19209	0,18699	0,18928	0,19032	0,19242
0,8	0,18768	0,18896	0,19131	0,19682	0,19366	0,19518	0,19878	0,20831	0,22885	0,20835
0,9	0,20891	0,20325	0,20774	0,20655	0,21315	0,21363	0,21357	0,21514	0,21822	0,22941
1	0,21706	0,21815	0,21625	0,20591	0,20752	0,20892	0,20904	0,20905	0,21061	0,21547

FIGURE 4.20 – Test 1 : $\theta = 1, \gamma = 50$

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,39	0,38	0,42	0,44	0,32	0,35	0,39	0,33	0,31	0,31
0,6	0,35	0,28	0,33	0,32	0,33	0,35	0,33	0,33	0,3	0,38
0,7	0,34	0,36	0,33	0,34	0,41	0,41	0,43	0,41	0,38	0,41
0,8	0,33	0,35	0,38	0,38	0,36	0,34	0,36	0,35	0,42	0,43
0,9	0,4	0,31	0,4	0,38	0,32	0,39	0,38	0,3	0,42	0,36
1	0,33	0,43	0,33	0,36	0,39	0,33	0,28	0,25	0,35	0,35
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,22795	0,23071	0,23309	0,23464	0,23634	0,23792	0,23979	0,24178	0,2431	0,24164
0,6	0,23525	0,24253	0,24275	0,24385	0,24525	0,24576	0,24778	0,24783	0,25015	0,25521
0,7	0,2493	0,25056	0,25342	0,25412	0,25768	0,2583	0,2599	0,26159	0,26437	0,26633
0,8	0,2608	0,2624	0,2622	0,2655	0,26693	0,26867	0,27128	0,27029	0,27178	0,27666
0,9	0,26956	0,27164	0,27346	0,27538	0,27735	0,2792	0,2817	0,28373	0,28622	47,22647
1	0,28588	0,28331	0,28159	0,28483	0,28676	0,28656	0,28828	0,29439	0,29468	0,29402

FIGURE 4.21 – Test 2 : $\theta = 1, \gamma = 100$

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,69	0,74	0,79	0,71	0,77	0,74	0,81	0,66	0,68	0,73
0,6	0,72	0,69	0,75	0,74	0,76	0,74	0,69	0,74	0,79	0,69
0,7	0,67	0,7	0,77	0,77	0,73	0,75	0,66	0,76	0,65	0,69
0,8	0,72	0,73	0,8	0,64	0,79	0,76	0,74	0,71	0,73	0,75
0,9	0,73	0,73	0,68	0,7	0,78	0,69	0,66	0,83	0,68	0,74
1	0,75	0,62	0,78	0,77	0,76	0,7	0,77	0,75	0,71	0,79
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,93228	0,9274	0,934	0,94279	0,94208	0,95131	0,95505	0,9588	0,96608	0,9766
0,6	0,94059	0,94941	0,95566	0,97002	0,97093	8,98577	1,00527	1,0051	1,00152	1,01105
0,7	0,97067	0,98689	0,99622	1,00155	1,00387	48,00896	1,00096	1,00339	1,00692	1,01107
0,8	0,9772	0,9862	0,99262	1,00368	1,00415	1,01246	1,01932	1,0146	1,02254	1,0269
0,9	0,99226	1,00806	1,01789	1,01666	1,02898	1,03377	1,02899	3,48216	1,0575	1,04819
1	1,01635	1,0244	1,03788	1,04318	1,04767	1,0542	1,06019	1,06767	1,07314	1,07064

FIGURE 4.22 – Test 3 : $\theta = 1, \gamma = 500$

Nous observons dans les trois tests précédents que la méthode RCGA donne une fiabilité très faible (environ 25%) pour 50 générations avec des probabilités respectives $P_c = 0.6$ et $P_m = 0.09$ avec un temps de calcul relativement faible 0.18395 seconde par itération. De plus elle possède une fiabilité moyenne (environ 44%) pour 100 générations avec des probabilités respectives $P_c = 0.5$ et $P_m = 0.04$. Pour 500 générations, la fiabilité optimale obtenue est 81% avec des probabilités respectives $P_c = 0.5$ et $P_m = 0.07$.

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,06	0,06	0,04	0,06	0,05	0,05	0,1	0,1	0,02	0,08
0,6	0,05	0,07	0,05	0,11	0,01	0,08	0,04	0,05	0,07	0,04
0,7	0,07	0,04	0,03	0,07	0,03	0,13	0,06	0,09	0,04	0,11
0,8	0,07	0,12	0,11	0,05	0,07	0,04	0,03	0,07	0,09	0,09
0,9	0,07	0,09	0,04	0,08	0,1	0,02	0,07	0,06	0,07	0,02
1	0,04	0,05	0,05	0,09	0,07	0,07	0,07	0,02	0,04	0,06
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,15224	0,15411	0,1541	0,15743	0,15848	0,16007	0,16086	0,16195	0,1646	0,16461
0,6	0,15994	0,16223	0,16237	0,16584	0,16551	0,16743	0,17109	0,17138	0,17154	0,17457
0,7	0,16854	0,16987	0,17233	0,17305	0,17691	0,17956	0,1805	0,18188	0,18267	0,18707
0,8	0,18048	0,18316	0,18298	0,18377	0,188	0,18785	0,19002	0,19092	0,18969	0,19221
0,9	0,18985	0,1902	0,1933	0,19407	0,19577	0,195	0,19594	0,1973	0,1989	0,20141
1	0,19653	0,19842	0,20032	0,20199	0,20377	0,2045	0,20677	0,20711	0,20935	0,21083

FIGURE 4.23 – Test 4 : $\theta = 10, \gamma = 50$

Le taux de fiabilité par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,38	0,36	0,36	0,49	0,4	0,44	0,4	0,4	0,41	0,4
0,6	0,46	0,39	0,47	0,42	0,4	0,39	0,38	0,36	0,35	0,34
0,7	0,37	0,46	0,44	0,3	0,36	0,37	0,39	0,41	0,49	0,35
0,8	0,47	0,46	0,39	0,42	0,31	0,36	0,38	0,43	0,39	0,47
0,9	0,39	0,35	0,41	0,39	0,37	0,41	0,41	0,33	0,35	0,38
1	0,46	0,35	0,36	0,33	0,4	0,36	0,48	0,36	0,42	0,34
Le temps d'exécution par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,22415	0,22574	0,22704	0,23229	6,24143	0,23574	0,23758	0,23913	0,24055	0,24117
0,6	0,23307	0,23586	0,23664	0,24185	0,24475	0,24666	0,24768	0,25098	0,25115	0,25416
0,7	0,24631	0,25033	0,25117	0,25319	0,25333	0,25711	0,25896	0,2602	0,26207	0,26475
0,8	0,25523	0,25675	0,25737	0,26321	0,26584	0,26503	0,26661	0,26719	0,27004	0,27462
0,9	0,26723	0,26786	0,26906	0,26984	0,27294	0,27787	0,28017	0,28082	0,28046	0,28267
1	0,27783	0,27956	0,28065	0,28377	0,28784	0,28595	0,28586	0,28893	0,29104	0,29527

FIGURE 4.24 – Test 5 : $\theta = 10, \gamma = 100$

Le taux de fiabilité par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,71	0,62	0,76	0,69	0,76	0,64	0,72	0,63	0,69	0,64
0,6	0,67	0,66	0,71	0,71	0,62	0,6	0,73	0,68	0,6	0,6
0,7	0,53	0,59	0,65	0,71	0,67	0,68	0,67	0,67	0,63	0,7
0,8	0,57	0,72	0,72	0,63	0,69	0,65	0,62	0,65	0,71	0,76
0,9	0,61	0,66	0,72	0,59	0,68	0,6	0,62	0,63	0,71	0,63
1	0,72	0,64	0,67	0,7	0,64	0,68	0,63	0,6	0,7	0,8
Le temps d'exécution par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,91168	0,92834	0,93632	0,93849	0,94288	0,95107	0,95605	0,96254	0,96443	0,97248
0,6	0,94115	0,95036	0,95756	0,96455	0,96652	0,97379	0,97903	0,98308	0,9843	0,99121
0,7	0,95504	0,96924	0,97273	0,9862	0,98695	0,99229	0,99652	16,01322	1,00921	1,06368
0,8	1,02414	1,02814	1,05694	1,05581	1,10004	1,06766	1,11136	1,12972	1,18143	1,13115
0,9	1,1229	1,17406	1,15071	1,13953	1,1171	1,06671	1,0842	1,08951	1,1103	1,1283
1	1,0697	1,0726	1,12292	1,09593	1,0863	1,0741	1,0778	1,0829	1,087	1,09321

FIGURE 4.25 – Test 6 : $\theta = 10, \gamma = 500$

En ce qui concerne les tests (4), (5) et (6), les résultats montrent que la RCGA pour $\theta = 10$ et $\gamma = 500$ générations offre un bon niveau de fiabilité avec une forte probabilité de croisement et de mutation mais avec un temps de calcul nettement plus grand.

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,11	0,09	0,04	0,03	0,05	0,05	0,04	0,05	0,07	0,06
0,6	0,04	0,08	0,04	0,07	0,07	0,07	0	0,04	0,03	0,04
0,7	0,09	0,06	0,04	0,07	0,02	0,06	0,03	0,05	0,03	0,07
0,8	0,04	0,05	0,06	0,06	0,1	0,07	0,1	0,06	0,02	0,07
0,9	0,06	0,02	0,08	0,05	0	0,08	0,08	0,04	0,03	0,03
1	0,07	0,06	0,09	0,06	0,09	0,04	0,03	0,03	0,03	0,06
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,155	0,1563	0,1565	0,1587	0,1609	0,1616	0,163	0,1651	0,166	0,1682
0,6	0,1643	0,1652	0,167	0,1691	0,1702	0,1721	0,1735	0,1746	0,178	0,1813
0,7	0,1757	0,1774	0,1789	0,1824	0,1837	0,1814	0,1841	0,1846	0,18581	0,1875
0,8	0,1837	0,1858	0,1859	0,1878	0,1901	0,1911	0,1918	0,1944	0,1958	0,1987
0,9	0,1943	0,1947	0,1969	0,1987	0,2006	0,2011	0,2045	0,2045	0,2129	0,21875
1	0,21189	0,21131	0,21426	0,21604	0,22367	0,22303	0,21517	0,21729	0,21894	0,22089

FIGURE 4.26 – Test 7 : $\theta = 100, \gamma = 50$

Le taux de fiabilité par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,34	0,42	0,37	0,41	0,37	0,36	0,41	0,28	0,31	0,31
0,6	0,35	0,39	0,36	0,4	0,33	0,37	0,37	0,36	0,34	0,4
0,7	0,36	0,35	0,27	0,39	0,41	0,38	0,32	0,43	0,32	0,37
0,8	0,43	0,35	0,33	0,31	0,43	0,39	0,38	0,3	0,33	0,39
0,9	0,39	0,41	0,43	0,32	0,38	0,38	0,37	0,36	0,39	0,29
1	0,42	0,37	0,38	0,38	0,48	0,41	0,37	0,51	0,27	0,41
Le temps d'exécution par 100 itérations										
$P_c \backslash P_m$	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,23364	0,23556	0,24387	0,24808	0,24368	0,24373	0,24671	0,25387	0,24818	0,2481
0,6	0,2398	0,2432	0,2455	0,2463	0,2487	0,2497	0,25221	0,2556	0,2554	0,2575
0,7	0,2511	0,2531	0,256	0,2568	0,2589	0,2612	0,262	0,2651	0,266	0,2688
0,8	0,2617	0,2649	0,2711	0,2721	0,32531	0,3048	0,3058	0,30589	0,30406	0,29042
0,9	0,28369	0,30141	0,28323	0,28971	0,29194	0,31316	0,29933	0,29843	0,29979	0,30175
1	0,29225	0,29337	0,29893	0,30465	0,30599	0,30483	0,30284	0,30878	0,31048	0,31181

FIGURE 4.27 – Test 8 : $\theta = 100, \gamma = 100$

Pour $\theta = 100$, les tests (7), (8) et (9) montrent que la méthode RCGA donne une fiabilité moindre (environ 77% de la fiabilité pour $\theta = 10$) avec des probabilités respectives de $P_c = 0.7$ et $P_m = 0.04$.

Nous concluons que la méthode de RCGA nous fournis simultanément avec un temps de calcul relativement faible une fiabilité élevée. De plus elle maintient ce niveau élevé de fiabilité même dans le cas les plus difficiles ($\theta = 100$).

Le taux de fiabilité par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,66	0,61	0,7	0,66	0,7	0,59	0,66	0,69	0,63	0,6
0,6	0,57	0,6	0,71	0,68	0,72	0,66	0,68	0,57	0,7	0,74
0,7	0,54	0,59	0,63	0,77	0,6	0,66	0,73	0,69	0,72	0,65
0,8	0,57	0,62	0,71	0,6	0,59	0,66	0,72	0,64	0,62	0,61
0,9	0,65	0,61	0,73	0,63	0,64	0,55	0,69	0,63	0,71	0,63
1	0,65	0,75	0,67	0,62	0,65	0,62	0,62	0,68	0,62	0,61
Le temps d'exécution par 100 itérations										
Pc \ Pm	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
0,5	0,97748	1,02257	0,99816	1,00419	1,0044	1,01606	1,02279	1,02835	1,02922	0,99573
0,6	0,9558	0,9693	0,9693	0,979	0,98651	0,9948	0,9989	1,0028	1,0081	1,0085
0,7	0,965	0,98291	0,9899	1,0016	1,0049	1,0072	59,3412	1,0251	1,0272	1,0335
0,8	0,9991	1,0097	1,01731	1,0234	1,021	1,034	1,0386	1,0442	1,0512	1,0488
0,9	1,0087	1,0277	87,10038	1,0489	1,0522	1,0522	1,0562	1,0635	1,06781	1,0753
1	1,0349	1,0477	1,0474	1,0611	1,0699	1,07481	1,0802	1,0832	43,40136	1,0977

FIGURE 4.28 – Test 9 : $\theta = 100, \gamma = 500$

Réglage des paramètres de la méthode RCGA

Quelle doit être la taille de cette population ? Une population trop petite évoluera probablement vers un optimum local peu intéressant. Une population trop grande sera inutile car le temps de convergence sera excessif. La taille de la population doit être choisie de façon à réaliser un bon compromis entre temps de calcul et qualité du résultat. Les tableaux ci-dessous, montre la performance de la méthode utilisée par rapport à la taille du population.

Pour $N = 100$

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	1	0.6	0.09	0.25	0.18395 s
100	1	0.5	0.04	0.44	0.23464 s
500	1	0.5	0.07	0.81	0.95505 s

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	10	0.7	0.06	0.13	0.17956 s
100	10	0.5	0.04	0.49	0.23229 s
500	10	1	0.1	0.8	1.09321 s

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	100	0.5	0.01	0.11	0.155 s
100	100	1	0.08	0.51	0.36878 s
500	100	0.7	0.04	0.77	1.0016s

Pour $N = 225$

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	1	0.7	0.01	0.29	0.33925 s
100	1	0.8	0.03	0.48	0.41902 s
500	1	0.5	0.06	0.84	1.70467 s

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	10	0.5	0.09	0.38	0.32749 s
100	10	1	0.04	0.48	0.46706 s
500	10	0.5	0.6	0.77	1.95732 s

γ	θ	P_C	P_M	Fiabilité	temps de calcul
50	100	0.9	0.06	0.33	0.39484 s
100	100	0.5	0.07	0.51	0.41777 s
500	100	0.5	0.06	0.76	1.91821s

Pour le problème qui nous intéressait et avec les moyens de calcul dont nous disposions, nous avons pu constater qu'une population de 100 individus constituait un bon compromis. Nous pouvions ainsi tester 100 à 225 générations en une nuit sur un Intel(R) core (TM) i5 2.40 GHZ. Mais il faut être conscient que cette taille de population dépend de la puissance de calcul dont on dispose, des méthodes utilisées (sélection, opérateurs génétiques...), du nombre de variables considérées et de la fonction d'adaptation. Si la fonction à optimiser comporte peu d'optima locaux et un optimum global net, la population nécessaire sera plus petite que dans le cas d'une fonction beaucoup plus compliquée comportant de nombreux optima locaux. Nous touchons là au délicat problème du réglage des paramètres de l'algorithme. Celui-ci doit être optimisé pour chaque type de problème traité, ce qui constitue une part importante du travail de l'utilisateur. Les caractéristiques de l'algorithme doivent être adaptées à la topologie du paysage dessiné par la fonction (fitness). L'ensemble problème-méthodes-paramètres constitue donc un tout. En témoigne certaines études où les paramètres d'un Algorithme Génétique sont réglés et optimisés par un autre Algorithme Génétique [29]. Dans la pratique, les méthodes et paramètres des AG sont tout d'abord réglés approximativement par tâtonnement avec des fonctions de n variables couramment utilisées pour tester les algorithmes d'optimisation. Le temps de calcul de ces fonctions étant minime, on peut ainsi régler rapidement les paramètres.

4.8 Conclusion

Les Algorithmes Génétiques à codage réel (RCGA), tels qu'ils sont décrits ci-dessus, s'avèrent plus pratique d'emploi, plus rapides, plus précis (surtout pour les larges domaines qui requièrent une représentation trop longue avec les Algorithmes Génétique discret ; ils sont plus consistants de générations en générations, ce qui est dû au fait que les opérateurs agissent directement sur le phénotype et non pas sur le génotype. En outre, ils permettent d'éviter certains problèmes dus au caractère artificiel du codage binaire (falaise de Hamming).

Annexe

4.9 Annexe A : Scaling

La technique de réajustement la plus répandue est la mise à l'échelle linéaire (Goldberg 1989). Celle-ci consiste à comprimer le domaine de variation de la fonction d'adaptation tout en gardant une valeur moyenne identique ($f'_{moy} = f_{moy}$). Au maximum de la performance correspond une nouvelle valeur $f'_{max} = \beta f_{moy}$ avec $1.2 \leq \beta \leq 2$ (typiquement $\beta = 2$). La fonction d'adaptation après mise à l'échelle prend la forme suivante :

$$f' = af + b$$

avec $a = \frac{\beta-1}{f_{max}-f_{moy}} f_{moy}$ et $b = \frac{f_{max}-\beta f_{moy}}{f_{max}-f_{moy}} f_{moy}$

Toutefois, il faut effectuer le test de non-négativité (4.4) avant la mise à l'échelle. En effet, si $f_{max} < \beta f_{moy}$ la compression du domaine de performance est déjà implicite et dans ce cas, la mise à l'échelle étend au contraire les valeurs de f , diminuant l'influence des individus trop "faibles".

$$f_{min} > \frac{\beta f_{moy} - f_{max}}{\beta - 1} \quad (4.5)$$

Si l'inéquation précédente n'est pas vérifiée, la fonction d'adaptation mise à l'échelle prendra des valeurs négatives. Il convient donc de limiter l'extension de la borne inférieure à $f'_{min} = 0$ tout en gardant $f'_{moy} = f_{moy}$. Les coefficients du réajustement linéaire s'expriment alors par : $a = \frac{f_{moy}}{f_{moy}-f_{min}}$ et $b = -\frac{f_{moy}}{f_{moy}-f_{min}} f_{min}$

Pour résumer, le mécanisme de la mise à l'échelle linéaire peut être représenté par la figure suivante. Il existe d'autres types de réajustement de la fonction d'adaptation notamment la tronqueur sigma qui supprime les individus de faible performance avant une éventuelle mise à l'échelle linéaire :

$$f' = f - (f_{moy} - c\sigma)$$

Où $1 \leq c \leq 3$ et σ est l'écart-type de l'adaptation, calculé sur la population. Les résultats négatifs sont forcés à zéro.

Notons enfin le changement d'échelle en puissance qui modifie la fonction d'adaptation de la façon suivante : $f' = f^\beta$ où la constante β peut être adaptée au cours des générations pour accentuer la sélection des individus situés à proximité des sommets (Goldberg 89).

4.10 Annexe B : Codage

Un nombre décimal peut être représenté par son équivalent dans un code différent tel que :

- (a) le code binaire,
- (b) le code octal,
- (c) le code hexadécimal

Définition Un code est un ensemble de règles de représentation de données qui peuvent être

- (a) numériques,
- (b) alphabétiques,
- (c) ou alphanumériques.

Outre le système décimal, les principaux systèmes de numération que l'on utilise dans le domaine du traitement de l'information sont : les systèmes binaire, octal et hexadécimal.

Définition Lorsqu'un code s'applique à la manière. d'énoncer les nombres, il définit un système de numération.

Exemple Le nombre décimal 12 est représenté :

- (a) par le nombre 1100 dans le code binaire,
- (b) par le nombre 14 dans le code octal,
- (c) par la lettre C dans le code hexadécimal.

4.11 Annexe C : Code Source des opérateurs de La RCGA

f. RWS

```

function[indice,Ln] = select(Ln,step,K) n = numel(Ln);
if step > 0 sigma = std(Ln); fmoy = mean(Ln); Ln = Ln - (fmoy - 3 * sigma); Ln = Ln.*
(Ln > 0); end
cumLn = cumsum(Ln);
f = cumLn/sum(Ln);
r = rand(n,1);
f = [0;f];
for i = 1 : n
    for j = 2 : n + 1
        if (r(i) <= f(j)) (r(i) > f(j - 1))
            indice(i) = j - 1;
        end
    end
end
Ln = Ln(indice);
end

```

g. Croisement Heuristique

```

function pop = croiser(pop,pc,K)
[n,m,p] = size(pop); r = rand(p,1);
for i = 1 : p
    crom = ceil(rand(2,1) * K^2);
    if r(i) <= pc
        pop(:, :, crom(1)) = pop(:, :, crom(1)) + rand(1) * (pop(:, :, crom(2)) - pop(:, :, crom(1)));
        pop(:, :, crom(2)) = pop(:, :, crom(2)) + rand(1) * (pop(:, :, crom(1)) - pop(:, :, crom(2)));
    end
end
end

```

h. Mutation non uniforme

```

function pop = mut(pop,pm,t,T,K)
[n,m,p] = size(pop); r = rand(p,1);
for i = 1 : p
    crom = ceil(rand * K^2);
    if r(i) <= pm
        xmin = ones(length(pop(:, :, crom)),1) * min(pop(:, :, crom));
        xmax = ones(length(pop(:, :, crom)),1) * max(pop(:, :, crom));
        if rand <= .5
            pop(:, :, crom) = pop(:, :, crom) + (xmax - pop(:, :, crom)) * (1 - rand^(1 - (t/T)^5));
        end
    end
end

```

```

elsepop(:, :, crom) = pop(:, :, crom) - (pop(:, :, crom) - xmin) * (1 - rand^(1 - (t/T)^5));
end
end
end
end

i. test function table=test(iter,K) N = [50100500];theta=100;for j = 1 : numel(theta)
    for i = 1 : numel(N)
        k = 1;
        for pc = .5 : .1 : .7
            l = 1;
            for pm = .01 : .01 : 0.1
                [param(k,l),time(k,l)] = test1(iter,N(i),theta(j),pc,pm,K);
                clc
                param

                l = l + 1;
            end
            k = k + 1;
        end
        t1 = size([paramtime],1);t2 = size([paramtime],2);
        table((i-1)*(t1+1)+1,1:t2) = [N(i)theta(j)zeros(1,t2-2)];
        table((i-1)*(t1+1)+2:i*(t1+1),:) = [paramtime];
    end
    xlswrite('C : xls',table,j);
end
end

function [param,time] = test1(iter,N,theta,pc,pm,K)
h = waitbar(0,['N=' num2str(N,' Pc=' num2str(pc,' for i = 1 : iter
[param1(i),time1(i)] = eva(N,1,theta,pc,pm,100,K);
waitbar(i/iter,h,['N=' num2str(N,' Pc=' num2str(pc,'end
close(h)
param = mean(param1);time = mean(time1);
end

```

CONCLUSION GÉNÉRALE

Dans ce mémoire nous avons proposé une adaptation de la méthode d'estimation des paramètres de dérive dépendants du temps d'une équation différentielle stochastique par l'algorithme génétique à codage réel (RCGA). Cette adaptation est motivée par le fait que les algorithmes génétique à codage réel représentent des outils d'optimisation adaptés a des fonctions réelles, des problème difficiles, complexe , irrégulier, par rapport aux algorithmes génétiques à codage binaire (BCGA), ce qui signifie qu'il existe un gain de temps de calcul à utiliser le premier algorithme lors des étapes de compilation.

Sur le plan pratique, l'étude dans le cas continu est très riche pour les résultats qui sont utilisés. Les démarches qui sont employées présentent une méthode intéressante pour le problème d'estimation des paramètres de dérive dépendant du temps d'une équation différentielle stochastique.

Cependant, les performances de cette méthode se trouvent pénalisées par le choix heuristique des opérateur de croisement et mutation cela explique le mode presque sur en théorie , nous proposons alors une nouvelle piste de recherche basée sur les probabilités de croisement et mutation qui varient par rapport au temps tels que :

Pour $t \in [0, 1]$

$$\begin{cases} 0.5 \leq (P_C(t) = tP_C) \leq 1 \\ 0.00001 \leq (P_M(t) = tP_M) \leq 0.1 \end{cases}$$

Bibliographie

- [1] Alden H.Wright, *Genetic Algorithms for Real parameter Optimization*, Missoula, Montana 1991.
- [2] Allen.E, *Modeling with Itô Stochastic Differential Equations*, Springer 2007.
- [3] Andrea.S.Foulkes, *Applied statistical Genetics with R*, Springer 2009.
- [4] Balding.D.J, Bishop.M, Cannings.C, *Handbook of statistical genetics*, Wiley 2007.
- [5] Bischwal Jaya.P.N, *Parameter Estimation in stochastic Differential Equation*, Springer 2004.
- [6] Boukhatala Kamel, *Estimation des paramètres de dérive d'une equation différentielle stochastique linéaire, contrôlée*, thèse de Magistère 1984.
- [7] Boukhatala Kamel, *Identification and simulation of a communication system*, Maghreb Math. Rev., Vol. 4, No 2, pp.55-79, 1995.
- [8] Brodeau.F- Le Breton.A, *Identification de paramètres pour un système excité par des bruits Gaussien et Poissonien*, Annals de l'institut Henri Poincaré- Section B- Vol. XV, n° 1-1979.
- [9] Capasso.V -Bakstein.D , *An Introduction to Continuous-Time Stochastic Processes*, Milian Research centre for industrial and applied Mathematics, pp. 127-203, 2003.
- [10] Chun-Liang Lin Ching-Huei Huang and Chin-Wei Tsai, *Structure-Specified Real Coded genetic Algorithms With Application*, TMRF e-book Advanced Knowledge based systems Model, Applications and Research (Eds.Sajja and Akerkar), pp. 160-187, 2010.
- [11] Comtet.L, *Advanced Combinatorics :The Art of Finite and Infinite Expansion*, D.Reidel Publishing, Dordrecht, The Netherlands, 1974.
- [12] Doleans-Dade Cathrine, Meyer Paul-Ander, *Equations différentielles stochastiques*, Séminaire de probabilités, Strasbourg, tome 11 pp. 376-382, 1977.
- [13] Dréo Johann, Pérowski Alain, *Métaheuristiques pour l'optimisation difficile*, Eyrolles 2003.
- [14] Henderson Douglas, Plaschko Peter, *Stochastic Differential Equation An Introduction Science And Engineering*, World Scientific Publishing 2006.
- [15] Janikow.Z Cezary, Michalewicz Zbigniew, *An Experimental Comparison of Binary and floating Point Representation in Genetic Algorithms*, computational statistic et data analysis 1992.

-
- [16] Kloeden.E Peter, *Stochastic Differential Equation And their Numerical solution*, Fachberich Mathematik, germany 2003.
 - [17] Kotelenez Peter, *Stochastic Ordinary and Stochastic Partial Differential Equations*, Springer 2008.
 - [18] Ksendal Brent, *Stochastic Differential Equation An introduction withe application*, Springer 2003.
 - [19] Morgerthaler Stephan, *Génétique Statistique*, Springer 2008.
 - [20] Protter.E Philip, *Stochastic Integration and Differential Equations*, Springer 2004.
 - [21] Rangel-Merino.A, López-Bonilla.L.J, Linares y Miranda.R , *Optimization Method based on Genetic Algorithms*, Apeiron, México. pp. 389-408, 2005.
 - [22] Renders Jean-Michel, *Algorithmes génétiques et réseaux de neurones*, Hermès, Paris, pp. 139-167, 1995.
 - [23] Ronglig Wu.Chang-Xing Ma George Sasella, *Statistical Genetics of quantitative traits*, Springer 2007.
 - [24] Tutkun Nedim, *Parameter estimation in mathematical models using the real coded genetic algorithms*, Expert Systems with Applications 2009.
 - [25] Gendreau Michel, Potvin Jean-Yves, *Handbook of Metaheuristics*, pp. 109-139. Springer 2010.
 - [26] Girsanov.I.V , *On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures*, Theory Probab. Appl. 5, pp. 285-301, 1960.
 - [27] Goldberg.E David, *Genetic Algorithms, Virtual Alphabets and Blocking*, Departement of general Engineering, University of Illinois at, 1989.
 - [28] Goldberg.E David, *Real-coded Genetic Algorithms, Virtual Alphabets, and Blocking*, Complex Sy stems 5, pp. 160-187, 1991.
 - [29] Goldberg.E David, *Genetic Algorithms*, University of Illinois, USA, pp. 98-125, 2003.
 - [30] Vallee Thomas, Y.Id.zo.glu Murat, *Présentation des algorithmes génétiques et de leurs applications en économie*, Université de Nantes, LEA-CIL 2001.

