



UNIVERSITE DJILLALI LIABES  
FACULTE DES SCIENCES  
SIDI BEL-ABBÈS

BP 89 SBA 22000 –ALGERIE–

TEL/FAX 048-54-43-44

# ***THESE***

*Présentée par:*  
HABIBECHE MUSTAPHA

*Pour obtenir le Diplôme de Doctorat*

*Spécialité : Mathématiques*  
*Option : Probabilités-Statistiques*

*Intitulée*

**Une étude de la modélisation et comparative entre  
Réseau de neurones artificiels, Algorithmes génétiques  
et Box-Jenkins en prévision de série chronologique  
chaotique**

*Thèse soutenue le 12 /03/2020*

*Devant le jury composé de :*

**Président :**

**Mr BENAÏSSA Samir**

**Professeur**

**à L'Université S.B.A**

**Directeur de thèse :**

**Mr BENCHIKH Tawfik**

**Professeur**

**à L'Université S.B.A.**

**Examineurs:**

**Mr GUANDOZI Toufik**

**Professeur**

**à L'Université de Saida**

**Mr BELGUERNA Abderrahmane Maître de Conférence A à C. U. de Nâama**

*Année universitaire 2019/2020*

---

~ Remerciements ~

*Avant tout, je remercie Dieu le tout puissant qui m'a donné la force et la patience pour mener à terme ce long et dur travail.*

*Je tiens à exprimer mes plus vifs remerciements à Benchikh Tewfik qui fut pour moi un directeur de thèse attentif et disponible malgré ses nombreuses charges. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont beaucoup appris. Ils ont été et resteront des moteurs de mon travail de chercheur.*

*Je voudrais aussi remercier chaleureusement chacun des membres du jury qui me font le grand honneur d'y participer.*

*Je remercie sincèrement Monsieur BENAÏSSA Samir pour l'honneur qu'il me fait en présidant le jury. Je remercie vivement Monsieur GUENDOUIZI Tawfik pour la confiance dont il me fait preuve en faisant parties de ce jury.*

*Je voudrais aussi remercier Monsieur pour l'intérêt BELGUERENA Abderrahmane qu'il a bien voulu accorder à mes travaux en acceptant de participer au jury. Je suis très honoré de sa présence.*

*Je remercie sincèrement Monsieur Dr. CHIKR El Mezouar Zouaoui pour l'énorme soutien scientifique et moral qu'il a su m'adresser pendant ces années , Je le remercie aussi pour ses qualités humaines et scientifiques et de m'avoir toujours encouragée à aller de l'avant.*

*Je remercie Monsieur MECHAB Boubaker qui m'a toujours encouragé et précieusement conseillé.*

*Je remercie aussi Monsieur HAIDAS Mohammed, pour leurs patientes et leurs disponibilités lorsque j'étais bloqué par des morceaux difficiles de ses programmes.*

*Je tiens à exprimer aussi ma reconnaissance à tous mes enseignants. Je remercie très amicalement tous mes amis et d'ailleurs de leurs sympathie et leurs aides de près ou de loin.*

*Enfin, j'adresse mes remerciements à l'ensembles des personnes présentes à cette soutenance.*

– Abstract –

Predicting the future behavior of chaotic time series in the last decade has become an important subject in statistical science. In this thesis, we are interested in the application of artificial neural networks and genetic algorithms to estimate the parameters of the chaos time series simulated (Logistics Map Template), to confirm the effectiveness of these new mechanisms, as well as the method Box; Jenkins then compare the results obtained.

**Keywords.** Artificial Neural Networks, Genetic Algorithms; Time Series Chaotic Prediction; Method of Box and Jenkins.

– Résumé –

Prédire le comportement futur des séries chronologiques chaotiques au cours de la dernière décennie est devenu un sujet important en statistique. Dans cette thèse, nous intéressons à l'application de réseaux de neurones artificiels et d'algorithmes génétiques pour estimer les paramètres de la série chronologique chaotique simulée (Modèle de carte logistique), afin de confirmer l'efficacité de ces nouveaux mécanismes, ainsi la méthode Box - Jenkins, Comparez ensuite les résultats obtenus.

**Mots clés :** Réseaux de neurones artificiels, Algorithmes génétiques; Modélisation de la séries chronologiques Chaotiques, Prévision; Méthode Box - Jenkins .

- مُلَخَّص -

تَوَقَّع السلوك المستقبلي لسلسلة زمنية فوضوية خلال العقد الماضي أصبحت موضوعًا مهمًا في الأبحاث. في هذه الأطروحة، نحن مهتمون بتطبيق الشبكات العصبية الاصطناعية والخوارزميات الجينية لتقدير معالمات السلسلة الزمنية الفوضوية المحاكاة (نموذج الخريطة اللوجيستية) من أجل تأكيد فعالية هذه الآلية الجديدة، مثلها مثل طريقة بوكس - جينكز، ثم نقارن النتائج المتحصل عليها.

**كلمات مفتاح .** نمذجة السلاسل الزمنية الفوضوية، الشبكات العصبية الاصطناعية، الخوارزميات الجينية، التنبؤ، نموذج *Box - Jenkins* .

# Table des matières

<b>1</b>	<b>Analyse série temporelle</b>	<b>14</b>
1.1	Introduction	14
1.2	Les concepts de base de séries chronologiques	15
1.2.1	Objectifs de l'analyse d'une série temporelle	16
1.2.2	Domaines d'application :	17
1.2.3	Eléments constitutifs d'une série temporelle	17
1.2.4	Tendance. Saisonnalité. Résidus :	18
1.3	Analyse de la tendance	19
1.3.1	Moyenne mobile	19
1.3.2	Différenciation	19
1.3.3	Modèle additif avec tendance paramétrique	20
1.3.4	Comparaison des différentes méthodes	21
1.3.5	Modélisation de la tendance	21
1.4	Analyse des composantes saisonnières	22
1.4.1	Calcul des moyennes saisonnières	22
1.4.2	Vérification du modèle	23
1.4.3	Différenciation	23
1.5	Estimation de la tendance et de la saisonnalité	23
1.5.1	Bruit blanc	24
1.5.2	Processus stationnaire	24
1.5.3	Une estimation paramétrique de la tendance (trend)	24
1.5.4	Estimation non paramétrique : moyenne mobile	25
1.5.5	Élimination de la tendance et de la saisonnalité par la méthode des différences	26
1.6	Série temporelle lineare	26
1.7	Modélisation des séries stationnaires	27
1.7.1	Corrélation, fonction d'auto corrélation et fonction d'auto corrélation partielle	27

1.7.2	Les processus auto-régressifs $AR_p$ . . . . .	28
1.7.3	Les processus en moyenne mobile $MA_q$ : . . . . .	30
1.7.4	Les processus mixtes $ARMA_{p,q}$ . . . . .	31
1.7.5	Illustrations . . . . .	32
1.8	Processus non stationnaire : . . . . .	33
1.8.1	Les processus $ARIMA$ et $SARIMA$ . . . . .	33
1.8.2	Estimation des paramètres du processus $ARMA_{(p;q)}$ . . . . .	35
1.8.3	Validation du processus $ARMA_{(p;q)}$ . . . . .	37
1.9	Tests sur les résidus . . . . .	37
1.9.1	Tests d'autocorrélation . . . . .	37
1.9.2	Tests de normalité . . . . .	38
1.9.3	Tests de stationnarité . . . . .	39
1.9.4	Critères de choix des modèles . . . . .	40
1.10	Conclusion . . . . .	42
<b>2</b>	<b>les séries temporelles chaotiques</b>	<b>43</b>
2.1	Introduction . . . . .	43
2.2	Historique . . . . .	46
2.3	Théorie du chaos . . . . .	47
2.4	Définition du chaos . . . . .	47
2.5	Les attracteurs . . . . .	48
2.6	Bifurcation et routes vers le chaos . . . . .	49
2.7	Les outils de détection du chaos . . . . .	49
2.8	la Conception de chaos mathématique . . . . .	52
2.9	Les modèles qui génèrent les séries chronologiques chaotique . . . . .	53
2.10	La prévision des séries chronologiques chaotiques . . . . .	59
2.11	Conclusion . . . . .	59
<b>3</b>	<b>Les méthodes de modélisation d'une série chronologique chaotique</b>	<b>60</b>
3.1	Les réseaux de Neurones Artificiels . . . . .	60
3.1.1	Introduction . . . . .	60
3.1.2	Définition . . . . .	61
3.1.3	Historique . . . . .	61
3.1.4	Applications . . . . .	62
3.1.5	Le modèle du neurone . . . . .	63

3.1.6	Neurone formel .(modèle mathématique)	63
3.1.7	Fonctions de transfert	66
3.1.8	Architectures des réseaux de neurones	68
3.1.9	Réseaux de neurones non bouclés	68
3.1.10	Réseaux de neurones bouclés	70
3.1.11	L'apprentissage	70
3.1.12	Définition	70
3.1.13	Type d'apprentissage	71
3.1.14	Règles d'apprentissage :	72
3.1.15	Les étapes de la conception d'un réseau de neurone	73
3.1.16	Détermination des entrées/sorties du réseau de neurones	73
3.1.17	Choix et préparation des échantillons	74
3.1.18	Elaboration de la structure du réseau	74
3.1.19	Apprentissage	75
3.1.20	Validation et Tests	75
3.1.21	Réseau de neurones à L couches	76
3.2	Les Algorithmes Génétiques	76
3.2.1	Introduction	76
3.2.2	Présentation des algorithmes génétiques (AG)	77
3.2.3	Principe	78
3.2.4	Fonctionnement général des algorithmes génétiques	79
3.2.5	Le codage	80
3.2.6	La fonction d'évaluation	83
3.2.7	Les opérateurs	84
3.2.8	Applications des algorithmes génétiques (AG)	87
3.2.9	Conclusion et discussion	88
<b>4</b>	<b>Application</b>	<b>90</b>
4.1	Modélisation de la série simulée (BJ)	90
4.2	Modélisation de la série simulée (RNA)	91
4.3	Modélisation de la série simulée (AG)	92
	<b>Bibliographie</b>	<b>96</b>

# Table des figures

1.1	les données du prix du cuivre . . . . .	15
1.2	:Cours de l'action Alcatel et Niveau du lac Huron . . . . .	16
1.3	Simulation d'un $ARMA_{2,2}$ , auto-corrélation et auto-corrélation partielle . . . . .	32
1.4	Méthodologie Box et Jenkins . . . . .	35
2.1	la diagramme de bifurcation de la fonction logistique. . . . .	49
2.2	Itération de la carte logistique. . . . .	53
2.3	Générer des Série chronologique chaotique . . . . .	54
2.4	la diagramme de bifurcation de Carte logistique . . . . .	54
2.5	Aspect aléatoire des états du système de Lorenz . . . . .	55
2.6	Attracteur étrange de Lorenz . . . . .	56
2.7	Exposants de Lyapunov du système chaotique de Lorenz . . . . .	56
2.8	Attracteur de Hénon pour $a = 1.4$ , $b = 0.3$ . . . . .	57
2.9	Diagramme de bifurcation de Hénon. . . . .	58
2.10	Les exposants de Lyapunov en fonction du temps pour l'application de Hénon. . . . .	58
3.1	Le neurone biologique . . . . .	63
3.2	modèle mathématique . . . . .	64
3.3	Représentation matricielle du modèle d'un neurone artificiel. . . . .	66
3.4	Fonction de transfert : (a) du neurone "seuil" ; (b) du neurone "linéaire", et (c) du neurone "sigmoïde". . . . .	68
3.5	Un réseau de neurones a une seule couche . . . . .	69
3.6	Réseau de neurones multicouche. . . . .	69
3.7	Réseau de neurones multicouche. . . . .	70
3.8	Apprentissage supervisé. . . . .	71

---

3.9	Apprentissage non-supervisé. . . . .	72
3.10	Organigramme de conception d'un réseau de neurones. . . . .	75
3.11	Chromosome . . . . .	77
3.12	Les principales étapes d'un algorithme génétique . . . . .	79
3.13	Exemple de Présentation d'AG . . . . .	81
3.14	Principe général de l'évolution d'une population d'un algorithme génétique . . . . .	84
3.15	Principe de la roulette . . . . .	85
3.16	Croisement . . . . .	87
3.17	Principe de la mutation . . . . .	87
4.1	série simulé à 205 données par carte logistique . . . . .	90
4.2	série simulé à 195 données. . . . .	90
4.3	Nombre : des entrées , des sorties , les fonctions de sommation et de transfert . . . . .	91
4.4	$Y_t$ et $Y_{tajuste}$ . . . . .	92
4.5	$y_t$ et $y_t$ ajus AG. . . . .	93
4.6	Performance des modèles $BJ$ , $RNA$ , $AG$ . . . . .	94

# Liste des tableaux

1.1	Différentes versions du test DFA. . . . .	39
4.1	Coefficient du meilleur model $ARIMA(4, 1, 0)$ de la série. . . . .	91
4.2	Tests sur résidus . . . . .	92
4.3	Les conditions pour obtenir le meilleur résultat $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5.)$ . . . . .	93
4.4	Tests sur résidus. . . . .	93
4.5	conclusion de comparaison . . . . .	94
4.6	Échantillons plus grands que 205. . . . .	95

## Introduction générale

La majorité des phénomènes naturels sont complexes Soit non- linéaires ou chaotiques, Par contre les séries temporelles chaotiques sont générées par les phénomènes chaotiques.

La prévision du comportement futur des séries chronologiques chaotiques est l'un des sujets importants en sciences statistiques, et ce, en fonction de l'importance de son utilisation dans les domaines de la vie, bien qu'il soit difficile de prévoir à long terme, la prévision de série chronologique chaotique peut être possible à court terme, sans savoir toutes les raisons d'influence (L'effet papillon de Lorenz).

L'étude de la prévision était difficile auparavant car les chaotiques sont des systèmes déterministes complexes en présence d'un grand nombre de variables, qui influent sur l'évolution du processus, ce qui rend impossible à l'humain de simuler avec toutes formes d'imprévisible. Comme il est impossible aussi de déterminer l'état initial du système en sachant juste la détermination de l'état final.

La plupart des systèmes que l'on trouve dans la nature impliquent l'interaction de nombreux facteurs, le chaos est satisfait dans : le dynamique du système solaire, l'évolution des populations, la météo, les réactions chimiques, etc. Cependant, le pouvoir de prévision dans le cas de systèmes chaotiques peut être amélioré et peut être illustré par le système météo pour de courtes périodes ayant atteint une très bonne précision.

Nous devons alors insister sur le fait que l'émergence et le développement de la théorie du chaos ne peuvent pas avoir lieu avant l'invention des ordinateurs puisque la simulation de systèmes complexes avec de nombreuses variables ne pouvait pas se faire sans leur aide.

Les prévisions des séries chronologiques peuvent généralement être classées dans la méthode de prévision linéaire et non linéaire. Les modèles linéaires typiques de prévision de séries temporelles incluent principalement autorégressif (AR), moyenne mobile (MA), auto régressif moyenne mobile (ARMA).

Cependant, ces modèles sont pour la plupart des modèles de prévision basés sur des méthodes statistiques et ont besoin d'une connaissance à priori qui sont difficilement prouvés de sorte que ces modèles peuvent difficilement obtenir des résultats de prévision efficaces pour les séries temporelles non linéaire ou chaotique en raison de la plupart des séries chronologiques possédant des caractéristiques non linéaires ou chaotiques dans la vie réelle et la méthode de l'intelligence

artificielle tenant une meilleure capacité de rapprochement pour les séries temporelles non linéaire ou chaotiques et séquences associées à un grand nombre des éléments D. Bosq et D. Blanke (2007)[5].

De nombreux chercheurs ont proposé d'utiliser la méthode de l'intelligence artificielle pour résoudre le problème de la prévision de séries non linéaire du temps, comme algorithme génétique (AG), réseau de neurones artificielle (RNA), et la logique floue, etc.

La plupart des modèles de la reproduction de l'intelligence humaine et génétique représente le but de la recherche en intelligence artificielle. Plusieurs modèles et processus ont été conçus à cette fin. Ces modèles permettent souvent de résoudre plusieurs problèmes dans différents domaines en simulant des processus de raisonnement du cerveau humain et génétique.

Chaque modèle de l'intelligence artificielle contient des points forts, mais aussi des points faibles. La puissance de l'informatique et le processus de raisonnement du cerveau humain et génétique, ceci dans le but d'obtenir des résultats plus performants avec une moindre incidence mathématique ou physique. Elles ont développé des modèles non-linéaires puis les méthodes modernes ont vu le jour, en particulier dans le domaine de l'informatique, par exemple les Réseaux de neurones artificiels, Algorithmes génétiques et logique floue ; cette méthode est capable d'apprendre et de s' auto-adapter avec n'importe quel modèle et ne nécessite pas d'hypothèses sur la nature de la série temporelle, tels que les modèles traditionnels.

Plusieurs modèles de séries chronologiques chaotiques ont été conçus pour la prévision, par exemple, la modélisation par programmation génétique (GPM) et l'optimisation de l'essaimage de particules (PSO).

Dans ce thèse nous essayons de modéliser des séries chronologiques chaotiques avec Trois modèles et comment prévoir le comportement futur, puis on comparez le résultat.

## **A ) Problématique :**

Il ya quatre décennies, le paradigme du chaos a été introduit [8].

Depuis la découverte du chaos, l'intérêt dans ce domaine de recherche a augmenté rapidement. Plusieurs tentatives ont été faites pour intégrer la théorie du chaos dans les programmes dans les premiers cycles .

Par conséquent, des expériences simples ont été proposées pour démontrer les caractéristiques de base du comportement chaotique, tels que l'irrégularité de mouvement et l'imprévisibilité de la sensibilité aux conditions initiales. Dans de nombreux cas, l'analyse du comportement chaotique observé expérimentalement a été confinée, pour des raisons de simplicité, à des simulations informatiques sur des modèles mathématiques.

Pour démontrer de façon convaincante qu'un système expérimental se comporte de façon chaotique, le chaos doit être identifié directement à partir des données expérimentales. A cet effet, les méthodes d'analyse des séries temporelles non-linéaires doivent être appliquées.

Les systèmes chaotiques qui génèrent notre série de temps chaotique sont en fait complexes du fait d'un grand nombre de variables qui influent sur l'évolution du processus qui rend impossible la simulation pour les humains et qui font paraître des systèmes inattendus. Aussi, est-il impossible de déterminer l'état initial du système pour nous permettre de connaître le statut final.

Il est connu que le chaos des systèmes complexes est inévitable néanmoins il existe des systèmes chaotiques évitables car prévisibles.

Cependant, le pouvoir de prévision dans le cas des systèmes chaotiques peut être amélioré sur de courtes périodes et peut atteindre une très haute précision en temps opportun.

Il faut rappeler que l'apparition et le développement de la théorie du chaos ne pouvaient pas avoir lieu avant l'invention des ordinateurs puisque la simulation de systèmes complexes avec de nombreuses variables ne pouvait pas se faire sans leur aide.

## **B ) Plan du travail :**

Ce travail est composé de quatre chapitres :

### **Le chapitre 1 :**

Cette partie est focalisée sur les séries temporelles, leur définition et leurs types. Nous commencerons par l'analyse des séries temporelles et leurs objectifs représentés par la description, la modélisation, la prévision .

### **Dans le chapitre 2 :**

concernera, dans la première partie, la théorie de chaos, les séries temporelles chaotiques, dans la deuxième partie quelques modèles qui génèrent ces dernières et enfin la prévision des séries chronologiques chaotiques.

### **Le chapitre 3 :**

Les méthodes de modélisation d'une série chronologique chaotique.

**Dans le chapitre 4** intitulé Application, nous présenterons l'application des méthodes BJ , RNA et AG à la série simulée chaotique (carte logistique) et nous effectuerons une rapide comparaison dans les prévisions de chaque méthode.

# Chapitre 1

## Analyse série temporelle

### 1.1 Introduction

Une série temporelle est une suite finie  $(x_1, \dots, x_n)$  de données indexées par se temps. L'indice temps peut être selon les cas : la minute, l'heure, le jour, l'année etc. La discipline qui traite ce type de quantités est l'analyse des séries temporelles. L'analyse des séries temporelles en question consiste en opération de prévision, de modélisation et caractérisation. L'objectif de prévision est de prévoir avec précision l'évolution à court termes de séries temporelles, et la modélisation vise la détermination des caractéristiques du comportement à long terme. La caractérisation à comme but de déterminer les propriétés fondamentales d'une série temporelle. La prévision est le sujet de ce travail.

Nous rappelons que la prévision des séries temporelles a principalement débuté dans les années 70, de nombreuses techniques de prévision de l'évolution des systèmes dynamiques ont été développées pour des domaines aussi variés que l'agriculture, la météorologie, l'automatique, la finance, etc. Une importante classe du modèle stochastique utilisé pour la prédiction décrite la relation entre la future valeur de série et les précédentes.

Le choix du modèle dépend de la nature de série (stationnaire, non stationnaire, linéaire, non linéaire). Pour les séries temporelles stationnaires Box et Jenkins (1970) ont proposé d'utiliser les fonctions d'auto-corrélation et d'auto-corrélation partielle empiriques. Leur méthodologie repose sur le fait que la fonction d'auto-corrélation théorique d'un modèle moyenne mobile MA  $q$  s'annule à partir du rang  $q+1$  et la fonction d'auto-corrélation partielle d'un modèle autorégressif AR  $p$  s'annule à partir du rang  $p+1$ . Une alternative pour la sélection de modèle a été introduite par Cleveland (1972). Sa méthode utilise la fonction d'auto-corrélation inverse  $p_i(h)(h \in z)$ , définie à l'aide de l'inverse de la den-

sité spectrale[37]. Bhansali (1980) fut le premier à publier une étude complète concernant la distribution asymptotique de la fonction d'auto-corrélation inverse estimée par les deux méthodes évoquées précédemment [18]. Dans la suite de ces travaux, Bhansali (1983) a développé une nouvelle approche basée sur les propriétés des espaces de Hilbert pour définir la fonction d'auto-corrélation partielle inverse [40]. Les séries non stationnaire subit un prétraitement avant d'utiliser le modèle ARIMA qui les transforment en séries stationnaire par l'utilisation d'un opérateur selon la nature de la non-stationnarité (non-stationnarité en moyenne, non-stationnarité en variance, saisonnalité), pour les séries non linéaire, les modèles ARCH et GARCH et EAR sont utilisées.

Le problème que ces prévisions sont construites sur des séries transformées qui peuvent perdre leur caractère optimal lorsqu'on les exprime de la même façon que les données initiales. Les propriétés des prévisions construites sur la série transformée ne se conservent pas forcément après l'opération de prétraitement inverse. Mais, ces modèles restent très difficiles à appliquer dans la plupart des séries temporelles réelles( cas pratiques ).

## 1.2 Les concepts de base de séries chronologiques

### Définition 1.2.1. (série temporelle)

Une série temporelle (ou série chronologique) à temps discret est une suite réelle finie  $(x_t)_{1 \leq t \leq n}$ , où  $t$  représente le temps (en minute, jour, année...).

Voici quelques exemples de séries temporelles :

**Exemples 1.2.1.** : les prix du cuivre (1800-1997) Les données (Martino, 1983) sont annuelles. Le graphe annoté est présenté dans la *figure 1.1*. On peut localiser la fin de la 1ère guerre mondiale, la grande crise des années '30, les crises pétrolières de 1973 et 1980.

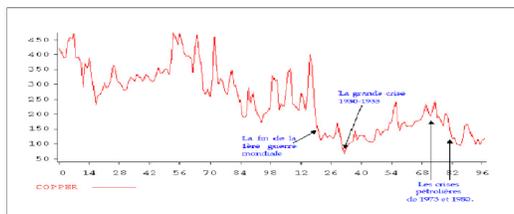


FIGURE 1.1 – les données du prix du cuivre

### 1.2.1 Objectifs de l'analyse d'une série temporelle

Décrire Quand on s'intéresse à une série temporelle, la première étape, comme pour toutes données est de décrire la série. On utilise pour ce faire un certain nombre de graphiques : - diagramme séquentiel (time plot)

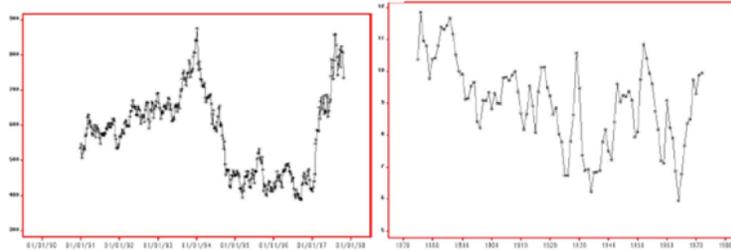


FIGURE 1.2 – :Cours de l'action Alcatel et Niveau du lac Huron

- histogramme pour avoir une idée de la distribution des valeurs  
 - histogramme des valeurs absolues pour apprécier l'hétéroscédasticité éventuelle Sur les graphiques on peut repérer les valeurs atypiques ou aberrantes. On calcule aussi les statistiques descriptives usuelles : moyenne, variance, coefficients d'aplatissement et d'asymétrie.

Une même série temporelle peut être analysée de différentes façons suivant l'objectif poursuivi. [25]

Résumer Dans certains cas on veut une vue synthétique débarrassée de détails de court-terme; c'est souvent un besoin des instituts de statistique officielle. Modéliser Expliquer le niveau ou parfois la variance du niveau, par des modèles à peu de paramètres

- Modèle sans variable explicative

$$Y_t = f(Y_{t-1}, Y_{t-2}, \dots) + u_t \quad (1.1)$$

- Modèle avec variable explicative

$$Y_t = f(X_t) + u_t \quad (1.2)$$

où  $u_t$  est une erreur

- statique :  $X_t$  ne contient pas de valeurs passées de  $Y_t$  et les  $u_t$  sont non corrélés entre eux - dynamique : les  $u_t$  sont auto-corrélés ou  $X_t$  contient des valeurs retardées de  $Y_t$  Prédire La prévision de valeurs à des dates futures connaissant le présent et le passé de la série peut être basée sur un modèle ou bien être construite sans référence à un modèle. [28]

### 1.2.2 Domaines d'application :

On trouve des exemples de séries chronologiques dans de très nombreux domaines.

La liste suivante n'est qu'un échantillon :

- Finance et économétrie : évolution des indices boursiers, des prix, des données économiques des entreprises, des ventes et achats de biens, des productions agricoles ou industrielles,
- Assurance : analyse des sinistres,
- Médecine/biologie : suivi des évolutions des pathologies, analyse d'électro-encéphalogrammes et d'électrocardiogrammes,
- Sciences de la terre et de l'espace : indices de marées, variations des phénomènes physiques (météorologie), évolution des taches solaires, phénomènes d'avalanches.
- Traitement du signal : signaux de communications, de radars, de sonars, analyse de la parole,
- Traitement des données : mesures successives de position ou de direction d'un objet mobile (trajectographie),
- Démographie : analyse de l'évolution d'une population
- Météorologie : analyse de données climatiques. . .
- Géophysique : analyse de données sismiques. . .
- Traitement d'images : analyse d'images satellites, médicales.
- Energie : prévision de la consommation d'électricité.
- Agriculture : production mensuelle de lait par vache, la quantité d'un produit .[31]

### 1.2.3 Eléments constitutifs d'une série temporelle

Une série temporelle est généralement constituée de plusieurs éléments.

- Tendance : représente l'évolution à long terme de la série (échelle interannuelle).

Exemples :

croissance économique, évolution climatologique à long terme (cyclique ou non)

- Saisonnalité : évolution se répétant régulièrement tous les ans.

Exemples :

En météorologie, température plus faibles en hiver qu'en été.

En économie, saisonnalité induite par les périodes de vacances, les périodes de fêtes, le climat...

- Composante stationnaire (ou résiduelle) : ce qui reste lorsque l'on a enlevé les autres composantes.

Décrit l'évolution à court terme de la série (échelle journalière).

Notion de série temporelle stationnaire définie plus précisément dans la suite. Cette hypothèse jouera un rôle fondamentale dans la suite, et remplacera l'hypothèse usuelle des v.a i.i.d. (ici, il peut exister une dépendance entre deux valeurs successives prises par la série observée).

Le modèle le plus usuel consiste à supposer que la série initiale s'écrit sous la forme (modèle additif)

$$Y_t = T_t + S_t + \varepsilon_t \quad \text{pour tout } t \in \{1, \dots, n\} \quad (1.3)$$

avec  $T_t$  la tendance,  $S_t$  la composante saisonnière (fonction périodique) et  $\varepsilon_t$  la composante stationnaire (ou résiduelle).

Les différentes étapes de la modélisation sont alors :

1. modéliser la tendance.
2. modéliser la composante saisonnière.
3. modéliser la série résiduelle. [\[32\]](#)

#### 1.2.4 Tendence. Saisonnalité. Résidus :

Il est classique de décomposer une série temporelle  $\{Y_t, t = 1, \dots, n\}$  en tendance  $T_t$  (trend), effet saisonnier  $S_t$ , erreur  $\varepsilon_t$ . On s'intéresse habituellement à un modèle additif :

$$Y_t = T_t + S_t + \varepsilon_t, \quad \text{Où } E(\varepsilon_t) = 0 \quad (1.4)$$

ou à un modèle multiplicatif :

$$Y_t = T_t \times S_t \times \varepsilon_t, \quad \text{Où } E(\varepsilon_t) = 1 \quad (1.5)$$

Par exemple, les séries montrant une saisonnalité qui a de plus en plus d'ampleur (cas des ventes de champagne), sont souvent mieux ajustées par un modèle multiplicatif que par un modèle additif.

## 1.3 Analyse de la tendance

### 1.3.1 Moyenne mobile

Si  $\{x_i\}_{i \in \{1, \dots, n\}}$  est une série temporelle, alors la moyenne mobile d'ordre  $p$  associée est la série temporelle définie pour  $t \in \{p + 1, \dots, n - p\}$  par

$$\hat{x}_t = \sum_{i=t-p}^p x_i \quad (1.6)$$

avec  $2p + 1$  la largeur de la fenêtre.

### 1.3.2 Différenciation

On peut éliminer la tendance par différenciation.

**Définition 1.3.1.** Notons  $\nabla$  l'opérateur de différenciation, c'est à dire l'opérateur défini par

$$\nabla x_t = x_t - x_{t-1} \quad (1.7)$$

pour tout  $t \geq 2$

**Définition 1.3.2.** On définit l'opérateur de différenciation d'ordre  $k$  par la formule de récurrence

$$\nabla^{(k)} x_t = \nabla(\nabla^{(k-1)} x_t) \quad (1.8)$$

Propriétés -

1. Soient  $u_t$  et  $v_t$  deux séries temporelles,  $\nabla(u_t + v_t) = \nabla u_t + \nabla v_t$ .
2. Soit  $x_t$  une série temporelle et  $\lambda \in \mathbb{R}$ ,  $\nabla(\lambda x_t) = \lambda \nabla x_t$ .
3. Soit  $y_t = a_0 + a_1 t + \dots + a_k t^k$ , alors  $\nabla^k y_t = k! a_k$  et  $\nabla^{(k+1)} y_t = 0$ .

La 3ème propriété implique que si  $X_t = T_t + Z_t$  avec  $T_t$  une tendance polynomiale, alors on peut supprimer la tendance en appliquant successivement plusieurs fois l'opérateur  $\nabla$ .

### 1.3.3 Modèle additif avec tendance paramétrique

**Définition 1.3.3.** *On dit qu'une série temporelle suit un modèle de tendance additif lorsqu'elle peut se décomposer sous la forme*

$$X_t = T_t + Z_t \quad (1.9)$$

avec  $Z_t$  une série temporelle sans tendance. On dit qu'une série temporelle a une tendance linéaire lorsque

$$T_t = a \times t + b \quad (1.10)$$

On dit qu'une série temporelle a une tendance polynomiale lorsque

$$T_t = a_0 + a_1t + a_2t^2 + \dots \quad (1.11)$$

#### Ajustement du modèle :

On suppose que les observations  $\{x_1, \dots, x_n\}$  suivent un modèle additif avec une tendance paramétrique représentée par une fonction  $f$  de paramètres  $\theta$ , c'est à dire vérifie :

$$x_t = f(t; \theta) + z_t \quad (1.12)$$

avec  $z_t$  une série temporelle sans tendance.

On cherche alors à estimer les paramètres inconnus  $\theta$ . On utilise généralement la méthode des moindres carrés.

Vérification du modèle et autres modèles paramétriques.

Afin de vérifier l'adéquation du modèle aux observations, on calcule les estimations  $\hat{\theta}$  puis on trace le graphique.

$$t \longrightarrow x_t - f(t; \hat{\theta})$$

Si le modèle est valable, il ne reste plus de tendance sur cette nouvelle série temporelle. Si-non on cherche un modèle plus adapté.

Il reste une tendance visible. Sur la série initiale, il semble que les fluctuations saisonnières sont proportionnelles à la tendance. Dans ce cas, un modèle additif n'est pas adapté. On peut alors tester un modèle multiplicatif.

**Définition 1.3.4.** - *On dit qu'une série temporelle suit un modèle de tendance multiplicatif lorsqu'elle peut se décomposer sous la forme :*

$$X_t = T_t \times Z_t \quad (1.13)$$

avec  $Z_t$  une série temporelle sans tendance.

En pratique, on se ramène à un modèle additif par passage au logarithme, puisqu'on a alors

$$\log(X_t) = \log(T_t) + \log(Z_t) \quad (1.14)$$

### 1.3.4 Comparaison des différentes méthodes

Les principaux avantages des méthodes paramétriques sur les deux autres méthodes sont

- qu'elles fournissent un modèle facilement interprétable et qui permet, par exemple, de comparer plusieurs séries temporelles entre elles.
- qu'elles peuvent être utilisées en prévision, ce qui n'est pas le cas des autres méthodes. L'avantage principal des deux premières méthodes (moyenne mobile et différenciation) est qu'elles s'ajustent à de nombreuses séries temporelles sans modifications, alors que pour les modèles paramétriques il peut être difficile de choisir le bon modèle.

### 1.3.5 Modélisation de la tendance

Pour modéliser la tendance dans une série temporelle, les différentes étapes sont :

1. Tracer la série temporelle
2. Utiliser ce graphique pour identifier le modèle en répondant aux questions suivantes :
  - a. Modèle additif ou multiplicatif? Pour cela, on regarde si les composantes saisonnières et stationnaires sont proportionnelles à la tendance ou non. Lorsque le modèle est additif, on travaille sur la série afin de se ramener à un modèle additif.
  - b. Méthode paramétrique ou non-paramétrique? Pour cela, on regarde si il existe un modèle simple permettant de décrire la tendance observée (modèle linéaire, polynomiale, exponentiel...). Si on ne trouve pas de tel modèle, on peut utiliser une méthode non paramétrique.

3. Vérifier le modèle en traçant la série "résiduelle", c'est à dire la série dans laquelle on a enlevé la tendance en utilisant la méthode choisie.

## 1.4 Analyse des composantes saisonnières

Dans ce paragraphe,  $X_t$  désigne une série temporelle sans tendance. En pratique, lorsqu'il existe une tendance sur la série initiale  $X_t$ , on commence par l'enlever en utilisant l'une des méthodes décrites dans la section précédente.

On supposera dans la suite de cette section, sauf mention contraire, que la série  $X_t$  suit un modèle de saisonnalité additif, c'est à dire que

$$X_t = S_t + Z_t \quad (1.15)$$

avec  $S_t$  une fonction périodique de période  $\tau$ , avec  $\tau$  le nombre de données par année, c'est à dire vérifiant  $S_t = S_{t+\tau}$  et  $Z_t$  une série temporelle stationnaire.

De nombreuses séries temporelles observées dans la nature suivent un modèle multiplicatif, c'est à dire vérifient

$$X_t = S_t \times Z_t \quad (1.16)$$

Dans ce cas, les fluctuations de la série autour de la composante saisonnière sont proportionnelles à celle-ci. On se ramène alors à un modèle additif par passage au logarithme.

Dans le cadre des modèles additifs, plusieurs méthodes peuvent être utilisées pour estimer la fonction  $t \rightarrow S_t$ . C'est l'objet des paragraphes ci-dessous.

### 1.4.1 Calcul des moyennes saisonnières

On renumérote la série  $X_t$  sous la forme  $X_{jk}, j = 1, \dots, T_{an}, k = 1, \dots, N_{an}$  avec  $T_{an}$  la longueur de l'année et  $N_{an}$  le nombre d'années de mesure. Le modèle se réécrit alors sous la forme :

$$X_{jk} = S_k + Z_{jk} \quad (1.17)$$

Un estimateur naturel de  $S_k$  est donné par  $\frac{1}{T} \sum_{j=1}^{N_{an}} Z_{jk}$ , c'est à dire par la moyenne des observations disponibles à la date  $k$  pour chaque année. On appelle généralement  $S_k$  les "moyennes saisonnières".

### 1.4.2 Vérification du modèle

Afin de vérifier l'adéquation du modèle, on trace la série résiduelle  $Z_k$ . Si le modèle convient, il ne reste plus de saisonnalité apparente sur cette série résiduelle, généralement appelée "série corrigée des variations saisonnières".

### 1.4.3 Différenciation

**Définition 1.4.1.** - On appelle différence d'ordre  $\tau$  de la série  $X_k$  les quantités

$$\nabla_t^{(\tau)} = X_{t+\tau} - X_t \quad \text{pour} \quad \tau \in \mathbb{N} \quad (1.18)$$

Si on suppose que la série  $\{X_t\}$  suit un modèle de saisonnalité additif, c'est à dire que

$$X_t = S_t + Z_t$$

avec  $S_t$  une fonction périodique de période  $\tau$ , avec  $\tau$  le nombre de données par année, c'est à dire vérifiant  $S_{t+\tau} = S_t$  et  $Z_t$  une série temporelle stationnaire, et  $\nabla_\tau x_t = x_t - x_{t-\tau} = z_t - z_{t-\tau}$  est une série stationnaire.

## 1.5 Estimation de la tendance et de la saisonnalité

Une série temporelle  $(x_t)_{1 \leq t \leq n}$  est l'observation des  $n$  premières réalisations d'un processus stochastique  $(X_t)_t$ . C'est ce processus que l'on cherche désormais à modéliser.

Pour cela, la démarche suivante doit être adoptée :

- représenter graphiquement la série afin de repérer les tendances et saisonnalités ;
- estimer et supprimer les tendances et saisonnalités (partie déterministe du processus stochastique) ;
- choisir un modèle pour les résidus (partie aléatoire du processus stochastique) et l'estimer ;
- prédire les réalisations futures à l'aide de ce modèle.

### 1.5.1 Bruit blanc

**Définition 1.5.1.** *Un processus de **bruit blanc faible** est une suite de variables aléatoires  $(X_t)_t$  indépendantes, d'espérance et de variance constantes. Si l'espérance est nulle, le bruit blanc est centré, et si les variables aléatoires sont gaussiennes, le bruit blanc est **gaussien**.*

### 1.5.2 Processus stationnaire

Un processus aléatoire  $(X_t)_t$  est dit **stationnaire** si et seulement si :

- 1)  $E(X_t) = \mu \forall t$ , c'est-à-dire que la moyenne du processus est de  $\mu$ , donc **constante** ;
- 2)  $var(X_t) < \infty, \forall t$ , c'est-à-dire que la variance est finie et invariable dans le temps ;
- 3)  $cov(X_t, X_{t+h}) = E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma(h) \forall t$  , la covariance est indépendante du temps.

Il apparaît, à partir de ces propriétés, qu'un processus de bruit blanc<sup>1</sup>  $\varepsilon_t$  dans lequel les  $\varepsilon_t$  et de même loi  $N(0, \sigma_\varepsilon^2)$  est stationnaire

### 1.5.3 Une estimation paramétrique de la tendance (trend)

Nous supposons que la série temporelle étudiée soit la réalisation d'un processus stochastique composé d'une tendance déterministe  $T_t$  et d'une partie aléatoire  $\varepsilon_t$  (supposée de moyenne nulle) :

$$X_t = T_t + \varepsilon_t \quad (1.19)$$

Une méthode simple consiste à supposer que cette tendance est linéaire :

$$T_t = a + bt, \quad (1.20)$$

et d'estimer les paramètres  $a$  et  $b$  par moindres carrés.

Ainsi, si on observe la série  $x_1, \dots, x_n$ , il faut trouver  $a$  et  $b$  qui minimisent la quantité :

$$\sum_{t=1}^n (x_t - a - bt)^2 \quad (1.21)$$

---

1. Bruit Blanc "White Noise" :Ce terme est emprunté à la physique faisant référence au spectre de la lumière blanche.

Les solutions de ce problème sont :

$$\hat{a} = \frac{6}{n(n-1)} \left( - \sum_{t=1}^n tx_t + \frac{2n+1}{3} n\bar{x} \right) \quad (1.22)$$

$$\hat{b} = \frac{12}{n(n^2-1)} \left( \sum_{t=1}^n tx_t + \frac{n+1}{2} n\bar{x} \right) \quad (1.23)$$

L'hypothèse de linéarité de la tendance convient très bien à certaines séries temporelles, mais ce n'est pas le cas de toutes les séries (voir la [figure\(1.1\)](#)). Il est alors possible de supposer que la tendance soit de forme polynomiale :

$$T_t = a + bt + ct^2 \quad (1.24)$$

et d'estimer les paramètres  $a, b$  et  $c$  par moindres carrés. Mais il est parfois difficile d'estimer le degré du polynôme, et lorsque le degré est trop important, le nombre de paramètres à estimer devient grand et les calculs fastidieux. Dans cette situation, on a recourt à une méthode d'estimation non paramétrique .

#### 1.5.4 Estimation non paramétrique : moyenne mobile

**Tendance** Supposons que la tendance  $T_t$  soit linéaire dans un petit intervalle  $[t - q, t + q]$  autour de  $t$ . Dans ce cas, un bon estimateur de la tendance est la moyenne sur cet intervalle :

$$\hat{T}_t = \frac{1}{2q+1} \sum_{k=-q}^q x_{t+k} \quad (1.25)$$

On peut donc estimer la tendance à chaque temps  $t$  en calculant la moyenne sur les observations étant dans une fenêtre de largeur  $2q + 1$  autour de  $t$  : c'est ce que l'on appelle une estimation par moyenne mobile.

Pour éviter les problèmes de bord, on suppose que  $x_t = x_1$  si  $t < 1$  et  $x_t = x_n$  si  $t > n$ .

**Tendance et saisonnalité.** Supposons désormais que le processus ne comporte pas uniquement une tendance, mais également une saisonnalité :

$$X_t = T_t + S_t + \varepsilon_t,$$

avec  $S_t$  une fonction  $T$ -périodique. Le principe d'estimation est (en simplifiant légèrement) le suivant : On estime la tendance moyenne sur une période, puis on estime la composante saisonnière en moyennant sur toutes les périodes les écarts à la tendance moyenne de la période.

### 1.5.5 Elimination de la tendance et de la saisonnalité par la méthode des différences

Cette méthode permet de supprimer les tendances et saisonnalité d'une série temporelle sans les estimer.

Soit  $\nabla_T$  l'opérateur qui associe  $(X_t - X_{t-T})$  à  $(X_t)$

$$\nabla_T = X_t - X_{t-T}.$$

On note  $\nabla$  l'opérateur  $\nabla_1$ , et  $\nabla_T^k$  l'opérateur  $\underbrace{\nabla_T \circ \dots \circ \nabla_T}_{k \text{ fois}}$

**Proposition 1.5.1.** *Soit un processus admettant une tendance polynomiale d'ordre  $k$  :*

$$X_t = \underbrace{\sum_{j=0}^k a_j t^j}_{m_t} + \varepsilon_t \quad (1.26)$$

Le processus  $\nabla X_t$  admet une tendance polynomiale d'ordre  $k - 1$ . Ainsi, en appliquant  $k$  fois  $\nabla$ , on élimine la tendance.

## 1.6 Série temporelle lineare

Une série temporelle  $Y_t$  est dite linéaire si elle peut s'écrire :

$$Y_t = \mu + \sum_{i=-\infty}^{\infty} \psi_i Z_{t-i} \quad (1.27)$$

où  $Z_t$  est un bruit blanc  $(0, \sigma_Z^2)$ ,  $\psi_0 = 1$  et la suite  $\{\psi_i\}$  est absolument sommable, c'est-à-dire  $\sum_i |\psi_i| < \infty$ . On admettra qu'une série linéaire est stationnaire.

Une série est dite linéaire et causale si elle est linéaire et  $\psi_i = 0, i < 0$ , autrement dit elle ne dépend que du bruit blanc passé. Pour une discussion sur le sens de l'écriture ci-dessus, on peut voir par exemple Hamilton.

Si  $Y_t$  est linéaire et causal on obtient :  $E(Y_t) = \mu, Var(Y_t) = \sigma_Z^2 \sum_{i=0}^{\infty} \psi_i^2$  L'autocovariance d'ordre  $k$  est :

$$\begin{aligned} \gamma_k &= COV(Y_t, Y_{t-k}) = E\left[\sum_{i=0}^{\infty} \psi_i Z_{t-i}, \sum_{j=0}^{\infty} \psi_j Z_{t-k-j}\right] \\ &= E\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j Z_{t-i} Z_{t-k-j}\right) \\ &= \sum_{j=0}^{\infty} \psi_{j+k} \psi_j E(Z_{t-k-j}^2) = \sigma_Z^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \end{aligned}$$

Si la série est linéaire et causale et si de plus  $\psi_i = 0$  pour  $i > q$  on dit que  $Y_t$  est une moyenne mobile d'ordre  $q$  ( $MA(q)$ ). Une série linéaire causale est un  $MA(\infty)$ .

## 1.7 Modélisation des séries stationnaires

Après avoir examiné des outils d'exploration, nous abordons la modélisation des séries temporelles. On considère qu'une série temporelle observée  $\{y_t, t = 1, \dots, T\}$  est la réalisation de v.a. (variables aléatoires)  $\{Y_t, t = 1, \dots, T\}$  qui forment elles-mêmes une portion d'un processus aléatoire  $\{Y_t, t = \dots, 0, 1, 2, \dots\}$ , c'est-à-dire d'une série infinie de v.a. Pour pouvoir faire de l'estimation (de moyenne, de variance,.....), il faut que le phénomène ait une certaine stabilité. La notion de stationnarité, est la clef de l'analyse des séries temporelles [35].

### 1.7.1 Corrélation, fonction d'auto corrélation et fonction d'auto corrélation partielle

**Fonction d'auto corrélation (ACF).** Considérons une série (faiblement) stationnaire  $X_t$ . On est souvent intéressé par décrire la dépendance de  $X_t$  par rapport à son passé, notamment pour expliquer le niveau actuel de la série par le niveau à une date précédente.

Si une dépendance est linéaire, elle est bien décrite par le coefficient d'auto corrélation. Par définition, le coefficient d'auto corrélation d'ordre  $l$  est :

$$\rho_t = \frac{Cov(X_t, X_{t-l})}{\sqrt{V(X_t)V(X_{t-l})}} \quad (1.28)$$

Mais,  $V(X_{t-l}) = V(X_t) = \gamma_0$  et  $E(X_t) = \mu$  donc :

$$\rho_t = \frac{Cov(X_t, X_{t-l})}{V(X_t)} = \frac{E[(X_t - \mu)(X_{t-l} - \mu)]}{E(X_t - \mu)^2} = \frac{\gamma_l}{\gamma_0} \quad (1.29)$$

**Fonction d'auto-corrélation partielle(ACPF)** de  $X_t$ , est définie par :

$$\begin{cases} \alpha(1) = corr(X_2, X_1) = \rho(1) \\ \alpha(k) = corr(X_{k+1} - P_{\bar{s}p\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{\bar{s}p\{1, X_2, \dots, X_k\}} X_1) \end{cases}$$

où  $P_{\bar{s}p\{1, X_2, \dots, X_k\}}$  est la projection sur le sous-espace engendré par  $1, X_2, \dots, X_k$ .  
 $\alpha(k) = \phi_{kk}, k \geq 1$  avec  $\phi_{kk}$  tel que :

$$\begin{pmatrix} \rho_0 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & \rho_2 & \dots & \rho_{k-2} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_0 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \rho_k \end{pmatrix} \quad (1.30)$$

### 1.7.2 Les processus auto-régressifs $AR_p$

Les premiers modèles que nous présentons sont les processus auto-régressifs, construits à partir de l'idée que l'observation au temps  $t$  s'explique linéairement par les observations précédentes.

**Définition 1.7.1.** On dit que  $(X_t)$  est un processus auto-régressif d'ordre  $p$  (centré) s'il s'écrit

$$X_t = \epsilon_t + \sum_{j=1}^p a_j X_{t-j}, \quad (1.31)$$

où  $\epsilon_t$  est un bruit blanc centré de variance  $\sigma^2$ .

L'observation  $X_t$  au temps  $t$  est alors la somme d'un choc aléatoire à l'instant  $t$ ,  $\epsilon_t$ , indépendant de l'historique, et d'une fonction linéaire de son passé  $\sum_{j=1}^p a_j X_{t-j}$ , qui peut être vue comme la prédiction de  $X_t$  à partir des  $p$  dernières observations passées.

$\epsilon_t$  est l'**innovation** contenue dans le processus au temps  $t$ . On dit que  $(\epsilon_t)$  est le processus d'innovation. Les coefficients  $a_j$  doivent vérifier la contrainte suivante pour assurer la stationnarité du processus : le polynôme caractéristique du processus (2),  $A(z) = 1 - a_1 z - \dots - a_p z^p$ , ne doit avoir que des racines (réelles ou complexes) de module strictement supérieur à 1.

**Remarque 1.7.1.** en prenant l'espérance de (2) et en utilisant la stationnarité du processus, on obtient que l'espérance  $\mu$  du processus vérifie  $\mu(1 - \sum_{j=1}^p a_j) = 0$ . Comme 1 ne peut être racine du polynôme  $A(z)$ , on a nécessairement  $\mu = 0$ .

**Remarque 1.7.2.** la Définition 1.6.1 nous ne considérons que des processus ARp centrés.

Un processus ARp non centré serait défini par  $X_t = \mu + \epsilon_t + \sum_{j=1}^p a_j X_{t-j}$ .

**Proposition 1.7.1.** La variance du processus ARp défini en 1.6.1 est :

$$\sigma(0) = \sigma^2 + \sum_{j=1}^p a_j \sigma(j) \quad (1.32)$$

et l'auto-covariance, pour tout  $h > 0$  :

$$\sigma(h) = \sum_{j=1}^p a_j \sigma(h-j). \quad (1.33)$$

L'auto-covariance d'un processus ARp vérifie donc la formule de récurrence d'ordre  $p$  suivante :

$$\sigma(h) - a_1 \sigma(h-1) - \dots - a_p \sigma(h-p) = 0. \quad (1.34)$$

Le polynôme caractéristique de cette relation de récurrence est :

$$z^p - a_1 z^{p-1} - \dots - a_{p-1} z - a_p = z^p \left( 1 - \frac{a_1}{z} - \dots - \frac{a_p}{z^p} \right) = z^p A \left( \frac{1}{z} \right). \quad (1.35)$$

Les racines  $\lambda_i$  de ce polynôme sont les inverses des racines du polynôme  $A$ , et sont donc de module strictement inférieur à 1. En supposant que  $A$  a toute ses racines distinctes, les solutions de l'équation de récurrence précédente sont de la forme :

$$\sigma(h) = \sum_{i=1}^p c_i \lambda_i^h. \quad (1.36)$$

Ainsi, puisque  $|\lambda_i| < 1$ , **l'auto-covariance décroît exponentiellement avec  $h$ .**

Ces résultats s'étendent immédiatement à l'auto-corrélation d'un processus  $AR_p$ . Quant à l'auto-corrélation partielle, elle est nulle à tout ordre strictement supérieur à  $p$ , et vaut  $a_p$  à l'ordre  $p$  :

$$r(h) = 0 \quad \forall h > p,$$

$$r(h) = a_p.$$

### 1.7.3 Les processus en moyenne mobile $MA_q$ :

La seconde catégorie de modèles classiques regroupe les processus en moyenne mobile.

**Définition 1.7.2.** On appelle *moyenne mobile (Moving Average)* d'ordre  $q$  un processus de la forme

$$X_t = \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (1.37)$$

où les  $\epsilon_j$  pour  $t - q \leq j \leq t$  sont des bruits blancs centrés de variance  $\sigma^2$ .

On notera parfois  $X_t = \sum_{j=0}^q b_j \epsilon_{t-j}$  en imposant  $b_0 = 1$ .

A noter que pour l'instant aucune condition n'est nécessaire sur les  $b_i$  pour que le processus stationnaire<sup>2</sup>.

Un processus moyenne mobile est nécessairement centré, et son auto-covariance vérifie la proposition suivante.

---

2. pour des questions de prévision, nous serons amené à supposer que le polynôme caractéristique  $B(z) = 1 + b_1 z + \dots + b_q z^q$  a toute ses racines de module strictement supérieurs à 1.

**Proposition 1.7.2.** *L'auto-covariance d'un processus moyenne mobile  $X_t = \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q}$  est*

$$\sigma(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} b_k b_{k+h} & \forall h \leq q, \quad \text{Où } b_0 = 1 \\ 0 & \forall h > q \end{cases} \quad (1.38)$$

L'auto-corrélation est donc également nulle au dessus de l'ordre  $q$ . On retrouve le même comportement que l'autocorrélation partielle d'un  $ARp$ . L'auto-corrélation partielle d'un processus moyenne mobile est compliquée à calculer, et sans grand intérêt. Néanmoins, il est important de savoir qu'elle tend vers 0 à vitesse exponentielle lorsque  $h$  tend vers l'infini.

-Ce comportement symétrique entre les processus moyenne mobile et auto-régressif est du à la propriété suivante :

**Proposition 1.7.3.** *Un processus auto-régressif est un processus moyenne mobile d'ordre infini, et réciproquement un processus moyenne mobile est un processus auto-régressif d'ordre infini.*

#### 1.7.4 Les processus mixtes $ARMAp, q$

Cette classe plus générale de modèles définit des processus sous la forme d'une récurrence auto-régressive avec un second membre de type moyenne mobile.

**Définition 1.7.3.** *Un processus auto-régressif moyenne mobile d'ordres  $p$  et  $q$  est de la forme :*

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \sum_{j=0}^q b_j \epsilon_{t-j}. \quad (1.39)$$

où les  $\epsilon_j$  pour  $t - q \leq j \leq t$  sont des bruits blancs centrés de variance  $\sigma^2$ .

La stationnarité d'un  $ARMAp, q$  est assurée lorsque toutes les racines du polynôme

$A(z) = 1 - a_1 z - \dots - a_p z^p$  sont de module strictement supérieur à 1. Ce polynôme forme avec  $B(z) = 1 + b_1 z + \dots + b_q z^q$  les deux polynômes caractéristiques du processus.

On supposera également que les polynômes  $A$  et  $B$  n'ont pas de racine commune, afin de s'assurer qu'il n'y a pas de représentation plus courte.

-On peut écrire le processus  $ARMA_{p,q}$  sous la forme suivante :

$$X_t - a_1X_{t-1} - \dots - a_pX_{t-p} = \epsilon_t - b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q} \quad (1.40)$$

Ainsi,

$$\begin{aligned} Cov(X_{t+h} - a_1X_{t+h-1} - \dots - a_pX_{t+h-p}, X_t) &= \sigma(h) - a_1\sigma(h-1) - \dots - a_p\sigma(h-p) \\ &= Cov(\epsilon_{t+h} - b_1\epsilon_{t+h-1} + \dots + b_q\epsilon_{t+h-q}, X_t) \end{aligned}$$

qui est nulle dès que  $h > q$ . L'auto-covariance d'un processus  $ARMA_{p,q}$  évolue selon la même récurrence qu'un  $AR_p$  (cf. équation (1.33)).

Ainsi, l'auto-covariance (et l'auto-corrélation) d'un processus  $ARMA_{p,q}$  vont tendre exponentiellement vers 0 lorsque  $h$  tend vers l'infini, à partir de l'ordre  $q + 1$ .

### 1.7.5 Illustrations

Voici une simulation d'un processus  $ARMA_{2,2}$  [figure \(1.3\)](#) ainsi que ses auto-corrélation et auto-corrélation partielle empiriques.

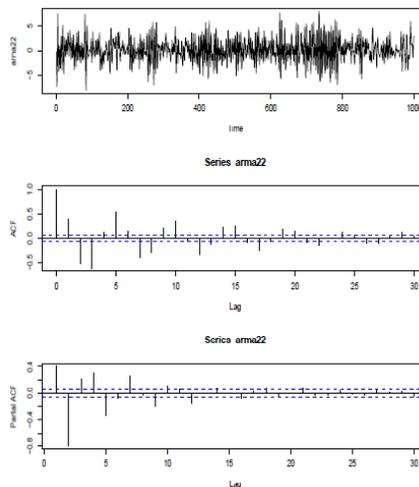


FIGURE 1.3 – Simulation d'un  $ARMA_{2,2}$ , auto-corrélation et auto-corrélation partielle

## 1.8 Processus non stationnaire :

Les modèles *ARMA* sont destinés à modéliser des processus stationnaires. En pratique, les séries temporelles sont généralement non stationnaires, et un pré-traitement est nécessaire pour supprimer les tendances et saisonnalités. Une fois la série stationnarisée analysée, et les valeurs futures prédites, il est ensuite nécessaire de revenir à la série initiale.

**Exemples 1.8.1.** *Considérons une série temporelle  $(x_t)_{1 \leq t \leq n}$  qui présente une saisonnalité de période 12. Afin de supprimer cette saisonnalité, c'est la série suivante que nous étudions :*

$$y_t = x_t - x_{t-12} \quad \text{pour } t = 13, \dots, n \quad (1.41)$$

*L'ajustement d'un modèle ARMA et les prévisions sont réalisées sur la série  $y_t$ . Il est donc nécessaire de revenir à la série initiale, car ce sont les prévisions de  $(x_t)$  qui nous intéressent.*

*De l'équation (1.41) on obtient :*

$$\begin{aligned} x_t &= y_t + x_{t-12} \\ &= y_t + y_{t-12} + x_{t-24} \\ &\cdot \\ &\cdot \\ &= y_t + y_{t-12} + y_{t-24} + \dots + y_{r(t)+12} + x_{r(t)} \end{aligned}$$

*où  $r(t)$  est le reste de la division euclidienne de  $t$  par 12.*

*Puisque l'on connaît les  $x_t$  pour  $t \leq n$  ainsi que les prévisions  $y_{n+h}$ , on peut en déduire les prévisions de  $x_{n+h}$ .*

### 1.8.1 Les processus ARIMA et SARIMA

Les processus *ARIMA* et *SARIMA* sont la généralisation des modèles *ARMA* pour des processus non stationnaires, admettant une tendance (*ARIMA*), ou encore une tendance et une saisonnalité (*SARIMA*). En pratique, et dans le logiciel R notamment, ce sont ces processus qui sont directement utilisés.

Soit  $\nabla$  l'opération de différenciation, qui associe au processus  $(X_t)_{t \in \mathbb{N}}$  le processus  $(X_t - X_{t-1})_{t \in \mathbb{N}}$ . Nous rappelons que le processus obtenu en différenciant deux fois de suite,  $(X_t - 2X_{t-1} + X_{t-2})_{t \in \mathbb{N}}$ , est noté  $\nabla^2$ . Et ainsi de suite.

**Définition 1.8.1.** *Un processus  $(X_t)$  est un processus  $ARIMA_{p,d,q}$  si le processus*

$$Y_t = \nabla^d X_t \quad (1.42)$$

*est un processus  $ARMA_{p,q}$ .*

*Les processus  $ARIMA_{p,d,q}$  sont donc bien adaptés aux séries temporelles présentant une tendance polynômiale de degré  $d - 1$ .*

Soit  $\nabla_T$  l'opérateur qui fait passer de  $(X_t)$  à  $(X_t - X_{t-T})$ .

**Définition 1.8.2.** *Un processus  $(X_t)$  est un processus  $SARIMA_{p,d,q,T}$  si le processus*

$$Y_t = \nabla_T \circ \nabla^d X_t \quad (1.43)$$

*est un processus  $ARMA_{p,q}$ .*

*Les processus  $SARIMA_{p,d,q,T}$  sont bien adaptés aux séries temporelles présentant une période de longueur  $T$  et une tendance polynômiale de degré  $d$ .*

**Remarque 1.8.1.** *en réalité, les processus  $SARIMA_{p,d,q,T}$  sont plus complexes et comportent d'autres paramètres. Nous ne considérerons que cette version  $SARIMA_{p,d,q,T}$  pour des compléments sur ces processus.*

\* **Box et Jenkins** ont introduit une méthodologie pour la construction d'un modèle linéaire qui s'ajuste au mieux à une série temporelle. Cette méthodologie se décompose en trois étapes : l'identification du modèle, l'estimation des paramètres et la validation du modèle. La figure.(1.4) résume la procédure pour un modèle  $ARIMA$  , [7].

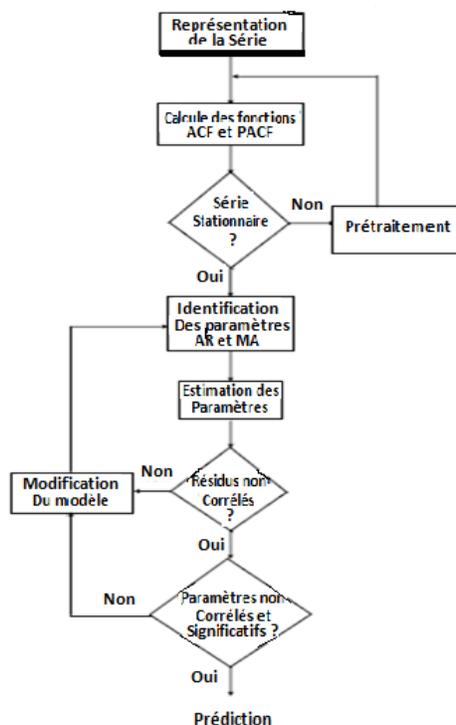


FIGURE 1.4 – Méthodologie Box et Jenkins

### 1.8.2 Estimation des paramètres du processus $ARMA_{(p;q)}$

L'estimation des coefficients du processus  $ARMA_{(p;q)}$  s'effectue principalement à l'aide de la méthode du maximum de vraisemblance. On suppose pour cela que

$$\epsilon_t \rightsquigarrow N(0, \sigma_\epsilon^2) \quad (1.44)$$

**Méthode d'estimation du maximum de vraisemblance :** La méthode du maximum de vraisemblance est couramment utilisée pour estimer les coefficients des modèles des séries temporelles car c'est une méthode simple à mettre en place pour estimer des modèles . Soit le modèle :

$$y_t = a_0 + a_1 x_t + \epsilon_t \quad (1.45)$$

La fonction de densité de la loi normale de la variable  $y_t$  s'écrit :

$$\begin{aligned} f(y_t) &= \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \exp \left[ \frac{-(y_t - E(y_t))^2}{2\sigma_\epsilon^2} \right] \\ &= \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \exp \left[ \frac{-(y_t - a_0 - a_1 x_t)^2}{2\sigma_\epsilon^2} \right] \end{aligned}$$

La fonction de vraisemblance est donnée par :

$$f(y_1, \dots, y_n; a_0, a_1, \sigma_\epsilon^2) = \prod_{t=1}^n f(y_t; \alpha, \beta, \sigma_\epsilon^2) \quad (1.46)$$

où  $f(y_t; \alpha, \beta, \sigma_\epsilon^2)$  est la fonction de densité de  $y_t$ .

On obtient alors :

$$\begin{aligned} f(y_1, \dots, y_n; a_0, a_1, \sigma_\epsilon^2) &= \prod_{t=1}^n \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \exp \left[ \frac{-(y_t - a_0 - a_1 x_t)^2}{2\sigma_\epsilon^2} \right] \\ &= \left( \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \right)^n \exp \left[ \sum_{t=1}^n \frac{-(y_t - a_0 - a_1 x_t)^2}{2\sigma_\epsilon^2} \right] \end{aligned}$$

Il vient alors :

$$\ln f(y_1, \dots, y_n; a_0, a_1, \sigma_\epsilon^2) = n \ln \left( \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \right) - \left[ \sum_{t=1}^n \frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\epsilon^2} \right]$$

Cette fonction est à maximiser. Les valeurs des coefficients qui permettent de maximiser la fonction sont issues des conditions du premier ordre suivantes :

$$\frac{\partial \ln f(\dots)}{\partial a_0} = 0, \quad \frac{\partial \ln f(\dots)}{\partial a_1} = 0, \quad \frac{\partial \ln f(\dots)}{\partial a_\epsilon} = 0.$$

### 1.8.3 Validation du processus $ARMA_{(p;q)}$

Lors de la détermination des ordres  $petq$  du processus  $ARMA_{(p;q)}$  à l'aide des corrélogrammes simple et partiel, on peut être amené à sélectionner plusieurs ordres possibles  $petq$  pour le processus  $ARMA_{(p;q)}$ . Après avoir estimé les différents processus  $ARMA_{(p;q)}$  possibles, il reste à les valider et à les départager. La validation des processus passe par un examen des coefficients estimés et par un examen des résidus.

#### Tests sur les coefficients

Parmi les processus  $ARMA$  estimés, on ne retiendra que ceux dont tous les coefficients ont un de  $t_{Student} > 1;96$  (pour un risque de 5% et pour une taille d'échantillon suffisamment grande :  $T > 30$ ).

## 1.9 Tests sur les résidus

### 1.9.1 Tests d'autocorrélation

Il existe un grand nombre de tests d'autocorrélation, les plus connus sont ceux de Box et Pierce (1970) et Ljung et Box (1978).

Soit une autocorrélation des erreurs d'ordre  $K$  ( $K > 1$ ) :

$$\varepsilon_t = \sum_{i=1}^K p_i \varepsilon_{t-i} + v_t, \quad \text{Où } v_t \rightsquigarrow N(0, \sigma_{v^2}) \quad (1.47)$$

Les hypothèses du test de Box-Pierce sont les suivantes :

$$\begin{cases} H_0 : p_1 = p_2 = \dots = p_K = 0 \\ H_1 : \text{il existe au moins un } p_1 \text{ significativement différent de } 0. \end{cases}$$

Pour effectuer ce test, on a recours à la statistique  $Q$  qui est donnée par :

$$Q = n \sum_{i=1}^K \hat{p}_i^2 \quad (1.48)$$

Sous l'hypothèse  $H_0$  vraie,  $Q$  suit la loi du Khi-Deux avec  $K$  degrés de liberté :  $Q \rightsquigarrow \chi^2(K)$  si  $Q > k^*$  où  $k^*$  est la valeur donnée par la table du Khi-Deux pour un risque fixé et un nombre  $K$  de degrés de liberté. Alors, on rejette  $H_0$  et on accepte  $H_1$  (autocorrélation des erreurs).

### 1.9.2 Tests de normalité

Pour calculer des intervalles de confiance prévisionnels et aussi pour effectuer les tests de Student sur les paramètres, il convient de vérifier la normalité des erreurs. Le test de Jarque et Bara (1984), fondé sur la notion de Skewness (asymétrie) et de Kurtosis (aplatissement), permet de vérifier la normalité d'une distribution statistique.

#### Les tests du Skewness et du Kurtosis

Soit  $\mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^k$  le moment centré d'ordre  $k$ , le coefficient de Skewness ( $\beta_1^{1/2}$ ) est égal à :  $(\beta_1^{1/2}) = \frac{\mu_3}{\mu_2^{3/2}}$  et le coefficient de Kurtosis :  $\beta_2 = \frac{\mu_4}{\mu_2^2}$ . Si la distribution est normale et le nombre d'observation grand ( $n > 30$ ) :  $\beta_1^{1/2} \rightarrow N(0; \sqrt{\frac{6}{n}})$  et  $\beta_2 \rightarrow N(3; \sqrt{\frac{24}{n}})$

$$v_1 = \frac{|\beta_1^{1/2} - 0|}{\sqrt{\frac{6}{n}}} \text{ et } v_2 = \frac{|\beta_2 - 3|}{\sqrt{\frac{24}{n}}} \quad (1.49)$$

que l'on compare à 1,96 (valeur de la loi normale au seuil de 5%). Si les hypothèses

$H_0 : v_1 = 0 \text{ et } v_2 = 0$  sont vérifiées, alors  $v_1 \leq 1,96$  et  $v_2 \leq 1,96$ ; dans le cas contraire, l'hypothèse de normalité est rejetée.

#### Le test de Jarque et Bara

Il s'agit d'un test qui synthétise les résultats précédents; si  $\beta_1^{1/2}$  et  $\beta_2$  obéissent à des lois normales alors la quantité  $s : s = \frac{n}{2}\beta_1 + \frac{n}{24}(\beta_2 - 3)^2$  suit un  $\chi^2$  à deux degrés de liberté. Donc si  $s > \chi_{1-\alpha}^2(2)$ , on rejette l'hypothèse  $H_0$  de normalité des résidus au seuil  $\alpha$ .

#### Le Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov est un test non paramétrique de conformité de deux populations régies par des variables quantitatives continues. Il est plus puissant que le test de  $\chi^2$  et ne nécessite pas de regroupement en classes ce qui le rend efficace dans le cas de petits échantillons. Le principe est de comparer les fréquences relatives cumulées (FRC) de l'échantillon  $E$  et celui, théorique, de la population  $P$ . La fonction discriminante est l'écart maximal observé (qui est aussi une distance entre deux probabilités) :

$$D_c = \max_{1 \leq i \leq n} |FRC_{obs}(x_i) - FRC_{th}(x_i)| \quad (1.50)$$

Les hypothèses relatives à un test bilatéral sont les suivantes :

$$\begin{cases} H_0 : FRC_{obs} = FRC_{th} & \text{pour tout } x \text{ réel;} \\ H_1 : FRC_{obs} \neq FRC_{th} & \text{pour au moins une valeur de } x \text{ réel.} \end{cases}$$

La distribution de la variable  $D$  est, sous l'hypothèse  $H_0$  connue et tabulée. Dès que  $n > 40$ , on peut considérer les valeurs seuils suivantes :

$$D_{5\%} = \frac{0.886}{\sqrt{n}} \quad D_{1\%} = \frac{1.031}{\sqrt{n}} \quad (1.51)$$

On rejettera  $H_0$  dès que  $D_c \leq D_s$  [4].

### 1.9.3 Tests de stationnarité

#### Tests de Dickey-Fuller augmenté

Soit  $Y_t$  notre série initiale centrée. Le test de Dickey-Fuller augmenté (ADF, 1981) consiste à effectuer la régression (1.52) et tester la significativité du coefficient  $p - 1$  à l'aide des seuils de MacKinnon [23].  $D_t = \mu + \alpha t$  est la partie déterministe de la série,  $\alpha$  et  $\mu$  pouvant être nuls. Selon les valeurs de  $\alpha$  et  $\mu$ , on obtient les différentes versions du test DFA du tableau (1.1)

$$\Delta Y_t = D_t + (p - 1)Y_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta Y_{t-1} + \varepsilon_t \quad (1.52)$$

TABLE 1.1 – Différentes versions du test DFA.

version du test		valeur critique à 5%
Tendance+ constante	$\alpha \neq 0, \mu \neq 0$	-3.454
Constante	$\alpha = 0, \mu \neq 0$	-2.89
$\phi$	$\alpha = 0, \mu = 0$	-1.944

DICKEY-FULLER augmenté : test de racine unitaire
$H_0$ : $\rho - 1 = 0$ dans la regression (1.52) . Racine unitaire, non stationnarité.
$H_1$ : $\rho - 1 < 0$ , stationnarité de la série.
statistique de test : statistique de STUDENT du coefficient ( $\rho - 1$ ).
rejet de $H_0$ : seuils de Mac KINNON, tableau 4.6.

**Remarque 1.9.1.** *Il faut aussi tester la normalité du résidu (au moins en dessinant un histogramme). Sans cette normalité, les statistiques de test sur les coefficients n'ont pas les lois asymptotiques habituelles et les intervalles de confiance sur les coefficients, appuyés sur le théorème de la limite centrale, sont très approximatifs.*

*Cependant, on montre qu'un modèle  $AR_{(p)}$  choisi par le critère AIC permet une prédiction avec une plus petite erreur  $\hat{\sigma}_\varepsilon^2(p)$  (on dit que ce critère est asymptotiquement efficace : le rapport entre l'erreur de prédiction évaluée au minimum théorique et l'erreur se basant sur la valeur choisie par le critère AIC tend vers 1 si  $T$  tend vers l'infini) [26], [21].*

#### 1.9.4 Critères de choix des modèles

Après examen des coefficients et des résidus, certains modèles sont écartés. Pour départager les modèles restants, on fait appel aux critères d'information et aux critères standards.

##### Critères d'information :

. Le critère Bayesian :  $BIC = \ln \sigma_\varepsilon^2 + (p + q)$

. Le critère d'Akaike :  $AIC = \ln \sigma_\varepsilon^2 + \frac{2(p+q)}{n}$

. Le critère de Schwarz :  $SIC = \ln \sigma_\varepsilon^2 + (p + q) \frac{\ln(n)}{n}$

. Le critère de Hannan - Quinn :  $HQ = \ln \sigma_\varepsilon^2 + \alpha(p + q) \ln \left[ \frac{\ln(n)}{n} \right]$

où  $\alpha (> 2)$  est une constante [22], où  $n$  est le nombre d'observations de la série.

Les modèles ayant la plus petite valeur du critère devront être choisis.

**Critères standards :**

. L'erreur absolue moyenne (Mean Absolute Error) :

$$MAE = \frac{1}{n} \sum_t |e_t|.$$

où  $e_t$  est le résidu du modèle *ARMA* étudié et  $n$  le nombre d'observation.

. l'erreur quadratique moyenne ( Mean Squared Error) :

$$MSE = \frac{1}{n} \sum_1^n e_t^2.$$

. la racine carrée de l'erreur quadratique moyenne(Root Mean Square Error) :

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n e_t^2}.$$

. Ecart absolu moyen en pourcentage (Mean Absolute Percent Error) :

$$MAPE = 100 \frac{1}{n} \sum_1^n \left| \frac{e_t}{X_t} \right|.$$

Plus la valeur de ces critères est faible, plus le modèle estimé est proche des observations.

## 1.10 Conclusion

L'analyse de série temporelle est un des objectifs fondamentaux de la statistique. Nous insistons sur le fait, que quelle que soit la méthode utilisée, il faut être vigilant sur les prévisions effectuées qui peuvent être dans certains cas totalement aberrantes.

Nous conseillons fortement aux praticiens de se limiter aux méthodes qu'ils connaissent lorsqu'ils effectuent leurs prévisions.

Le logiciel qu'ils utilisent doit être sûr, contrôlé : il nous est arrivé d'obtenir des résultats différents sur les mêmes données et par la même méthode en employant deux logiciels différents, ou d'aboutir des prévisions manifestement fausses puisque différentes des valeurs données dans les ouvrages de référence.

# Chapitre 2

## les séries temporelles chaotiques

### 2.1 Introduction

Les séries temporelles chaotiques sont des séries temporelles des phénomènes chaotiques et les phénomènes chaotiques sont des systèmes dynamiques. Les systèmes dynamiques chaotiques sont omniprésents dans la nature, comme la tornade, le marché boursier, la turbulence et la météo.

Leurs fonctions sont différentes dans des situations différentes. Par exemple, dans le cas de la tornade, le comportement chaotique est nocif pour les êtres humains et doit être évité ou contrôlé. Mais dans le cas des activités dans le cerveau humain, les comportements chaotiques sont utiles et nécessaires pour maintenir les fonctions normales du cerveau . Ainsi, il est une tâche importante de comprendre le chaos et le laisser servir la société humaine.[13]

Mathématiquement, le chaos peut être produit par les deux équations discrètes et continues.

- Les systèmes discrets peuvent être exprimés en tant que :

$$X_{n+1} = f(X_n) \tag{2.1}$$

tels que le plan logistique, la carte Henon, la carte standard, la tente carte, la cercle carte, et la carte ikeda.

- Les systèmes continus peuvent être exprimées par l'équation différentielle

$$\frac{dX(t)}{dt} = F(X(t)) \tag{2.2}$$

avec trois ou plusieurs degrés de liberté  $X(t) = [x_1(t), x_2(t), x_3(t), \dots, x_m(t)]$ .  
chaotique typique flux sont l'équation de Lorenz, équation "Rössler", l'équation

de "Duffing's", circuit de "Chua's", et ainsi de suite.

La carte discrète et flux continu ne sont pas étrangères, mais ont une relation étroite. La carte discrète (2.1) peut être considérée comme une projection de flux (2.2) sur une surface spécifique de l'article Poincaré. Laisser  $\delta x = 1$  et en prenant la limite, (2.1) sera transformée en (2.2). Sur D'autre part, (2.2) peut être transformé en (2.1), par intégration, c'est-à-dire  $X_{n+1} = X_n + \int_t^{t+T_n} F(X(t))dt$ ,  $T_n$  étant le temps de retourner à la section Poincaré.

Un point commun entre les systèmes discrets et continus est que les deux sont déterministes. Autrement dit, chaque condition initiale détermine de façon unique l'évolution du temps. Ainsi, le chaos est généralement appelé "déterministe chaos" et peut être considéré comme une sorte de hasard produit par le système déterministe.

Une condition nécessaire pour (2.1) et (2.2) à montrer le chaos est que les fonctions  $f$  et  $F$  doivent être non-linéaire. Il ya une infinité orbites périodiques instables dans le chaos, qui forment des ensembles invariants (2.2).

Un ensemble invariant est l'image de lui-même en vertu de l'évolution dans le temps. Une fois que la trajectoire se trouve dans ces invariants, il vont y rester pour toujours. En comparant avec les orbites chaotiques denses, les invariants sont rarement et sa mesure est nulle. Dans les situations expérimentales ou dans les simulations numériques, car il ya toujours du bruit ou d'erreurs de coupure, une trajectoire arbitraire ne sera jamais rester à ces invariants. Les invariants ne sont donc pas respectées en général. Compte tenu d'un temps assez long, une trajectoire perdra la mémoire sur son état initial et finalement s'installer sur une géométrie restreinte, appelé attracteur stable.

Une fois une trajectoire entre l'attracteur, il restera à tout jamais s'il n'y a pas de perturbation externe. La trajectoire va passer par tous les points de phase là avec sa mesure de nature à l'exception des orbites périodiques instables. Autrement dit, l'attracteur est également un ensemble invariant. Cette propriété est appelée "ergodicité" dans attracteur chaotique. [9]

- Mathématiquement, il ya beaucoup de mesures telles que la mesure de Lebesgue, séparé mesure, comptage mesure, et mesure naturelle. On suppose que la probabilité relative de  $x$  variable continue à apparaître à un emplacement est décrit par la densité de probabilité  $p(x)$ . La mesure  $\mu(D)$  est la probabilité de trouver  $x$  dans une zone  $D$ , à savoir,  $\mu(D) = \int_D p(x)dx$ . La moyenne d'une

fonction  $f(x)$  est :

$$\langle f(x) \rangle = \int f(x)p(x)dx = \int f(x)d\mu(x). \quad (2.3)$$

Lorsque la mesure  $\mu$  est une mesure invariante et la moyenne de la fonction  $f(x)$  sur  $\mu$  est égal à la moyenne sur le temps,  $\mu$  devient une mesure ergodique

$$\langle f(x) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x)dt = \int f(x)d\mu(x). \quad (2.4)$$

Cette mesure ne comprend pas les orbites périodiques instables dans l'attracteur et est appelé en tant que mesure naturelle. Pour utiliser cette mesure, on attend jusqu'à ce que tous les transitoires ont été rayés et attend pour une mesure de probabilité invariante décrivant la distribution des orbites typiques. Ainsi, dans la mesure naturelle, nous disons que les trajectoires chaotiques sont ergodique. Sur la base de cette propriété, de nombreuses quantités sont plus facilement définis comme des moyennes sur la mesure naturelle dans l'espace des phases.

En raison de la non périodicité de la trajectoire chaotique, l'attracteur est généralement appelé " attracteur étrange ", à la différence des attracteurs de point fixe et les orbites périodiques.

Pratiquement, nous faisons généralement pas les équations dynamiques (2.1) et (2.2), comme le tremblement de terre, le cerveau et le marché boursier, et ainsi de suite, donc les détails des équations du système dans l'espace phases et l'ensemble invariant asymptotique qui détermine ce qui peut être observé à travers des sondes expérimentales sont inconnus.

Ce que nous pouvons obtenir une certaine série de temps d'une ou, au mieux, quelques-unes des variables dynamiques du système, ce qui pose un grand défi pour caractériser les systèmes chaotiques. Une question intéressante est de savoir comment on d' un ou de quelques séries chronologiques de l'espace d'état ou de phase multi variée qui est nécessaire pour mouvements chaotiques de se produire ?

Heureusement, ce problème a été étudié intensivement au cours des dernières décennies et a maintenant formé un sujet appelé l'extraction de données ou l'analyse des données.

L'hypothèse de base est que la série de temps mesuré provient de l'attracteur du système inconnu avec ergodicity, qui contient les informations de l'attracteur. Ensuite, on peut utiliser la série de temps mesuré pour déterminer les propriétés

de l'attracteur tels que sa dimension, son squelette dynamique, et son degré de sensibilité aux conditions initiales.

Le retard coordonnée intégration technique établi par Takens fournit une solution pratique à cette tâche[17]. Supposons qu'une expérience est menée et série une fois  $s_n = s(n\Delta t)$  est mesurée, où  $\Delta t$  est l'intervalle d'échantillonnage. Une telle série de temps peut être, par exemple, un signal de tension provenant d'une expérience physique ou biologique, ou la concentration d'une substance dans une réaction chimique, ou la quantité de trafic instantané à un point / noeud dans l'Internet, et ainsi de suite. En principe, la série de temps mesuré provient d'un système dynamique sous-jacente avec (2.1) et (2.2), Avec ou sans l'influence du bruit. Le plus souvent, la série temporelle est une conséquence de mesures scalaires d'une certaine quantité qui est fonction de l'état actuel du système, prise à un temps d'échantillonnage fixe :

$$s_n = s(x(n\Delta t)) + \eta_n \quad (2.5)$$

où, le plus souvent, la fonction de mesure  $s$  est inconnu comme  $F$ ,  $x$  est le  $m$  dimensionnelle  $\eta_n$  est variable et le bruit de mesure.

Laissez-nous négligeons l'effet du bruit à ce niveau de la présentation. Équation(2.5) devient

$$s_n = s(x(n\Delta t)), \quad n = 1, \dots, N \quad (2.6)$$

La durée totale d'enregistrement est alors  $T = N\Delta t$ . En général,  $T$  doit être suffisamment grand pour que toute la dynamique soit exposée dans la série chronologique. C'est ce que nous avons en main.

## 2.2 Historique

Le chaos a été étudié pendant une longue période. Il est généralement estimé que Poincaré est le premier qui a étudié le chaos, a la fin du 19<sup>em</sup> siècle.

poincaré a étudié la trois corps ( problème restreint) où une petite masse est négligeable par rapport aux deux autres. Poincaré a trouvé que la solution de ce système simple est très compliquée et ne peut être donnée avec précision.

Puis en 1963, Lorenz a révélé "l'effet papillon" dans l'étude de la prévision du temps et est donc reconnu comme le père du chaos. Mais l'utilisation officielle de chaos est du papier de "période trois implique chaos" byLi et Yorke en 1975. Après cela, le chaos a été largement étudié et beaucoup de concepts importants

ont été mis en place, tels que les dimensions, les exposants de Lyapunov, Transformée de Fourier et transformée de Hilbert, et la reconstruction de l'attracteur. La caractéristique la plus frappante du chaos est l'imprévisibilité de son avenir à long terme. Cette fonction est généralement appelé la "dépendance de sensibilité aux conditions initiales" ou "effet papillon".

Par exemple, si deux conditions initiales ont un petit  $\delta x$  différence, leur différence après un temps  $t$  sera être  $\delta x e^{\lambda t}$  avec  $\lambda > 0$ , c'est-à-dire, la séparation exponentielle.

Ainsi, une petite différence ou même une coupure erreur sera soufflé rapidement et les résultats dans une grande différence dans l'avenir proche. "similaires les causes ont des effets similaires" n'est pas valide dans les systèmes chaotiques, sauf pour de courtes périodes.

Un système est appelé système chaotique, à condition que son maximum exposant de Lyapunov soit positif.

## 2.3 Théorie du chaos

La théorie du chaos est une théorie qui explique qu'il est impossible de prévoir l'état d'un système à long terme malgré les lois physiques connues parce que nous ne connaissons pas les conditions de départ du système [Physique].

## 2.4 Définition du chaos

Un système chaotique est un système déterministe non-linéaire, possédant un attracteur sensible Aux condition initiales. l'attracteur d'un système est défini comme le sous-ensemble de points vers lequel convergent toutes les trajectoires de ce système, les attracteurs les plus connus sont le point fixe, le cycle limite et le tore : ces attracteurs sont décrits périodiquement par les trajectoires représentatives du système étudié.

Le comportement de ces systèmes est prévisible : connaissant l'état, il est possible de prévoir les états futurs même si cette connaissance comporte une marge d'erreur, celle-ci reste du même ordre de grandeur pour toute détermination future. En d'autres termes, deux trajectoires initialement voisines le resteront toujours au cours du temps. Cette propriété a pour conséquence que de tels attracteurs ne peuvent correspondre aux attracteurs de systèmes chaotique .

La propriété fondamentale des systèmes chaotique est en effet la sensibilité aux

conditions initiales selon laquelle deux trajectoires initialement très proches vont diverger, à un taux exponentiel, au cours du temps<sup>1</sup>

pour rendre compte de ce phénomène, RUELLE et TAKENS [1971] ont introduit le concept d'attracteur étrange en tant qu'élément clé dans la compréhension des systèmes dynamiques chaotique. le terme étrange étant quelque peu ambigu<sup>2</sup>, nous parlerons d'attracteur chaotique pour décrire un attracteur qui affiche de la sensibilité aux conditions initiales.

Un système chaotique est ainsi caractérisé par une stabilité globale ( convergence des trajectoires vers l'attracteur) et une instabilité locale (divergence des trajectoires sur l'attracteur). par conséquent, un système qui possède un attracteur chaotique, c'est-à-dire un attracteur qui affiche de la sensibilité aux condition initiales .[27]

## 2.5 Les attracteurs

La région de l'espace de phases vers laquelle convergent les trajectoires d'un système dynamique dissipatif s'appelle "attracteur". Les attracteurs sont des formes géométriques qui caractérisent l'évolution à long terme des systèmes dynamiques. Il en existe quatre types distincts : un point, un cycle limite, un tore ou avoir une structure encore plus complexe de type fractale.

**1.** L'attracteur "point fixe" est un point de l'espace de phase vers lequel tendent les trajectoires, c'est donc une solution stationnaire constante.

**2.** L'attracteur "cycle limite" est une trajectoire fermée dans l'espace des phases vers laquelle tendent les trajectoires. C'est donc une solution périodique du système.

**3.** L'attracteur "tore" représente les mouvements résultant de deux ou plusieurs oscillations indépendantes que l'on appelle parfois "mouvements quasi périodiques".

**4.** Les attracteurs étranges sont bien plus complexes que les autres.

---

1. Une conséquence de la sensibilité aux conditions initiales est donc l'impévisibilité à long terme .

2. Il existe en effet des attracteurs étranges (fractals) non chaotiques et des attracteurs chaotiques non fractals (Cf.GREBOGI et al. [1984] et ECKMANN et RUELLE [1985]).

## 2.6 Bifurcation et routes vers le chaos

la theorie de bifurcation est l'étude mathématique des changements qualitatifs ou topologiques de la structure d'un système dynamique une bifurcation survient qualitatif des propriétés d'un système telles que la stabilité ,le nombre de points d'équilibre ou la nature des régimes permanents . au moment du changement sont appelées valeurs de bifurcation. dans les systèmes dynamiques, un diagramme de bifurcation montre les comportements possibles d'un système, à long terme, en fonction des paramètres de bifurcation[29] - par exemple la diagramme de bifurcation de la fonction logistique [figure\(2.1\)](#)

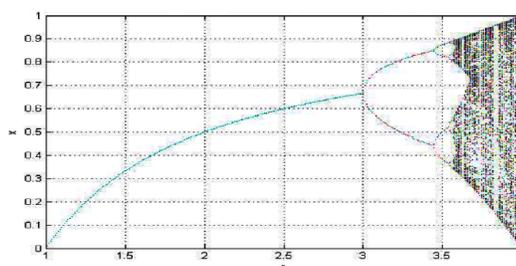


FIGURE 2.1 – [la diagramme de bifurcation de la fonction logistique.](#)

## 2.7 Les outils de détection du chaos

La découverte empirique du chaos nécessite l'utilisation d'outils spécifiques tels que la dimension fractale et mes exposants de lyapunov . En effet ,les méthodes traditionnelles basées sur le calcul d'auto corrélations ou l'analyse spectrale ne sont pas adaptées à l'étude des caractéristiques des systèmes chaotiques. En particulier, ces techniques ne permettent pas de distinguer un processus aléatoire d'un processus déterministe . Après avoir énoncé le théorème de takens [1981], nous présenterons deux outils de détection du chaos :les exposants de lyapunov et la dimension de corrélation .

### • Reconstruction de la dynamique dans l'espace des phases

la justification de l'utilisation des outils de détection du chaos réside dans le théorème de TAKENS [1981] -embedding theorem- supposons que l'on observe une série temporelle  $x_t$

$t = 1, \dots, T$ , générée par un processus inconnu . Comment comprendre la dy-

namique du système sous-jacent en analysant un seul signal ?

Ceci est possible car ,d'une part ,en pratique , on se restreint à l'étude de la dynamique sur l'attracteur et ,d'autre part ,on peut générer différents signaux à partir de la série initiale  $x_t$  . ainsi , afin de déterminer les propriétés du processus sous-jacent à cette série ,on va chercher à reconstruire son attracteur à partir de la série temporelle observée .pour cela , on forme  $(t - m + 1)$  vecteurs à  $m$  dimensions  $\{x_t^m\}_{t=1,\dots,T-m+1}$ ,appelés m-historiques , dont les composantes sont les valeurs consécutives de la série observée<sup>3</sup> décalées d'un retard  $\tau$  fixé<sup>4</sup>

$$x_t^m = (x_t, x_{t+\tau}, x_{t+2\tau}, x_{t+3\tau}, \dots, x_{t+(m-1)\tau})$$

Avec  $t = 1, ..T - m + 1$ .

$M$  est la dimension de plongement (embedding dimension) :c'est la dimension de l'espace des phases dans lequel l'attracteur est reconstruit . En d'autres termes , la dimension de plongement est le nombre d'axes nécessaire pour représenter l'attracteur .

TAKENS [1981] établit que si  $m \geq 2n + 1$  , où n est la dimension inconnue du vrai système , le comportement des m-historique imitera celui du système dynamique sous-jacent inconnu. En d'autres termes, l'attracteur ainsi reconstruit aura les même propriétés topologiques que le système initial . Grâce au théorème de TAKENS [1981] , il est alors possible de déterminer si l'attracteur du vrai système sous-jacent (inconnu) est chaotique, à partir uniquement de l'observation d'une série temporelle. a cette fin convient d'appliquer les divers outils de détection du chaos sur la série observée.

Ces différents instruments d'analyse - exposants de lyapynov et dimension de corrélation -sont exposés ci-après. - **Exposants de lyapynov**

Les exposants de lyapynov constituent un moyen direct de détection du chaos à partir de l'observation d'une série temporelle .ils permettent de quantifier la sensibilité aux conditions initiales puisqu'il constituent une mesure de la vitesse moyenne avec laquelle deux trajectoires s'écartent exponentiellement l'une de l'autre sur l'attracteur. Plus particulièrement, lorsqu'ils sont négatifs, ils mesurent le taux moyen de contraction et, lorsqu'ils sont positifs, le taux moyen de dilatation des orbites sur l'attracteur.

par conséquent, la positivité des exposants de lyapynov est révélatrice de l'instabilité locale des trajectoires sur l'attracteur, i.e. de la sensibilité aux condition

3. La construction des m-historiques est appeléeopération de reconstruction.

4. Ce retard est souvent choisi comme le premier retard tel qu'il n'y ait plus de dépendance temporelle dans la série.

initiales . dès lors , pour qu'un système soit chaotique , il faut qu'il y ait au moins un exposant de Lyapynov positif.

Formellement ,soit une application  $f$  de dimension 1,

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad \text{telle que } x_{n+1} = f(x_n), \quad t = 1, \dots, T.$$

L'Exposants de Lyapynov  $\lambda$  est défini par :

$$\lambda = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \ln \left| \frac{df(x_t)}{dx_t} \right| \quad (2.7)$$

On notera que l'exposant de Lyapynov est défini en termes de limite asymptotique ( $T \rightarrow \infty$ ) de manière à ne prendre en compte que le comportement de long terme du système , c'est -à-dire le comportement sur l'attracteur .

#### • La dimension de corrélation

La dimension est une notion importante dans l'étude des systèmes dynamiques non linéaires. En particulier, elle permet, en théorie, de distinguer un processus stochastique d'un processus déterministe au sens où la dimension d'un processus stochastique est élevée alors que celle d'un processus déterministe est faible.

L'intérêt d'étudier la dimension de l'attracteur d'un système provient également du fait qu'elle indique :

- La quantité d'information nécessaire pour spécifier la position d'un point sur l'attracteur,
- La limite inférieure du nombre de variables nécessaires pour décrire le comportement du système dynamique à l'intérieur de l'attracteur<sup>5</sup>,
- La densité relative des points sur l'attracteur, c'est-à-dire la façon dont sont répartis les points sur l'attracteur.

#### • Limites statistiques des outils et solutions apportées

Les exposants de Lyapunov et la dimension de corrélation ne constituent pas des tests au sens statistique du terme : leur distribution asymptotique est en effet inconnue. ainsi, comment savoir si le plus grand exposant de Lyapunov est significativement positif? Comment déterminer si la dimension de corrélation estimée est significativement inférieure à celle d'un processus purement aléatoire?

---

5. Contrairement à une idée couramment admise (Cf. notamment PETERS [1991],[1994]), la dimension ne fournit pas le nombre de variables indépendantes entrant dans le système , mais le nombre de dimension que remplit l'intérieur. Ces deux nombres sont par ailleurs indépendantes l'un de l'autre, le nombre de variables endogènes pouvant être nettement supérieur à la dimension de l'intérieur (Cf. ECKMANN [1992]).

Afin de répondre à ces interrogations, deux tests de validation des résultats fournis par les exposants de Lyapunov et la dimension de corrélation ont été proposés : le test résiduel de BROCK [1986] et le test du mélange aléatoire de SCHEINKMAN et LEBARON [1989].

L'idée sous-jacente au test résiduel de BROCK [1986] réside dans le fait que les propriétés d'un système déterministe restent inchangées sous une transformation linéaire. BROCK montre ainsi que si une série temporelle a une explication déterministe et que l'on estime un modèle autorégressif sur cette série, alors les résidus de ce modèle doivent avoir le même plus grand exposant de Lyapunov et la même dimension de corrélation que la série originelle. Si les résultats concernant les deux séries sont différents, le processus générant les données est plutôt de nature stochastique que déterministe.

Le test du "mélange aléatoire" de SCHEINKMAN et LEBARON [1989] consiste à mélanger de façon aléatoire les valeurs de la série étudiée. Si la structure sous-jacente à la série est déterministe chaotique, le mélange aura pour conséquence une destruction de la structure interne du système et par conséquent un accroissement de la dimension estimée. À l'inverse, si le processus sous-jacent est stochastique, le mélange n'aura qu'un très faible impact et la dimension estimée sur la série mélangée sera identique à la dimension estimée sur la vraie série.

## 2.8 la Conception de chaos mathématique

La conception de la théorie du chaos est considérée comme l'un des interdisciplines, et liée à une certaine connaissance de la discipline, comme la physique, les mathématiques et l'informatique et ainsi de suite, et pendant ce temps appliquée à de nombreux domaines des sciences naturelles ou sociales, par exemple des informations, de la physique et de la météorologie etc. D'une manière générale, le chaos est considéré comme un comportement apparemment aléatoire qui possède une certaine inhérente la régularité.

Et simultanément mouvement chaotique générée par le système déterministe est un comportement non-répétitif non périodique et détient sensibilité pour des conditions initiales. Jusqu'à présent, nous n'avons pas encore cherché une définition mathématique précise largement acceptée sur le chaos, mais la définition de chaos proposée par Devaney dans la référence est comprise et acceptée facilement par les gens, sa définition formelle de la manière suivante :

**Définition 2.8.1.** *Soit  $V$  un espace métrique. Pour une application continue*

$F : X \rightarrow X$ ,  $F$  est chaotique sur  $V$  si elle satisfait à trois conditions :

- 1)  $F$  possède dépendance sensible aux conditions initiales :  $\exists \delta > 0$  si  $\forall \varepsilon > 0, \forall x \in V$ , puis  $y$  et naturel numéro  $n$  exister dans le quartier  $\varepsilon$  de  $x$ , et satisfait la condition  $d(F^n(x), F^n(y)) > \delta$  ;
- 2)  $F$  est transitive topologiquement : pour toutes les paires de consigne ouvert  $X, Y$  on  $V$ , il existe  $k > 0$  et faire  $F^k(x) \cap Y \neq \phi$  ;
- 3) Le point périodique mis en  $U$  de  $F$  est dense sur  $X$ .

Selon les phénomènes chaotiques découvertes et les théories chaotiques recherches, leurs résultats de recherche montrent :

- 1) la théorie du chaos a proposé une analyse chaotique en fusionnant analyse déterministe traditionnelle et l'analyse stochastique qui sont brisés, les mouvements de chose objectif existent mouvement chaotique non ordonnée ;
- 2) la manifestation extérieure du mouvement chaotique est généralement considéré comme un phénomène de trouble, mais la forme de mouvement interne comprend des structures hiérarchiques prolifiques et ordonnée ;
- 3) pour tous les deux états systématiques de petites différences dans les conditions initiales existantes, leurs résultats sont très différence en raison du mouvement chaotique, et même plus pertinentes.

## 2.9 Les modèles qui génèrent les séries chronologiques chaotique

• **modèle Carte logistique** La carte logistique est une équation non linéaire très simple, donnée par l'équation (2.8)

$$x_{n+1} = rx_n(1 - x_n) \quad (2.8)$$

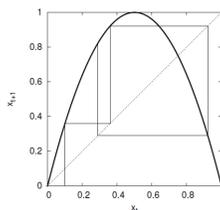


FIGURE 2.2 – **Itération de la carte logistique.**

La [figure\(2.2\)](#) génèrent les séries chronologiques chaotique pour  $0 < r < 4$  Fait intéressant, le comportement chaotique très complexe peut être observée lorsque,

(a)  $r = 2$ , (b)  $r = 3.5$ , (c)  $r = 4$  [figure\(2.3\)](#)

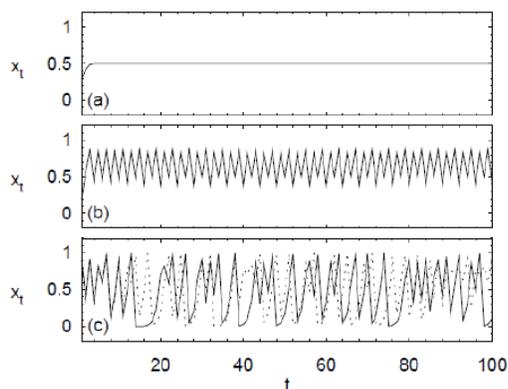


FIGURE 2.3 – Générer des Série chronologique chaotique .

la diagramme de bifurcation de Carte logistique dans La [figure\(2.4\)](#)

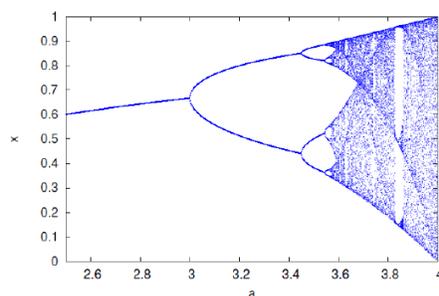


FIGURE 2.4 – la diagramme de bifurcation de Carte logistique .

### • Modèle de Lorenz

Edward Lorenz fut un météorologue qui, le premier, mit en évidence le caractère chaotique de la météorologie. Le couplage de l'atmosphère avec l'océan est décrit par le système d'équations aux dérivées partielles couplées de Navier-Stokes de la mécanique des fluides. En 1963, Lorenz eut l'idée de chercher un modèle d'équations pour étudier l'instabilité de Rayleigh-Bénard [\[34\]](#).

Ce modèle a joué un rôle historique important puisque son évolution temporelle fait apparaître un comportement chaotique.

De plus, il constitua le premier et le célèbre système différentiel dissipatif permettant d'observer un attracteur étrange pour certaines valeurs des paramètres.

Dans sa version exprimée en paramètres et variables réduits, le système de trois équations différentielles couplées s'écrit :

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = rx - y - xz \\ \dot{z} = xy - bz \end{cases} \quad (2.9)$$

$\sigma$  ou  $P_r$  désigne le nombre de Prandtl( égal à 10 dans les simulations) ,  $b = 8/3$  et  $r$  est le nombre de Rayleigh réduit.

Le système de Lorenz est un système d'équation non linéaires à cause des termes  $xy$  et  $xz$ . Ce système n'est pas intégrable dans le cas général. La détermination de ce système doit se faire à l'aide des méthodes d'approximation.

Lorsque les paramètres réels  $\sigma$ ,  $r$  et  $b$  prennent les valeurs suivantes :

$\sigma = 10$ ,  $r = 28$  et  $b = 8/3$ , avec les condition initiales

$x(0) = y(0) = z(0) = 0.01$ , le système (2.9) est chaotique.

Dans ce qui suit, nous vérifions quelques propriétés du système chaotique (2.9)

### - Aspect aléatoire

La figure (2.5) illustre l'aspect aléatoire des états du système (2.9)

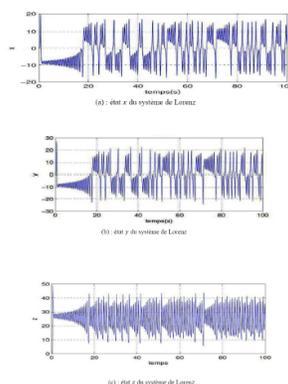


FIGURE 2.5 – Aspect aléatoire des états du système de Lorenz .

### - Attracteur étrange

Le système chaotique (2.9) présente un superbe attracteur étrange en forme

d'ailes de papillon, représenté sur la figure(2.6). La trajectoire commençant par s'enrouler sur une aile, puis sautant pour commencer à s'enrouler sur l'autre aile, et ainsi de suite.

On observe que la dynamique du système de Lorenz donné par le système (2.9) est indépendante du temps  $t$ , par conséquent ce type de système est qualifié d'être autonome [39].

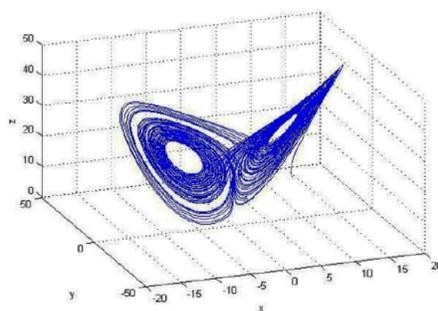


FIGURE 2.6 – **Attracteur étrange de Lorenz .**

### - Les exposants de Lyapunov

Après calcul des exposants de Lyapunov (voir figure (2.7)), nous avons obtenu les valeurs suivantes :

$$\lambda_1 = 0.85922, \lambda_2 = -0.0015763, \lambda_3 = -14.5208$$

Nous constatons bien qu'il y a un exposant de Lyapunov, ce qui signifie que le système est chaotique.

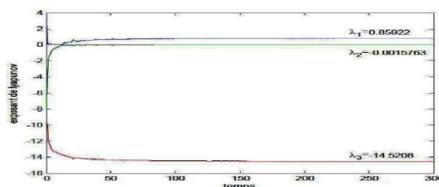


FIGURE 2.7 – **Exposants de Lyapunov du système chaotique de Lorenz .**

### • Le modèle de Hénon

L'astronome Michel Hénon exploitera la suggestion de Pomeau et Ibanez , pour obtenir un système très simple. Celui-ci est présent au séminaire de Nice (sur la turbulence), en janvier 1976 lorsque Pomeau expose l'idée de réaliser des systèmes dynamiques plus simples encore que celui de Lorenz, mais présentant des

caractéristiques similaires, et qui permettraient de prouver plus clairement des "évidences" mises en lumière par les calculs numériques. Puisque le raisonnement repose sur la section de Poincaré il propose de produire une application du plan dans lui-même, plutôt qu'une équation différentielle, imitant le comportement de Lorenz et son attracteur étrange[24]. La familiarité de Hénon avec les applications du plan dans lui-même, lui permet de réagir rapidement et de produire un système très simple donnant un attracteur étrange. Il est connu depuis sous le nom de système de Hénon et correspond à l'équations (2.10), un problème très général avec les attracteurs étranges, amené par la proposition de Ruelle et non résolu depuis : il n'existe aucune preuve mathématique de leur existence, seulement des indications numériques (graphiques), même pour les plus étudiés d'entre eux. Cela signe une certaine "limite" aux expériences numériques dans le domaine.[41]

### - Attracteur de Hénon

L'application de Poincaré correspondante au modèle de Lorenz étant fortement dissipative, donc difficile à calculer précisément. M. Hénon en 1976, dans le but de simplifier le système de trois équations différentielles du modèle de Lorenz, proposa l'équations suivante du plan sur lui même :

$$\begin{cases} X_{t+1} = 1 - aX_t^2 + Y_t \\ Y_{t+1} = bX_t \end{cases} \quad (2.10)$$

a et b étant deux paramètres réels, où la valeur de la constante a contrôle la nonlinéarité de l'itération, et celle de b traduit le rôle de la dissipation. Les valeurs habituellement utilisées pour a; b sont a = 1.4 et b = 0.3. L'attracteur de Hénon contient deux points fixes. Pour a = 1.4, on ne distingue plus les cycle ; le système est chaotique.

L'attracteur de Hénon pour a = 1.4 ; b = 0.3 est représenté dans la figure(2.8) :

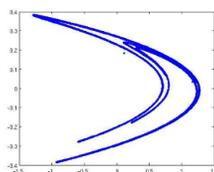


FIGURE 2.8 – **Attracteur de Hénon pour a = 1.4 , b = 0.3 .**

### - Diagramme de bifurcation

La construction de diagramme de bifurcation est faite en faisant varier le paramètre  $a$  de 0 à 2 avec un pas de 0.0005;  $b$  est égale à 0.3. Ce diagramme est de type de bifurcation de doublement de période.

Le diagramme de bifurcation obtenu est représenté par la [figure \(2.9\)](#).

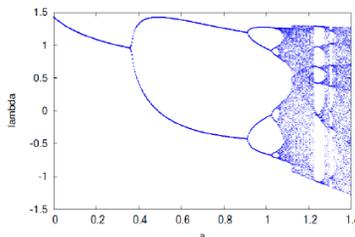


FIGURE 2.9 – [Diagramme de bifurcation de Hénon.](#)

### - Dimension de l'attracteur de Hénon

On peut aussi calculer la dimension de l'attracteur de Hénon par la dimension de Lyapunov. On a pour  $a = 1.4$ ,  $b = 0.3$ , l'application de Hénon a deux exposants de Lyapunov

$\lambda_1 = 0.42205$ ,  $\lambda_2 = -1.626$  la dimension de Lyapunov par définition est égale à  $DL = 1.2596$ . **Les exposants de Lyapunov** en fonction du temps pour l'application de Hénon est représenté par la [figure \(2.10\)](#).

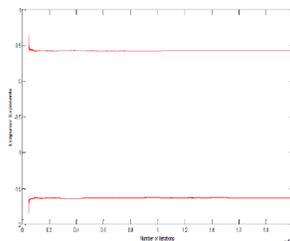


FIGURE 2.10 – [Les exposants de Lyapunov en fonction du temps pour l'application de Hénon.](#)

## 2.10 La prévision des séries chronologiques chaotiques

La prédiction du comportement futur des séries chronologiques chaotiques est l'un des sujets importants en sciences statistiques, en raison de la nécessité d'utiliser dans les domaines de la vie, Bien qu'il soit difficile de prédire à long terme.[11]

la prédiction de série chronologique chaotique à court terme, que de ne pas savoir toutes les raisons d'influence (L'Effet papillon de Lorenz), et a été de plus en plus d'intérêt dans le sujet de prédire ces dernières années par exemple, les prévisions météorologiques, les marchés financiers ,catastrophe, .....

Où il a développé de modèles l'inéluclabilité non-linéaires et les méthodes modernes ont vu le jour, en particulier dans le domaine de l'informatique, par exemple les réseaux de neurones artificiels et la logique floue et algorithme génétique, que ces méthodes sont capables d'apprendre et de s'adapter auto avec n'importe quel modèle et ne nécessite pas d'hypothèses sur la nature de la série temporelle, tels que les modèles traditionnels.[20]

## 2.11 Conclusion

La théorie du chaos et de l'analyse des séries chronologiques chaotiques a été bien étudié au cours des dernières décennies.

Dans ce chapitre, nous avons fourni un bref résumé de la notion de chaos et d'analyser la série de temps de chaos.[10]

En expliquant : conditions initiales sensibilité, les dimensions, les exposants de Lyapunov, attracteurs, Bifurcation et routes vers le chaos, Les outils de détection du chaos et Les modèles qui génèrent les séries chronologiques chaotique et il ya beaucoup des méthodes pour modélisé les séries chronologiques chaotique et prévision par exemple les réseaux de neurones artificiels, logique floue, algorithme génétique,.....

Dans ce thèse nous appliquons les méthodes des Réseaux de neurones artificiels,Algorithmes génétiques et Box-Jenkins pour modélisé les séries chronologiques chaotiques ( L'estimation des paramètres des séries simulées ) en fin La prévision.[38]

# Chapitre 3

## Les méthodes de modélisation d'une série chronologique chaotique

### 3.1 Les réseaux de Neurones Artificiels

#### 3.1.1 Introduction

L'évolution technologique durant les dernières années a permis aux scientifiques d'élaborer et de perfectionner des méthodes pour différents domaines. L'évolution des ordinateurs en particulier et la capacité d'intégration de composants formidable atteintes à nos jours ont permis une grande vitesse de calcul et une grande capacité mémoire.

Parmi ces méthodes, il existe une méthode qui est utilisée dans plusieurs domaines de recherches et de différentes manières, ainsi elle peut être utilisée d'une manière complètement Soft en utilisant uniquement l'ordinateur ou d'une manière Hard en utilisant les circuits intégrés. Cette méthode est celle des réseaux de neurones artificiels (RNA).

Les réseaux de neurones artificiels sont des outils puissants capables d'être utilisés dans près que tous les domaines technologiques, et on peut citer : Le traitement du signal, vision, parole, prévision, modélisation, aide à la décision, robotique, évaluation des écosystèmes, identification des bactéries, commande des processus, modélisation des systèmes physiques, reconnaissance des formes, mesure, instrumentation. Applications présente l'historique de l'application d'ANN aux problèmes chimiques et biologiques, donne un guide sur les architectures de réseau, la formation et l'extraction de règles à partir de réseaux formés et couvre de nombreux exemples d'application d'ANN à la chimie et à la biologie [6].

### 3.1.2 Définition

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau[12].

### 3.1.3 Historique

#### Les premiers succès

- 1957 : F. Rosenblatt développe le modèle du Perceptron. Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes. Notons qu'à cet époque les moyens à sa disposition sont limités et c'est une prouesse technologique que de réussir à faire fonctionner correctement cette machine plus de quelques minutes.

- 1960 : B. Widrow, un automaticien, développe le modèle Adaline (Adaptative Linear Element).

Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétro-propagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches. Les réseaux de type Adaline restent utilisés de nos jours pour certaines applications particulières. B. Widrow a créé dès cette époque une des premières firmes proposant neuro-ordinateurs et neuro-composants, la "Memistor Corporation".

Il est aujourd'hui le président de l'International Neural Network Society

- 1969 : M. Minsky et S. Papert publient un ouvrage qui met en exergue les limitations théoriques du perceptron. Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Ils étendent implicitement ces limitations à tous modèles de réseaux de neurones artificiels. Leur objectif est atteint, il y a abandon financier des recherches dans le domaine (surtout aux U.S.A.), les chercheurs se tournent principalement vers l'IA et les systèmes à bases de règles[14].

L'ombre - 1967-1982 : Toutes les recherches ne sont, bien sûr, pas interrompues. Elles se poursuivent, mais déguisées, sous le couvert de divers domaines comme : le traitement adaptatif du signal, la reconnaissance de formes, la modélisation en neurobiologie, etc.

De grands noms travaillent durant cette période tels : S. Grossberg, T. Kohonen,

....

**La levée des limitations** - 1983 : La Machine de Boltzmann est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du perceptron. Mais l'utilisation pratique s'avère difficile, la convergence de l'algorithme étant extrêmement longue (les temps de calcul sont considérables). - 1985 : La rétro-propagation de gradient apparaît. C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés Perceptrons multicouches). Sa découverte réalisée par trois groupes de chercheurs indépendants indique que "la chose était dans l'air".

Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétro-propagation de gradient reste le modèle le plus étudié et le plus productif au niveau des applications. [19]

**La situation actuelle (1992)** : En France, elle est à l'image du congrès Neuro-Nîmes qui a pour thème les réseaux neuromimétiques et leurs applications. Créé en 1988, le chiffre de ses participants croit chaque année et reflète bien l'intérêt que le monde scientifique et industriel (50% des participants).

### 3.1.4 Applications

Les réseaux de neurones servent aujourd'hui à toutes sortes d'applications dans divers domaines. Par exemple, on a développé un auto-pilote pour avion, ou encore un système de guidage pour automobile, on a conçu des systèmes de lecture automatique de chèques bancaires et d'adresses postales, on produit des systèmes de traitement du signal pour différentes applications militaires, un système pour la synthèse de la parole, des réseaux sont aussi utilisés pour bâtir des systèmes de vision par ordinateur, pour faire des prévisions sur les marchés monétaires, pour évaluer le risque financier ou en assurance, pour différents processus manufacturiers, pour le diagnostic médical, pour l'exploration pétrolière ou gazière, en robotique, en télécommunication, et j'en passe ! Bref, les réseaux de neurones ont aujourd'hui un impact considérable et, il y a fort à parier, que leur importance ira grandissant dans le futur.

### 3.1.5 Le modèle du neurone

#### - Neurone biologique

En biologie, un neurone est une cellule nerveuse dont la fonction est de transmettre un signal électrique dans certaines conditions. Il agit comme un relai entre une couche de neurones et celle qui la suit. Les caractéristiques des neurones sont encore mal connues (et font l'objet de recherches) mais on connaît leur principe d'action. Le corps d'un neurone est relié d'une part à un ensemble de dendrites (entrées du neurone) et d'autre part à un axone, partie étirée de la cellule, qui représentera pour nous sa sortie. Le neurone étudié est connecté aux neurones qui l'environnent : il reçoit au niveau de ses dendrites les signaux électriques des neurones "en amont", propagés par les axones de ces derniers. Les charges électriques s'accumulent dans le neurone jusqu'à dépasser un certain seuil : à ce moment la transmission du signal électrique se déclenche via son axone vers d'autres neurones "en aval".

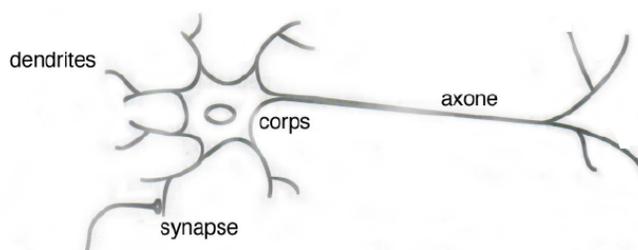


FIGURE 3.1 – **Le neurone biologique**

On remarque que les liaisons axone/dendrite entre deux neurones (connexions synaptiques) ne sont pas toutes de la même efficacité. Ainsi l'entrée associée à une certaine dendrite du neurone pourra avoir plus d'influence qu'une autre sur la valeur de sortie. On peut représenter la qualité de la liaison par un poids, sorte de coefficient s'appliquant au signal d'entrée. Le poids sera d'autant plus grand que la liaison est bonne. Un poids négatif aura tendance à inhiber une entrée, tandis qu'un poids positif viendra l'accentuer.

#### 3.1.6 Neurone formel .(modèle mathématique)

Le modèle mathématique d'un neurone artificiel est illustré à la (figure 3.2). Un neurone est essentiellement constitué d'un intégrateur qui effectue la somme pondérée de ses entrées.

Le résultat  $n$  de cette somme est ensuite transformée par une fonction de transfert  $f$  qui produit la sortie  $a$  du neurone.

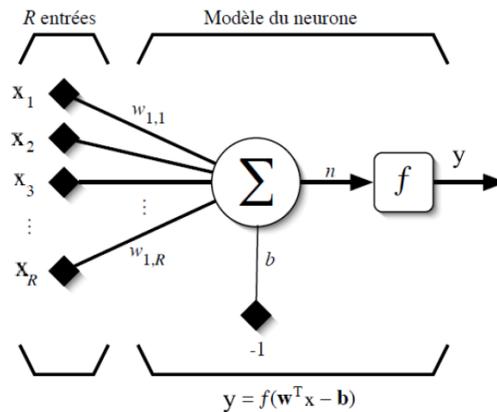


FIGURE 3.2 – **modèle mathématique**

En suivant les notations présentées à la section précédente, les  $R$  entrées du neurone correspondent au vecteur  $X = [X_1 X_2 \dots X_R]^T$ , alors que  $W = [w_{1,1} w_{1,2} \dots w_{1,R}]^T$  représente le vecteur des poids du neurone. La sortie  $n$  de l'intégrateur est donnée par l'équation suivante :

$$\begin{aligned} n &= \sum_{j=1}^R w_{1,j} X_j - b \\ &= w_{1,1} X_1 + w_{1,2} X_2 + \dots + w_{1,R} X_R - b, \end{aligned}$$

que l'on peut aussi écrire sous forme matricielle :

$$n = W^T X - b \quad (3.1)$$

Cette sortie correspond à une somme pondérée des poids et des entrées moins ce qu'on nomme le biais  $b$  du neurone. Le résultat  $n$  de la somme pondérée s'appelle le niveau d'activation du neurone. Le biais  $b$  s'appelle aussi le seuil d'activation du neurone. Lorsque le niveau d'activation atteint ou dépasse le seuil  $b$ , alors l'argument de  $f$  devient positif (ou nul). Sinon, il est négatif.

On peut faire un parallèle entre ce modèle mathématique et certaines informations que l'on connaît (ou que l'on croit connaître) à propos du neurone biologique. Ce dernier possède trois principales composantes :

les dendrites, le corps cellulaire et l'axone (voir [figure 3.1](#)). Les dendrites forment un maillage de récepteurs nerveux qui permettent d'acheminer vers le corps du neurone des signaux électriques en provenance d'autres neurones. Celui-ci agit comme un espèce d'intégrateur en accumulant des charges électriques. Lorsque le neurone devient suffisamment excité (lorsque la charge accumulée dépasse un certain seuil), par un processus électrochimique, il engendre un potentiel électrique qui se propage à travers son axone<sup>1</sup> pour éventuellement venir exciter d'autres neurones. Le point de contact entre l'axone d'un neurone et le dendrite d'un autre neurone s'appelle le synapse. Il semble que c'est l'arrangement spatial des neurones et de leur axone, ainsi que la qualité des connexions synaptiques individuelles qui détermine la fonction<sup>2</sup> précise d'un réseau de neurones biologique. C'est en se basant sur ces connaissances que le modèle mathématique décrit ci-dessus a été défini.

Un poids d'un neurone artificiel représente donc l'efficacité d'une connexion synaptique. Un poids négatif vient inhiber une entrée, alors qu'un poids positif vient l'accentuer. Il importe de retenir que ceci est une grossière approximation d'un véritable synapse qui résulte en fait d'un processus chimique très complexe et dépendant de nombreux facteurs extérieurs encore mal connus. Il faut bien comprendre que notre neurone artificiel est un modèle pragmatique qui, comme nous le verrons plus loin, nous permettra d'accomplir des tâches intéressantes. La vraisemblance biologique de ce modèle ne nous importe peu. Ce qui compte est le résultat que ce modèle nous permettra d'atteindre.

Un autre facteur limitatif dans le modèle que nous nous sommes donnés concerne son caractère discret. En effet, pour pouvoir simuler un réseau de neurones, nous allons rendre le temps discret dans nos équations. Autrement dit, nous allons supposer que tous les neurones sont synchrones, c'est-à-dire qu'à chaque temps  $t$ , ils vont simultanément calculer leur somme pondérée et produire une sortie  $a(t) = f(n(t))$ . Dans les réseaux biologiques, tous les neurones sont en fait asynchrones.

---

1. Un axone peut être plus ou moins long selon le type de neurone.

2. Notez bien, cependant, que des théories récentes remettent en cause cette hypothèse. Mais ceci sort du cadre du cours!

Revenons donc à notre modèle tel que formulé par l'équation (3.1) et ajoutons la fonction d'activation  $f$  pour obtenir la sortie du neurone :

$$Y = f(n) = f(W^T X - b). \quad (3.2)$$

En remplaçant  $W^T$  par une matrice  $\mathbf{W} = W^T$  d'une seule ligne, on obtient une

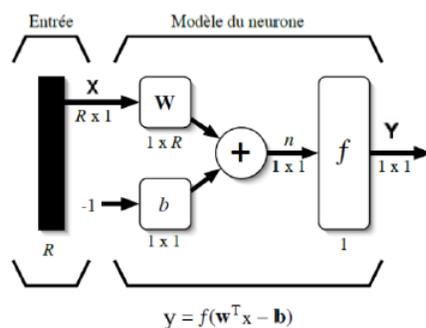


FIGURE 3.3 – Représentation matricielle du modèle d'un neurone artificiel.

forme générale que nous adopterons tout au long de cet ouvrage :

$$Y = f(\mathbf{W}X - b). \quad (3.3)$$

L'équation (3.3) nous amène à introduire un schéma de notre modèle plus compact que celui de la figure (3.2). La figure (3.3) illustre celui-ci. On  $K$  représente les  $R$  entrées comme un rectangle noir (le nombre d'entrées est indiqué sous le rectangle). De ce rectangle sort le vecteur  $\mathbf{X}$  dont la dimension matricielle est  $R \times 1$ . Ce vecteur est multiplié par une matrice  $\mathbf{W}$  qui contient les poids (synaptiques) du neurones. Dans le cas d'un neurone simple, cette matrice possède la dimension  $1 \times R$ . Le résultat de la multiplication correspond au niveau d'activation qui est ensuite comparé au seuil  $b$  (un scalaire) par soustraction. Finalement, la sortie du neurone est calculée par la fonction d'activation  $f$ . La sortie d'un neurone est toujours un scalaire.

### 3.1.7 Fonctions de transfert

Jusqu'à présent, nous n'avons pas spécifié la nature de la fonction d'activation de notre modèle. Il se trouve que plusieurs possibilités existent. Différentes fonctions de transfert pouvant être utilisées comme fonction d'activation du neurone sont énumérées au tableau (3.1).

Les trois les plus utilisées sont les fonctions " seuil " (en anglais "hard limit"), "linéaire" et "sigmoïde".

TABLE 3.1 – Fonctions de transfert  $Y = a = f(n)$ .

Nom de la fonction	Relation d'entrée/sortie	Icône	Nom Matlab
seuil	$a = 0$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlim
seuil symétrique	$a = -1$ si $n < 0$ $a = 1$ si $n \geq 0$		hardlims
linéaire	$a = n$		purelin
linéaire saturée	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$		satlin
linéaire saturée symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$		satlins
linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$		poslin
sigmoïde	$a = \frac{1}{1+\exp^{-n}}$		logsig
tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
compétitive	$a = 1$ si $n$ maximum $a = 0$ autrement		compet

Comme son nom l'indique, la fonction seuil applique un seuil sur son entrée. Plus précisément, une entrée négative ne passe pas le seuil, la fonction retourne alors la valeur 0 (on peut interpréter ce 0 comme signifiant faux), alors qu'une entrée positive ou nulle d'épasse le seuil, et la fonction retourne 1 (vrai). Utilisée dans le contexte d'un neurone, cette fonction est illustrée à la [figure \(3.4\(a\)\)](#). On remarque alors que le biais  $b$  dans l'expression de  $a = \text{hardlim}(W^T X - b)$  (équation (3.3)) détermine l'emplacement du seuil sur l'axe  $W^T X$ , où la fonction passe de 0 à 1. Nous verrons plus loin que cette fonction permet de prendre des décisions binaires. La fonction linéaire est très simple, elle affecte directement son entrée à sa sortie :

$$Y = a = n \quad (3.4)$$

Appliquée dans le contexte d'un neurone, cette fonction est illustrée à la [figure \(3.4\(b\)\)](#). Dans ce cas, la sortie du neurone correspond à son niveau d'activation dont le passage à zéro se produit lorsque  $W^T X = b$ . La fonction de transfert sigmoïde est quant à elle illustrée à la [figure \(3.4\(c\)\)](#).

Son équation est donnée par :

$$Y = a = \frac{1}{1 + \exp^{-n}}. \quad (3.5)$$

Elle ressemble soit à la fonction seuil, soit à la fonction linéaire, selon que l'on est loin ou près de  $b$ , respectivement. La fonction seuil est très non-linéaire car il y a une discontinuité lorsque  $W^T X = b$ . De son côté, la fonction linéaire est tout à fait linéaire. Elle ne comporte aucun changement de pente. La sigmoïde est un compromis intéressant entre les deux précédentes. Notons finalement, que la fonction "tangente hyperbolique" est une version symétrique de la sigmoïde.

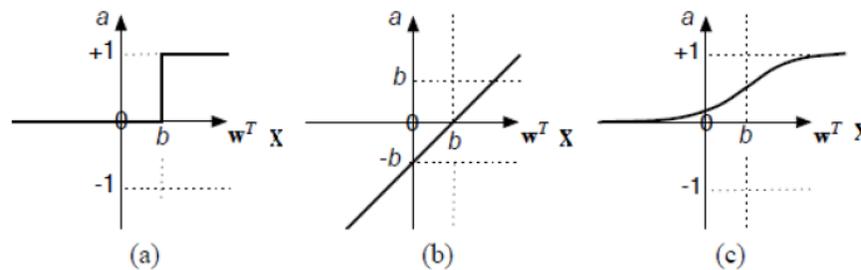


FIGURE 3.4 – **Fonction de transfert : (a) du neurone "seuil" ; (b) du neurone "linéaire", et (c) du neurone "sigmoïde".**

### 3.1.8 Architectures des réseaux de neurones

On distingue deux structures de réseau, en fonction du graphe de leurs connexions, c'est-à-dire du graphe dont les nœuds sont les neurones et les arêtes les "connexions" entre ceux-ci :

- Les réseaux de neurones statiques (ou acycliques, ou non bouclés).
- Les réseaux de neurones dynamiques (ou récurrents, ou bouclés).

### 3.1.9 Réseaux de neurones non bouclés

Un réseau de neurones non bouclé définit une fonction statique non linéaire. C'est-à-dire que les sorties du réseau à un instant  $t$  donné ne dépendent que des entrées appliquées au même instant.

Le comportement d'un réseau de neurones statique peut alors être décrit par la représentation suivante :

$$Y(t) = h(X(t), W) \quad (3.6)$$

Où  $X(t)$  est le vecteur des entrées,  $Y(t)$  le vecteur de sortie,  $h$  représente la fonction non linéaire réalisée par le réseau et  $W$  est la matrice des poids des connexions.

### - les réseaux simples à une seule couche

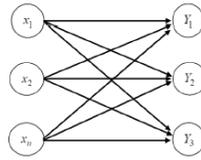


FIGURE 3.5 – **Un réseau de neurones a une seule couche .**

### - Réseau de neurones multicouche.

La **figure (3.6)** présente un réseau de structure particulière, très couramment utilisé, appelé réseau en couches ou perceptron multicouche. On trouve les entrées du réseau, une ou plusieurs couches de neurones intermédiaires, dits neurones cachés et des neurones de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux, ils rendent possible la résolution de problèmes non séparables linéairement. La fonction d'activation des cellules de la couche de sortie est généralement linéaire.

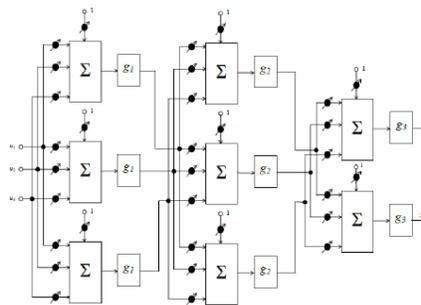


FIGURE 3.6 – **Réseau de neurones multicouche.**

Les réseaux non bouclés sont généralement utilisés pour l'approximation de fonctions non linéaires, c'est le cas de notre application en fiabilité.

### 3.1.10 Réseaux de neurones bouclés

Un réseau de neurones bouclé à temps discret définit une, ou plusieurs, équations aux différences non linéaires, par composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions. En d'autres termes, les sorties du réseau à un instant donné dépendent non seulement des entrées appliquées au même instant mais aussi des entrées et des sorties passées. De tels réseaux sont alors être utilisés pour la modélisation de systèmes dynamiques non linéaires.

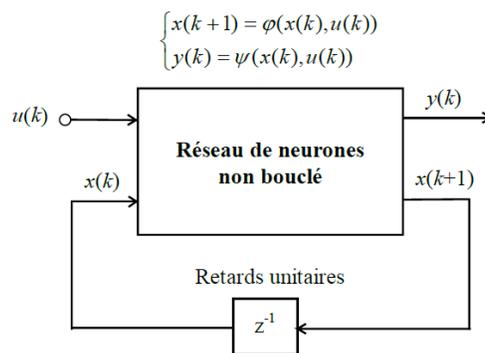


FIGURE 3.7 – Réseau de neurones multicouche.

On montre que tout réseau de neurones bouclé, aussi complexe soit-il, peut toujours être mis sous la forme canonique de la figure (3.7), c'est à dire, comportant un réseau de neurones non bouclé, dont certaines sorties, appelées variables d'état, sont ramenées aux entrées par des bouclages unitaires.

### 3.1.11 L'apprentissage

### 3.1.12 Définition

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

L'apprentissage d'un réseau de neurones peut être défini comme la phase durant laquelle les divers paramètres le caractérisant sont remis à jour jusqu'à ce qu'ils permettent au réseau d'approximer au mieux la fonction qu'il a à réaliser. Selon l'application dans laquelle le réseau va être intégré, la fonction à approcher peut être connue ou inconnue analytiquement.

En effet, le phénomène physique, chimique, économique, biologique, etc. qui nécessite l'emploi d'un réseau de neurones peut être parfaitement modélisé par l'intermédiaire d'équations analytiques ou simplement observables, mettant en jeu un nombre fini de valeurs expérimentales de sorties fonction de valeurs d'entrées bien choisies, sans que les relations qui les relient ne soient connues.

### 3.1.13 Type d'apprentissage

On distingue deux grandes classes d'algorithmes d'apprentissage :

#### - Apprentissage supervisé :

Un apprentissage est dit supervisé lorsque l'on force le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif. Ce genre d'apprentissage est réalisé à l'aide d'une base d'apprentissage, constituée de plusieurs exemples de type entrées-sorties (les entrées du réseau et les sorties désirées ou encore les solutions souhaitées pour l'ensemble des sorties du réseau). La procédure usuelle dans le cadre de la prévision est l'apprentissage supervisé (ou à partir d'exemples) qui consiste à associer une réponse spécifique désirée à chaque signal d'entrée. La modification des poids s'effectue progressivement jusqu'à ce que l'erreur (ou l'écart) entre les sorties du réseau (ou résultats calculés) et les résultats désirés soient minimisés. **Figure (3.8)**

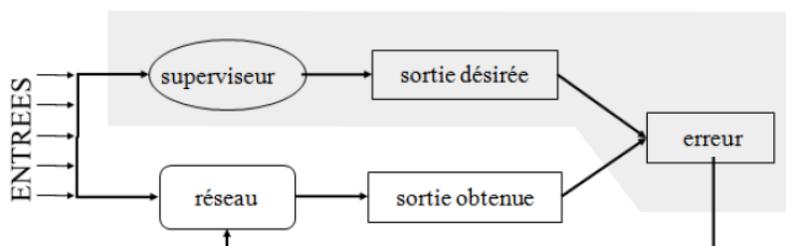


FIGURE 3.8 – Apprentissage supervisé.

#### - Apprentissage non-supervisé :

Ce type d'apprentissage est choisi lorsqu'il n'y pas de connaissances à priori des sorties désirés pour des entrées données. En fait, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente des entrées. **Figure (3.9)**

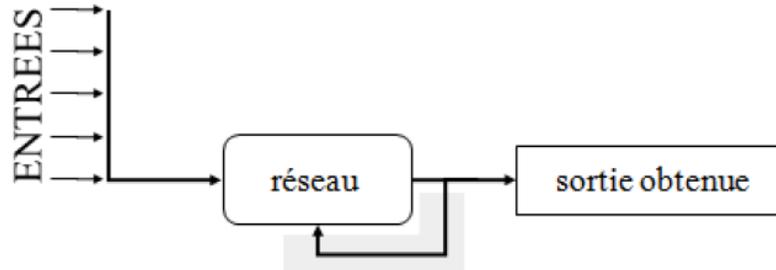
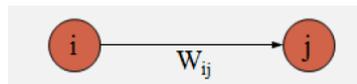


FIGURE 3.9 – Apprentissage non-supervisé.

### 3.1.14 Règles d'apprentissage :

L'apprentissage consiste à modifier le poids des connexions entre les neurones.



Il existe plusieurs règles de modification :

- Loi de Hebb :  $\Delta W_{ij} = R a_i a_j$
- Règle de Widrow-Hoff (delta rule) :  $\Delta W_{ij} = R(d_i - a_i) a_j$
- Règle de Grossberg :  $\Delta W_{ij} = R(a_j - W_{ij}) a_i$

- **Loi de Hebb** : Si deux unités connectées sont actives simultanément, le poids de leur connexion est augmenté ou diminué.  $R$  est une constante positive qui représente la force d'apprentissage (learning rate).

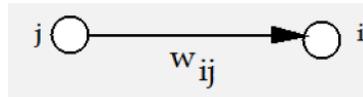


	$a_i = -1$	$a_i = 1$
$a_j = -1$	$\Delta W_{ij} = R$	$\Delta W_{ij} = -R$
$a_j = 1$	$\Delta W_{ij} = -R$	$\Delta W_{ij} = R$

$$\Delta W_{ij} = R a_i a_j$$

**- Loi de Widrow-Hoff (delta rule) :**

ai activation produite par le réseau di réponse désirée par l'expert humain Par exemple si la sortie est inférieure à la réponse désirée, il va falloir augmenter le poids de la connexion à condition bien sûr que l'unité j soit excitatrice (égale à 1). On est dans l'hypothèse d'unités booléennes  $\{0, 1\}$ .



$$\begin{array}{lll} a_i = 0 & & a_i = 1 \\ d_i = 0 & \Delta W_{ij} = 0 & \Delta W_{ij} = -R \\ d_i = 1 & \Delta W_{ij} = R & \Delta W_{ij} = 0 \end{array}$$

$$\Delta W_{ij} = R(d_i - a_i)a_j$$

**- Loi de Grossberg :**

On augmente les poids qui entrent sur l'unité gagnante  $a_i$  s'ils sont trop faibles, pour les rapprocher du vecteur d'entrée  $a_j$ . C'est la règle d'apprentissage utilisée dans les cartes auto-organisatrices de Kohonen [2].

$$\Delta W_{ij} = R(a_j - W_{ij})a_i$$

**3.1.15 Les étapes de la conception d'un réseau de neurone**

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées. Leurs mises en oeuvre nécessitent :

**3.1.16 Détermination des entrées/sorties du réseau de neurones**

Pour toute conception de modèle, la sélection des entrées doit prendre en compte deux points essentiels :

- Premièrement, la dimension intrinsèque du vecteur des entrées doit être aussi petite que possible, en d'autres termes, la représentation des entrées doit être la plus compacte possible, tout en conservant pour l'essentiel la même quantité d'information, et en gardant à l'esprit que les différentes entrées doivent être indépendantes.

- En second lieu, toutes les informations présentées dans les entrées doivent être pertinentes pour la grandeur que l'on cherche à modéliser : elles doivent donc avoir une influence réelle sur la valeur de la sortie.

### 3.1.17 Choix et préparation des échantillons

Le processus d'élaboration d'un réseau de neurones commence toujours par le choix et la préparation des échantillons de données. La façon dont se présente l'échantillon conditionne le type de réseau, le nombre de cellules d'entrée, le nombre de cellules de sortie et la façon dont il faudra mener l'apprentissage, les tests et la validation. Il faut donc déterminer les grandeurs qui ont une influence significative sur le phénomène que l'on cherche à modéliser. Lorsque la grandeur que l'on veut modéliser dépend de nombreux facteurs, c'est-à-dire lorsque le modèle possède de nombreuses entrées, il n'est pas possible de réaliser un "pavage" régulier dans tout le domaine de variation des entrées : il faut donc trouver une méthode permettant de réaliser uniquement des expériences qui apportent une information significative pour l'apprentissage du modèle. Cet objectif peut être obtenu en mettant en oeuvre un plan d'expériences. Pour les modèles linéaires, l'élaboration de plans d'expériences est bien maîtrisée, par ailleurs, ce n'est pas le cas pour les modèles non linéaires.

Afin de développer une application à base de réseaux de neurones, il est nécessaire de disposer de deux bases de données, une pour effectuer l'apprentissage et l'autre pour tester le réseau obtenu et déterminer ses performances.

### 3.1.18 Elaboration de la structure du réseau

La structure du réseau dépend étroitement du type des échantillons. Il faut d'abord choisir le type de réseau : un perceptron standard, un réseau de Hopfield, un réseau à décalage temporel (TDNN), un réseau de Kohonen, etc...

Dans le cas du perceptron multicouches, il faudra aussi bien choisir le nombre de couches cachées que le nombre de neurones dans cette couche.

- **Nombre de couches cachées** : Mis à part les couches d'entrée et de sortie, il faut décider du nombre de couches intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation.

- **Nombre de neurones cachés** : Chaque neurone peut prendre en compte des profils spécifiques de neurones d'entrée. Un nombre plus important permet donc de mieux "coller" aux données présentées mais diminue la capacité de généralisation du réseau. Il faut alors trouver le nombre adéquat de neurones cachés nécessaire pour obtenir une approximation satisfaisante.

### 3.1.19 Apprentissage

L'apprentissage est un problème numérique d'optimisation. Il consiste à calculer les pondérations optimales des différentes liaisons, en utilisant un échantillon. La méthode la plus utilisée est la rétropropagation, qui est généralement plus économe que les autres en terme de nombres d'opérations arithmétiques à effectuer pour évaluer le gradient.

### 3.1.20 Validation et Tests

Alors que les tests concernent la vérification des performances d'un réseau de neurones hors échantillon et sa capacité de généralisation, la validation est parfois utilisée lors de l'apprentissage. Une fois le réseau de neurones développé, des tests s'imposent afin de vérifier la qualité des prévisions du modèle neuronal. Cette dernière étape doit permettre d'estimer la qualité du réseau obtenu en lui présentant des exemples qui ne font pas partie de l'ensemble d'apprentissage. Une validation rigoureuse du modèle développé se traduit par une proportion importante de prédictions exactes sur l'ensemble de la validation. Si les performances du réseau ne sont pas satisfaisantes, il faudra, soit modifier l'architecture du réseau, soit modifier la base d'apprentissage.

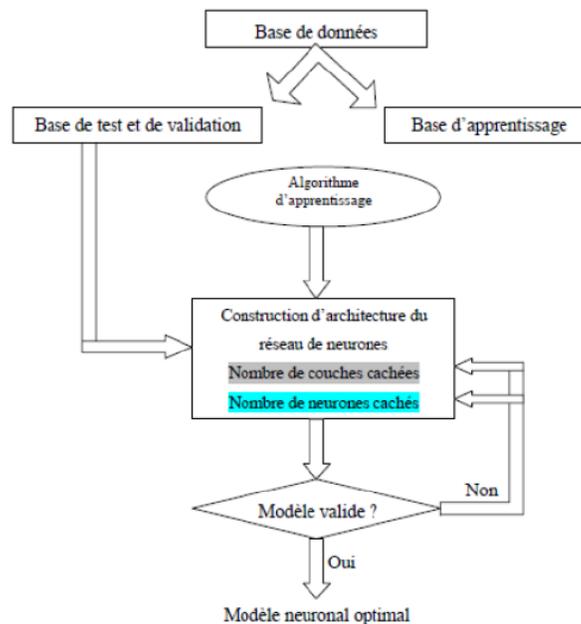
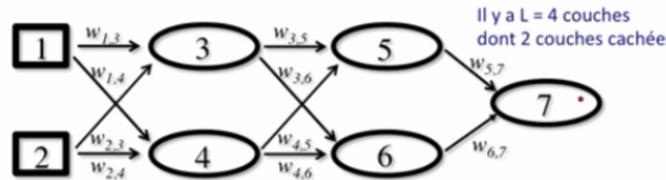


FIGURE 3.10 – Organigramme de conception d'un réseau de neurones.

### 3.1.21 Réseau de neurones à L couches



- On note  $a_j$  l'activité du K neurone, incluant les neurones d'entrée et de sortie. Donc on aura  $a_i = x_i$
- On note  $in_j$  l'activité du K neurone avant la non-linéarité logistique, c'est à dire

$$a_j = \text{logistic}(in_j) = \text{logistic}(\sum_i w_{i,j} a_i)$$

## 3.2 Les Algorithmes Génétiques

### 3.2.1 Introduction

Les principes de base des algorithmes évolutionnaires (en court AE), dont les plus connus sont les algorithmes génétiques, s'inspirent de l'observation de phénomènes biologiques, plus précisément de la capacité de populations d'organismes vivants, à s'adapter à leur environnement à l'aide de mécanismes de sélection et d'héritage génétique. En d'autres termes, ces algorithmes évolutionnaires représentent une version artificielle, informatique, de la théorie de l'évolution selon Darwin. Les algorithmes génétiques (AG) sont des algorithmes d'optimisation stochastique fondés sur les mécanismes de la sélection naturelle et de la génétique. Ils ont été initialement développés par John Holland (1975). Leurs champs d'application sont très vastes. Outre l'économie, ils sont utilisés pour l'optimisation de fonctions ,Le volume de preuves requises a également augmenté(2000)[16] en programmation génétique<sup>3</sup> (Koza (1992))[33], pour le contrôle de pipeline(Goldberg (1981)), en théorie du contrôle optimal (Krishnakumar et Goldberg (1992) et plus récemment Jamshidi et al. (2003)), ou encore en théorie des jeux répétés (Axelrod (1987)) et dynamiques (Özyildirim (1996, 1997) et Özyildirim et Alemdar (1998)).

3. AG : population de solutions, PG : population de programmes.

Les raisons de ce grand nombre d'applications sont la simplicité et l'efficacité de ces algorithmes. Les AG s'inspirent des mécanismes de l'évolution biologique (voir la Figure(3.11)) pour les transposer à la recherche de solutions adaptées au problème qu'on cherche à résoudre.

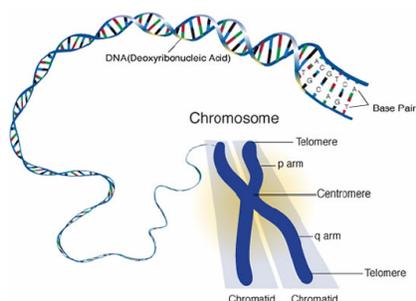


FIGURE 3.11 – **Chromosome**

Les génotypes qui sont les mieux adaptés à leur environnement ont une plus grande facilité à se reproduire et la reproduction sexuée assure le croisement des gènes les plus performants dans la population. Les erreurs de copies de chromosomes pendant la reproduction introduisent par ailleurs de la nouveauté sous la forme de mutations au niveau des gènes et de leurs conséquences au niveau du phénotype correspondant. La sélection continue, dans la population, des génotypes avec la performance relative la plus élevée conduit à une amélioration de l'adaptation de la population à son environnement (cela correspond alors à une augmentation de la performance moyenne de la population)[36].

### 3.2.2 Présentation des algorithmes génétiques (AG)

Un algorithme génétique implémente une version très simplifiée et très schématique des mécanismes de l'évolution biologique(principes de l'évolution « darwinienne »). A partir d'un problème qu'on cherche à résoudre, un algorithme génétique est défini, selon Lerman et Ngouenet (1995) , par la donnée des quatre éléments de base suivants :

1. *Individu/chromosome/séquence* : une solution potentielle du problème qui correspond à une valeur codée de la variable (ou des variables) en considération ;
2. *Population* : un ensemble de chromosomes ou de points de l'espace de recherche (donc des valeurs codées des variables) ;

3. *Environnement* : l'espace de recherche (caractérisé en termes de performance correspondant à chaque individu possible) ;
4. *Fonction de performance* : la fonction - positive - que nous cherchons à maximiser car elle représente l'adaptation de l'individu à son environnement.

### 3.2.3 Principe

Un algorithme génétique permet de trouver une solution en cherchant un extremum sur un ensemble de solutions donnés. Cet algorithme est construit selon :

- Le codage des éléments de la population. C'est à ce moment que l'on détermine les structures de données utiles. Bien entendu cette étape suit la modélisation du problème posé. La réussite des algorithmes génétiques est issue de la qualité de leur codage. Au début, un codage binaire était très utilisé mais maintenant, on travaille avec des codages réels plus pratiques pour des problèmes à variables réelles comme les problèmes d'optimisation.
- La génération de la population initiale. Ce mécanisme doit produire une population non homogène qui évoluera pour permettre de trouver une solution. La convergence ou la divergence peut être plus ou moins rapide en fonction de la qualité de cette population initiale.
- Une fonction qui retourne le fitness du brin d'ADN pour pouvoir le comparer.
- Des outils qui permettront de créer de l'aléatoire comme la mutation afin de garantir l'exploration des solutions.
- Des paramètres afin de stopper le programme comme un critère de convergence ou divergence ou encore un nombre d'itérations pour effectuer le programme.

Les algorithmes évolutionnaires constituent une approche originale : on ne cherche pas à trouver une solution exacte mais d'obtenir un résultat satisfaisant, le plus possible, les critères, souvent contradictoires, des problèmes. Les solutions trouvées ne sont peut être pas optimales mais sont souvent meilleures que les solutions obtenues grâce à des algorithmes plus conventionnels qui donneront, pour un temps identique, un résultat plus éloigné des besoins.[1]

Afin d'expliciter le fonctionnement des algorithmes génétiques, nous présentons dans la suite de cette section les différentes étapes d'un algorithme génétique simple illustrées sur la ( Figure 3.12).

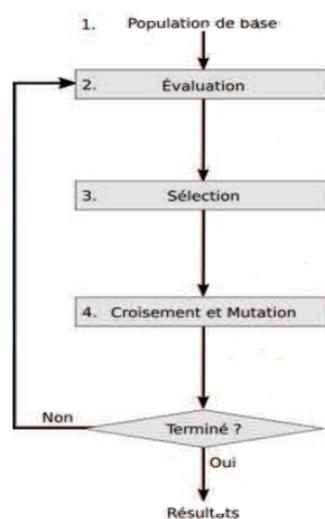


FIGURE 3.12 – Les principales étapes d'un algorithme génétique

### 3.2.4 Fonctionnement général des algorithmes génétiques

Les algorithmes génétiques sont employés afin de répondre aux problèmes n'ayant pas de solutions calculables en temps raisonnable de façon analytique ou algorithmique.

L'avantage des algorithmes génétiques, est qu'il n'est pas nécessaire de connaître la façon dont le problème est résolu. Il suffit juste de fournir une fonction ayant des paramètres permettant de coder l'information sous forme de gène et une autre qui permet d'évaluer la pertinence de la solution.

En conséquence les algorithmes génétiques sont efficaces dans divers domaines :

- problèmes de « magie noire », tels que la synthèse automatisée des circuits électriques analogues, des contrôleurs, des antennes, des réseaux des réactions chimiques, et d'autres secteurs de conception ;
- Reconnaissance d'images (Robinson et al.), classification d'images (Zao), traitement d'images satellites (Daïda), ...

- Prédiction de séries temporelles (Lee) , génération d'arbres de décisions (Koza), data-mining (Freitas), ...
- Classification de segments d 'ADN (Handley), de protéines (Koza et al.), ... .

Les critères d'évaluation pour tester la convergence sont divers :

- Taux minimum désiré pour l'adaptation de la population au problème
- Temps de calcul à ne pas dépasser
- Combinaison des points précédents

### 3.2.5 Le codage

Chaque paramètre d'une solution est assimilé à un gène, toutes les valeurs qu'il peut prendre sont les allèles de ce gène, on doit trouver une manière de coder chaque allèle différent de façon unique. Pour continuer avec les parallèles entre le codage et le domaine de la génétique, on parle de chromosome pour une suite de gène. On peut regrouper dans un chromosome des paramètres similaires. Dans ce cas, on identifiera chaque gène par sa position : son locus sur le chromosome. Au final chaque individu est représenté par un ensemble de chromosomes, et une population est un ensemble d'individus. On notera qu'il existe 3 principaux types de codages utilisables, et on peut passer de l'un à l'autre relativement facilement :

- Le codage binaire : les gènes sont représentés par des entiers longs
- Le codage réel
- Le codage de Gray <sup>4</sup> : les gènes voisins n'ont qu'un seul bit de différence.

1. **Initialisation** (Figure 3.13 -étape (a) pour  $t = 0$ ). Une population initiale de chromosomes est tirée aléatoirement.
2. **Évaluation** (premier élément de l'étape (b)). Chaque chromosome est décodé, puis évalué.
3. **Sélection** (les derniers éléments de l'étape (b) et (c)). Création d'une nouvelle population de  $N$  chromosomes par l'utilisation d'une méthode de sélection appropriée.

---

4. Le code de Gray :est un type de codage binaire,permettant de ne modifier qu'un seul bit à la fois quand un nombre est augmenté d'une unité Le nom du code vient de l'ingénieur américain Frank Gray qui déposa un brevet sur ce code

4. **Reproduction.** Possibilité de croisement (étape **(d)**) et de mutation (étape **(e)**) au sein de la nouvelle population.
5. **Retour** à la phase d'évaluation (étape **(f)**) tant que la condition d'arrêt du problème n'est pas satisfaite.

## - Codage et population initiale

Premièrement, il faut représenter les différents états possibles de la variable dont on cherche la valeur optimale sous une forme utilisable par un AG : c'est le codage. Cela permet d'établir une connexion entre les valeurs de la variable et les individus de la population de manière à imiter la connexion qui existe en biologie entre *le génotype et le phénotype*<sup>5</sup>

Il existe principalement deux types de codage : le codage binaire (représentation sous forme de chaîne binaire) et le codage réel (représentation directe des valeurs réelles de la variable). Le codage initialement retenu par John Holland est le codage binaire. Nous allons par conséquent d'abord présenter ce cas.

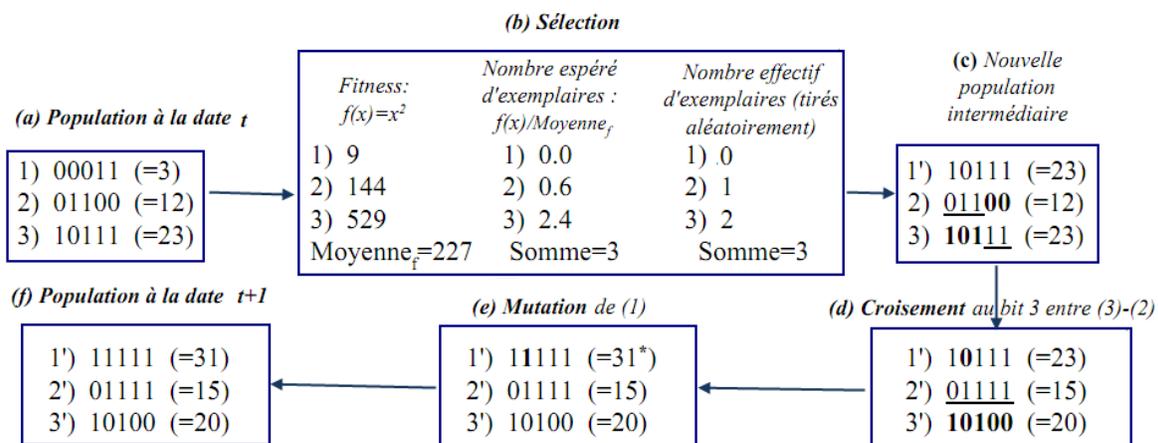


FIGURE 3.13 – Exemple de Présentation d'AG

5. le génotype pour ce qui concerne la représentation binaire d'un individu, et de phénotype pour ce qui est de sa valeur réelle correspondante dans l'espace de recherche (i.e. une solution potentielle du problème).

"Une application des AG à l'optimisation de la fonction  $f(x)=x^2$  sur le domaine formé par des entiers entre 0 et 31. Les entiers sont codés avec une chaîne binaire de cinq bits : 00001=1, 11111=31 . L'exemple utilise une population initiale aléatoire de trois individus (chromosomes) et l'AG construit une nouvelle population à travers une reproduction sélective (**étape (b)** ), des croisements (combinaison de solutions existantes - **étape (d)**) et des mutations (expériences aléatoire introduction de la nouveauté - **étape (e)** ). L'AG atteint en une étape la solution optimale ( $x=31$ ) dans cet exemple "

**Définition 3.2.1.** *Séquence/Chromosome/Individu (Codage binaire)*

*Nous appelons une séquence (chaîne, chromosome, individu)  $A$  de longueur  $l(A)$  une séquence*

$$A = \{a_1, a_2, \dots, a_l\} \text{ avec } \forall i \in \{1, \dots, l\}, a_i \in \{0, 1\}$$

Un chromosome est donc une suite de bits (formée de zéros et de uns), appelé aussi chaîne binaire. Dans le cas d'un codage non binaire, tel que le codage réel, la séquence ne contient qu'un point, nous avons  $A = \{a\}$  avec  $a \in \mathbb{R}$

Une chaîne binaire de longueur  $l$  peut représenter  $2^l$  états de la variable. L'exemple de la [Figure \(3.13\)](#) suppose  $2^5 = 32$  états de la variable  $x$ . Une fonction de décodage  $d$  doit permettre le passage de la représentation binaire(cas particulier) vers la représentation en termes d'états de la variable initiale.  $d : \{0, 1\}^l \rightarrow \mathbb{N}$   
Le décodage le plus souvent retenu est un simple changement de base. Ainsi la chaîne  $A = \{0, 0, 0, 1, 1\}$  peut être décodée de manière à donner

$$0 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 3.$$

Plus générale, toute chaîne binaire  $A$  peut donc être décodée en une valeur entière  $x$  selon la règle :

$$x = d(A) = \sum_{i=1}^l a_i 2^{l-i} \quad (3.7)$$

Cette approche peut être généralisée pour représenter les états d'une variable réelle. Supposons que nous cherchons à maximiser  $f$  en fonction d'une variable réelle  $x$  . Soit  $D = [x_{min}, x_{max}]$ , avec  $D \subset \mathbb{R}$ . Soit  $prec$  la précision (chiffre après la virgule) avec laquelle nous cherchons  $x$  . Soit  $ld = x_{max} - x_{min}$  la longueur de l'intervalle  $D$ . Nous devons alors diviser cet intervalle en  $n_i = ld \times 10^{prec}$  sous-intervalles égaux afin de respecter la précision.

Soit  $D = [-1, 2]$ , nous avons donc  $ld = 3$  si nous voulions une précision  $prec = 6$ , alors il nous faut diviser cet intervalle en  $n_i = 3000000$  sous-intervalles. Si  $s$  est l'entier naturel le plus petit tel que  $2^s > n_i$  la chaîne binaire de longueur  $l = s$  peut alors représenter tous les états de  $x$  correspondant à cette précision (dans notre exemple,  $s = 22$  car  $2^{21} = 2097152 < 3000000 < 2^{22}$ ). La transformation d'une chaîne binaire  $A = \{a_1, a_2, \dots, a_l\}$  en un nombre réel  $x$  peut alors s'exécuter en deux étapes :

1. conversion (de base 2 en base 10) :

$$\hat{x} = \sum_{i=1}^s a_i 2^{s-i}$$

2. recherche du nombre réel correspondant :

$$x = x_{min} + \hat{x} \frac{x_{max} - x_{min}}{2^s - 1} \quad \text{ou, } x = x_{min} + \frac{ld}{2^s - 1} \sum_{i=1}^s a_i 2^{s-i}$$

**Exemple 3.2.1.** Soit  $A = \underbrace{\{1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1\}}_{s=22}$ .

Alors  $\hat{x} = \sum_{i=1}^{s=22} a_i 2^{s-i} = 1 \times 2^{21} + 1 \times 2^{20} + 0 \times 2^{19} + \dots + 1 \times 2^0 = 3311359$

$$\text{et } x = -1 + 3311359 \frac{2 - (-1)}{2^{22} - 1} = 1.368469 \in [-1, 2].$$

L'AG démarre alors avec une population composée par un tirage aléatoire de  $N$  individus (chaînes, chromosomes) dans le codage retenu. Avec un codage binaire, selon la taille  $l$  de la chaîne, on effectue alors pour chaque individu  $l$  tirages dans  $\{0, 1\}$  selon une loi de probabilité (*la loi uniforme est couramment retenue*). La performance de chacun de ces individus doit alors être évaluée.

### 3.2.6 La fonction d'évaluation

La fonction de performance -qu'on appelle aussi fonction d'adaptation, fonction objectif ou fonction fitness -associe une valeur de performance à chaque individu ce qui offre la possibilité de le comparer à d'autres individus et permet à l'algorithme génétique de déterminer qu'un individu sera sélectionné pour être reproduit ou pour déterminer s'il sera remplacé .

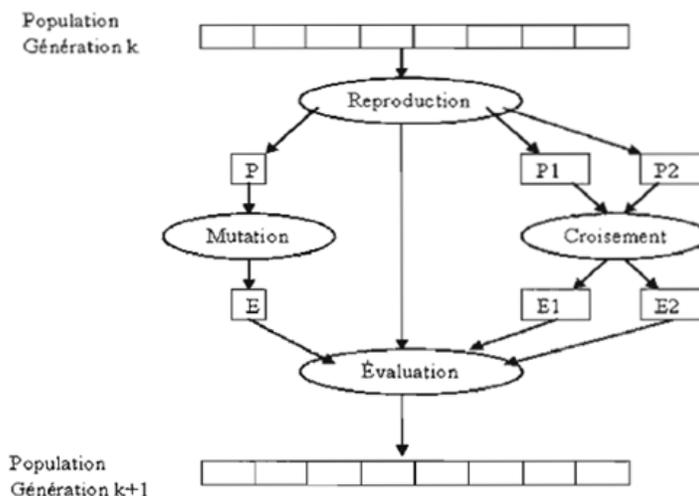


FIGURE 3.14 – Principe général de l'évolution d'une population d'un algorithme génétique

De manière plus générale :

**Définition 3.2.2.** *Performance -fitness- d'une séquence* Nous représentons la performance d'une séquence  $A$  par une fonction à valeurs **positives**,  $f : A \rightarrow f(A) \in \mathbb{R}^+$

Dans le cas du codage binaire, la fonction de performance doit affecter une valeur positive au codage correspondant (phénotype) à chaque chaîne binaire (génotype) :  $\forall A \in \{0, 1\}^l, f(d(A)) \in \mathbb{R}^+$ . Le but d'un algorithme génétique est alors simplement de trouver la chaîne qui maximise cette fonction  $f$ .

Une fois que la performance de chaque individu dans la population actuelle est évaluée, les mécanismes évolutionnaires entrent en jeu pour procéder à l'exploration de l'espace des valeurs. Les opérateurs de l'AG cherchent à imiter ces mécanismes.

### 3.2.7 Les opérateurs

Trois opérateurs jouent un rôle prépondérant dans la possible réussite d'un AG : l'opérateur de **sélection**, l'opérateur de **croisement** et l'opérateur de **mutation**. Si le principe de chacun de ces opérateurs est facilement compréhensible, il est toutefois difficile d'expliquer l'importance isolée de chacun de ces opérateurs dans la réussite de l'AG. Pour partie, cela tient au fait que chacun de ces opérateurs agit selon divers critères qui lui sont propres (valeur sélective des individus, probabilité d'activation de l'opérateur, etc.).

### - Opérateur de sélection

Cet opérateur permet de sélectionner les individus de  $P$  qui vont être sauvegardés dans la nouvelle famille composée  $P'$  et vont servir de parents (application de l'opérateur de croisement). On sélectionne toujours autant d'individus qu'on en supprime.

Cet opérateur est très important puisqu'il gère la population en permettant à certains éléments de rester ou en supprime d'autres. Il est préférable que la survie d'un individu soit en accord avec son efficacité à résoudre le problème. On trouve le plus souvent quatre types de méthodes de sélection différentes :

#### - La sélection par roulette

La sélection de la roulette ou wheel se base sur le principe de la roue de la fortune. On associe à chaque individu un quartier de disque dont l'angle de secteur est proportionnel à son fitness. Tout comme le jeu dont la méthode s'inspire, une "bille" est lancée et le chromosome sur lequel elle s'arrête sera sélectionné.

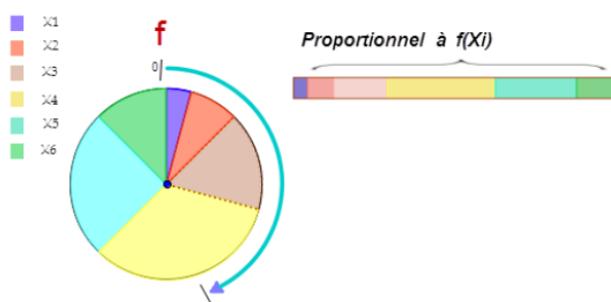


FIGURE 3.15 – Principe de la roulette

#### - L'élitisme

La méthode "élitiste" : cette méthode sélectionne les  $n$  individus dont on a besoin pour la nouvelle génération  $P'$  en prenant les  $n$  meilleurs individus de la population  $P$ .

Cette méthode de sélection permet de mettre en avant les meilleurs individus de la population. Ce sont donc les individus les plus prometteurs qui vont participer à l'amélioration de notre population.

Cette méthode a l'avantage de permettre une convergence (plus) rapide des solutions, mais au détriment de la diversité des individus. On prend en effet le risque d'écartier des individus de piètre qualité, mais qui auraient pu apporter de quoi créer de très bonnes solutions dans les générations suivantes.

#### - La sélection par tournois

Cette technique utilise la sélection proportionnelle sur des paires d'individus, puis choisit parmi ces paires l'individu qui a le meilleur score d'adaptation. Cette méthode sélectionne deux individus de  $P$  et on compare leur taux de satisfaction. Le plus intéressant, est sélectionné avec une probabilité  $p$  comprise entre 0,5 et 1. Cette méthode est très intéressante car elle présente une variance très élevée (Miller et Goldberg, 1995) [3].

Les individus sélectionnés, proportionnellement à leur fitness ou par tournoi, sont ensuite recombinaisonnés pour générer de nouveaux individus (enfants).

#### - La sélection universelle stochastique

Cette méthode peut être vue comme le fait de prendre un segment découpé en autant de sous-segments qu'il y a d'individus. Les individus sélectionnés sont désignés par un ensemble de points équidistants. Mais cette méthode a une variance très faible et est donc très peu utilisée .

#### - Opérateur de croisement

L'opérateur de croisement permet la création de nouveaux individus selon un processus fort simple. Il permet donc l'échange d'information entre les chromosomes (individus) par le biais de leur combinaison. La population qui résulte de la sélection est divisée en deux sous-populations de taille  $N/2$  et chaque couple formé par un membre provenant de chaque sous-population participe à un croisement avec une probabilité donnée (la probabilité de croisement,  $p_c$ , souvent supérieure à 60 %). Si le croisement a lieu, sa localisation entre la position 1 et la position  $l$ , dans le cas du codage binaire, est tirée selon *une loi uniforme* et les deux individus échangent leurs gènes des deux cotés de cette localisation. Dans notre exemple un croisement localisé à la troisième position a eu lieu entre les individus 2 et 3. Il est tout à fait possible de faire des croisements aléatoires.

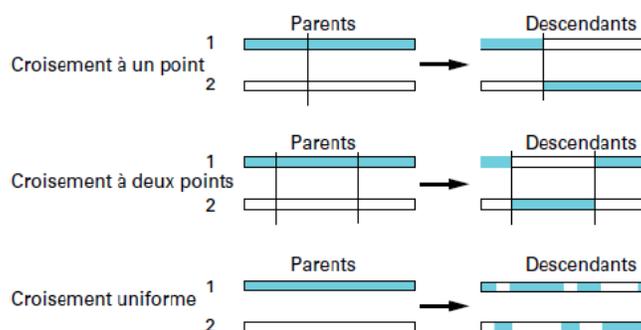


FIGURE 3.16 – Croisement

### - Opérateur de mutation

Une mutation consiste en l'inversion d'un bit se trouvant en un locus bien particulier et déterminé de manière aléatoire. L'opérateur de mutation modifie donc aléatoirement les gènes d'un enfant, ce qui permet de conserver la diversité au sein de la population. On peut dire que l'opérateur joue un rôle perturbateur : il introduit un bruit au sein de la population. En permettant la diversité de la population, la mutation permet d'éviter une convergence prématurée.



FIGURE 3.17 – Principe de la mutation

### 3.2.8 Applications des algorithmes génétiques (AG)

Cette section présente des applications des AG notamment en économie et en finance. Sans prétendre à l'exhaustivité, on peut classer ces applications autour de deux grands thèmes : l'utilisation des AG comme outil d'optimisation et de prévision, et l'utilisation des AG comme représentation de l'apprentissage .

#### - Optimisation et prévision

Dans cette sous-section, nous allons nous intéresser principalement à l'application des AG dans trois domaines : la résolution numérique, l'économétrie et la finance. Le point commun est l'utilisation des AG comme un simple  $n$  algorithme de calcul  $z$ .

### –Résolution numérique

A l'inverse des méthodes traditionnelles de résolutions numériques de type gradient, les AG ne sont pas fondés sur une approche analytique mais sur une approche itérative. En cela, peu d'information est nécessaire pour leur utilisation : l'espace de recherche possible et un critère d'efficacité. Les ingénieurs et les spécialistes en recherche opérationnelle ont très vite vu le potentiel de leur utilisation comme outil de résolution numérique, depuis le début des années 90, une liste importante d'articles sur le sujet comparant les AG à d'autres méthodes de résolution numérique, qu'elles soient analytiques ou aléatoires.

La complexité des problèmes en économie ne cessant de se développer, c'est tout naturellement que les économistes ont importé ces algorithmes dans leur domaine. Ainsi Dorsey et Mayer (1995) ont montré la potentialité des AG pour résoudre numériquement certains problèmes difficiles d'optimisation présentant des non-différentiabilités, des multi-modalités ou encore des discontinuités.

#### 3.2.9 Conclusion et discussion

- Un réseau de neurones artificiels est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des neurones biologiques. Dans ce chapitre, nous avons présenté le contexte de notre travail où nous avons vu un aperçu général sur les réseaux de neurones. Nous avons présenté les propriétés fondamentales des réseaux de neurones qui justifient l'intérêt croissant qui leur est accordé et que sont capable d'intervenir dans la résolution de nombreux problèmes de la modélisation ainsi de prévision.

Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type probabiliste, en particulier bayésien. Ils sont placés d'une part dans la famille des applications statistiques, qu'ils enrichissent avec un ensemble de paradigmes permettant de créer des classifications rapides (réseaux de Kohonen en particulier), et d'autre part dans la famille des méthodes de l'intelligence artificielle auxquelles ils fournissent un mécanisme perceptif indépendant des idées propres de l'implémenteur, et fournissant des informations d'entrée au raisonnement logique formel.

Les réseaux de neurones servent aujourd'hui à toutes sortes d'application dans divers domaines.

Par exemple ; un système de guidage pour automobile, conçu des systèmes de lecture automatique de chèques bancaires et d'adresses postales, applications militaires, pour faire des prévisions sur les marchés monétaires, pour évaluer le risque financier ou en assurance, pour la diagnostic médicale, etc. Que leur importance ira grandissant dans le futur.

- On sait que les applications des algorithmes génétiques sont multiples : optimisation de fonctions numériques difficiles, traitement d'image, optimisation d'emplois du temps, contrôle de systèmes industriels [Beasley, 1993a], cryptographie, etc.

Nos exemples d'application nous ont permis de nous rendre compte que le codage des données pour modéliser un problème est complexe. D'autre part, nous nous sommes aussi aperçus des difficultés pour choisir pertinemment de bons paramètres pour les divers opérateurs ( mutation , croisement ,sélection, remplacement ). Des choix par rapport aux opérateurs eux-mêmes sont aussi à gérer, sachant que certains sont plus appropriés au problème et qu'ils permettent d'optimiser.

# Chapitre 4

## Application

Commençons par une série simulée d'un carte logistique qui génèrent la série chronologique chaotique à 205 observations par l'équation (4.1),(figure.(4.1)).

$$y_{n+1} = ry_n(1 - y_n) \quad \text{pour } r = 3.7 \quad \text{et } y_0 = 0.4 \quad (4.1)$$

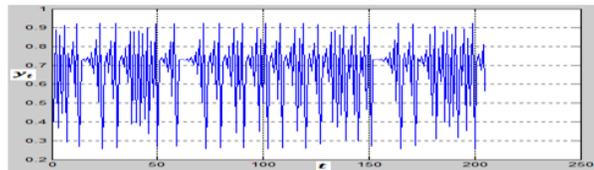


FIGURE 4.1 – série simulé à 205 données par carte logistique .

### 4.1 Modélisation de la série simulée (BJ)

Le graphe figure (4.2) de la série simulé qui génèrent la série chronologique chaotique à 195 observations par l'équation (4.1).

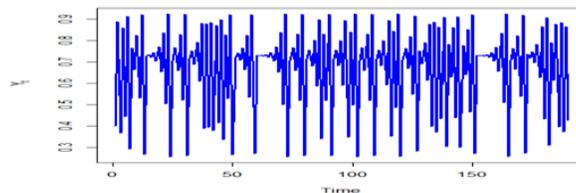


FIGURE 4.2 – série simulé à 195 données.

- Le meilleur modèle est  $ARIMA(4,1,0)$  avec moyenne nulle.

Tableau (4.1) donne les coefficients de ce modèle [15][30].

TABLE 4.1 – Coefficient du meilleur model  $ARIMA(4, 1, 0)$  de la série.

Coefficient(s)	Estimate	Std. Error
Ar1	-1.3380	0.0699
Ar2	-0.8459	0.1113
Ar3	-0.6130	0.1107
Ar4	-0.2627	0.0701

Les résidus forment un bruit blanc ( $BB$ ), d'où la validation du modèle. Finalement  $Y_t$  est ajustée par un  $ARIMA(4, 1, 0)$ , d'équation :

$$\nabla y_t = -1.3380\nabla y_{t-1} - 0.8459\nabla y_{t-2} - 0.6130\nabla y_{t-3} - 0.2627\nabla y_{t-4} + \varepsilon_t, \varepsilon_t \in (BB)$$

$$y_t = -0.338y_{t-1} + 0.4921y_{t-2} + 0.2329y_{t-3} + 0.3503y_{t-4} + 0.2627y_{t-5} + \varepsilon_t, \varepsilon_t \in (BB)$$

On peut donc avoir confiance en nos prédictions.

- Il y a 5 paramètres, que nous prenons dans RNA ; 5 entrées et en GA Nous prenons ;  $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$ .

## 4.2 Modélisation de la série simulée (RNA)

Dans ce thèse, les Réseaux de Neurones Artificiels sont appliqués pour 5 entrées, 1 sortie, un fonction de sommation (combinaison) et un fonction de transfert (d'activation) : (Figure (4.3)).

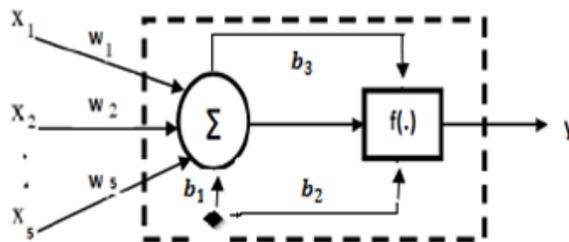


FIGURE 4.3 – Nombre : des entrées , des sorties , les fonctions de sommation et de transfert .

Les graphes(figure (4.4)) de la série  $Y_t$  et la série  $Y_{tajuste}$  ajustée par ce model(RNA) après l'apprentissage, montre l'absence d'une nette différence entre les deux courbes (on remarque un bon ajustement).

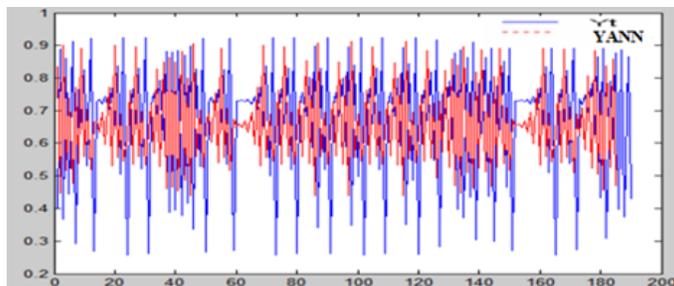


FIGURE 4.4 –  $Y_t$  et  $Y_{tajuste}$ .

$$Meilleure(w_1, w_2, w_3, w_4, w_5) = (0.6294, 0.8116, -0.7460, 0.8268, 0.2647). \quad (4.2)$$

$$Meilleure(b_1, b_2, b_3) = (-0.4430, 0.9298, 0.0938). \quad (4.3)$$

TABLE 4.2 – Tests sur résidus

Test	Estimate	Std. Error
Augmented Dickey Fuller ( Stationnarité )	Dickey-Fuller = -7.2016 Lag order= 5	0.01
Kolmogorov-Smirnovles ( normalité )	lillie test(residus)	0.7765
Ljung-Box ( correlation )	Box.test(residues ; lag = 50 type = "LjungBox")	0. 8231

Ce qui nous permet de conclure que les résidus forment un bruit blanc( $BB$ ) Tableau (4.2), on peut donc avoir confiance en nos prédictions.

### 4.3 Modélisation de la série simulée (AG)

On varions le nombre d'observation, la taille de la population initiale(les solutions  $\phi_i$ ), nombre d'iteration de l'algorithme génétique, probabilité de croisement(de 1% à 100%), probabilité de mutation(de 1% à 100%) et critère d'arrêt ( MSE soit minimum).

Le Meilleur  $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ , qu'on a obtenu était (selon le Tableau (4.3)) :

TABLE 4.3 – **Les conditions pour obtenir le meilleur résultat**  $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$  .

$NBOBSERVATIONS = 205$	<i>nombre d'observations</i>
$NBPOPULATION = 195$	<i>taille de la population initiale</i>
$NBITERATIONS = 100$	<i>nombre d'iteration de l'algorithme génétique</i>
$PROBCROISEMENT = 90$	<i>probabilité de croisement</i>
$PROBMUTATION = 10$	<i>probabilité de mutation</i>
$EPSILON = 0.000001$	<i>critère d'arrêt : <math>mse &lt; EPSILON</math></i>

La série ajusté par ce modél(AG) est stationnaire de plus on voit pas la différence entre les deux chroniques Graphe(4.5).

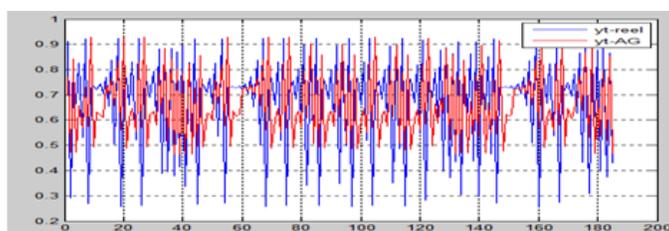


FIGURE 4.5 –  $y_t$  et  $y_t$  ajus AG.

Meilleure  $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5)$ , a été obtenu dans les conditions de Table (4.3),  
 Meilleure  $(\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) = (-0.3711, 0.3944, 0.1770, 0.4499, 0.3449)$

Avec ces paramètres nous obtenons l'équation du processus

$$y_t = -0.3711y_{t-1} + 0.3944y_{t-2} + 0.1770y_{t-3} + 0.4499y_{t-4} + 0.3449y_{t-5} + \varepsilon_t, \varepsilon_t \in (BB). \quad (4.4)$$

TABLE 4.4 – **Tests sur résidus.**

Test	Estimate	Std. Error
Augmented Dickey Fuller ( Stationnarité )	Dickey-Fuller = -7.2016 Lag order= 5	0.01
Kolmogorov-Smirnovles ( normalité )	lillie test(residus)	0.8812
Ljung-Box ( correlation )	Box.test(residues ; lag = 50 type = "LjungBox")	0.9477

Ce qui nous permet de conclure que les résidus forment un bruit blanc(BB) Tableau (4.4), on peut donc avoir confiance en nos prédictions.

Le graphique(4.6) résume la qualité de prévision et par suite la performance des trois modèles.

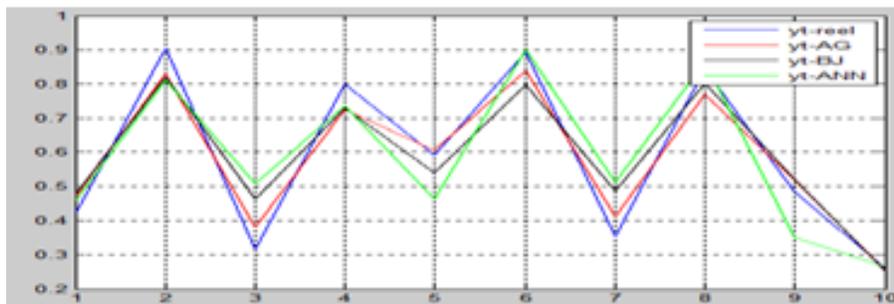


FIGURE 4.6 – Performance des modèles *BJ*, *RNA*, *AG*

Visuellement ,on voit pas de différence entre les trois modèles. Mais les mesures figurant dans le tableau (4.5)donnent la préférence au modèle de *RNA*.

TABLE 4.5 – conclusion de comparaison

	MAPE	MSE
<i>B – J</i>	40,3544	0,0868
<i>RNA</i>	20,1762	0,0216
<i>AG</i>	24,0402	0,0267

Nous avons obtenu un échantillon de 205 valeurs et retrouvé les résultats précédents. Nous augmentons maintenant les valeurs d'échantillon pour trouver le tableau suivant (4.6).

TABLE 4.6 – Échantillons plus grands que 205.

Nombre d'observations	Méthode	les paramètres	MAPE	MSE
405	B-J (ARIMA(4,1,0))	ar1=-1,3455, ar2=-0,8701 ar3=-0,6361, ar4= -0,2715	29,3096	0,0235
	AG	$\phi_1=-0,2197, \phi_2=0,4572$ $\phi_3=0,0683, \phi_4=0,3863$ $\phi_5=0,3009$	29,1113	0,0219
	RNA	$W_1=0,6294, W_2=0,8116$ $W_3=-0,7460, W_4=0,8268$ $W_5=0,2647$ $b_1=-0,4430, b_2=0,9298$ $b_3=0,0938$	24,1459	0,0197
805	B-J (ARIMA(4,1,0))	ar1=1,3636, ar2=0,9381 ar3=-0,7331, ar4= -0,2827	100,6512	0,3416
	AG	$\phi_1=0,0937, \phi_2=0,6505$ $\phi_3=0,0002, \phi_4=0,2036$ $\phi_5=0,0501$	27,1545	0,0216
	RNA	$W_1=0,7386, W_2=0,1357$ $W_3=-0,7597, W_4=0,9407$ $W_5=0,6799$ $b_1=-0,4430, b_2=0,9298$ $b_3=0,0938$	24,3612	0,0195
1505	B-J (ARIMA(4,1,0))	ar1=-1,3597, ar2=-0,9294 ar3=-0,7247, ar4= -0,2810	25,1576	0,0271
	AG	$\phi_1=-0,1299, \phi_2=0,4569$ $\phi_3=0,0861, \phi_4=0,3729$ $\phi_5=0,2255$	23,3751	0,0247
	RNA	$W_1=-0,0482, W_2=-0,2634$ $W_3=0,3112, W_4=0,8764$ $W_5=0,2409$ $b_1=-0,5896, b_2=0,7524$ $b_3=-0,1217$	21,6513	0,0225
2005	B-J (ARIMA(4,1,0))	ar1=-1,3600, ar2=-0,9230 ar3=-0,7168, ar4= -0,2828	24,7354	0,0240
	AG	$\phi_1=-0,3290, \phi_2=0,4829$ $\phi_3=0,2814, \phi_4=0,3803$ $\phi_5=0,1790$	22,1741	0,0206
	RNA	$W_1=-0,6649, W_2=0,0044$ $W_3=0,9987, W_4=-0,2892$ $W_5=-0,9058$ $b_1=-0,2043, b_2=0,8722$ $b_3=-0,3327$	21,3215	0,0192

## Conclusion Général

La théorie du chaos a été utilisée pour étudier et analyser des systèmes chaotiques. Les systèmes chaotiques génèrent des données aléatoires ; Cependant, dans l'espace de phase d'intégration, ils présentent un comportement déterministe. La prévision de séries chronologiques chaotiques est devenue un sujet de plus en plus populaire et stimulant.

Bon nombre des méthodes de prévision proposées dans la littérature sont soit inefficaces lorsqu'elles sont appliquées à des séries chronologiques chaotiques, soit difficiles à mettre en oeuvre. La motivation pour mener la présente étude était de développer une méthode plus efficace, facile à utiliser et pratique pour prévoir des séries chronologiques chaotiques.

La première inspiration et le principal moteur de cette étude ont été la conviction que le partage d'expériences et de connaissances entre différents domaines de la connaissance est un élément essentiel de l'enrichissement de cette connaissance. Cet thèse (dans la partie expérimentation) montre que la modélisation des séries chronologiques chaotiques (générée à partir de la simulation) en utilisant les capacités de réseaux de neurones artificiels (*RNA*), ainsi que les algorithmes génétiques (*AG*) et (*BJ*) dans la prédiction à court terme.

Le résultat montre que nous avons donc pris un nombre différent d'échantillons et de paramètres pour obtenir une précision de prévision. La méthode des réseaux de neurones artificiels est meilleure que la méthode des algorithmes génétiques et des modèles *ARIMA* pour la prédiction de séries chronologiques chaotiques.

Les résultats de la simulation montrent la capacité de cette méthode à fonctionner. Dans la suite, on peut envisager d'utiliser les réseaux de neurones artificiels à la prédiction de tout système chaotique (dynamique, rayan, déterministe,..... ), et d'autre part, de comparer cette méthode avec d'autre méthode comme la méthode de Swarm.

# Bibliographie

- [1] A. Lemamou, 2009 Ordonnancement de projet sous contraintes de ressources à l'aide d'un algorithme génétique à croisement hybride de type.
- [2] B. Philippe , 2006 Apprentissage Statistique and Data mining.
- [3] C. Bloch et P. Chatonnay, 2010 Différentes opérateurs évolutionnaires de permutation : sélections, croisements et mutations.
- [4] C. Gouriéroux et A. Montfort , 2006 Initiation à l'analyse des séries temporelles et à la prévision.
- [5] D. Bosq et D. Blanke, 2007 Inference and Prediction in Large Dimensions.
- [6] D.J. Lvingstone, 2009 Artificial Neural Networks Methods and Applications, Artificial Neural Networks in Biology and Chemistry, Springer 1.978-718.
- [7] D. M. Bates et D. G. Watts (1988), Nonlinear Regression Analysis and Its Applications.
- [8] D. R. Brillinger, 1975 Time Series : Data Analysis and Theory, Holt, Rinehart and Winston.
- [9] D. Ruelle et F. Takens, 1971 On the nature of turbulence. Commun Math Phys. 20, 167-192.
- [10] E. Lorenz, 1993 The Essence of Chaos.
- [11] F. Takens, 2006 Detecting strange attractors in turbulence, 366-381, Springer.
- [12] G.D. Magoulas et M.N. Vrahatis, 1999 Improving the convergence of the backpropagation algorithm using learning rate adaptation methods, Neural Computation . 1769-1796.
- [13] G. J. Rodgers 1992 de l'ordre dans le chaos Phys. Educ. 27 14-7.
- [14] G. Kevin, 2003 An introduction to neural networks.

- 
- [15] H. Akaike, 1969 Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics* 21, 243-247 Springer.
- [16] H. Chapman, 2000 *The Practical Handbook of Genetic Algorithms : Applications*.
- [17] H.D.I. Abarbanel, 1996 *Analysis of observed chaotic data*. Springer.
- [18] H. G. Funkhauser, 1936 A note on a tenth century graphe, *Osiris* 1 January.
- [19] H.H. John, 1992 *Adaptation in Natural Artificial Systems An Introductory Analysis with Application to Biology Control*.
- [20] H. Zhang, 2010 *Chaos synchronization and Its Application to Secure Communication*.
- [21] J.B. Peter et A.D. Richard, 2001 *Introduction to Time Series and Forecasting* Springer.
- [22] J. Commandeur, 2007 *An Introduction to State Space Time Series Analysis*.
- [23] J.D. Hamilton, 1994 *Time Series Analysis*, Princeton Press. 39- 63.
- [24] K. Kim, 1988 *Equilibrium Business Cycle Theory in Historical Perspective*, Cambridge.
- [25] L. J. Cheryl, 2008 *Introduction to Time Series Analysis and Forecasting*.
- [26] M. Gardener, 2012 *the statistical programming language Indianapolis*.
- [27] M. Hénon, 1976 A Two Dimensional Mapping with a Strange Attractor. 50,69-76.
- [28] M. José et M. Carnegie, 2008 *The Theory of Linear Prediction*.
- [29] M. Valérie, 1997 La dynamique des marchés boursiers est-elle chaotique, *journal de la société statistique de paris*, 63-81.
- [30] R. Engle, 1982 Autoregressive conditional heterocedasticity with estimates of the variance of uk inflation. *Econometrica* 50, 987-1008.
- [31] R. J. Bhansali, 1980 Autoregressive and window estimates of the inverse correlation function. *Biometrika* 66, 551-566.
- [32] R. J. Bhansali, 1983 The inverse partial correlation function of a time series and its applications, *journal of Multivariate Analysis* 13, 310-327.
- [33] R. K. John, 1998 *On the Programming of Computers by Means of Natural Selection*.
- [34] R. L. Devaney, 1985 *An Introduction to Chaotic Dynamical Systems*, CA.

- 
- [35] S. Ruey ,2010 Analysis of Financial Time Series.
- [36] T. Vallée et M. Yildizo,2003 Présentation des algorithmes génétiques et de leurs applications en économie.
- [37] W. S. Cleveland, 1972 The inverse autocorrelations of a time series and their applications, *Technometrics* 14, 277-293.
- [38] W. Zhang et Y. Gen-ke, 2002 Genetic programming-based chaotic time series modeling, *Journal of Zhejiang*, 1009-3095.
- [39] Y.Moussa, 2012 Elaboration d'algorithmes de masquage pour les systèmes de communication chaotique. Thèse de doctorat, université mentouri-constantine.
- [40] Z. Hongtao , 2010 Chaos Synchronization and Its Application to Secure Communication.
- [41] Z. X. Frank et C. Erik , 2017 Predicting Evolving Chaotic Time Series with Fuzzy Neural Networks.