Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

Université Abderahmane Mira de Béjaïa

Faculté des Sciences Exactes

Département de Recherche Opérationnelle

Mémoire de fin d'étude

En vue de l'obtention du Diplôme de

MASTER en Recherche Opérationnelle

Option : Modélisation Mathématique et Evaluation de Performance des Réseaux

Thème

Sur les modèles de séries chronologiques à valeurs entières

Présenté par : **GOUIRI Syphax**

Devant le jury composé de :

Présidente	AMROUN Sonia	M.C.B	Univ. de Béjaia.
Rapporteur	TOUCHE Nassim	M.C.A	Univ. de Béjaia.
Examinatrice	ZIANE Yasmine	M.C.A	Univ. de Béjaia.
Examinatrice	DJEROUD Lamia	M.C.B	Univ. de Béjaia.

Béjaia, Septembre 2021.

REMERCIEMENTS

D'abord, je remercie le bon Dieu tout puissant de m'avoir donné le courage, la force, la santé ét la volonté pour mener à terme ma formation de Master et pouvoir réaliser ce travail.

Je tiens à remercier mon promoteur Monsieur le docteur, Touche Nassim qui m'a proposé le théme de ce mémoire et pour sa disponibilité, son aide précieux et ses conseils qui m'accompagnait tout au long de ce travail. Je suis très reconnu d'avoir cru en mes capacités et d'avoir accordé sa confiance, et aussi pour ses qualités scientifiques, pédagogiques et humaine.

J'exprime tous mes remerciements à l'ensemble des membres de jury qui ont accepté d'évaluer mon travail.

J'adresse aussi mes remerciements à tous les enseignants de la filière Recherche Opérationnelle.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute ma gratitude à tous mes amis et à toutes les personnes qui m'ont aidé dans la réalisation de ce travail et en particulier à mes parents pour leur soutien qui m'a été bien utile durant mes études.

Merci pour tout.

Dédicace

Je *dédie ce modeste travail*

A *ma très chère mère, aucune dédicace ne saurait être à la hauteur pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis mon plus jeune âge, et même à l'âge adulte. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout-puissant, te préserver et t'accordée santé, longue vie et bonheur.*

Au *meilleur des pères, les mots ne suffisent guère pour exprimer l'attachement et le respect que je te porte. Ce travail est le fruit de tes sacrifices que tu as consentis pour mon éducation et ma formation.*

A *mes sœurs.*

A *mon oncle Rafik.*

A *mes amis.*

J'exprime enfin un humble geste de remerciement et de reconnaissance à une personne pour sa patience infinie, sa compréhension, son aide inestimable, ses conseils et ses encouragements tout au long de ce travail.

BVB

Table des matières

Table des figures

Liste des tableaux

Notation et Abréviation

$\xrightarrow{p.s}$ Convergence presque sûre .

$\mathbb{N}$ Ensemble des entiers naturels.

$\mathbb{N}^*$ Ensemble des nombres entiers positifs.

$\mathbb{R}$ Ensemble des nombres réels.

$\mathbb{Z}$ Ensemble des nombres entiers.

$\mathbb{G}$ Une distribution de probabilité discrète qui appartient à une famille de lois paramétrique.

$\mathbb{E}$ Espace d'etats.

$\mathbb{A}$ Un événement.

Ω Ensemble des résultats possibles.

$\mathbb{P}$ Probabilité.

$X \sim B(\alpha)$ La variable aléatoire X suit une distribution Bernoulli.

$X \sim P(\lambda)$ La variable aléatoire X suit une distribution poissonienne.

$X \sim G(\mu)$ La variable aléatoire X suit une distribution géométrique.

$X \sim DP(\lambda, \gamma)$ La variable aléatoire X suit une distribution double Poisson.

$X \sim GP(\lambda, \mu)$ La variable aléatoire X suit une distribution conditionnelle de Poisson généralisée.

$X \sim NB(r, p)$ La variable aléatoire X suit une distribution binomiale négative.

$X \sim COM - P(\lambda, v)$ La variable aléatoire X suit une distribution COM-Poisson.

$X \sim ZIP(\lambda, \nu)$ La variable aléatoire X suit une distribution ZIP.

$\circ$ Un opérateur d'amincissement binomial.

$*$ Un opérateur d'amincissement binomial négatif.

$\star$ Un opérateur d'amincissement binomial généralisé.

AA AL-Osh et Alzaïd.

ACF La fonction d'autocorrélation.

ACP Autoregressive Conditional Poisson.

AR AutoRegressive - AutoRégressif.

ARMA Modèle Moyen Mobile Autorégressive.

ARIMA Modèle Moyen Mobile Intégré Autorégressive.

BL Modèle Bilinéaire.

CAIDA Center for Applied Internet Data Analysies.

CLS Estimation des Moindres Carrés Conditionnels.

CML Maximum des Vraisemblance Conditionnelle.

CNN Convolutional Neural Network.

COM-INGARCH Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières de Poisson Conway-Maxwell.

Cov la convariance.

DINARCH Modèle Autorégressif Conditionnellement Hétéroscédastique à valeurs entières dispersés.

DL Du et Li.

DNN Deep Neural Networks.

DP-INGARCH Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières de Double Poisson.

E(X) l'espérance d'une série temporelle.

FARIMA Modèle moyen mobile intégré fractionné autorégressive.

GARCH Modèle Autorégressif Conditionnellement Hétéroscédastique Généralisé.

GINAR Modèle autorégressif Généralisé à valeurs entières.

GQL Quasi-Vraisemblance Réelles.

i.i.d Indépendant et identiquement distribué.

IPv4 La version 4 du protocole Internet (IP).

INAR Modèle Autorégressif à valeurs entières.

INARCH Modèle Autorégressif Conditionnellement Hétéroscédastique à valeurs entières.

INARMA Modèle Moyen Mobile Autorégressive à valeurs entières.

INBL Modèle Bilinéaire à valeurs entières.

INGARCH Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières.

INMA Modèle Moyen Mobile à valeurs entières.

IP Protocole Internet.

LSTM Un réseau Long short-term memory (réseau récurrent à mémoire court et long terme).

MA Moyenne mobile.

MGINAR Modèle Autoregréssif basé sur l'opérateur Matriciel Généralisé à valeurs entières.

ML Maximum Likelihood - Maximum de vraisemblance.

MLE Maximum Likelihood Estimate - Estimateur du maximum de vraisemblance.

MSE Erreur quadratique.

NMAE Erreur Absolue Moyenne Normalisée.

NMSE Erreur Carrée Moyenne Normalisée.

NN Réseaux Neuronaux.

NGINAR Modèle autorégressif stationnaire à valeurs entières avec des marges Géometriques.

PACF Fonction d'autocorrélation Partielle.

PARMA Modèles périodiques autorégressifs à moyenne mobile.

PIBL Modèle Bilinéaire Périodique à valeurs entières.

PINAR Modèle autorégressif à valeurs entière périodique.

PINARMA Modèle Moyen Mobile Autorégressive à valeurs entières Périodiques.

PINGARCH Autorégressif Conditionnellement Hétéroscédastique Généralisé Périodique à valeurs entières.

POINAR INAR d'innovation poissonienne.

PSER Rapport Signal/ Erreur de Prediction.

QML Quasi-Maximum de Vraisemblance.

RINAR Modèle autorégressif Arrondi à valeurs entières.

RNN Recurrent Neural Network.

t Temps.

V(X) La variance d'une série temporelle.

VAR Modèle Autorégressif Vectoriel standard à valeurs entières.

YW Yule-Walker.

ZIP-INGARCH INGARCH de Poisson Zero-Inflated.

Introduction générale

Au cours de ces trois dernières décennies, l'étude dynamique des données de comptage est impliquée dans de nombreuses applications de modélisation et de prévision (nombre de patients infectés par une maladie au cours du temps, nombre quotidien de certaines transactions financières, nombre mensuel d'entreprises en défaut de paiement, ... etc.). Donc il est nécessaire de développer un modèle mathématique de séries temporelles à la fois entier et suffisamment parcimonieux pour un seul objectif est de mieux expliquer les données de comptage observées. Cette double contrainte nécessite une approche spécifique et il est souvent difficile d'identifier des modèles similaires à ceux utilisés dans l'étude des séries temporelles à valeurs réelles [62] .

Pour cela, les séries chronologiques à valeurs entières ont suscité un intérêt grand et croissant dans de nombreux domaines (scientifique, médical, financier, économique, télécommunications,... etc.) où on s'intérèsse au nombre des occurrences d'un événement particulier dans un intervalle de temps spécifie. Plusieurs modèles ont été proposés dans la littérature pour modéliser ce type des séries. Ces modèles sont classifiés par (Cox et al, 1981) en deux catégories principales : les modèles "parameter-driven" (pour lesquels le paramètre d'intérêt dépend d'un processus latent) et les modèles "observation-driven" (pour lesquels le paramètre d'intérêt ne dépend que des observations). Une des premières approches proposées est le modèle INAR (1) (Integer valued Autoregressive) introduit par (McKenzie, 1985) et (Al-Osh and Alzaid, 1987). Ce modèle utilise l'opérateur d'amincissement introduit par (Steutel and al, 1979). Le modèle INAR appartient à la classe des modèles "parameter-driven". Un autre modèle populaire est le modèle INGARCH (p, q) (Integer valued Generalized Autoregressive Conditional Heteroscedastic) qui a été introduit par (Ferland and al, 2006). Il est aussi appelé ACP (Autoregressive Conditional Poisson) dans (Heinen ,2003). Le modèle INGARCH (p, q) est classifié comme un modèle "observation-driven" [1].

Dans ce travail, nous essaierons de présenter la classification mentionnée pour les

modèles de séries chronologiquesà valeurs entières dans la plupart de la littérature, qui est représentée en deux catégories principales qui jouent un rôle majeur dans l'étude de nombreux exemples en modélisation et en prévision. La première est la catégorie des modèles basés sur les équations aux différences stochastiques faisant intervenir l'opérateur d'amincissement dont l'exemple typique est le modèle AR entier (INAR) et le modèle MA entier (INMA).
Quant à la seconde catégorie, elle concerne les modèles basés sur la régression discrète tels que les modèles autorégression de Poisson et en particulier le fameux modèle autorégressif conditionnellement hétéroscédastique généralisé à valeurs entières (INGARCH).

Notre mémoire intitulé : "Sur les modèles de séries chronologiques à valeurs entières" est composé d'une introduction générale, de trois chapitres et une conclusion générale.
Au chapitre 1, avant de commencer à définir la première classe de modèles de séries chronologiques à valeurs entières, plusieurs notions de base qui caractérisent les distributions de comptage sont abordées. Ensuite, quelques modèles de séries chronologiques à valeurs entières basées sur l'opérateur d'amincissement sont présentés où nous nous intéressons principalement à la représentation des structures probabilistes pour ces modèles, tels que la stationnarité, l'existence de premier et deuxième moments et l'autocorrélation. Et on terminera ce chapitre par citer quelques méthodes d'estimation considérées dans la littérature pour les paramètres de chaque modèle. Au chapitre 2, une seconde classe de modèles de séries chronologiques à valeurs entières basées sur la régression discrète a été expliqué, où quelques modèles sont présentés avec leurs aspects probabilistes tels que : la distribution marginale, les conditions de stabilité et les structures de corrélation associées sont soulignées et accompagnées avec certaines méthodes d'estimation considérées dans la littérature, alors qu'au chapitre 3, l'objectif est d'expliquer en détail l'étude récente de Kim [55] où le modèle INGARCH a été proposé comme modèle prédictif pour le trafic réseau dans lequel les paramètres sont estimés à l'aide d'un algorithme classique (MLE), tandis que les paramètres du processus de Poisson sont prédits à pas de temps futurs sur la base d'un algorithme de prédiction sur un ensemble de données fourni par le Center for Applied Internet Data Analysis (CAIDA). Enfin, nous achèverons notre travail, par une conclusion et quelques perspectives.

1

Modèles de séries chronologiques à valeurs entières basés sur l'opérateur d'amincissement

1.1 Introduction

L'analyse des séries chronologiques à valeurs entières a connu ces dernières années de multiples développements et extensions, car lorsqu'une série prend qu'un nombre limité de valeur entières elle ne peut être approximée correctement par un modèle de série classique à valeur réelle.

Ce chapitre a pour but de fournir une vue sur l'ensemble des développements dans le domaine de la modélisation des séries temporelles à valeurs entières, en accordant une attention particulière aux modèles basés sur le concept d'amincissement.

Nous nous intéressons essentiellement la présentation de la structure probabiliste pour certains modèles basés sur l'amincissement pour l'analyse des séries temporelles à valeurs entières telles que la stationnarité, l'existence des moments de premier ordre et de seconde ordre, ainsi que les autocorrélations. Par la suite nous mettrons l'accent sur quelques approches d'estimations considérées dans la littérature.

1.2 Notations et propriétés de bases

Notre objectif consiste à décrire brièvement les caractéristiques des distributions de comptage, qui seront utiles pour identifier et déterminer les modèles appropriés pour un scénario ou un jeu de données donné.

Les données de comptage expriment le nombre de certaines unités ou événements dans un contexte, les résultats possibles sont contenus dans l'ensemble non négatifs entiers $\mathbb{N}$. Ces résultats ne sont pas seulement utilisés comme étiquettes ; elles ou ils proviennent du comptage et sont donc quantitatifs.

En conséquence, nous faisons référence à une variable aléatoire quantitative X en tant que variable aléatoire de comptage si sa réalisation est contenue dans l'ensemble des entiers non négatifs. On peut citer quelques exemples de phénomènes de comptage aléatoire :

• Le nombre de chambres occupées dans un hôtel de n chambres (n fini).

• Le nombre d'essais jusqu'à ce qu'un certain événement se produit.

Une manière courante d'exprimer la localisation (moyenne) et la dispersion (variance) d'un comptage aléatoire de la variable X, notées :

$$E[X] = \mu = \sum_x x.P(X = x).$$

$$V[X] = \sigma^2 = E[(X - E[X]^2)].$$

Pour obtenir une corrélation analogue à celle de séries classiques à valeurs réelles. Généralement, les modèles autorégressifs, moyens mobiles à valeurs entières, autorégressifs moyens mobiles à valeurs entières et les modèles bilinéaires à valeurs entières, sont basés essentiellement sur l'opérateur d'amincissement introduit par *Steutel et Van Harn* [90]. Ce dernier, noté par " $\circ$ ", est connu sous le nom de "opérateur d'amincissement binomial", il est défini comme suit

Définition 1.2.1 (Opérateur d'amincissement binomial)

Soit X une variable aléatoire à valeurs entières positives. Alors, l'opérateur d'amincissement binomial "$\circ$" est défini comme suit

$$\alpha \circ X = \begin{cases} \sum_{j=1}^{X} Y_j & si \;\; X > 0, \\ 0 & sinon. \end{cases} \tag{1.1}$$

Où $\{Y_j, j \in \mathbb{N}\}$ est une suite de variables aléatoires de Bernoulli de paramètre $\alpha \in [0, 1]$ indépendantes et identiquement distribuées (i.i.d) et indépendante de X avec

$$P(Y_j = 1) = \alpha,$$

et

$$P(Y_j = 0) = 1 - \alpha.$$

Cette suite est appelée série de comptage de $\alpha \circ X$. Par conséquent à X la variable aléatoire $\alpha \circ X$ suit une loi binomiale de paramètres X et α [52] [9].

***Remarque* 1.2.1**

L'opérateur "$\circ$" de stuetel parfois est présenté sous le nom d'un sous échantillon binomial [79].

Propriétés 1.2.1

- $0 \circ X = 0$,
- $1 \circ X = X$,
- $E(\alpha \circ X) = \alpha E(X)$,
- $V(\alpha \circ X) = \alpha^2 V(X) + \alpha(1 - \alpha)E(X)$,
- $Cov(\alpha \circ X, X) = \alpha V(X)$.

Pour une preuve explicite de ces dernières et bien d'autres propriétés nous nous référons à Freeland [36] , Da Silva [89] et Weiss [19].

***Remarque* 1.2.1**

L'interprétation de l'opérateur d'amincissement binomial, considérons une population de taille X à un certain moment t. Si nous observons la même population à un moment postérieur dit $t+1$, alors la population peut être diminuée, car une partie de ces individus sont morts entre les temps t et $t+1$. Si les individus survivent indépendamment les uns des autres, et si la probabilité de survivre de t à $t+1$ est égale à $1-\alpha$ pour tous les individus, alors le nombre de survivants est donné par $\alpha \circ X$.

1.3 Le modèle INAR (1)

Le modèle autorégressif à valeurs entières d'ordre 1 (INAR (1)) a été proposé par McKenzie [65], [66] et étudié par la suite par Al-Osh et Alzaid [3] est construit en utilisant l'opérateur d'amincissement binomial proposé par *Steutel et Van Harn* [90].

1.3.1 Structure probabiliste du modèle INAR (1)

1.3.1.1 Définitions

Définition 1.3.1 (Modèle INAR (1))

Un processus du second ordre $\{X_t, t \in \mathbb{Z}\}$ vérifie un modèle autorégressif à valeurs entières d'ordre 1 **INAR(1)** [52], s'il est de la forme :

$$\underbrace{X_t}_{\text{Population au temps t}} = \underbrace{\alpha \circ X_{t-1}}_{\text{Survivants à partir du temps t-1}} + \underbrace{\varepsilon_t}_{Immigration}, \qquad t \in \mathbb{Z}, \tag{1.2}$$

tel que

$$\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} Y_j, \qquad Y_j \sim \mathbb{B}(X_{t-1}, \alpha). \tag{1.3}$$

Où le bruit (ε_t) est une suite de variable aléatoire indépendante et identiquement distribuées (i.i.d) à valeurs dans $\mathbb{N}$, avec :

$$E[\varepsilon_t] = \mu_\varepsilon, \qquad V[\varepsilon_t] = \sigma_\varepsilon^2. \tag{1.4}$$

et (Y_j) est une suite de variables aléatoires i.i.d, suivant une loi de Bernoulli de paramètre α indépendante de X_{t-1} et (ε_t) est indépendante de (Y_j). Par conséquent, $(\alpha \circ X_t | X_t)$ suit une loi binomiale de paramètres X_t et α .
Le processus INAR (1) est une chaîne de Markov homogène avec la transition en 1 étape de probabilités données par Al-Osh et Alzaid [3] comme suit

$$P_{ij} = P(X_t = j | X_{t-1} = i) = \sum_{k=0}^{min(i,j)} \begin{pmatrix} i \\ k \end{pmatrix} \alpha^k (1-\alpha)^{i-k} . P(\varepsilon_t = j - k).$$

La prévision à un pas a l'instant T basée sur l'esperance conditionnelle comme pour un AR (1) et donnée par

$$\hat{X_{T+1}} = E[X_{T+1}/F_T] = \alpha X_T + \mu_\varepsilon.$$

avec $F_T = \sigma\{X_T, X_{T-1}, ...\}$.

1.3.1.2 Condition de stationnarité stricte

Le modèle INAR (1) est strictement stationnaire sous la condition (1.5) et lorsque (ε_t) et indépendantes et identiquement distribuées .

$$0 \leq \alpha < 1. \tag{1.5}$$

1.3.1.3 Les moments de premier et second ordre du processus INAR (1)

Sous l'hypothèse (1.5), les moments de ce modèle existent et sont tels que

$$E(X_t) = \frac{\mu_\varepsilon}{1-\alpha}, \qquad V(X_t) = \frac{\sigma_\varepsilon^2 + \alpha\mu_\varepsilon}{1-\alpha^2}. \tag{1.6}$$

1.3.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un processus INAR (1) stationnaire vaut

$$\rho(1) = \alpha, \tag{1.7}$$

$$\rho(k) = Corr(X_t, X_{t-k}) = \alpha^k, \qquad \forall k \in \mathbb{N}. \tag{1.8}$$

Beaucoup de propriétés sont similaire à celle d'un modèle AR (1) standard, en particulier la fonction d'autocorrélation.

1.3.1.5 Un exemple de données réelles

Pour illustrer les modèles INAR (1), considérons l'ensemble de données présentées par Weiß (2008a) [19].
Il s'agit d'une série chronologique exprimant le nombre quotidien de téléchargements d'un éditeur TEX pour la période allant de juin 2006 à février 2007 (T(jours) = 267).
Le graphique de la figure (1.1) montre que ces nombres quotidiens varient entre 0 et 14, sans tendance ni saisonnalité visible.

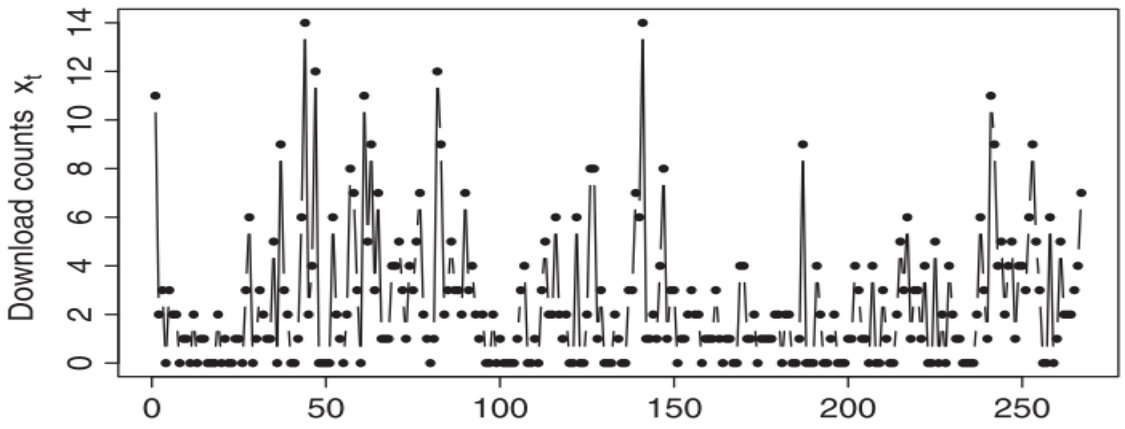

FIGURE 1.1 – Le nombre quotidien de téléchargements de TEX pour la période allant de juin 2006 à février 2007.

Un modèle INAR (1) semble également plausible au vu de l'interprétation qui dise que certains téléchargements au jour (t) peuvent être initié sur la recommandation d'utilisateurs du jour précédent (t-1)("survivants"), les autres téléchargements étant dus à des utilisateurs qui se sont intéressés au programme de leur propre initiative ("immigrants").

1.3.2 Estimation des paramètres du modèle INAR (1)

Le modèle INAR (1) est déterminé par le paramètre d'amincissement α, d'une part, et par des paramètres caractérisant la distribution marginale des observations, d'autre part. Dans cette sous-section, nous citerons des méthodes d'estimation pour les paramètres inconnus du processus INAR (1). Compte tenu des données de série chronologique $(X_1, X_2, ..., X_n)$, il s'agit d'estimer les valeurs de ces paramètres.

1.3.2.1 Méthode de Yule-Walker

L'estimateur de Yule-Walker pour le paramètre α n'est autre que le coeffient d'autocorrélation empirique de premier ordre (1.7), C.à.d.

$$\hat{\alpha_{YW}} = \rho(1). \tag{1.9}$$

1.3.2.2 Méthode des moments

Soit $(X_1, ..., X_n)$ une série chronologique issue d'un processus INAR (1) stationnaire. L'idée est de sélectionner les relations de moment appropriées de sorte que les vrais pa-

ramètres du modèle puissent être obtenus en résolvant le système d'équations obtenu. Pour l'estimation des paramètres, les moments sont remplacés par les exemples d'échantillons correspondants.

On sélectionne généralement au moins la moyenne marginale ainsi que l'autocorrélation de premier ordre, ce dernier conduit immédiatement à un estimateur de la méthode des moments défini par :

$$\alpha_{\hat{MM}} = \hat{\rho(1)} = \frac{\hat{\gamma(1)}}{\hat{\gamma(0)}}, \tag{1.10}$$

avec :

$$\hat{\gamma(k)} = \frac{1}{T}\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X}), k \in \mathbb{N}.$$

Si nous devons adapter un modèle INAR (1) avec des observations plus générales, des relations des moments supplémentaires sont nécessaires.

1.4 Le modèle POINAR (1)

L'instance la plus populaire de la famille INAR (1) est le modèle autorégressifs à valeurs entières d'innovations poissonienne **POINAR (1)**, qui a été introduit par McKenzie [65], [66] et aussi par Al-Osh et Alzaid [4].

1.4.1 Structure probabiliste du modèle POINAR (1)

1.4.1.1 Définitions

Définition 1.4.1 (Modèle POINAR(1))

Un processus du second ordre $\{X_t, t \in \mathbb{Z}\}$ vérifie qu'il appartient à la famille INAR (1) d'innovations poissonienne **POINAR (1)** [97], s'il est de la forme :

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1.11}$$

tel que

$$\alpha \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} Y_j, \qquad Y_j \sim \mathbb{B}(X_{t-1}, \alpha). \tag{1.12}$$

Et que les observations (ε_t) suivant une distribution de Poisson de paramétre (λ) (c.à.d) :

$$\varepsilon_t \rightsquigarrow \mathbb{P}(\lambda),$$

alors

$$E[\varepsilon_t] = \mu_\varepsilon = V[\varepsilon_t] = \sigma_\varepsilon^2 = \lambda$$

.

1.4.1.2 Condition de stationnarité stricte

Un processus POINAR (1) est une chaîne de Markov irréductible et apériodique donc il est strictement stationnaire, avec une marge stationnaire distribuée.
Il est bien connu que cette distribution marginale est aussi une loi de poisson d'une distribution, $\mathbb{P}(\mu)$ avec :

$$\mu = \frac{\lambda}{1-\alpha}.$$

***Remarque* 1.4.1**

Notons que les définitions précédente nous donnent deux propriétés importantes sur la distribution de Poisson :

1. L'invariance vis-à-vis de l'opérateur d'amincissement binomial, si

   $$X \sim \mathbb{P}(\mu)$$

   alors

   $$a \circ X \sim \mathbb{P}(\alpha\mu).$$

2. L'additivité, c'est-à-dire que si

   $$Z \sim \mathbb{P}(\mu), \;\; \varepsilon \sim \mathbb{P}((1-\alpha)\mu).$$

   et les deux sont indépendantes, alors

   $$Z + \mathbb{P}(\alpha\mu + (1-\alpha)\mu) = \mathbb{P}(\mu).$$

1.4.1.3 Les moments de premier et second ordre du processus

Les moments de ce modèle existent et sont tels que :

$$E(X_t) = \frac{\mu}{1-\alpha}, \qquad V(X_t) = \frac{\mu}{1-\alpha}. \tag{1.13}$$

1.4.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un processus POINAR (1) stationnaire vaut

$$\rho(k) = Corr(X_t, X_{t-k}) = \alpha^k, \qquad \forall k \in \mathbb{N}. \tag{1.14}$$

1.4.1.5 Un exemple de simulation

Les figures (1.2), (1.3) présentent deux exemples de chemins pour des processus simulés de POINAR (1) [19] . Les deux modèles ont été calibrés pour donner la même moyenne d'observation. Mais le paramètre d'autocorrélation se diffère par conséquent, les innovations signifient $\lambda = \mu(1 - \alpha)$.

Dans la figure (1.2), nous avons $\alpha = 0.5$ et $\lambda = 1.5$ et ce niveau modéré d'autocorrélation devient visible. la situation dans la figure (1.3) est plus extrême $\lambda = 0.15$ implique que ce n'est que rarement est une innovation vraiment positive. $\alpha = 0.95$ conduit à $\alpha \circ X$ égale à X la plupart du temps.

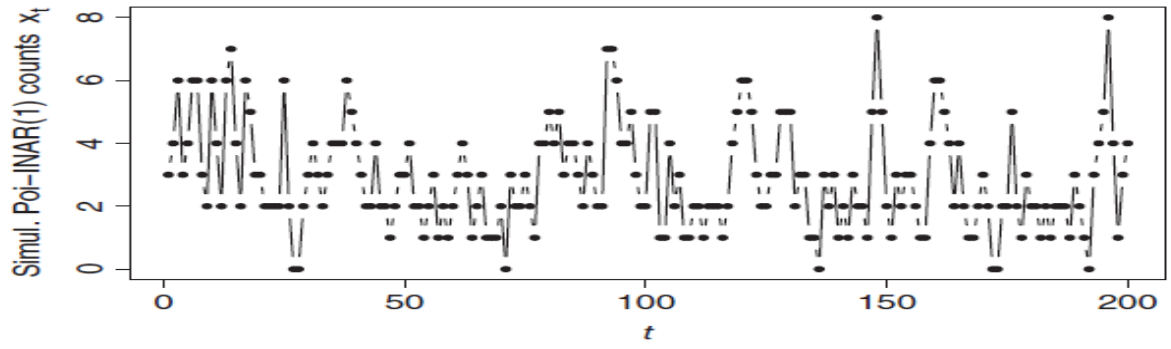

FIGURE 1.2 – Exemples de trajets simulés de processus INAR (1) Poissonien avec $\mu = 3$ et $\alpha = 0.5$.

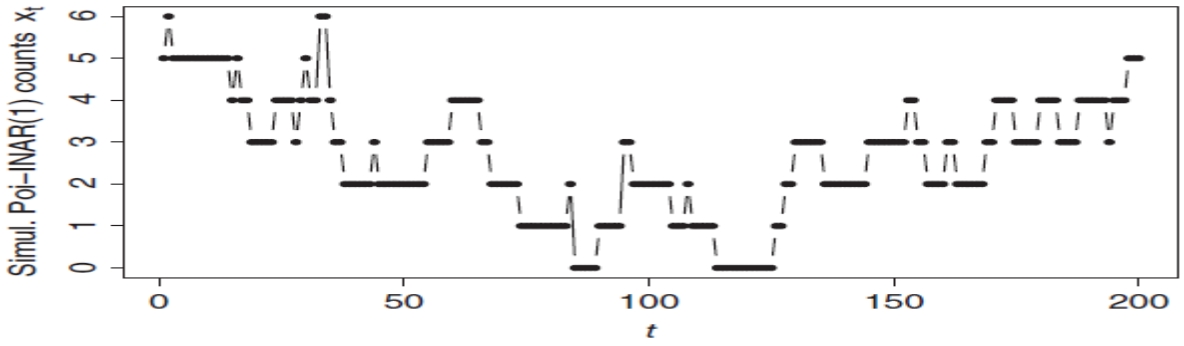

FIGURE 1.3 – Exemples de trajets simulés de processus INAR (1) Poissonien avec $\mu = 3$ et $\alpha = 0.95$.

1.4.2 Estimation des paramètres du modèle POINAR (1)

Le modèle POINAR(1) est déterminé par le paramètre d'amincissement α d'une part et par le paramètre qui caractérise la distribution marginale des observations λ, d'autre part.
Dans cette sous-section, nous citerons les méthodes d'estimation pour les paramètres inconnus du processus POINAR (1) Compte tenu des données de série chronologique $(X_1, X_1, ..., X_n)$, il s'agit d'estimer les valeurs de ces paramètres.

1.4.2.1 Méthode de Yule-Walker

L'estimateur de Yule-Walker pour le paramètre α n'est autre que l'autocorrélation empirique de premier ordre (1.7), C.à.d.

$$\hat{\alpha_{YW}} = \rho(1). \tag{1.15}$$

L'estimation de Yule-Walker de λ est basée sur le moment du premier ordre, C.à.d

$$\hat{\lambda_{YW}} = \bar{X}(1 - \hat{\alpha}). \tag{1.16}$$

Où $\hat{X}$ est la moyenne empirique.

1.4.2.2 Méthode des moments

Soit $(X_1, ..., X_n)$ une série chronologique issue d'un processus POINAR (1) stationnaire. Pour l'estimation des paramètres, les moments sont remplacés par les exemples d'échantillons correspondants.

Pour un modèle INAR (1) poissonnier, nous avons un paramètre supplémentaire en plus, qui est la moyenne des observations. Nous définissons donc les estimateurs de la méthode des moments par :

$$\hat{\alpha_{MM}} = \hat{\rho(1)} = \frac{\hat{\gamma(1)}}{\hat{\gamma(0)}}, \tag{1.17}$$

avec :

$$\hat{\gamma(k)} = \frac{1}{T} \sum_{t=k+1}^{T} (X_t - \bar{X})(X_{t-k} - \bar{X}), k \in \mathbb{N},$$

et

$$\hat{\lambda_{MM}} = \bar{X}(1 - \hat{\alpha}). \tag{1.18}$$

1.4.2.3 Méthode du maximum de vraisemblance

La fonction de vraisemblance d'un échantillon de (n + 1) observations du modèle POINAR (1) peut être écrite comme suit :

$$L(x, \alpha_{ML}, \lambda_{ML}) = (\prod_{t=1}^{n} P_t)(x_t) \frac{[\frac{\lambda}{(1-\alpha)}]^{x_0}}{x_0!} exp[\frac{-\lambda}{(1-\alpha)}], \tag{1.19}$$

où

$$P_t(x) = exp(-\lambda) \sum_{i=0}^{min(x_{t-1}, x_t)} \frac{\lambda^{x-1}}{(x-i)!} (x_{t-1}, i)^T \alpha^i (1-\alpha)^{x_{t-1}-i}.$$

pour $t = 1, 2, ..., n$ et $x = (x_0, x_1, .., x_n)$.

***Remarque* 1.4.3**

L'étude de simulation par Al-Osh et Alzaid (1987) a montré que l'estimateur des moindres carrés conditionnels et l'estimateur de Yule-Walker ne sont pas efficaces pour le modèle INAR (1) quand la distribution marginale de l'innovation est Poissonienne. Par ailleurs, l'estimateur du maximum de vraisemblance (ML) est difficile à calculer, notamment lorsque l'ordre du modèle est grand. Ce problème a incité certains auteurs à proposer des techniques et algorithmes alternatifs pour estimer ce modèle [32], [92].

1.5 Le modèle NGINAR (1)

Bien que le modèle POINAR (1) soit largement utilisé, il présente deux limites principales dans la pratique.
Premièrement, l'opérateur d'amincissement binomial du modèle POINAR (1) n'est pas approprié lorsque l'unité observée peut générer plus d'objets de comptage ou produire plus de nouveaux événements aléatoires. Deuxièmement, la distribution de Poisson souffre de l'exigence d'équidispersion qui ne peut expliquer la sous-dispersion et la surdispersion.
Alors un nouveau processus autorégressif stationnaire d'ordre 1 à valeur entière (NGINAR(1)) avec des distributions marginales géométriques utilisant l'opérateur d'amincissement binomial négatif a été introduit.

1.5.1 Structure probabiliste du modèle NGINAR (1)

1.5.1.1 Définitions

Définition 1.5.1 (Modèle NGINAR (1))

Soit $\{X_t, t \in \mathbb{Z}\}$, une suite de variables aléatoires à valeurs entières positives, un processus autorégressif stationnaire à valeurs entières du premier ordre avec des marges géométriques **(NGINAR (1))** [14] est défini comme suit

$$X_t = \alpha * X_{t-1} + \varepsilon_t, \qquad t \geq 1, \tag{1.20}$$

le cas ou (ε_t) sont des variables aléatoires (i.i.d) avec des distributions géométrique, Ristic et al [69] ont utilisé un opérateur d'amincissement binomial négatif "$*$" qui est défini comme suit :

$$\alpha * X = \sum_{i=1}^{x} W_i, \tag{1.21}$$

où

$$W_i \sim \mathbb{G}(\frac{\alpha}{\alpha+1}), \tag{1.22}$$

et

- $\{X_t, t \in \mathbb{Z}\}$ est un processus stationnaire avec des marginales géométriques $(\mu/(1+\mu))$.
- W_i est une séquence de variables aléatoires géométriques $G(\alpha/(\alpha+1))$ indépendantes de X.
- ε_t est une séquence de variables aléatoires indépendantes, identiquement distribuées, indépendantes de W_i et X_{t-1}.

1.5.1.2 Condition de stationnarité stricte

La stationnarité stricte du modèle NGINAR (1) est assurée car nous pouvons voir que le processus NGINAR (1) est markovien (irréductible et apériodique) avec des probabilités de transition

$$P(X_n = j|X_{n-1} = 0) = (1 - \frac{\alpha\mu}{\mu-\alpha})\frac{\mu^j}{(1+\mu)^{j+1}} + \frac{\alpha\mu}{\mu-\alpha}.\frac{\alpha^j}{(1+\alpha)^{j+1}} \tag{1.23}$$

$$P(X_n = j|X_{n-1} = i) = \frac{\mu\alpha^{j+1}}{(\mu-\alpha)(1+\alpha)^{j+i+1}}\begin{pmatrix} j+i \\ j \end{pmatrix} + (1-\frac{\alpha\mu}{\mu-\alpha})\frac{\mu^j}{(1+\alpha)^i(1+\mu)^{j+1}}\sum_{k=0}^{j}\begin{pmatrix} j+k-1 \\ i-1 \end{pmatrix}(\frac{\alpha(1+\mu)}{\mu(1+\alpha)})^k, i \geq 1.$$

Puisque $\{X_t, t \in \mathbb{Z}\}$ est un processus de Markov avec les mêmes marginales, et le processus NGINAR (1) est un processus strictement stationnaire. Alors, nous pouvons voir que le processus NGINAR (1) est un processus ergodique.

1.5.1.3 Les moments de premier et second ordre du processus NGINAR (1)

Il est a noter que la moyenne et la variance de la variable aléatoire X_t sont

$$E(X_t) = \mu. \tag{1.24}$$

$$V(X_t) = \mu(1+\mu). \tag{1.25}$$

L'autocovariance du processus NGINAR(1) vaut

$$\gamma_k = Cov(X_{t+k}, X_t) = Cov(\alpha * X_{t-1} + \varepsilon_t, X_{t-k}) = \alpha^k \gamma_0. \tag{1.26}$$

***Remarque* 1.5.1**

Si $\alpha = (\mu)/(1+\mu)$, alors (ε_t) a une distribution géométrique avec le paramètre $(\frac{\alpha}{\alpha+1})$, notez que dans ce cas particulier, le modèle a un seul paramètre.
A partir de la fonction de probabilité de la variable aléatoire (ε_t) , on obtient que la moyenne et la variance de la variable aléatoire vaut :

$$E[\varepsilon_t] = \mu_\varepsilon = (1-\alpha)\mu.$$

$$V[\varepsilon_t] = \sigma_\varepsilon^2 = (1+\alpha)\mu((1+\mu)(1-\alpha) - \alpha).$$

1.5.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un processus NGINAR (1) stationnaire vaut

$$\rho(k) = \alpha^k, \qquad \alpha \in [0, \frac{\mu}{1+\mu}] \ et \ k \geq 0. \tag{1.27}$$

La fonction d'autocorrélation décroît exponentiellement lorsque $(k \to \infty)$.

***Remarque* 1.5.2**

Le modèle NGINAR (1) capture la surdispersion, au cours des dix dernières années le modèle NGINAR (1) est devenu populaire dans certains domaines tels que la théorie de la fiabilité, la médecine et la théorie des réservoirs.

1.5.1.5 Exemple de simulation

Dans la figure (1.4), nous présentons 200 valeurs simulées du processus $\{X_t\}$ (NGINAR(1)) pour différentes valeurs des paramètres μ et α tel que : (a) $\mu = 0.5$, $\alpha = 0{,}1$, (b) $\mu = 0.5$, $\alpha = 0.2$, (c) $\mu = 0.5$, $\alpha = 0.3$, (d) $\mu = 1$, $\alpha = 0.1$, (e) $\mu = 1$, $\alpha = 0.2$, (f) $\mu = 1$, $\alpha = 0.4$, (g) $\mu = 5$, $\alpha = 0.2$, (h) $\mu = 5$, $\alpha = 0.4$, (i) $\mu = 5$, $\alpha = 0.6$.

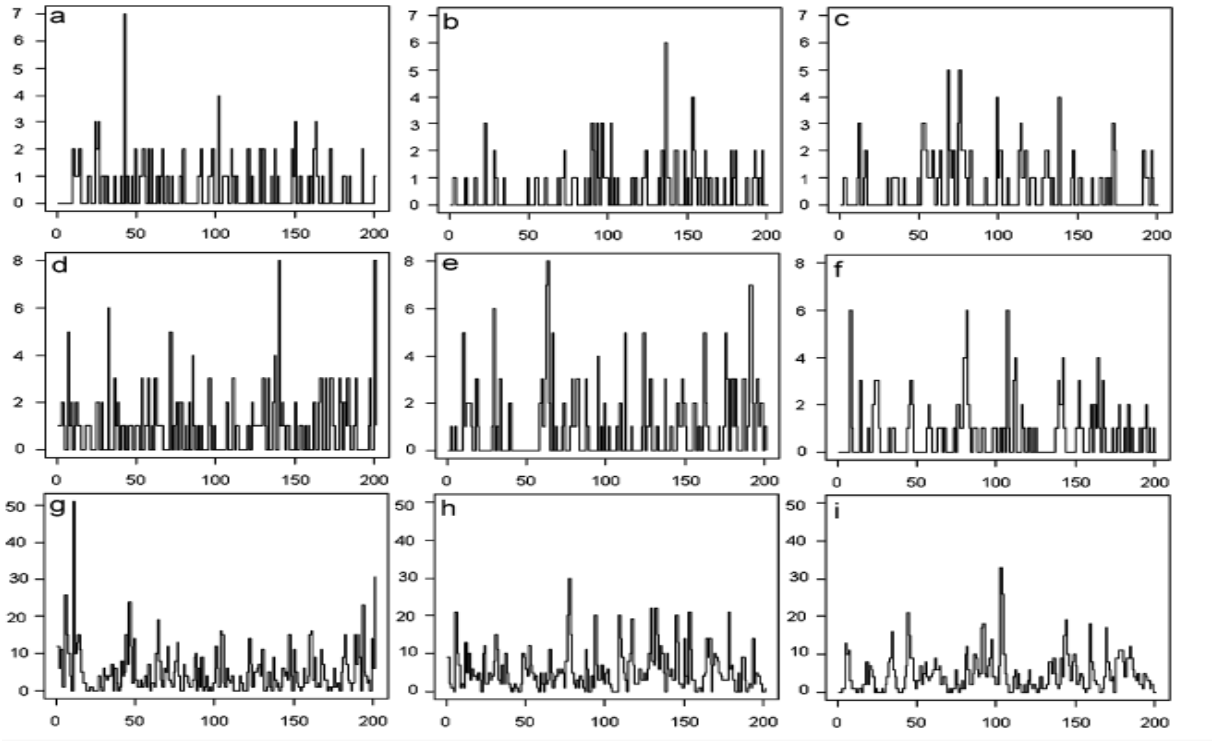

FIGURE 1.4 – Exemple de chemin du processus X_t, avec le modèle NGINAR(1) pour différentes valeurs des paramètres μ et α.

1.5.2 Estimation des paramètres du modèle NGINAR (1)

Dans cette sous-section, nous allons voir comment estimer les paramètres inconnus du processus NGINAR (1). Supposons que nous ayons une série chronologique $(X_1, X_2, ..., X_N)$ issue d'un processus NGINAR (1).

1.5.2.1 Méthode des moindres carrés conditionnels

Les estimateurs des moindres carrés conditionnels avec les paramètres α et μ sont obtenus en minimisant la fonction Q.

$$Q_N(\alpha, \mu) = \sum_{i=2}^{N} (X_i - \alpha X_{i-1} - \mu(1-\alpha))^2. \tag{1.28}$$

Les estimateurs sont donnés par

$$\hat{\alpha_{CLS}} = \frac{\sum_{i=2}^{N} X_i X_{i-1} - \frac{1}{N-1}\sum_{i=2}^{N} X_i \sum_{i=2}^{N} X_{i-1}}{\sum_{i=2}^{N} X_{i-1}^2 - \frac{1}{N-1}(\sum_{i=2}^{N} X_{i-1})^2}, \hat{\mu_{CLS}} = \frac{\sum_{i=2}^{N} X_i - \hat{\alpha}\sum_{i=2}^{N} X_{i-1}}{(N-1)(1-\hat{\alpha})}. \tag{1.29}$$

1.5.2.2 Méthode du maximum de vraisemblance

La fonction de vraisemblance d'un échantillon d'observations du modèle NGINAR (1) peut être écrite comme suit :

$$\log L(x_1, x_2, ..., x_n, \alpha, \mu) = x_1 \log \mu - (x_1 + 1) \log(1 + \mu) + \sum_{i=2}^{N} \log P_n(x_n, x_{n-1}; \alpha, \mu), \quad (1.30)$$

où

$$\log P_n(x_n, x_{n-1}; \alpha, \mu) = P(X_n = x_n | X_{n-1} = x_{n-1}).$$

On peut donc obtenir les estimateurs du maximum de vraisemblance en résolvant le système d'équations

$$\partial \log L / \partial \mu = 0,$$

et

$$\partial \log L / \partial \alpha = 0.$$

1.5.2.3 Méthode de Yule-Walker

Les estimateurs de Yule-Walker pour les paramètres α et μ du modèle NGINAR (1) sont donnés par rapport a $\mu = E(X)$ et $\alpha = \gamma(1)/\gamma(0)$, nous pouvons dériver des estimateurs de μ et α comme suit

$$\hat{\mu_{YW}} = \hat{X} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad \hat{\alpha_{YW}} = \frac{\sum_{i=2}^{N} (X_i - \hat{X}_i)(X_{i-1} - \hat{X}_i)}{\sum_{i=1}^{N} (X_i - \hat{X}_i)^2}. \quad (1.31)$$

1.6 Le modèle INAR (p)

Le modèle INAR (p) est une extension directe aux modèles autoregressifs à valeurs entières d'ordre 1 (INAR (1)), généralisant ainsi les résultats donnés par Al-Osh et Al-Zaid [3] , [4]. Aussi il est l'alternative conceptuellement plus proche de l'AR (p) gaussien est la proposition de Du et Li [27].

1.6.1 Structure probabiliste du modèle INAR (p)

1.6.1.1 Défnitions

Définition 1.6.1 (Modèle INAR (p))

Soit $\{X_t, t \in \mathbb{Z}\}$, une suite de variables aléatoires à valeurs entières positives et $\{\varepsilon_t, t \in \mathbb{Z}\}$ une suite de variables aléatoires indépendantes et identiquement distribuées

à valeurs entières positives, tel que :

$$E(\varepsilon_t) = \mu_\varepsilon, \qquad V(\varepsilon_t) = \sigma_\varepsilon^2. \tag{1.32}$$

Alors, $\{X_t, t \in \mathbb{Z}\}$ est un processus INAR (p) [6] s'il est de la forme :

$$X_t = \sum_{j=1}^{p} \alpha_j \circ X_{t-j} + \epsilon_t, \qquad \forall t \in \mathbb{Z}. \tag{1.33}$$

avec $p \in \mathbb{N}^*$ et $\{\alpha_j\}_{j\in\{1,...,p\}}$ une suite constant telles que :

$$0 \leq \alpha_j < 1, \qquad \forall j \in \mathbb{N}^*.$$

***Remarque* 1.6.1**

Notons qu'il en existe deux structures particulières pour ce modèle ; l'une est proposée par Al-Osh et Alzaïd (AA) [6] et l'autre étudiée par Du et Li (DL) [27] : Dans la premiére approche (AA), les auteurs supposent que la distribution conditionnelle du vecteur aléatoire $(\alpha_1 \circ X_t, \alpha_2 \circ X_t, ..., \alpha_p \circ X_t)$ étant donné $X_t = x_t$ est multinomiale de paramètres $(\alpha_1, ..., \alpha_p, X_t)$, et indépendante du processus passé, ç'est à dire indépendante de X_{t-k} et de tous les amincissements de celui-ci, $\forall k > 0$.
Par contre, dans la deuxième approche (DL), les auteurs supposent de plus que les séries de comptage sont mutuellement indépendantes.

1.6.1.2 Condition de stationnarité stricte

Comme pour le deux modéles INAR (p)-(AA) et INAR (p)-DL, les moments du processus INAR (p) sont faciles à dériver sous la condition de la stricte stationnarité, car le processus INAR (p) possède la structure de corrélation classique AR (p) [27].

$$\sum_{j=1}^{p} \alpha_j < 1. \tag{1.34}$$

***Remarque* 1.6.2**

Pour le processus $\{X_t, t \in \mathbb{Z}\}$ de INAR (p) si $\sum_{j=1}^{p} \alpha_j = 1$ alors le cas est instable et si $\sum_{j=1}^{p} \alpha_j > 1$ le cas est un cas explosif sinon il est stable [88].

1.6.1.3 Les moments de premier et second ordre du processus INAR (p)

Sous l'hypothèse (1.34), les moments des deux types de ce modèle (INAR (p)-AA et INAR (p)-DL) [52] existent et sont tels que

$$E(X_t) = \frac{\mu_\varepsilon}{1 - \sum_{j=1}^{p} \alpha_j}, \qquad V(X_t) = \frac{\mu_\varepsilon}{1 - \sum_{j=1}^{p} \alpha_j}. \tag{1.35}$$

***Remarque* 1.6.3**

la variance $V(X_t)$ n'est pas en général égale à (1.35).

1.6.1.4 Structure d'autocorrélation

La fonction d'autocorrélation de modèle INAR (p)-AA ressemble à celle obtenue pour ARMA (p, q)

$$\rho(1) = \alpha. \tag{1.36}$$

$$\rho(k) = \sum_{j=1}^{k} \alpha_j \rho(k-j), \qquad k \geq 2. \tag{1.37}$$

La strucutre de corrélation (ACF) du processus INAR (p)-DL est identique à celle du processus $AR(p)$ réel [9].

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \qquad k \in \mathbb{N}, \tag{1.38}$$

où $\gamma(k)$ est la « fonction d'autocovariance » donnée par $\gamma(h) = Cov(X_{t+1}, X_t)$.

***Remarque* 1.6.4**

Cependant, ces deux différentes formulations impliquent diverses structures du second ordre du processus : sous la première approche, (AA), le modèle INAR (p) a la même structure de covariance que celle du modèle ARMA (p, p-1) tandis que sous la deuxième approche, (DL), cette structure du second ordre est similaire à celle du modèle AR (p).

1.6.2 Estimation des paramètres du modèle INAR (p)

Dans cette sous-section, nous citerons des méthodes d'estimation pour les paramètres inconnus du processus INAR (p), supposons que nous ayons une série chronologique $(X_1, X_2, ..., X_N)$ issue d'un processus INAR (p).

1.6.2.1 Méthode de Yule-Walker

A. Estimation des paramètres du modèle INAR (p)-AA

L'estimateur de Yule-Walker pour le paramètre α.

$$\hat{\alpha_1} = \rho(1). \tag{1.39}$$

$$\hat{\alpha_k} = \sum_{j=1}^{k} \rho(k)\rho(k-j)^2. \tag{1.40}$$

Ainsi, l'estimation de Yule-Walker de λ est

$$\hat{\lambda} = \bar{X}(1 - \sum_{j=1}^{k} \hat{\alpha_j}). \tag{1.41}$$

Où $\hat{X}$ est la moyenne empirique.

B. Estimation des paramètres du modèle INAR(p)-DL

L'estimateur de Yule-Walker pour le paramètre α.

$$\hat{\alpha_1} = \rho(1)\frac{1-\rho(2)}{1-\rho(1)^2}. \tag{1.42}$$

$$\hat{\alpha_k} = \frac{\sum_{j=1}^{k} \rho(k) - \rho(k-j)^2}{1-\rho(1)^2}. \tag{1.43}$$

Ainsi, l'estimation de Yule-Walker de λ est

$$\hat{\lambda} = \bar{X}(1 - \sum_{j=1}^{p} \hat{\alpha_j}). \tag{1.44}$$

Où $\hat{X}$ est la moyenne empirique.

1.6.2.2 Méthode des moindres carrés conditionnels

La méthode d'estimation des moindres carrés conditionnels, proposée par Klimko et Nelson [57] a été adoptée pour l'estimation des processus INAR-AA et INAR-DL [78]. Soit $\Theta = (\alpha_1, ..., \alpha_p, \lambda)^T$ le vecteur des paramètres inconnus. L'estimateur conditionnel des moindres carrés de Θ est défini comme

$$\hat{\theta_{CLS}} = \underset{\theta \in \Theta}{argmin} Q_n(\theta), \tag{1.45}$$

où

$$Q_n(\theta) = \sum_{i=p+1}^{n} (X_i - \alpha_1 X_{i-1} - ... - \alpha_p X_{i-p} - \lambda)^2.$$

1.6.2.3 Méthode du maximum de vraisemblance

L'estimation par maximum de vraisemblance nécessite des hypothèses distributionnelles et peut être difficile à mettre en oeuvre.
L'estimateur du maximum de vraisemblance de Θ est obtenu en maximisant la fonction de logarithme de vraisemblance conditionnelle [78].

$$\hat{\theta_{ML}} = \underset{\theta\in\Theta}{argmax} l(\theta), \tag{1.46}$$

où

$$l(\theta) = \sum_{i=1+p}^{n} logP(X_t = x_t | X_{t-1} = x_{t-1}, ..., X_{t-p} = x_{t-p}).$$

En général, la maximisation de $(l(\theta))$ est assez lourde, en raison des sommations impliquées dans son calcul, et des difficultés numériques qui peuvent survenir lors de l'addition de nombreuses petites probabilités.

1.7 Le modèle PINAR (p)

De nombreuses séries temporelles à valeurs entières économiques, financières et environnementales rencontrées aujourd'hui dans la pratique présentent une structure d'autocorrélation périodique, cette caractéristique ne peut pas être correctement prise en compte et décrite par des modèles de séries chronologiques à valeurs entières avec un paramètres invariants.
Toutes ces raisons donnent une motivation pour étendre la classe de modèles INAR invariante dans le temps à une classe $\mathbf{PINAR}_S$ [86] variant dans le temps de façon périodique.

1.7.1 Structure probabiliste du modèle PINAR (p)

1.7.1.1 Défnitions

Définition 1.7.1 (Modèle PINAR (p))

Un processus $\{X_t, t \in \mathbb{Z}\}$, est dit satisfaisant à un modèle autorégressif à valeurs entière périodique d'ordre p, dans le sens de Gladyshev [39] avec une période S ($S \geq 2$), est noté $\mathbf{PINAR}_S$ **(p)** [86] si

$$X_t = \sum_{i=1}^{p} \varphi_{t,i} \circ X_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.47}$$

Où le processus d'innovation $\{\varepsilon_t; t \in \mathbb{Z}\}$, représente une suite de variables aléatoires indépendantes à valeurs entières non-négative, suivant une certaine distribution de probabilité discrète appartenant à une famille de lois paramétrique de la forme $\{\mathbb{G}_\alpha | \alpha_t = (\alpha_{t,1}, \alpha_{t,2}, ..., \alpha_{t,q})^T \in A \subset \mathbb{R}_+^{\shortparallel}\}$.
Les vecteurs de paramètres $\varphi_t = (\varphi_{t,1}, \varphi_{t,2}, ..., \varphi_{t,p})^T$ et α_t sont périodiques, concernant t, avec une période S ($S \geq 2$), où S est le plus petit entier positif satisfaisant la relation $(\varphi_{t+rS} = \varphi_t)$ et $(\alpha_{t+rS} = \alpha_t)$.

***Remarque* 1.7.1**

Les processus $(PINAR_S)$ ont été introduits pour modéliser les phénomènes à valeurs entières non-négatives qui évoluent dans le temps affectés par des perturbations saisonnières.

Définition 1.7.2 (Modèle PINAR (1))

Un processus $\{X_t, t \in \mathbb{Z}\}$, est dit satisfaisant à un modèle autorégressif à valeurs entière périodique d'ordre 1, et est noté $\mathbf{PINAR}_S$ **(1)** [85] si

$$X_t = \varphi_{t,1} \circ X_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.48}$$

Où le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$, représente une suite de variables aléatoires indépendantes à valeurs entières non-négative, suivant une certaine distribution de probabilité discrète appartenant à une famille de lois paramétrique de la forme $\{\mathbb{G}_\alpha | \alpha_t = (\alpha_{t,1}, \alpha_{t,2}, ..., \alpha_{t,q})^T \in A \subset \mathbb{R}_+^{\shortparallel}\}$.

1.7.1.2 Condition de stationnarité stricte

La stationnarité stricte pour le processus $\{X_t, t \in \mathbb{Z}\}$ peut être obtenu via l'analyse du processus autorégressif multivarié à valeurs entières introduit par Latour (1997) où $X_t = (x_{1+tT}, x_{2+tT}, ..., x_{T+tT})^T$ si les modules de toutes les valeurs propres de la matrice A sont inférieurs à 1 et que (ε_t) est indépendante de $X_s, (s < t)$.

$$A = \begin{Bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_{p-1} & \varphi_p \\ I_1 & 0 & \dots & 0 & 0 \\ 0 & I_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_p & 0 \end{Bmatrix}$$

1.7.1.3 Les moments de premier et second ordre du processus $PINAR_S(1)$

Sous la condition précédente, les moments de ce modèle vaut

$$E[X_t] = V[X_t] = \frac{\sum_{k=0}^{j-1} \beta_{j,k}\mu_{j-k} + \beta_{j,j} \sum_{i=0}^{T-j-1} \beta_{T,i}\mu_{T-i}}{1 - \beta_{T,T}}. \tag{1.49}$$

pour $j = 1, ..., T$ et $T \in \mathbb{N}$, t=j+kT ($k \in \mathbb{N}^*$).

La fonction d'autocovariance de ce modèle, $\gamma(k)$ n'est pas symétrique et elle vaut

$$\gamma_t(k) = \gamma_{t+k}(-k) = \beta_{T,T}\beta_{j+i,i}\mu_j$$

1.7.1.4 Exemple numérique

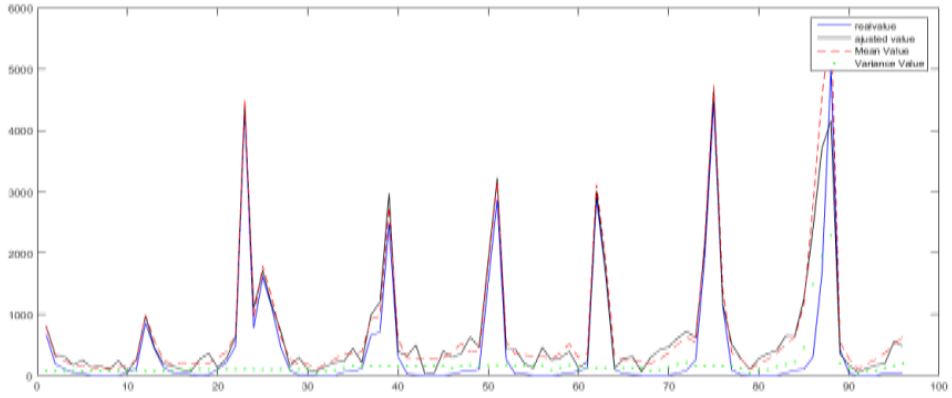

FIGURE 1.5 – Différentes trajectoires ajustées du processus X_t opposées aux vraies données .

Dans cet exemple, nous considérons une série de données saisonnière composée de 96 observations, cette dernière représente le nombre mensuel de personnes diagnostiqués en état de grippe au niveau des services hospitaliers de la région de la Catalogne (Espagne) entre 2009 et 2016 [86].
Cette série chronologique est illustrée dans la figure (1.5) ainsi que différentes trajectoires ajustées du processus $\{X_t, t \in \mathbb{Z}\}$ générées à partir d'un modèle $PINAR_{12}$ conduit par une distribution marginale géométrique.

1.7.2 Estimation des paramètres du modèle PINAR (p)

Dans la plupart des cas, la méthode d'estimation du maximum de vraisemblance n'est pas directement applicable dans les modèles semi-paramétriques, pour ces modèles PINAR avec une structure périodique, en raison de la discrétion de $\mathbb{G}_t$ dans ce cas pour tout $t, r \in \mathbb{Z}^+$ où $t = s + rS$ $(S > 2)$ alors pour tout valeur fixe de s tel que $s \in \{1, 2, ..., S\}$, l'estimateur de maximum de vraisemblance est faisable [85].

1.8 Le modèle GINAR (p)

Dans le but d'enrichir la classe des modèles autorégressifs permettant l'analyse des séries chronologiques à valeurs entières, Dion et al [50] et Latour [60] ont considéré une version plus générale du modèle INAR(p) noté **GINAR (p)** [52].

1.8.1 Structure probabiliste du modèle GINAR (p)

1.8.1. Défnitions

Dans la littérature, l'une des principales approches pour la modélisation des séries chronologiques de comptage est basée sur l'opérateur d'amincissement binomial " $\circ$ " de *Steutel & Van Harn* (Définition 1.2.1).
Par la suite , Latour (1998, [60]) a proposé l'opérateur d'amincissement généralisé "$\star$" et aussi un autorégressif d'ordre p généralisé à valeurs entières (GINAR (p)). Rappelons d'abord la définition de ce processus comme suit.

Définition 1.8.1 (Modèle GINAR (p))

Soit $\{X_t, t \in \mathbb{Z}\}$ une suite de variables aléatoires à valeurs entières positives
$\{\varepsilon_t, t \in \mathbb{Z}\}$, une suite de variables aléatoires i.i.d. à valeurs entières positives de moyenne finie et de variance finie

$$E(\varepsilon_t) = \mu_\varepsilon, \qquad V(\varepsilon_t) = \sigma_\varepsilon^2. \tag{1.50}$$

Alors, $\{X_t, t \in \mathbb{Z}\}$ est un processus **GINAR (p)** [52] [9] s'il est de la forme :

$$X_t = \sum_{i=1}^{p} \alpha_i \star X_{t-i} + \varepsilon_t, \qquad \forall t \in \mathbb{Z}, \tag{1.51}$$

où l'opérateur d'amincissement généralisé "$\star$" [47] est défini comme suit

$$\alpha_i \star X_{t-i} = \sum_{j=1}^{X_{t-i}} Y_j. \tag{1.52}$$

Avec $p \in \mathbb{N}^*$ et $\{\alpha_i\}_{i\in\{1...p\}}$ une suite de constantes telles que :

$$\forall i, \qquad 0 \leq \alpha_i < 1. \tag{1.53}$$

Ici, toutes les séries de comptage $(Y_j)_{k\in\mathbb{N}}$ associées à $\alpha_i\star$ pour $i = 1, 2, ..., p$ sont indépendantes entre elles et indépendantes de (ε_t) : Elles sont de moyenne finie α_i et de variance finie λ_i.

Remarque **1.8.1**

On a deux séries de comptage (Y_k) et (σ_k) qui sont mutuellement indépendantes, alors les deux opérateurs d'amincissement généralisé pour chaque série $\alpha\star$ et $\beta\star$ sont aussi indépendants l'un de l'autre.

1.8.1.2 Condition de stationnarité stricte

Dion et al [50] ont également établi la stationnarité stricte du modèle
GINAR (p), avec l'utilisation de la théorie des processus de branchement multi-type.

$$0 \leq \sum_{j=1}^{p} \alpha_j < 1. \tag{1.54}$$

Alors il existe un unique processus $\{X_t\}$ strictement stationnaire et ergodique qui satisfait (1.51) (Zhang, [100]).

1.8.1.3 Les moments de premier et second ordre du processus GINAR (p)

Le processus GINAR (p) est strictement stationnaire sous l'hypothèse (1.54) et les moments du premier et du second ordre sont tels que

$$E(X_t) = \frac{\mu_\varepsilon}{1 - \sum_{j=1}^{p} \alpha_j} = \mu_x, \qquad Var(X_t) = \mu_x \sum_{j=1}^{p} \lambda_j + \sum_{j=1}^{p} \alpha_j \gamma(j) + \sigma_\varepsilon^2. \tag{1.55}$$

Où

$$\gamma(j) = Cov(X_t, X_{t+j}), \qquad \forall t \in \mathbb{Z}.$$

1.8.1.4 Structure d'autocorrélation

Pour tout $(k \in \mathbb{N}^*)$ la structure d'autocorrélation pour le processus GINAR (p) est la même que le processus AR (p) réel pour les retards $k \geq p$.

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \qquad k \in \mathbb{N}, \tag{1.56}$$

où $\gamma(k)$ est la fonction d'autocovariance donnée par $\gamma(h) = Cov(X_{t+1}, X_t)$.

1.8.2 Estimation des paramètres du modèle GINAR (p)

Dans cette sous-section, nous citerons les méthodes d'estimation pour les paramètres inconnus du processus GINAR (p). Supposons que nous ayons une série chronologique $(X_1, X_2, ..., X_N)$ issue d'un processus GINAR (p).

1.8.2.1 Méthode des moindres carrés conditionnels

La méthode d'estimation des moindres carrés conditionnels (CLS), proposée par Fan et Li [30] a été adoptée pour l'estimation des processus GINAR (p) [47].
Soit $\Theta = (\alpha_1, ..., a_p, \mu_\varepsilon)^T$ le vecteur des paramètres inconnus. L'estimateur $\hat{\theta_{CLS}}$ est défini comme suit

$$\hat{\theta_{CLS}} = \underset{\theta \in \Theta}{argmin} Q_n(\theta), \tag{1.57}$$

où

$$Q_n(\theta) = \frac{1}{2}\sum_{t=1}^{n}(X_t - \sum_{i=1}^{p}\alpha_i X_{t-1} - \mu_\varepsilon)^2 + n\sum_{i=1}^{p+1} P_\lambda(|\theta_i|),$$

tel que $P_\lambda(-)$ est une fonction de pénalité [100], [30].

1.8.2.2 Méthode du maximum de vraisemblance

Sous l'hypothèse de stationnarité (1.54) la vraisemblance (ML) d'un modèle GINAR (p) [43] est donnée par :

$$L = \prod_{i=i_{start}}^{n} f_{X_t|X_{t-1},...,X_{t-p}}(x_t|x_{t-1}, ..., x_{t-p}; \alpha_1...\alpha_p, \gamma, \theta_{innov}). \tag{1.58}$$

Où $\alpha_1...\alpha_p \in [0, 1]$, avec $0 \leq \alpha_1 + ... + \alpha_p < 1$, sont les paramètres autorégressifs , et θ_{innov} est le vecteur des paramètres des variables aléatoires d'innovation. S'il existe des covariables, alors θ_{innov} est constitué de paramètres (de régression) reliant les covariables

aux paramètres de la distribution de l'innovation, (Par exemple, θinnov $= \lambda$ ¿ 0 pour les innovations de Poisson).

***Remarque* 1.8.2**

Il est bien connu que la méthode (CLS) est statistiquement inefficace et que la méthode (ML) est difficile à mettre en oeuvre, principalement en raison de la complication du calcul de probabilité de transition implique dans la fonction de log-vraisemblance lorsque l'ordre est augmenté.

1.9 Le modèle MGINAR (p)

Latour [59] a introduit le modèle **MGINAR (p)** [52], pour analyser les séries chronologiques à valeurs entières multivariées. Ce dernier est basé sur l'opérateur matriciel généralisé de *Steutel & Van Harn* [90].

1.9.1 Structure probabiliste du modèle MGINAR (p)

1.9.1.1 Défnitions

Définition 1.9.1 (Opérateur matriciel d'amincissement généralisé)

Soit $A = \{a_{ij}\}_{1\leq i,j\leq d}$ une matrice $(d \times d)$ et Z un vecteur de variables aléatoires à valeurs entières positives de dimension d. $A \star Z$ est défini par :

$$A \star \begin{pmatrix} Z_1 \\ . \\ . \\ . \\ Z_d \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{d} a_{1j} \star Z_j \\ . \\ . \\ . \\ \sum_{j=1}^{d} a_{dj} \star Z_j \end{pmatrix} \tag{1.59}$$

Ici, tous les opérateurs $\{a_{ij}\star\}_{1\leq i,j\leq d}$ sont mutuellement indépendants.

Définition 1.9.2 (Modèle MGINAR (p))

Soit $\{X_t, t \in \mathbb{Z}\}$, une suite de vecteurs de variables aléatoires à valeurs entières positives de dimension d ; $p \in \mathbb{N}$ et $A_j\star_{j\in\{1,2,\ldots,p\}}$, une suite d'opérateurs matriciels mutuellement indépendantes.
$\{\varepsilon_t\}_{t\in\mathbb{Z}}$, une suite de vecteurs de dimension d de variables aléatoires i.i.d. à valeurs entières

positives, de carré integrable et indépendantes de tous les opérateurs. Alors $\{X_t, t \in \mathbb{Z}\}$ est un processus **MGINAR (p)** [52] si

$$X_t = \sum_{j=1}^{p} Aj \star X_{t-j} + \varepsilon_t, \qquad \forall t \in \mathbb{Z}. \tag{1.60}$$

Pedeli et Karlis [77] ont défini des processus MGINAR(1) multivariées, où la matrice A est une matrice diagonale.

1.9.1.2 Condition de stationnarité strite

Un processus $\{X_t, t \in \mathbb{Z}\}$ d'un modèle MGINAR (p) est stable si les modules de toutes les valeurs propres de la matrice compagnon A sont inférieurs à 1, c'est-à-dire

$$det(I_{Kp} - \mathbf{A}z) \neq 0, \qquad \forall |z| \leq 1. \tag{1.61}$$

Où de façon équivalente, si

$$det(I_K - A_1 z - A_2 z^2 - ... - A_p z^p) \neq 0, \qquad \forall |z| \leq 1 \tag{1.62}$$

Où $det(I_{Kp})$ est appelé le polynôme caractéristique inverse et la matrice A vaut

$$A = \begin{Bmatrix} \alpha_1 & \alpha_2 & ... & \alpha_{p-1} & \alpha_p \\ I_1 & 0 & \dots & 0 & 0 \\ 0 & I_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_p & 0 \end{Bmatrix}$$

En plus on note que (X_t) est une chaîne de Markov avec des états dans $\mathbb{N}^d$ avec des probabilités de transition

$$P_{ij} = P(X_t = j | X_{t-1} = i) = \sum_{k=0}^{X_t} P(A \star X_{t-1} = X_t - k) P(\varepsilon_t = k).$$

le faite que la chaîne de Markov est irréductible et apériodique alors il existe une solution strictement stationnaire [34].

***Remarque* 1.9.1**

- Latour [59] [61] a donné les conditions de stationnarité et de causalité d'un tel processus et a montré que la fonction d'autocovariance du MGINAR (p) est identique à celle du processus autorégressif vectoriel standard à valeurs réelles, noté VAR (p) et il en déduit que le processus MGINAR (P) n'est autre qu'un processus VAR (p).
- Un processus stationnaire n'est pas obligatoirement stable.

1.9.1.3 Le moments de premier ordre du processus MGINAR (p)

La moyenne μ d'un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ est l'espérance de ce processus

$$E(X_t) = \mu, \qquad t \in \mathbb{Z}. \tag{1.63}$$

La fonction d'autocovariance $\gamma(h)$ d'un processus stationnaire $\{X_t, t \in \mathbb{Z}\}$ est

$$\gamma(h) = E[(X_t - \mu)(X_{t-h} - \mu)^T]. \tag{1.64}$$

Proposition 1.9.1

La fonction d'autocovariance d'un processus X_t stationnaire à l'horizon (h) notée $\gamma(h)$ verifier la propriété suivante

$$\gamma(h) = \gamma(-h)^{'}, \qquad \forall h = 0, 1, ... \tag{1.65}$$

et elle dépend uniquement de l'écart de temps h.

1.9.1.4 Structure d'autocorrélation

La fonction d'autocorrélation $\rho(h)$ d'un processus X_t stationnaire (stable) est

$$\rho(h) = D^{-1}\gamma(h)D^{-1}. \tag{1.66}$$

Où D est une matrice diagonale composée des racines carrées des éléments de la diagonale de $\gamma(0)$.

1.9.1.5 Un exemple de données réelles

Pour illustrer les modèles MGINAR (p), considérons l'ensemble de données concernent le nombre quotidien d'accidents de la route de jour et de nuit dans la région de Schiphol (Pays-Bas) au cours de l'année 2001 (voir Pedeli et Karlis [77]).

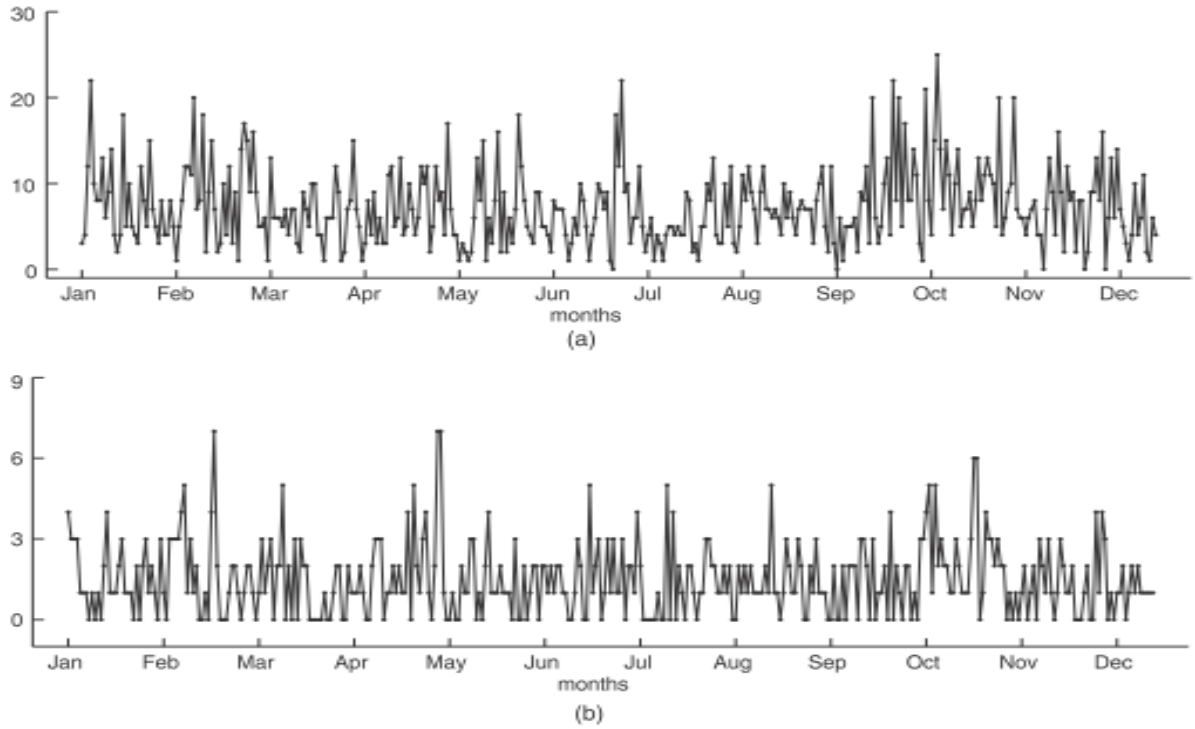

FIGURE 1.6 – (a)-(b) Nombre quotidien d'accidents de la route de jour et de nuit pour l'année 2001 (Pays-Bas).

1.9.2 Estimation des paramètres du modèle MGINAR (p)

Supposons que $(X_p, ..X_0, X_1, ..., X_n)$ soit observé à partir du processus MGINAR (p) défini par (1.60). Alors, l'estimateur conditionnel des moindres carrés (CLS) $\hat{\theta}_{CLS}$ pour θ est défini par

$$\hat{\theta_{CLS}} = \underset{\theta \in \Theta}{argmin} Q_n(\theta), \tag{1.67}$$

où

$$Q_n(\theta) = \sum_{t=0}^{n} (X_t - g_t(\theta))^T (X_t - g_t(\theta)),$$

et

$$g_t(\theta) = E[X_t|F_{T-1}] = \alpha + \sum_{k=0}^{p} A^k X_{t-k}.$$

***Remarque* 1.9.3**

L'autour [59] a considéré l'estimateur des moindres carrés conditionnels pour estimer les paramétres du modèle et a montré la consistance et la normalité asymptotique de l'estimateur (La méthode des moindres carrés multivariée (GLS)).

1.10 Le modèle RINAR (p)

La plupart des modèles autorégressifs à valeurs entières existants dans la littérature ont pour but de modéliser des séries chronologiques à valeurs entiéres positives, en particulier les séries de comptage.
Le modèle arrondi à valeurs entières autorégressif d'ordre p, **RINAR (p)** a été étudié par Kachoour [51] pour but de récolter des observations entières à partir des données réelles (par exemple, les données météorologique).

1.10.1 Structure probabiliste du modèle RINAR (p)

1.10.1.1 Défnitions

Définition 1.10.1 (Modèle RINAR (p))

Le processus $\{X_t, t \in \mathbb{Z}\}$ est dit un processus **RINAR (p)** [51] [52] *(Rounded INteger-valued AutoRegessif)* s'il possède la présentation suivante :

$$X_t = (\sum_{j=1}^{p} \alpha_j \cdot X_{t-j} + \lambda_t) + \varepsilon_t. \tag{1.68}$$

Où $(\cdot)$ représente l'opérateur d'arrondi à l'entier le plus prés, (λ_t) est une suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{Z}$, définie sur un espace probabilisé $(\Omega, A, \mathbb{P})$, λ et les α_j sont des paramétres réels.

***Remarque* 1.10.1**

L'opérateur d'arrondi peut être interprêté comme une fonction de censure sur le modéle AR (p) réel et λ comme la moyenne du bruit non-centré ($\varepsilon_t^{'} = \varepsilon_t + \lambda$).

Définition 1.10.2 (Modèle RINAR (1))

Le processus arrondi à valeurs entières autorégressif de premier ordre (RINAR (1)) a été introduit par Kachour et Yao [53], il posséde la présentation suivante :

$$X_t = (\alpha \cdot X_{t-j} + \lambda) + \varepsilon_t = f(X_{t-1}, \theta) + \varepsilon_t. \tag{1.69}$$

Où $(\cdot)$ représente l'opérateur d'arrondi à l'entier le plus prés, (λ_t) est une suite de variables aléatoires i.i.d. centrées à valeurs dans $\mathbb{Z}$, définie sur un espace probabilisé $(\Omega, A, \mathbb{P})$, λ et les α sont des paramétres réels.
Le bruit (ε_t) est une suite de variables aléatoires i.i.d et le processus (X_t) définie par (1.69), forme une chaine de Markov homogéne avec un espace d'états $E = \mathbb{Z}$ et une probabilité de transition.

$$\pi(x, y) = P\{\varepsilon_1 = y - f(x, \theta)\}, \qquad x, y \in E. \tag{1.70}$$

Où la fonction de régression $f(x, \theta) = (\alpha x + \lambda)$, pour tout $x \in E$ et $\theta = (\alpha, \lambda) \in \Theta$, l'espace des paramétres.

1.10.1.2 Condition de stationnarité stricte

L'étude de processus RINAR (p) [51] revient à étudier le processus vectoriel suivant

$$Y_t = \begin{pmatrix} X_1 \\ X_{t-1} \\ . \\ . \\ X_{t-p+1} \end{pmatrix} = \begin{pmatrix} \langle \sum_{j=1}^{p} \alpha_j \cdot X_{t-j} + \lambda \rangle + \varepsilon_t. \\ X_{t-1} \\ . \\ . \\ X_{t-p+1} \end{pmatrix} \tag{1.71}$$

Le processus Y_t forme une chaine de markov homogène.
Pour assurer la stationnarité et l'ergodicité du modèle RINAR (p), nous imposons que la somme des valeurs absolues des coefficients de régression soit strictement inférieur à 1.
Et pour assurer la stationnarité et l'ergodicité du processus RINAR (1), nous supposons que la valeur absolue du coefficient de régression est strictement inférieur à 1. Cette hypothèse est similaire à celle qui garantit la stationnarité du modèle AR(1) réel.

Proposition 1.10.1 (Kachour, 2009)

Soit $\theta = (\alpha, \lambda) \in \Theta$ fixé. Supposons que :

1. La chaïne de Markov (X_t) est irréductible ;

2. pour un certain $k > 1$, $\mathbb{E}|\varepsilon_t|^k$ ¡ $+\infty$;
3. $|\alpha| < 1$.

Alors

1. (X_t) posséde une unique mesure de probabilité invariante, notée μ. De plus, μ posséde un moment d'ordre k.
2. Pour tout $x \in E$ et $f \in L^1(\mu)$ et $\mathbb{P}_x$ représente la probabilité conditionnelle, nous avons
$$\frac{1}{n}\sum_{k=1}^{n} f(X_t) \xrightarrow{p.s} \mu(f), \quad \mathbb{P}_x \quad p.s.$$

D'aprés l'hypothèse (1), (X_t) est une chaïne de Markov récurrente positive et par conséquent μ est unique.
La proposition est une conséquence directe du théorème ergodique classique pour les chaïnes de Markov.
l'hypothése (1) concernant l'irréductibilité de la chaïne est assurée. Notons que l'hypothèse 2 est équivalente à la condition qui assure la stationnarité du modèle AR (1) réel.

1.10.1.3 Le moment de premier ordre du processus RINAR (1)

On supposons que les hypothèses de la Proposition (1.10.1) sont satisfaites. Par conséquent le processus RINAR (1) défini par (1.69) est strictement stationnaire. Alors

$$E[X_t] = m, \qquad \forall t \in \mathbb{Z}, \tag{1.72}$$

où

$$|m - \frac{\lambda}{1-\alpha}| \leq \frac{1}{2(1-\alpha)},$$

et $(\frac{\lambda}{1-\alpha})$ n'est autre que la moyenne d'un modèle AR (1) réel ayant les mêmes paramètres que RINAR (1) .

1.10.1.4 Structure d'autocorrélation

Sous les hypothèses de la Proposition (1.10.1) la fonction d'autocorrélation vaut

$$\rho(k) = \frac{Cov(X_t, X_{t+k})}{V(X_t)}, \qquad \forall t \in \mathbb{Z} \quad et \quad k \in \mathbb{N}. \tag{1.73}$$

Remarque **1.10.2**

A cause de l'opérateur d'arrondi l'étude de la loi stationnaire et du corrélogramme du modèle RINAR (1) est compliquée.

1.10.2 Estimation des paramètres du modèle RINAR (p)

1.10.2.1 Estimation des paramètres du modèle RINAR (P)

Soient $(X_0, X_1, ..., X_n)$ des obsevations du processus RINAR (1) [52] [51]. Pour l'estimation du paramétre θ, nous considérons l'estimateur des moindres carrés défini par :

$$\hat{\theta}_n = \underset{\theta \in \Theta}{argmin} Q_n(\theta). \tag{1.74}$$

Où

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} [X_t - f(X_{t-1}; \theta)]^2. \tag{1.75}$$

Et

$$f(x; \theta) = f(x_1, ..., x_p) = \langle \sum_{j=1}^{p} \alpha_j \cdot X_{t-j} + \lambda \rangle. \tag{1.76}$$

***Remarque* 1.10.3**

Nous notons $\theta_0 = (\alpha_1^*, ..., \alpha_p^*, \lambda^*)$ la vraie valeur des paramètres du modèle et $\mathbb{P}_0$ représente la distribution de probabilité sous θ_0.
Par la suite, nous supposons que, sous $\mathbb{P}_0$, les hypothèses suivantes sont vérifiées :
(Y_t) est irréductible ; $\mathbb{E}|\varepsilon_t|^k < +\infty$ où $k \geq 2$; $\sum_{j=1}^{p} |\alpha_j^*| < 1$ et Θ est compact.
L'estimateur des moindres carrés du modèle RINAR (p) par l'algorithme de la recherche dichotomique est assez proche de celle de l'estimateur de Yule-Walker du modèle AR (p) qui représente le point de départ de cet algorithme.

1.10.2.2 Exemple numérique

Dans cette partie, une application d'un modèle autorégressif arrondi à valeurs entières de premier ordre sur des données réelles a été étudié par [52]. Il s'agit de 70 observations représentant des résultats consécutifs d'un processus chimique, source : O'Donovan [26].

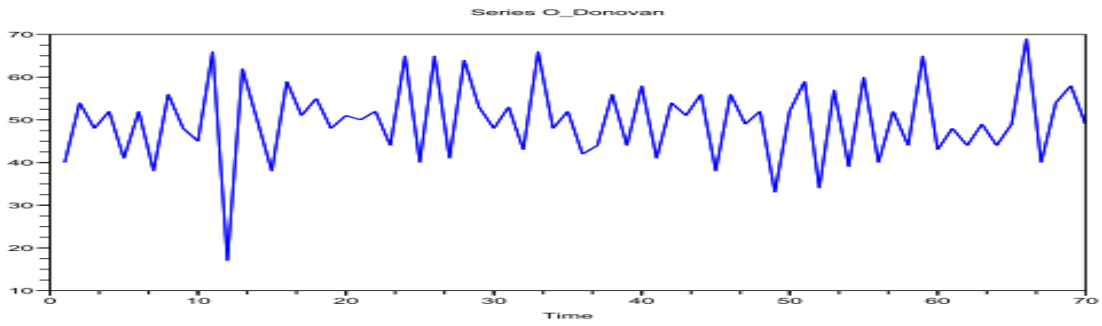

FIGURE 1.7 – Observations provenant des résultats consécutifs d'un processus chimique (O'Donovan).

Toutes les observations sont positives, elles varient entre 17 et 66 , et la moyenne et la variance empirique vaut

$$\bar{x} = 49,6857, \qquad \sigma^2 = 84,7403.$$

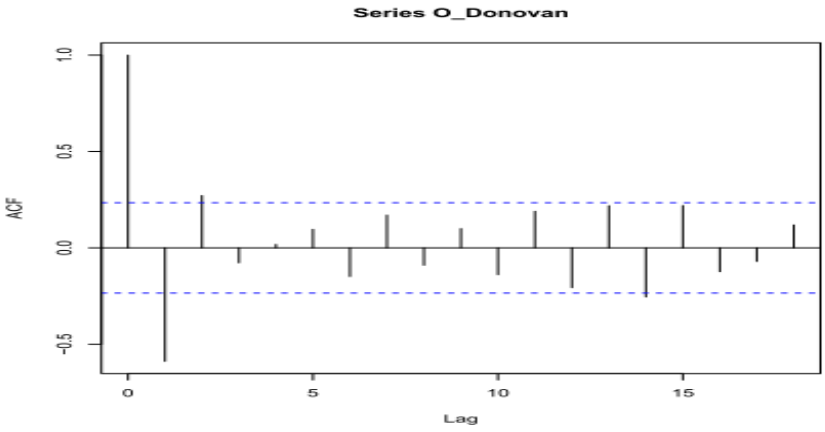

FIGURE 1.8 – ACF des 70 observations d'O'Donovan.

D'aprés l'ACF, le coefficient d'autocorrélation empirique du premier ordre est significatif et le coefficient d'autocorrélation empirique du second ordre est presque significatif $\bar{\rho}(1) = -0.588$ et $\bar{\rho}(2) = 0.272$. Aussi des signes entre les coefficients d'autocorralations empiriques du premier et second ordre, et les coefficients d'ordre supèrieur tendent vers zèro rapidement.

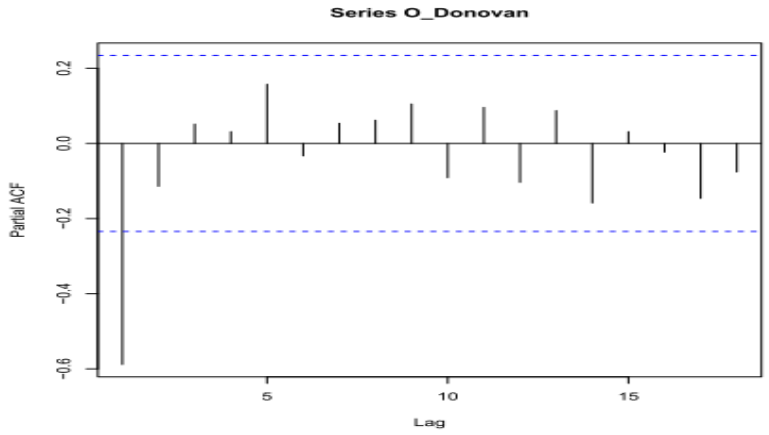

FIGURE 1.9 – PACF des 70 observations d'O'Donovan.

D'aprés le PACF, nous pouvons voir que seulement le coefficient d'autocorrélation partiel du premier ordre est siginificatif.
Une comparaison a été faite entre le modèle proposé par O'Donovan AR (1) et un processus RINAR (1). Pour analyser les données, les 60 premières observations sont faites pour l'estimation des paramètres et les 10 dernières pour comparer les performances de la prévision.

Le tableau suivant represente les résultats de prévision sur les 10 derniéres observations de la sèrie en utilisant AR (1). Rappelons qu'à l'instant 60, nous avons $X_{60} = 43$.

Instant	61	62	63	64	65	66	67	68	69	70
Vraie valeur	48	44	49	44	49	69	40	54	58	49
Valeur de prèvision	54	53	51	50	53	50	37	56	47	44

TABLE 1.1 – Les résultats de prévision sur les 10 derniéres observations de la sèrie en utilisant AR (1).

Le tableau suivant represente les résultats de prévision en utilisant le processus RINAR (1) sur les 10 dernières observations de la série.

Instant	61	62	63	64	65	66	67	68	69	70
Vraie valeur	48	44	49	44	49	69	40	54	58	49
Valeur de prèvision	54	51	53	50	53	50	38	56	47	44

TABLE 1.2 – Les résultats de prévision sur les 10 derniéres observations de la sèrie en utilisant RINAR (1).

Interprétation des résultats

L'analyse basée sur le modèle stochastique doit considérer la nature entière de la série observée. Généralement, le processus AR (1) standard ne garantit pas cette fonctionnalité. En revanche, le coefficient d'autocorrélation du premier ordre étant négatif, le processus INAR (1) ne peut pas modéliser la séquence courante. De plus,une remarque a été conclue que sur la valeur calculée de l'estimateur des moindres carrés $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\lambda}_n)$ d'après la recherche dichotomique égale exactement à l'estimateur de Yule-Walker du modèle AR (1) .

Ceci explique les performances de prédiction presque identiques des deux modèles (les prédictions sont les mêmes, sauf que le $7^{\text{ème}}$ RINAR (1) augmente la prédiction d'une unité).

1.11 Le modèle INMA (q)

Des équivalents discrets des processus classiques de moyenne mobile d'ordre q basés sur l'opérateur d'amincissement binomial ont également été proposés dans la littérature par Al-Osh et Alzaid [5] et aussi par McKenzie [67], **INMA (q)**, le modèle INMA est un cas particulier du modèle INARMA ou dans ce modèle INMA (q) remplace la multiplication dans le modèle MA (q) par l'opération d'amincissement binomial (indépendamment de toute variable aléatoire disponible à l'instant t).

1.11.1 Structure probabiliste du modèle INMA (q)

1.11.1.1 Défnitions

Définition 1.11.1 (Modèle INMA (q))

Un processus, à valeurs entières positives $\{X_t, t \in \mathbb{Z}\}$ est un processus moyen mobile à valeurs entières, d'ordre q, **INMA (q)** [12] [96] (voir Al-Osh et Alzaid [6], Al-Osh et

Alzaid [5], ...) s'il est donné comme suit

$$X_t = \sum_{i=1}^{q} \beta_i \circ \varepsilon_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.77}$$

Où le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corrélées, à valeurs entières positives, avec une moyenne μ_ε et une variance finies σ_ε^2 et où "$\circ$ " est un opérateur d'amincissement de "*Steutel-Van Harn*".

***Remarque* 1.11.1**

Jusqu'à présent, un total de cinq modèles INMA (q) [75] différents ont été proposés dans la littérature (INMA, BINMA, VINMA, BINFIMA et INARFIMA), chacun ayant des interprétations et des propriétés probabilistes légèrement differentes. (voir Al-Osh et Alzaid [5], McKenzie [67], Brännäs et Hall [15], Weiss [19],...).
Les modèles INMA, BINAM, VINMA, INARFIMA et BINFIMA sont appliqués aux données tick-by-tick des marchés financiers, Chaque tick représente une modification, par exemple, d'une cotation ou une transaction.

Définition 1.11.2 (Modèle INMA (1))

Le modèle INMA d'ordre 1, **INMA (1)** est introduit par Al-Osh et Alzaid [5] et sous une forme légèrement differente par McKenzie . Ces études ont supposé une distribution de Poisson pour les séries temporelles. Alors, l'INMA (1) d'Al-Osh et Alzaid est défine comme suit

$$X_t = \beta_1 \circ \varepsilon_{t-1} + \varepsilon_t. \tag{1.78}$$

Où le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corrélées, à valeurs entières positives, avec une moyenne μ_ε et une variance finies σ_ε^2 et où "$\circ$ " est un opérateur d'amincissement de "*Steutel-Van Harn*".

***Remarque* 1.11.2**

Le modèle (1.78) est similaire dans sa forme au processus standard MA (1) dans lequel la multiplication scalaire est remplacée par opération "$\beta \circ$ ", de la même manière que l'INAR(1) a été défini.
Le modèle INMA (q) peut se modalisé pour le nombre de comptages pendant le temps n est composé de deux éléments :

- (i) les survivants des arrivées pendant le temps n-l ($\beta \circ \varepsilon_{t-1}$).
- (ii) les arrivées pendant le temps n (ε_t) .

Une réalisation de nombreux comptages comme le nombre de patients séjournant dans un hôpital ou le nombre de clients dans un grand magasin, admet une telle interprétation.

1.11.2 Condition de stationnarité stricte

Par définition contrairement à l'INAR (1), le modèle INMA (1) [13] est explicite et le processus $\{X_t, t \in \mathbb{Z}\}$ est strictement stationnaire et dépendant de 1, c'est-à-dire que X_t et X_{t-k} sont indépendants si $k > 1$.

1.11.1.3 Les moments de premier et seconde ordre du processus INMA (1)

La moyenne et la variance des observations X_t sont données par

$$E[X_t] = \mu_\varepsilon(1+\beta), \qquad V[X_t] = \beta(1-\beta)\mu_\varepsilon + (1+\beta^2)\sigma_\varepsilon^2. \tag{1.79}$$

La fonction d'autocovariance pour (q = 1) est

$$Cov(X_{t-1}, X_t) = \gamma(1) = \beta\sigma_\varepsilon^2. \tag{1.80}$$

1.11.1.4 Structure d'autocorrélation

La fonction d'autocorrélation pour (q= 1) vaut

$$\rho(1) = \frac{\beta\sigma_\varepsilon^2}{\beta(1-\beta)\mu_\varepsilon + (1+\beta^2)\sigma_\varepsilon^2}. \tag{1.81}$$

L'indice de dispersion I peut être calculé à partir de l'indice de dispersion I_ε de ε_t.

$$I = \frac{\sigma^2}{\mu}, \qquad I_\varepsilon = I - \frac{\theta(1-\theta)}{\theta+1}\left(\frac{1+\theta^2}{\theta+1}\right). \tag{1.82}$$

***Remarque* 1.11.3**

Si $(k > 1)$ la fonction (ACF) égle a 0 $(\rho(k) = 0)$, Il s'agit d'une condition nécessaire et suffisante pour qu'un processus $\{X_t, t \in \mathbb{Z}\}$ soit un INMA(q) .

1.11.1.5 Un exemple de données réelles

Pour illustrer les modèles INMA, considérons l'ensemble de données présentant les nombres de transactions sur des intervalles de 30 minutes de négociation d'AstraZeneca [81].Chaque numéro d'observation correspond à une minute de temps. Ce type de série de données comprend une fréquence nulle et elle est motivée vers un modèle de données de comptage.

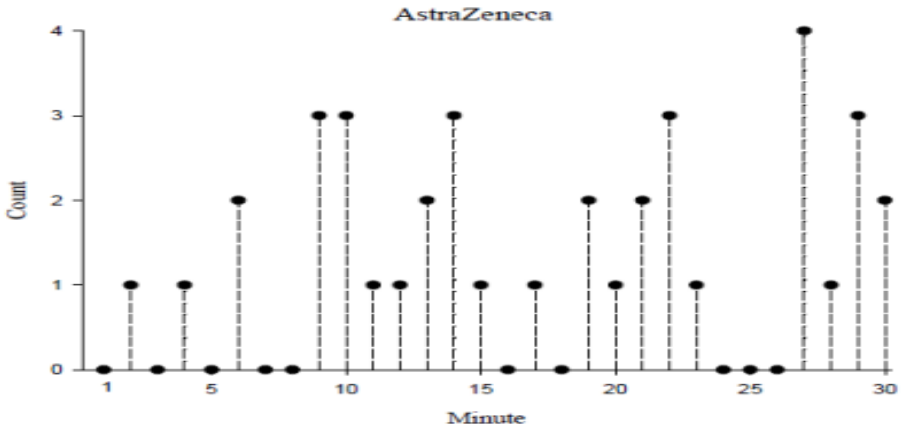

FIGURE 1.10 – Les données relatives au nombre de transactions sur des intervalles de 30 minutes de négociation d'AstraZeneca.

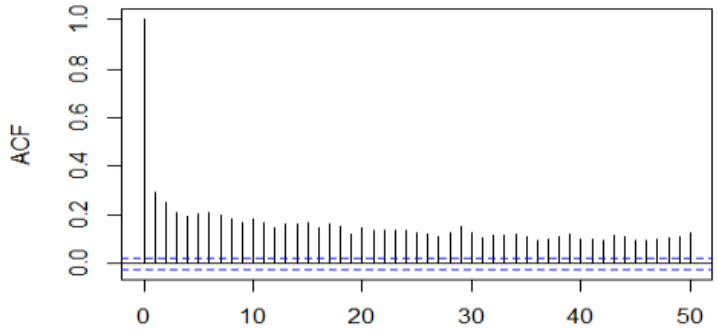

FIGURE 1.11 – La fonction d'autocorrélation pour les données de transactions boursières agrégées sur un intervalle de temps de cinq minutes pour AstraZeneca.

1.11.2 Estimation des paramètres du modèle INMA (p)

Brännäs et Quoreshi [16] fournissent des propriétés de moyenne et de variance conditionnelles et appliquent ces propriétés dans les estimateurs des moindres carrés conditionnels (CLS) lorsque les données sont générées selon un modèle INMA à décalage infini. Dans l'étude de Monte Carlo [81] l'estimateur des moindres carrés est le meilleur choix, pour le modèle VINMA et le modèle à mémoire longue. Les auteurs n'ont pas considéré le maximum de vraisemblance en raison des distributions sous-jacentes inconnues des innovations. Récemment, d'autres recherches sur la modélisation INMA ou INMA bivariée ont montré qu'une approche d'estimation alternative et robuste, appelée Quasi-Vraisemblance Généralisée (GQL), peut être utilisée pour estimer les paramètres inconnus. L'équation d'estimation GQL est issue de l'équation d'estimation de la vraisemblance basée sur la famille de dispersion exponentielle [64], l'approche GQL produit des estimations asymptotiquement aussi efficaces que l'approche basée sur le maximum de vraisemblance et l'approche CLS [70], [98].

1.12 Le modèle INARMA (p, q)

le modèle INARMA pour les séries chronologiques de comptage de type ARMA , **INARMA (p, q)**, a été proposé pour la première fois indépendamment par Mckenzie [67] et Al-Osh et Alzaid [3]. Et aussi une récente étude a été proposés par Christian Weiss.
Cette classe de modèles INARMA est basés sur l'idée est de modifier la récursion ARMA ordinaire en remplaçant les multiplications par des opérateurs d'amincissement binomial de Steutel et Van Harn , qui combine une variable aléatoire (X_t) à valeur entière non négative avec une probabilité α_i, ces opérateurs sont les plus fréquemment utilisés par les analystes des séries temporelles à valeurs entières positives.

1.12.1 Structure probabiliste du modèle INARMA (p, q)

1.12.1.1 Défnitions

Définition 1.12.1 (Modèle INARMA (p, q))

Un processus à valeurs entières positives $\{X_t, t \in \mathbb{Z}\}$ est un processus autoregressif moyen mobile, d'ordre p et q, **INARMA (p,q)** [12] (voir : Alzaid (1990), Benjamin et al

(2003), ...) s'il est défini comme suit

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \sum_{j=1}^{q} \beta_j \circ \varepsilon_{t-j} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.83}$$

Où le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corrélées, à valeurs entières positives, avec une moyenne μ_ε et une variance finies σ_ϵ^2 et "$\circ$" est un opérateur d'amincissement dont le fréquemment utilisé est celui de "*Steutel-Van Harn*" avec $1 \leq i \leq p$ et $1 \leq j \leq q$.

Définition 1.12.2 (Modèle INARMA (1, 1))

Un processus à valeurs entières positives $\{X_t, t \in \mathbb{Z}\}$ est un processus autoregressif moyen mobile, d'ordre 1 et 1 (INARMA (1, 1)) [19] [91] s'il est défini comme suit

$$X_t = \alpha_1 \circ X_{t-1} + \beta_1 \circ \varepsilon_{t-1} + \varepsilon_t. \tag{1.84}$$

Où les paramètres $\alpha_1, \beta_1 \in (0, 1)$ sont utilisés au sein des opérations d'amincissement binomial, et celles-ci sont supposées être exécutées indépendamment les unes des autres.

De nombreuses séries temporelles non négatives à valeurs entières rencontrées dans des différents domaines tels que (l'épidémiologie, l'économie, la criminologie, l'environnement,...) qui révèlent la caractéristique de périodicité dans leurs structures d'autocovariance et la plupart des résultats existants concernant la modélisation des séries temporelles de comptage linéaires périodiques sont principalement consacrés à l'étude des modèles autorégressifs périodiques à valeurs entières (PINAR) [95], [63].
La classe des modèles de moyenne mobile autorégressive à valeurs entières PINARMA(p, q) elle est présenter comme suit

Définition 1.12.3 (Modèle PINARMA (p, q))

Un processus $\{X_t, t \in \mathbb{Z}\}$ à valeurs entières corrélé périodiquement au sens de Gladyshev [39], de période S (où $S \geq 2$), est dit satisfaire un modèle de moyenne mobile autorégressive à valeurs entières périodique, d'ordre p et q, noté **PINARMA$_S$(p, q)** [10] [95] si

$$X_t = \sum_{i=1}^{p} \alpha_{i,t} \circ X_{t-i} + \sum_{j=1}^{q} \beta_{j,t} \circ \varepsilon_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.85}$$

Où le processus à valeurs entières positives $\{X_t, t \in \mathbb{Z}\}$, est périodiquement corrélé, de période entières positive S ($S \geq 2$).

Le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corrélées à valeurs entières positives, avec une moyenne périodique "$\mu_{\varepsilon,t}$", et une variance périodique finies "$\sigma^2_{\varepsilon,t}$" et où les paramètres $\alpha_t, \beta_t, \mu_{\varepsilon,t}, \sigma^2_{\varepsilon,t}$ sont périodiques, par rapport à t, de période S ($S \geq 2$) i.e.

$$\alpha_{t+Sr} = \alpha_t, \qquad \beta_{t+Sr} = \beta_t, \qquad \mu_{\varepsilon,t+Sr} = \mu_{\varepsilon,t}, \qquad \sigma^2_{\varepsilon,t+Sr} = \sigma^2_{\varepsilon,t}, \qquad \forall t, r \in \mathbb{Z}$$

En particulier, nous avons, pour (p = q = 1), le modèle de moyenne mobile autorégressive à valeurs entières périodique d'ordre 1 et 1, **PINARMA$_S$(1, 1)** est défini comme suit

$$X_t = \alpha_t \circ X_{t-1} + \beta_t \circ \varepsilon_{t-1} + \varepsilon_t, \qquad t \in \mathbb{Z} \tag{1.86}$$

où les paramètres α_t et $\beta_t \in [0, 1]$, sont périodiques en t, avec une période S, c'est-à-dire

$$\alpha_{t+Sr} = \alpha_t, \qquad \beta_{t+Sr} = \beta_t, \qquad \forall t, r \in \mathbb{Z}.$$

***Remarque* 1.12.1**

Les modèles de moyenne mobile autorégressive périodique à valeurs entières **PINARMA** peut réduire sensiblement le nombre de paramètres à prendre en compte dans un modèle autorégressif périodique pur à valeurs entières.

1.12.1.2 Condition de stationnarité de second ordre

Pour garantir que le processus INARMA (p, q) [71] est stationnaire (de second ordre), nous exigeons que $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_p\}$ soient les racines du polynôme d'ordre p sont à l'intérieur du cercle unitaire (voir Latour, 1997) [59].

$$X^p - \alpha_1 X^{p-1} - ... - \alpha_{p-1} X - \alpha_p = 0$$

Nous utiliserons cependant le critère plus fort, mais plus facile à vérifier, selon lequel

$$\sum_{i=1}^{p} \alpha_i < 1,$$

pour garantir que le processus (1.83) est stationnaire (de second ordre).
La contrainte correspondante sur les θ pour l'inversibilité de la série temporelle est que $\sum_{j=1}^{q} \beta_j < 1$.

1.12.1.3 Les moments de premier ordre et seconde ordre du processus INARMA(1,1)

On supposons que le processus INARMA (1, 1) défini par (1.84) est stationnaire (de second ordre) . Alors

$$E[X_t] = \frac{(1+\beta_1)\mu_\varepsilon}{1-\alpha_1}, \quad V[X_t] = \frac{[\alpha_1(1+\beta_1)+\beta_1(1-\beta_1)\mu_\varepsilon]+[\beta_1(\beta_1+2\alpha_1)+1]\mu_\varepsilon}{1-\alpha_1^2}. \tag{1.87}$$

***Remarque*1.12.2**

Pour $\beta_1 = 0$, le modèle (1.84) se réduit au modèle INAR (1) de McKenzie [65] et pour $\alpha_1 = 0$, il se réduit au modèle INMA (1) récemment étudié par Aleksandrov et Weiss [13]. Les applications des modèles INARMA (p, q) sont fréquentes ; par exemple, l'étude du nombre de crises d'épilepsie Franke et Seligmann, (1993) et les applications à l'économie Brännäs et Hellström, (2001) ; Brännäs et Shahiduzzaman, (2004) ; Böckenholt, (1999b) ; Böckenholt, (2003) ; Rudholm, (2001) ; Freeland et McCabe, (2004).

1.12.2 Estimation des paramètres du modèle INARMA

L'estimation du maximum de vraisemblance (MLE) des paramètres (α, β, μ) est irréalisable compte tenu des données X, une possibilité est l'augmentation des données. L'augmentation des données dans le cadre de MLE est généralement réalisée à l'aide de l'algorithme EM (voir Dempster et al. (1977)). Alors les MLE pour $\Theta = (\alpha, \beta, \mu)$ sont

$$\hat{\alpha}_i = \sum_{t=1}^{n} y_{t,i} / \sum_{t=1}^{n} X_{t-1} \tag{1.88}$$

$$\hat{\beta}_i = \sum_{t=1}^{n} v_{t,i} / \sum_{t=1}^{n} \varepsilon_{t-1} \tag{1.89}$$

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t. \tag{1.90}$$

***Remarque* 1.12.3**

Un algorithme (EM) de Monte-Carlo peut être utilisé pour l'estimation des ces paramètres, mais il est très lent en raison de la lenteur de la convergence des paramètres et du nombre élevé de rejets.

1.13 Le modèle INBL (p, q, m, n)

En utilisant le concept de l'amincissement binomial, les modèles bilinéaires conventionnels peuvent aussi être adaptés au cas des entiers, ce qui conduit à la classe des modèles bilinéaires à valeurs entières (**INBL**) similaire au processus bilinéaire à valeurs réelles.

1.13.1 Structure probabiliste du modèle INBL (p)

1.13.1.1 Défnitions

Définition 1.13.1 (Modèle INBL (p, q, m, n))

Un processus stochastique, à valeurs entières, $\{X_t, t \in \mathbb{Z}\}$ du second ordre, est dit satisfaisant un modèle Bilinéaire INBL super diagonal à valeurs entières qui été introduit par Drost et al [31] qui est une classe particulière de la classe plus générale
INBL (p, q, m, n) s'il est défini comme suit

$$X_t = \sum_{i=1}^{p} \alpha_i \circ X_{t-i} + \sum_{j=1}^{q} \beta_i \circ \varepsilon_{t-j} + \sum_{r=1}^{m}\sum_{s=1}^{n} \gamma_{r,s} \circ (X_{t-r}\varepsilon_{t-s}) + \varepsilon_t, \quad t \in \mathbb{Z}. \tag{1.91}$$

Où $\gamma_{r,s}$= 0 si $r < s$ et (ε_t) est une séquence de variables aléatoires i.i.d. de valeurs non négatives et à valeurs entières positives avec une moyenne (μ_ε) et une variance (σ_ε^2) finies, indépendant des opérateurs.

Définition 1.13.2 (Modèle INBL (1, 0, 1, 1))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ du second ordre, est dit satisfaisant un modèle de séries chronologiques Diagonal Bilinéaire du premier ordre à valeurs entières de coefficients invariant dans le temps, **INBL (1, 0, 1, 1)** qui a été étudié par Doukhan et al [76] s'il est une solution de l'équation linéire au différence stochastique suivante

$$X_t = \alpha_1 \circ X_{t-1} + \gamma_{1,1} \circ (X_{t-1}\varepsilon_{t-1}) + \varepsilon_t, \qquad t \in \mathbb{Z}. \tag{1.92}$$

Où le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corrélées à valeurs entières positives, avec une μ_ε et une variance finie σ_ε^2.

Remarque **1.13.1**

La classe de modèles INBL (p, q, m, n) est particulièrement adaptée à la modélisation de processus qui prennent des valeurs faibles avec une forte probabilité, mais qui présentent en même temps des explosions soudaines de valeurs importantes.

Le fait que plusieurs séries rencontrées à valeurs entières exhibent une caractéristique de périodicité dans leur structures d'autocovariance qui ne peuvent pas être prises en compte et décrites dans la classe des modèles de séries entières à paramètres invariants dans le temps, a donné une bonne raison et une grande motivation à l'extension d'une classe de modèles de bilinear invariables dans le temps à une classe de modèles à paramètres variants dans le temps, et le premier papier traitant la modélisation du processus périodiquement corrélé, dans le sens de Gladyshev (1963) est celui de Monteiro et al [63].

Définition 1.13.3 (Modèle PIBL (p))

Le modèle Bilinéaire du premier ordre périodique à valeurs entières non négatives $\mathbf{PINBL}_S(1,0,1,1)$ [12] est défini comme suit

$$X_{t+Sr} = \alpha_t \circ X_{t-1+Sr} + \varepsilon t + Sr + \beta_t \circ X_{t-1+Sr}\varepsilon_{t-1+Sr}, \quad t = 0,1,...,S.et \quad r \in \mathbb{Z}. \quad (1.93)$$

Où le processus à valeurs entières positives $\{X_t, t \in \mathbb{Z}\}$, est périodiquement corrélés, de période entiéres positive S ($S \geq 2$), et le processus d'innovation $\{\varepsilon_t, t \in \mathbb{Z}\}$ est une suite de variables aléatoires non corréles à valeurs entières positives, avec une moyenne périodique "$\mu_{\varepsilon,t}$", et une variance périodique finies "$\sigma^2_{\varepsilon,t}$" et où les paramètres $\alpha_t, \beta_t, \mu_{\varepsilon,t}, \sigma^2_{\varepsilon,t}$ sont périodiques par rapport à t de période S ($S \geq 2$) i.e.

$$\alpha_{t+Sr} = \alpha_t, \qquad \beta_{t+Sr} = \beta_t, \qquad \mu_{\varepsilon,t+Sr} = \mu_{\varepsilon,t}, \qquad \sigma^2_{\varepsilon,t+Sr} = \sigma^2_{\varepsilon,t}, \qquad \forall t, r \in \mathbb{Z}.$$

1.13.1.2 Condition de stationnarité stricte

La stationnarité stricte pour les processus INBL (p, q, m, n) pour ($\gamma_{r,s}$= 0 si $r < s$) et avec l'utilisation des techniques de chaînes de Markov, nous obtenons l'existence et l'unicité d'une solution strictement stationnaire de (1.91) lorsque [31]

$$\sum_1 \alpha_i + \mu_\varepsilon \sum_{r,s} \gamma_{r,s} < 1. \quad (1.94)$$

Sous la condition $(\alpha+\beta\mu_\varepsilon)^2+\beta^2\sigma^2_\varepsilon < 1$ alors il existe un unique processus du second ordre, strictement stationnaire $\{X_t, t \in \mathbb{Z}\}$ suivant le modèle (1.92) et tel que $\{\varepsilon_t\}$(est distribué par P(μ)) est indépendant de $\{X_s\}$,$s < t$ [76].

1.13.1.3 Les moments de premier ordre et seconde ordre du processus INBL (p, q, m, n)

Le fait que le processus $\{X_t, t \in \mathbb{Z}\}$ est strictement stationnaire, on obtient l'existence des moments des processus INBL (1, 0, 1, 1)

$$E[X_t] = \frac{\beta\sigma_\varepsilon^2 + \mu_\varepsilon}{1 - (\alpha + \beta\mu_\varepsilon)}. \tag{1.95}$$

soit $\gamma(0)$ la variance du processus et la fonction autocovariance vaut si $k = 1$

$$\gamma(1) = (\alpha + \beta\mu_\varepsilon)\gamma(0) + (\alpha + \beta\mu_\varepsilon)E[X_t]^2 - E[X_t]^2 + 2\beta\mu_\varepsilon E[X_t] + \beta\mu_\varepsilon + \mu_\varepsilon E[X_t].$$

si $k \geq 2$

$$\gamma(k) = (\alpha + \beta\mu_\varepsilon)\gamma(k-1) + (\alpha + \beta\mu_\varepsilon)E[X_t]^2 - E[X_t]^2 + \beta\mu_\varepsilon E[X_t] + \mu_\varepsilon E[X_t]. \tag{1.96}$$

Remarque **1.13.2**

Ici la distribution des variables aléatoires de la séquence $\{\varepsilon_t, t \in \mathbb{Z}\}$ est $P(\mu)$, suit la distribution de Poisson avec le paramètre μ_ε, alors $\mu_\varepsilon = \sigma_\varepsilon^2$.

1.13.2 Estimation des paramètres du modèle INBL (1, 0, 1, 1)

Le problème d'estimation associé au processus INBL (1, 0, 1, 1) est plus compliqué que celui associé au processus BL (1, 0, 1, 1). Cependant, seule la méthode des moments a été étudiée jusqu'à présent comme technique d'estimation [76].

Étant donné les observations $(X_1, ..., X_n)$.

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_t \tag{1.97}$$

et

$$\hat{\gamma(k)} = \frac{1}{n}\sum_{t=1}^{n-k}(X_{t+k} - \bar{X})(X_t - \bar{X}). \tag{1.98}$$

A partir de (1.97) et (1.98), nous obtenons les estimateurs de moment $\hat{\mu}$, $\hat{\beta}$ et $\hat{\alpha}$ des paramètres correspondants μ, β, et α comme suit :

$$\hat{\mu} = \bar{X_n}(1 - \hat{A}) - \hat{B}, \qquad \hat{\beta} = \frac{\hat{B}}{\hat{\mu}}, \qquad \hat{\alpha} = \hat{A} - \hat{B}, \tag{1.99}$$

où

$$\hat{A} = \hat{\gamma(2)}/\hat{\gamma(1)}$$

et

$$\hat{B} = (\hat{\gamma(1)} - \hat{A}\hat{\gamma(0)})/(\bar{X} + 1).$$

Remarque **1.13.3**

Les estimateurs de moment $\hat{\mu}$, $\hat{\beta}$ et $\hat{\alpha}$, définés dans (1.99) sont fortement cohérents.

Conclusion

La classe de séries temporelles à valeurs entières basées sur l'opérateur d'amincissement est une classe assez large, et ce chapitre vise à limiter l'étude pour quelques modèles de cette classe et son contenu a donné différentes définitions et propriétés de base de l'opérateur d'amincissement, de plus, les structures probabilistes qui caractérisent les modèles proposés ci-dessus basés sur le concept d'amincissement, en plus de certains principes d'estimation mentionnés dans la littérature pour chaque modèle.

2

Modèles de séries chronologiques à valeurs entières basés sur la régression discrete

2.1 Introduction

Les séries chronologiques à valeurs entières non négatives sont rencontrées dans divers domaines, tels que l'épidémiologie, l'économie, l'environnement, la criminologie, ...etc. Cela a incité l'introduction d'une classe de modèles linéaires et non linéaires avec des valeurs entières linéaires dans la littérature sur les séries chronologiques. Les modèles examinés dans le chapitre précédent ont utilisé des types d'opérateur d'amincissement pour transférer le modèle ARMA à l'état des données de comptage. Une autre approche populaire pour modéliser de tels comptages stationnaires qui contient plusieurs modèles proposés dans la littérature qui appartiennent à la catégorie des modèles de régression.

Ainsi, dans ce chapitre, on envisage à étudier ces modèles de régression pour les séries temporelles à valeurs discrètes (entières) ainsi leurs aspects probabilistes tels que : la distribution marginale, les conditions de stabilité et les structures de corrélation associées sont également soulignées. Enfin, nous nous concentrerons sur certaines des méthodes d'estimation considérées dans la littérature.

2.2 Modèle INGARCH (1, 1)

La modélisation de séries chronologiques à valeurs entières a vu le développement de quelques modèles de régression non linéaire à valeurs entières qui n'utilisent pas l'opérateur d'amincissement. Parmi ceux-ci, Un modèle autorégressif conditionnel hétéroscédastique généralisé à valeurs entières qui a été étudié par Ferland et al [82] et précédemment par Rydberg et Shephard [84], noté INGARCH (1, 1).

2.2.1 Structure probabiliste du modèle INGARCH (1, 1)

2.2.1.1 Définitions

Définition 2.2.1 (Modèle INGARCH (1, 1))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières, d'ordre 1 et 1, noté **INGARCH (1, 1)**, s'il est donné par :

$$\begin{cases} X_t|F_{t-1} \sim P(\lambda_t) \qquad t \in \mathbb{Z}, \\ \lambda_t = \lambda_t(\theta) = \omega + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1}(\theta), \end{cases} \tag{2.1}$$

où F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$ " l'historique du processus" et $\lambda_t(\theta)$ est la moyenne conditionnelle avec $\theta = (\omega, \alpha_1, \beta_1)^T \in \Theta \subset R^m$ et X_t est non dégénéré.
Nous considérons le changement du bruit suivant $\varepsilon_t = X_t - \lambda_t$ pour lequel l'équation (2.1) devient

$$X_t = \lambda_t + (X_t - \lambda_t) = \omega + (\alpha_1 + \beta_1)X_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1}, \tag{2.2}$$

sachant que $\{\varepsilon_t\} = \{X_t - \lambda_t\}$ est considérée comme bruit blanc

1. D'une moyenne :

$$E(\varepsilon_t) = E(X_t - \lambda_t) = E\left(E(X_t - \lambda_t|F_{t-1})\right) = 0.$$

2. Sa variance :

$$V(\varepsilon_t) = Var\left(E(\varepsilon_t|F_{t-1})\right) + E\left(Var(\varepsilon_t|F_{t-1})\right) = E\left(Var(X_t|F_{t-1})\right) = E(\lambda_t) = \mu.$$

3. Convariance :

$$Cov(\varepsilon_t, \varepsilon_{t+h}) = E\left(\varepsilon_t E(\varepsilon_{t+h}|F_{t+h-1})\right) = 0.$$

2.2.1.2 Condition de stationnarité et d'ergodicité

Pour le modèle INGARCH (1, 1) si les paramètres sont positifs et inférieurs à 1 alors ce processus est strictement stationnaire et ergodique car le processus est une chaîne de Markov géométriquement ergodique avec des moments finis donc tous les moments existent (Condition suffisante) [58].

$$0 < \alpha_1 + \beta_1 < 1. \tag{2.3}$$

Exemple

La figure suivante représente la série temporelle du modèle INGARCH(1, 1) de Ferland et al [82]. Les graphes (a), (b) représente le processus INGARCH(1, 1) avec $\omega = 0.5, \alpha = 0.3$, et $\beta = 0.5$, qui est un cas stationnaire tandis que les graphes (c), (d) avec $\omega = 0.1, \alpha = 0.5$, et $\beta = 0.5001$, qui est un cas non stationnaire.

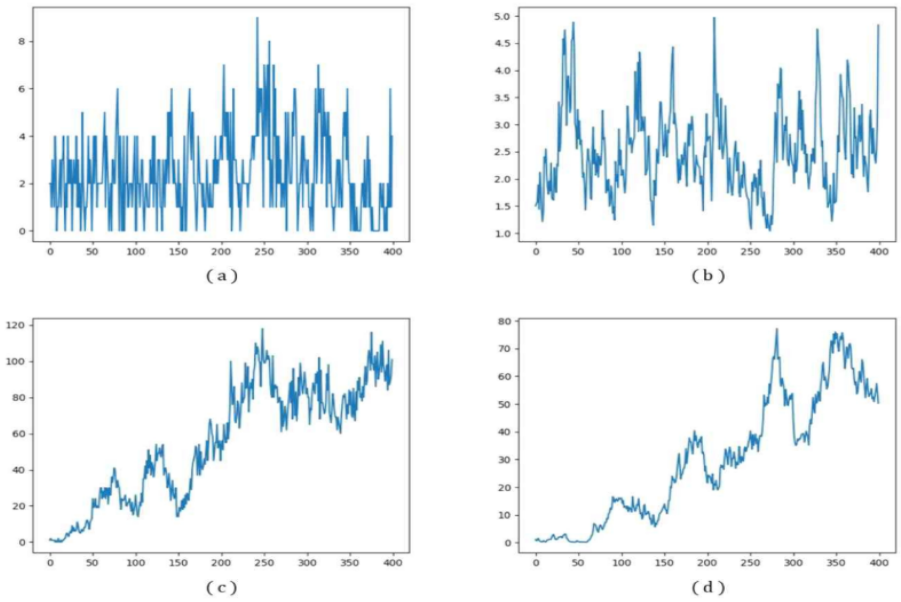

FIGURE 2.1 – Processus X_t du modèle INGARCH (1, 1).

2.2.1.3 Les moments de premier et second ordre du processus

Les moments d'un INGARCH (1, 1) sont tous finis si, et seulement si la condition (2.3) est verifie, dans ce cas

$$E[X_t] = E[\lambda_t] = \mu_X = \frac{\omega}{1 - (\alpha_1 + \beta_1)}, \tag{2.4}$$

$$V[X_t] = V[\varepsilon_t] + V[\lambda_t] = \frac{(1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2)}{1 - (\alpha_1 + \beta_1)^2}\mu_X. \tag{2.5}$$

De plus, sa fonction d'autocovariance est

$$\gamma(k) = \frac{\alpha_1(1-\beta_1(\alpha_1+\beta_1))(\alpha_1+\beta_1)^{k-1}\mu_X}{1-(\alpha_1+\beta_1)^2} \qquad \forall k \geq 1, \tag{2.6}$$

si (k=0) alors $\gamma(0) = V[X_t]$.

2.2.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un INGARCH (1, 1) vaut

$$\rho(k) = (\alpha_1+\beta_1)^{k-1}\frac{\alpha_1(1-\beta_1(\alpha_1+\beta_1))}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}, \qquad \forall k \geq 1. \tag{2.7}$$

***Remarque* 2.2.1**

Le modèle INGARCH (1, 1) offre une manière parcimonieusement paramétrée de représenter une longue mémoire.
Malgré le fait que la distribution conditionnelle de Poisson est équidispersée (la variance égale à la moyenne), la distribution inconditionnelle présente une surdispersion.

2.2.2 Estimation des paramètres du modèle INGARCH (1, 1)

La procédure d'estimation des paramètres du modèle INGARCH est similaire à la procédure utilisée dans le modèle GARCH traditionnel. le modèle INGARCH (1, 1) est déterminé par des paramètres qui caractérise la distribution marginale des observations $\Theta = (\omega, \alpha_1, \beta_1)$. Compte tenu des données d'une série chronologique $(X_1, X_2, ..., X_n)$, il s'agit d'estimer les valeurs de ces paramètres.

2.2.2.1 Méthode du maximum de vraisemblance

La fonction de vraisemblance conditionnelle des n observations $(X_1, ..., X_n)$ d'un échantillon, est donnée par

$$L(\theta) = \prod_{t=1}^{n} = \frac{e^{-\lambda_t}\lambda_t^{X_t}}{X_t!}, \tag{2.8}$$

où $\Theta = (\omega, \alpha_1, \beta_1)$ et $\lambda_t = \omega + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1}$.
Il est impossible de trouver des estimations de cette fonction de vraisemblance. En fait un système d'équations non linéaires est obtenu. Pour cette raison l'utilisation d'une méthode numérique d'optimisation pour trouver la valeur optimale de θ. Comme d'habitude, avec l'utilisation de la fonction logarithme de vraisemblance :

$$\log L(\theta) = \sum_{t=1}^{n}[X_t \log \lambda_t - \lambda_t - \log(X_t!)], \tag{2.9}$$

2.2.2.2 Méthode du quasi-maximum de vraisemblance

Ahmad et Francq [1] ont proposé un estimateur QML (quasi-maximum de vraisemblance) pour le paramètre θ de la moyenne conditionnelle, qui est défini comme étant une solution mesurable au problème suivant :

$$\hat{\theta_n} = argmax_{\theta\in\Theta}(\tilde{L}_n(\theta)), \tag{2.10}$$

où

$$\tilde{L}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(-\tilde{\lambda}_t(\theta) + X_t \log\tilde{\lambda}_t(\theta)\right).$$

Pour le processus INGARCH (1, 1) poissonien, nous définissons l'espace de stationnarité par :

$$\Theta_1 = \theta \in \mathbb{N}^2, \alpha + \beta < 1.$$

Si $\theta \in \Theta$, alors ce processus est strictement stationnaire et ergodique [1]. Nous pouvons énnoncer la proposition suivante

Proposition 2.2.1 (Ahmed, 2016)

Soit $\theta_0 \in \Theta$ un sous ensemble compact de Θ_1 et X une solution stationnnaire, alors :

1. L.EQMV poissonien est fortement consistent $\hat{\theta}_p \rightarrow^{p.s}_{n\to\infty} \theta_0$.
2. $\sqrt{n}(\hat{\theta}_p - \theta_0) \rightarrow^{\iota}_{n\to\infty} N(0, J_p^{-1})$, où

$$J_p = E\left(\frac{1}{\lambda_t(\theta_0)}\frac{\partial\lambda_t(\theta_0)}{\partial\theta}\frac{\partial\lambda_t(\theta_0)}{\partial\acute{\theta}}\right).$$

Dans ce cas, la matrice de variance asymptotique est telle que

$$V(\hat{\theta}_p) \simeq n^{-1}J_p^{-1} = \left(\sum_{t=1}^{n}\frac{1}{\tilde{\lambda}_t(\hat{\theta}_p)}\frac{\partial\tilde{\lambda}_t(\hat{\theta}_p)}{\partial\theta}\frac{\partial\tilde{\lambda}_t(\hat{\theta}_p)}{\partial\acute{\theta}}\right)^{-1}.$$

2.3 Modèle INGARCH (p, q)

Un modèle couramment utilisé pour les séries temporelles de comptages avec une dispersion excessive est le modèle autorégressif conditionnellement hétéroscédastique généralisé à valeurs entières (INGARCH (p, q)) d'ordre q et p où la distribution conditionnelle est souvent supposée être poissonienne.

2.3.1 Structure probabiliste du modèle INGARCH (p, q)

2.3.1.1 Définitions

Définition 2.3.1 (Modèle INGARCH (p, q))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières avec une surdispersion dont la distribution est supposée poissonienne, d'ordre q et p ;
noté **INGARCH (p, q)** ; s'il est donné par :

$$\begin{cases} X_t|F_{t-1} \sim P(\lambda_t) \qquad t \in \mathbb{Z}, \\ E[X_t|F_{t-1}] = \lambda_t = \lambda_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}(\theta), \end{cases} \tag{2.11}$$

où F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$ et $\omega \geq 0$, $\alpha_i \geq 0$ et $\beta_j \geq 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$ avec une moyenne conditionnelle $\lambda_t = \lambda_t(\theta)$ où : $\theta = (\omega, \alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q)^T \in \Theta \subset R^m$ et X_t est supposée non dégénéré.

Définition 2.3.2 (Prévision à n étapes)

La prévision à n étapes de la série temporelle au temps t X_t est donnée par $E[X_{t+n}|F_t], (n \in \mathbb{N})$ qui est l'espérance conditionnelle de X_t.
Notez que l'estimation de $E[X_{t+n}|F_t]$ est équivalente à l'estimation de $\hat{\lambda_t}(n) = E[\lambda_{t+n}|F_t]$.

$$\hat{\lambda_t}(n) = \omega + \sum_{i=1}^{p} \alpha_i E[\lambda_{t+n-i}|F_t] + \sum_{j=1}^{q} + \sum_{j=1}^{q} \beta_j E[X_{t+n-j}|F_t] = \omega + \sum_{i=1}^{p} \alpha_i \hat{\lambda_t}(n-i) + \sum_{j=1}^{q} + \sum_{j=1}^{q} \beta_j \hat{X_t}(n-j), \tag{2.12}$$

où

$$E[\lambda_{t+n-i}|F_t] = \begin{cases} \lambda_{t+n-i} & si\ i > n, \\ \hat{\lambda_t}(n-i) & si\ n \geq i, \end{cases} \tag{2.13}$$

$$E[X_{t+n-j}|F_t] = \begin{cases} X_{t+n-j} & si\ j > n, \\ \hat{\lambda_t}(n-j) & si\ n \geq j, \end{cases} \tag{2.14}$$

***Remarque* 2.3.1**

Il est bien connu qu'un processus GARCH standard est un processus ARMA en second ordre. Ceci est encore vrai dans le cas où le modèle INGARCH (p,q) donne un modèle INARMA (max{p,q}, q) [41].
Le modèle INGARCH a une variance qui varie avec le temps en raison de la nature du la moyenne conditionnelle. Il offre plusieurs avantages par rapport aux modèles de séries chronologiques continues pour l'analyse des données de comptage. Premièrement, leur similitude avec les modèles ARMA permet une analyse facile. Deuxièmement, ils semblent être plus naturels que d'autres modèles de séries chronologiques de comptage, tels que l'amincissement binomial. Un autre avantage de ces modèles est qu'ils sont capables de gérer à la fois des corrélations positives et négatives.

2.3.1.2 Condition de stationnarité et d'ergodicité

Pour le modèle INGARCH (p, q), Ferland et al [82] ont montré qu'il existe un processus strictement stationnaire $\{X_t, t \in \mathbb{Z}\}$ sous la condition suivante

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{2.15}$$

Le processus $\{X_t, t \in \mathbb{Z}\}$ est une chaîne de Markov avec une probabilité de transition

$$P(X_t|X_{t-1}, X_{t-2}...X_{t-p}) = P(X_t|X_{t-1}).$$

Selon le théorème (4.3.3) de Ross [83], $\{X_t, t \in \mathbb{Z}\}$ est irréductible et apériodique, alors ce processus qui définit le modèle (2.11) est ergodique.

2.3.1.3 Les moments de premier et second ordre du processus

Les moments du processus INGARCH (p, q) sont tous finis si et seulement si la condition de la stricte stationnarité (2.15) est verifie et sont tels que [82]

$$E[X_t] = E[\lambda_t] = \mu_X = \frac{\omega}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j)}, \tag{2.16}$$

$$V[X_t] = E[X_t] + V[\lambda_t], \tag{2.17}$$

où $V[\lambda_t] = \sum_{i=1}^{p} \alpha_i \gamma_X(i) + \sum_{j=1}^{q} \beta_j \gamma_\lambda(i)$ [20],
tel que les autocovariances vaut

$$\begin{aligned} \gamma_X(k) &= Cov[X_t, X_{t-k}]. \\ \gamma_\lambda(k) &= Cov[\lambda_t, \lambda_{t-k}]. \end{aligned}$$

2.3.1.5 Un exemple de simulation

Pour mieux voir le comportement de ce modèle, une simulation d'un processus INGARCH (2, 5) avec les paramètres : $\omega = 1$, $\alpha_i = (0.05, 0.05)$ et $\beta_j = (0.3, 0.2, 0.1, 0, 0.2)$ et comme dim (β_j) = 5 , les 5 premières valeurs de λ sont données, i.e, $(\lambda_0, ..., \lambda_4)$, tel que $\lambda_i = 1$, $\forall i = 0, ..., 4$.
La figure ci-dessus présente l'histogramme de 10000 valeurs simulées auprès le modèle (2.11) suivie par la fonction densité poissoniénne en rouge.

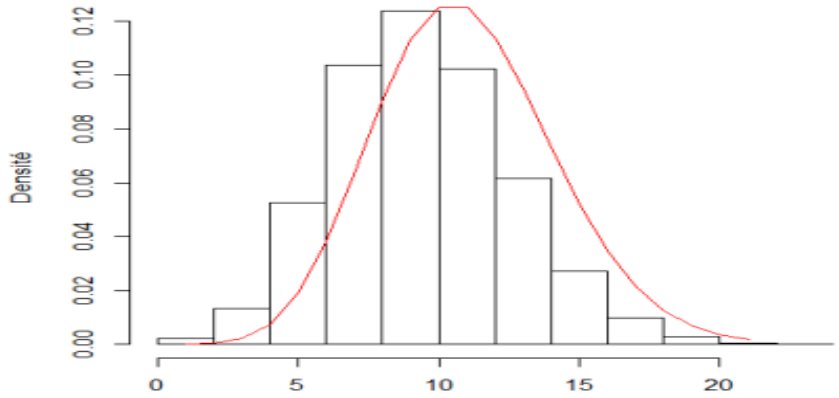

FIGURE 2.2 – Histogramme du processus INGARCH (5, 2) simulé avec la fonction densité poissoniénne en rouge.

Cette derniére exhibe que le processus INGARCH (5, 2) suit une loi poissoniénne qui tend vers la loi Normale. Dans notre cas, la valeur moyenne de λ est autour de 10. Ainsi, la distribution marginale en rouge ressemble à celle de la loi normale.

***Remarque* 2.3.2**

Le modèle INGARCH a une variance qui varie avec le temps en raison de la nature de la moyenne conditionnelle. Il offre plusieurs avantages par rapport aux modèles de séries chronologiques continues pour l'analyse des données de comptage. Premièrement, leur similitude avec les modèles ARMA permet une analyse facile. Deuxièmement, ils semblent être plus naturels que d'autres modèles de séries chronologiques de comptage, tels que l'amincissement binomial. Un autre avantage de ces modèles est qu'ils sont capables de gérer à la fois des corrélations positives et négatives.

2.3.2 Estimation des paramètres du modèle INGARCH (p, q)

Soit $\theta = (\theta_1, \theta_2, ..., \theta_n)^T = (\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)^T$, où l'espace des paramètres est Θ. Pour $\theta \in \Theta$, on définе le processus stationnaire et ergodique

$$\lambda_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}(\theta).$$

Alors la fonction logarithme de vraisemblance est donnée par

$$\log L(\theta) = \sum_{t=1}^{n} [X_t \log \lambda_t(\theta) - \lambda_t(\theta) - \log(X_t!)]. \tag{2.18}$$

Ensuite, le CMLE de θ, $\hat{\theta}$ est défini par

$$\hat{\theta_n} = argmax_{\theta \in \Theta} \log L(\theta). \tag{2.19}$$

***Remarque* 2.3.3**

Si $(X_1, ..., X_t)$ est une série temporelle d'un processus INGARCH (p , q), alors les paramètres peuvent être estimés selon l'approche de Yule-Walker, c'est-à-dire que les autocovariances des échantillons sont insérées dans les équations d'autocovariance (2.3.1.3) [20].

2.4 Modèle GP-INGARCH (p, q)

Le modèle INGARCH (p, q) ne peut traiter que la surdispersion dans les séries temporelles de comptages, mais la sous-dispersion peut également être rencontrée dans des applications réelles.
Zhu [105] a proposé un INGARCH avec une distribution conditionnelle de Poisson généralisée (GP) peut tenir compte à la fois de la surdispersion et de la sous-dispersion.

2.4.1 Structure probabiliste du modèle GP-INGARCH (p, q)

2.4.1.1 Définitions

Définition 2.4.1 (Modèle GP-INGARCH (p, q))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle Autorégressif Conditionnellement Hétéroscédastique Généralisé à valeurs entières avec une

surdispersion dont la distribution est supposée poissonienne généralisée, d'ordre p et q : noté **GP-INGARCH (p, q)** ; s'il est donné par :

$$\begin{cases} X_t|F_{t-1} \sim GP(\lambda_t^*, k) & t \in \mathbb{Z}, \\ \frac{\lambda_t^*}{1-k} = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2.20}$$

où $k < 1$ et F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$ et $\omega \geq 0$, $\alpha_i \geq 0$ et $\beta_j \geq 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$.
Le modèle GP-INGARCH (1, 1) est le modèle le plus simple de la famille GP-INGARCH, avec (p = q = 1), c'est-à-dire

$$\begin{cases} X_t|F_{t-1} \sim GP(\lambda_t^*, k) & t \in \mathbb{Z}, \\ \frac{\lambda_t^*}{1-k} = \lambda_t = \omega + \alpha_1 X_{t-1} + \beta_1 \lambda_{t-1}. \end{cases} \tag{2.21}$$

***Remarque* 2.4.1**

Lorsque $q = 0$, le modèle ci-dessus est dénoté par **GP-INARCH (p)**. Clairement, avec un $k = 0$, le modèle (2.20) se réduit au modèle (2.11).

2.4.1.2 Condition de stationnarité et d'ergodicité

Pour le modèle GP-INGARCH (p, q) Ferland et al [82] ont montré qu'il existe un processus strictement stationnaire et ergodique $\{X_t, t \in \mathbb{Z}\}$ satisfaisant ce modéle, sous la condition suivante

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{2.22}$$

2.4.1.3 Les moments de premier et second ordre du processus

Les moments du processus GP-INGARCH (p, q) sont tous finis si et seulement si la condition de la stricte satationnarité (2.22) est verifie et sont tels que

$$E[X_t] = \mu_X = \frac{\omega}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j)}, \tag{2.23}$$

$$V[X_t] = E[\phi^2 \lambda_t] + V[\lambda_t] = \phi^2 \mu_X + V[\lambda_t], \tag{2.24}$$

où $\phi = 1/(1-k)$.

2.4.1.4 Structure d'autocorrélation

Considérons le modèle GP-INGARCH (1, 1). Les autocorrélations sont défini comme suit :

$$\rho_X(k) = (\alpha_1+\beta_1)^{k-1}\frac{\alpha_1(1-\beta_1(\alpha_1+\beta_1))}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}, \qquad \forall k \geq 1, \tag{2.25}$$

$$\rho_\lambda(k) = (\alpha_1+\beta_1)^k, \qquad \forall k \geq 0. \tag{2.26}$$

2.4.1.5 Un exemple de simulation

Un exemple réel a été traité pour montrer l'utilité et la bonne performance du modèle $GP-INGARCH(p,q)$ dans la modélisation de données de comptage surdispersées, Il s'agit d'une série de recensements annuels des séismes majeurs (+7) pour les années 1900-2006.

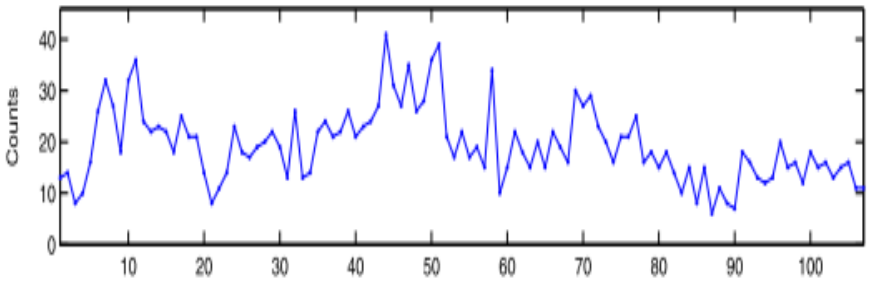

FIGURE 2.3 – La courbe temporelle de la série des grandes séismes durant (1900-2006).

Ces données ont été initialement analysées par Zucchini et MacDonald [58]. La série présente une forte dépendance positive car la moyenne égale (19,3645) et la variance vaut (51,5734) ce qui indique que la distribution marginale est surdispersée.
Le résultat de cette étude montre que le modèle proposé a non seulement une bonne performance dans la modélisation des données surdispersées mais a également la capacité de modéliser le phénomène de sous-dispersion.

2.4.2 Estimation des paramètres du modèle GP-INGARCH (p, q)

Soit $\theta^* = (\omega, \alpha_1, \beta_1)^T$, $\theta = (\phi, \theta^*)^T$ où $\phi = 1/(1-k)$, alors l'ecriteure de la vraie valeur de θ est $\theta = (\phi, \omega, \alpha_1, \beta_1)^T$. Supposons que l'observation $X = (X_1, ..., X_n)$ soit générée à partir du modèle (2.21). Donc la fonction de vraisemblance conditionnelle est

$$L(\phi, \omega, \alpha_1, \beta_1) = \prod_{t=2}^{n} \frac{\lambda_t [\lambda_t + (\phi-1)X_t]^{X_t - 1} \phi^{-X_t} e^{-[\lambda_t + (\phi-1)X_t]/\phi}}{X_t!}, \tag{2.27}$$

alors la fonction de vraisemblance logarithmique $\ln L(\phi, \omega, \alpha_1, \beta_1)$ est donnée par

$$\ln L(\phi, \omega, \alpha_1, \beta_1) = \sum_{t=2}^{n} (\ln \lambda_t + (X_t - 1) \ln(\lambda_t + (\phi-1)X_t) - X_t \ln \phi - \frac{(\lambda_t + (\phi-1)X_t)}{\phi} - \ln X_t!).$$

La solution de l'équation de score $S_n(\theta) = 0$, si elle existe elle donne la (MLE) conditionnelle de θ, notée $\hat{\theta}$ tel que

$$S_n(\theta) = \frac{\partial \ln L(\theta)}{\partial \theta} = \sum_{t=2}^{n} \frac{\partial L(\theta)}{\partial \theta}.$$

Une présentation plus précise pour la méthode d'estimation (MLE) a été cité par [105] avec une autre étude qui établit les propriétés asymptotiques du MLE $\hat{\theta}$.

2.5 Modèle NB-INGARCH (p, q)

Les séries chronologiques de comptage non négatif sont souvent trop dispersées. C'est-à-dire que sa variance est supérieure à la moyenne et la principale raison de la surdispersion est une corrélation positive entre les événements observés.
Pour expliquer cette surdispersion et la traiter, de nombreux chercheurs se sont tournés vers les modèles de régression de Poisson binomiale. En particulier, la distribution binomiale négative (NB) qui représente une extension normale de la distribution de Poisson qui est assez élastique permettant une surdispersion.

2.5.1 Structure probabiliste du modèle NB-INGARCH (p, q)

2.5.1.1 Définitions

Définition 2.5.1 (Modèle NB-INGARCH (p, q))

Un prossecus $\{X_t, t \in \mathbb{Z}\}$ est un modèle binomial négatif$_k$ NB_k-INGARCH (p, q) si sa distribution conditionnelle est une binomiale négative [17].

$$\begin{cases} X_t|F_{t-1} \sim NB(r_t, p_t), & t \in \mathbb{Z}, \\ \frac{1-p_t}{p_t} = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2.28}$$

où F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_t\}$ et $\omega > 0, \alpha_i > 0, \beta_j > 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$ et $\lambda_t = \lambda_t(\theta_0)$ satisfait la représentation INGARCH (p, q).
Avec $r \in \mathbb{N}$ est le nombre de succès et p_t est la probabilité de succès de la variable dans le temps t où

$$r_t = r\lambda_t^{2-k},$$

et

$$p_t = \frac{r\lambda_t^{2-k}}{r\lambda_t^{2-k} + \lambda_t}.$$

Considérant le modèle $NB_1 - INGARCH(p, q)$ qui correspond à $k = 1$

$$X_t|F_{t-1} \sim NB(r\lambda_t, \frac{r\lambda_t}{r\lambda_t + \lambda_t}) \equiv NB(r\lambda_t, \frac{r}{r+1}), \tag{2.29}$$

ce modèle est strictement stationnaire et ergodique, avec un moment d'ordre deux fini, sous la même condition de stationnarité (2.15) pour le modèle INGARCH poissonien [9].

Considérant maintenant le modèle $NB_2 - INGARCH(p, q)$ correspondant à (2.28) avec $k = 2$

$$X_t|F_{t-1} \sim NB(r, \frac{r}{r+\lambda_t}). \tag{2.30}$$

Ce modèle (2.30) a été considéré par Davis et Liu [24] ainsi Christou et Fokianos [21] a proposé pour $p = q = 1$ la condition de stationnarité stricte du modèle (2.30).

***Remarque* 2.5.1**

Si $r = 1$, alors la distribution (NB) devient une distribution géométrique et dans ce cas, le modèle NB-INGARCH peut être appelé modèle INGARCH géométrique [104].

2.5.1.2 Condition de stationnarité et d'ergodicité

Le modèle binomial négatif INGARCH (p, q) est strictement stationnaire et ergodique sous la condition suivante :

$$1 - \sum_{j=1}^{q}(r\alpha_i + \beta_j)z^{-i} - \sum_{i=q+1}^{p} r\alpha_j z^{-i} = 0, \tag{2.31}$$

tel que toutes les racines $z_1, ..., z_p$ de l'équation (2.31) se trouvent à l'intérieur du cercle unité Goldberg [40].
D'après l'auteur la condition stationnaire de second ordre pour le modèle
NB-INGARCH (1, 1) est

$$r^2\alpha_1 + (r\alpha_1 - \beta_1)^2 < 1. \tag{2.32}$$

Davis et Liu [24] ansi Christou et Fkianos [21] qui ont donné pour (p = q = 1) la condition de stationnarité stricte pour le modèle (2.30).

$$\alpha^2\left(1 + \frac{1}{r}\right) + 2\alpha\beta + \beta^2 < 1,$$

avec un moment d'ordre deux fini.

2.5.1.3 Les moments de premier et second ordre du processus

Il est a noter que la moyenne de la variable aléatoire X_t sous la condition (2.31) vaut

$$E[X_t] = \omega r + \sum_{i=1}^{p} r\alpha_i\lambda_{t-i} + \sum_{j=1}^{q}\beta_j\lambda_{t-i}. \tag{2.33}$$

Supposons maintenant que le processus $\{X_t, t \in \mathbb{Z}\}$ suivant le modèle
NB-INGARCH (1, 1) est stationnaire au premier ordre, alors on a

$$E[X_t] = \mu_x = \frac{\omega r}{1 - r\alpha_1 - \beta_1}. \tag{2.34}$$

$$V[X_t] = \frac{1 - (r\alpha_1 + \beta_1)^2 + r^2\alpha_1^2}{1 - (r\alpha_1 + \beta_1)^2 - r\alpha_1^2}(\mu_x + \frac{\mu_x^2}{r}). \tag{2.35}$$

De plus, sa fonction d'autocovariance est

$$\gamma(k) = (r\alpha_1 + \beta_1)^{k-1}\frac{r\alpha_1[1 - \beta_1(r\alpha_1 + \beta_1)]}{1 - (r\alpha_1 + \beta_1)^2 - r\alpha_1^2}(\mu_x + \frac{\mu_x^2}{r}), \qquad k \geq 1. \tag{2.36}$$

2.5.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un NB-INGARCH (1, 1) vaut

$$\rho(k) = (r\alpha_1 + \beta_1)^{k-1} \frac{r\alpha_1[1 - \beta_1(r\alpha_1 + \beta_1)]}{1 - (r\alpha_1 + \beta_1)^2 + r^2\alpha_1^2}, \qquad k \geq 1. \tag{2.37}$$

2.5.2 Estimation des paramètres du modèle NB-INGARCH (p, q)

Zhu [104] a étudié les estimateurs de Yule Walker pour le paramètre $\alpha = (\alpha_1, ..., \alpha_p)^T$ et ω d'un modèles NB-INGARCH (p, 0), également appelés modèles NB-INARCH (p), et pour estimer les paramètres des modèles NB-INGARCH (p , q) les deux méthodes suivantes sont utilisées : la méthode des moindres carrés conditionnels et la méthode du maximum de vraisemblance conditionnelle avec certaines valeurs de r.
Soit $\alpha = (\alpha_1, ..., \alpha_p)^T$, $\beta = (\beta_1, ..., \beta_p)^T$ et $\theta = (\omega, \alpha, \beta)^T = (\theta_1, ..., \theta_{p+q})^T$. Dans ce qui suit, nous nous concentrerons sur l'estimation de θ en supposant que le paramètre r est connu. En pratique, r peut être estimé (la méthode donnée dans la section 5 [104]).
Supposons que l'on observe $(X_{1-p}, ..., X_0, X_1, ..., X_n)$ à partir de modèle (2.28).

2.5.2.1 Méthode de Yule-Walker

Supposons que (q = 0). L'estimation YW des paramètres $(\omega, \alpha_1, ...\alpha_p)^T$ consiste à insérer des autocovariances d'échantillons dans l'équation (2.38), puis à résoudre les paramètres. Pour ω , il peut être estimé par la méthode des moments à l'aide de $E(X_t)$ [29].

$$\gamma(k) = \sum_{i=1}^{p} r\alpha_i \gamma(|k - i|), \qquad k \geq 1. \tag{2.38}$$

2.5.2.2 Méthode du maximum de vraisemblance

Pour décrire l'approche du maximum de vraisemblance (ML), notez d'abord que la fonction de log-vraisemblance conditionnelle a la forme suivante

$$\ln L(\theta) = \sum_{t=1}^{n} [X_t - (r + X_t) \ln(1 + \lambda_t) + \sum_{v=1}^{X_t} \ln(v + r - 1) - \ln X_t!]. \tag{2.39}$$

Il est alors naturel d'estimer θ en maximisant $\ln L(\theta)$ donné ci-dessus.

2.6 Modèle PINGARCH (p, q)

Les modèles de séries chronologiques (linéaires et non linéaires) avec coefficients périodiques sont plus appropriés et ils ont plus de flexibilité pour modéliser certain processus périodiquement corrélés, une motivation à étendre à la classe de modèles INGARCH (p, q) invariants dans le temps à celle des modèles périodiques à valeurs entières PINGARCH (p, q).
En effet, beaucoup de problèmes (l'identification, l'estimation, stationnarité périodique, inversibilité...) liés à ces modèles périodiques ont été intensivement étudiés par plusieurs auteurs [35].

2.6.1 Structure probabiliste du modèle PINGARCH (p, q)

2.6.1.1 Définitions

Définition 2.6.1 (Modèle PINGARCH (p, q))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle Autorégressif Conditionnellement Hétéroscédastique Généralisé périodique à valeurs entières, de période S, dont la distribution est supposée poissonienne, d'ordres p et q, noté **PINGARCH$_S$(p, q)**, s'il est donné par

$$\begin{cases} X_t|F_{t-1} \sim P(\lambda_t) \qquad t \in \mathbb{Z}, \\ E[X_t|F_{t-1}] = \lambda_t = \omega_t + \sum_{i=1}^{p} \alpha_{i,t} X_{t-i} + \sum_{j=1}^{q} \beta_{j,t} \lambda_{t-j}, \end{cases} \tag{2.40}$$

où F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$.Les paramètres $\alpha_{i,t}$, $i = 1, 2, ..., p$ et $\beta_{j,t}$, $j = 1, 2, ..., q$, sont périodiques en t, de S, c'est à dire, $\alpha_{i,t+rS} = \alpha_{i,t}$ et $\beta_{j,t+rS} = \beta_{j,t}$, $t, r \in \mathbb{Z}$.
Afin d'éliminer la possibilité d'avoir une moyenne conditionnelle nulle ou négative, nous imposons les conditions suivantes sur les paramètres : $\omega_t \geq 0$, $\alpha_{i,t} \geq 0$ et $\beta_{j,t} \geq 0$ pour tout $i = 1, 2, ..., p$, $j = 1, 2, ..., q$ et $t \in \mathbb{Z}$.

Définition 2.6.2 (Modèle PINGARCH (1, 1))

Particulièrement, un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle périodique d'ordres (p = q = 1), **PINGARCH$_S$(1, 1)** s'il est défini comme suit

$$\begin{cases} X_t|F_{t-1} \sim P(\lambda_t) \qquad t \in \mathbb{Z}, \\ \lambda_t = \omega_t + \alpha_{1,t} X_{t-i} + \beta_{1,t} \lambda_{t-j}, \end{cases} \tag{2.41}$$

Posons $t = s + \tau S$, $s = 1, 2, ..., S$ et $\forall \tau \in \mathbb{Z}$, le dernier modèle peut être réécrit sous la forme équivalente suivante

$$\begin{cases} X_{s+\tau S}|F_{s+\tau S-1} \sim P(\lambda_{s+\tau S}) \qquad t \in \mathbb{Z}, \\ \lambda_{s+\tau S} = \omega_{s+\tau S} + \alpha_{1,s+\tau S} X_{s+\tau S-i} + \beta_{1,s+\tau S}\lambda_{s+\tau S-j}, \end{cases} \tag{2.42}$$

Ce modèle est une extension du modèle non périodique INGARCH (1, 1) donnée par (2.1) étudié par Ferland et al [82].

2.6.1.2 Condition de stationnarité

Il existe deux conditions de la stationnarité périodiques, à savoir la stationnarité périodique en moyenne et la stationnarité périodique au second ordre.
Le processus périodiquement corrélé à valeurs entières $\{X_t, t \in \mathbb{Z}\}$, satisfaisant le modèle INGARCH (1, 1) périodique donné par (2.41), est périodiquement stationnaire en moyenne et périodiquement stationnaire au second ordre [11], si et seulement si ,

$$\prod_{i=1}^{S}(\alpha_{1,i} + \beta_{1,i}) < 1. \tag{2.43}$$

***Remarque* 2.6.1**

En effet cette condition assure l'existence de tous les moments d'ordres supérieurs. Ce fait a été prouvé, par Ferland et al [82], dans le modèle classique INGARCH (1, 1).

2.6.1.3 Les moments de premier et second ordre du processus

Pour le modèle $PINGARCH_s(1, 1)$ les moments non conditionnelle d'ordre supérieurs, existe si et seulement si la condition (2.43) est verifier :

$$E[X_s] = \mu_{X,s} = (1 - \prod_{i=1}^{S}(\alpha_{1,i} + \beta_{1,i}))^{-1} \sum_{j=0}^{S-1}(\prod_{i=1}^{j}(\alpha_{1,i} + \beta_{1,i}))\omega_{s-j}. \tag{2.44}$$

$$V[X_s] = \gamma_X^s(0) = (1 - \prod_{i=1}^{S}\psi_s^2)^{-1} \sum_{j=0}^{S-1}(\prod_{i=1}^{j}\psi_{s-i+1}^2)F_{s-j}, \tag{2.45}$$

où $F_s = \mu_{X,s} - \psi_s^2\mu_{X,s-1} + \alpha_{1,s}^2\mu_{X,s-1}$ et $\psi_s = \alpha_{1,s} + \beta_{1,s}$.
Avec la condition $\prod_{i=1}^{j} x_i = 1$ si $j < 1$.

2.6.1.4 Structure d'autocorrélation

Les fonctions d'autocorrélation des processus périodiquement corrélés à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ et $\{\lambda_t, t \in \mathbb{Z}\}$ satisfaisant le modèle (2.41) sont, sous la condition (2.43), données comme suit

$$\rho_X^s(v+kS) = (\prod_{i=1}^{S}\psi_i)^k(\prod_{i=1}^{v-1}\psi_{s-i+1})\sqrt{\frac{\gamma_X^{(s-v)}(0)}{\gamma_X^{(s)}(0)}}(\psi_{s-v+1} - \frac{\beta_{s-v+1}\mu_{X,s-v}}{\gamma_X^{(s-v)}}), \tag{2.46}$$

$$\rho_\lambda^s(v+kS) = (\prod_{i=1}^{S}\psi_i)^k(\prod_{i=1}^{v}\psi_{s-i+1})\sqrt{\frac{\gamma_\lambda^{(s-v)}(0)}{\gamma_\lambda^{(s)}(0)}}, \tag{2.47}$$

où $v = 1, 2, ..., S, \quad k \in \mathbb{N}$ et $\psi_s = \alpha_{1,s} + \beta_{1,s}$ et $\mu_{X,s}$ est donné dans (2.44).
Avec la condition $\prod_{i=1}^{j} x_i = 1$ si $j < 1$.

***Remarque* 2.6.2**

Le modèle (2.41), peut être écrit sous la forme d'un modèle PARMA [11].

2.6.1.5 Un exemple d'application

Une étude à été faite sur la série chronologique qui représente le nombre d'infections par la bactérie Campyloboacteriosis récupérés chaque 28 jours à partir du mois de Janvier 1990 jusqu'au mois d'Octobre 2000. Cette série, qui contient 140 observations a été modélisée par un modèle INGARCH (1, 13) [82].

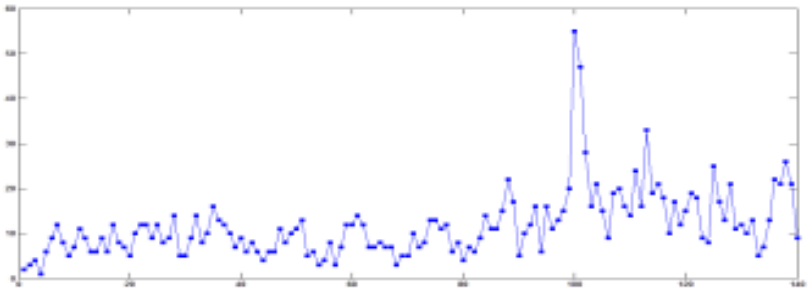

FIGURE 2.4 – le nombre d'infections par la bactérie chaque 28 jours de (1990-200).

A partir de la fonction d'autocorrélation de la série nous remarquons que cette série présente une structure d'autocorrélation périodique (S = 13).

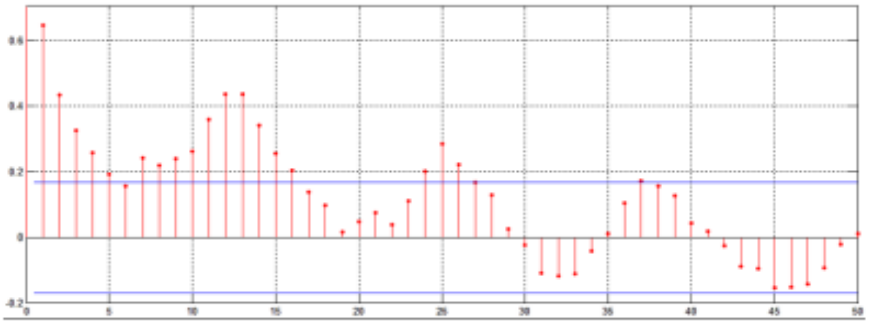

FIGURE 2.5 – La fonction d'autocorrélation

2.6.2 Estimation des paramètres du modèle PINGARCH (p, q)

Dans cette sous-section, nous nous intéressons à citer brièvement les méthodes d'estimation des paramètres du modèle (2.40) utilisé dans la littérature , où Bentarzi [12] a fait une étude en utilisant la méthode d'estimation de Yule-Walker (YW) et la méthode d'estimation du maximum de vraisemblance conditionnelle (CML). Comme nous le savons tous, la première méthode n'est pas aussi efficace que la deuxième méthode. Cependant, bien que la méthode de Yule-Walker ne soit pas aussi efficace que la méthode du maximum de vraisemblance, elle est encore utilisée aujourd'hui car elle est simple et produit des estimateurs cohérents qui peuvent être utilisés comme valeurs initiales dans des méthodes complexes comme la méthode (MLE).

2.7 Modèle INARCH (p)

Une sous-famille de la classe INGARCH (p, q), qui présente un intérêt pratique particulier, sont des modèles purement régressifs d'INGARCH, c'est-à-dire où (q = 0). Dans cette section, nous présentons brièvement la définition et les propriétés de base connues du modèle INARCH (p) qui est un cas particulier du modèle défini dans la section 2.

2.7.1 Structure probabiliste du modèle INARCH (p)

2.7.1.1 Définitions

Définition 2.7.1 (Modèle INARCH (p))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle Autorégressif Conditionnellement Hétéroscédastique à valeurs entières avec une surdispersion dont la distribution est supposée poissonienne, d'ordres p ; noté **INARCH (p)**, s'il est donné par :

$$\begin{cases} X_t|F_{t-1} \sim P(\lambda_t) & t \in \mathbb{Z}, \\ E[X_t|F_{t-1}] = \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}. \end{cases} \tag{2.48}$$

Où $\{X_t\}$ est l'observation à l'instant t, $t \in \mathbb{Z}$, $\omega > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$ et $P(\lambda_t)$ représente une distribution de Poisson avec une moyenne λ_t.

Le modèle INARCH (p) peut également être présenté comme un modèle GINAR(p) particulier voir (wieb [19]) il a des probabilités de transition de Poisson simples.

$$P(X_t = i|X_{t-1} = j) = \exp\left(-\omega - \sum_{i=1}^{p} \alpha_i X_{t-i}\right) \frac{(\omega + \sum_{i=1}^{p} \alpha_i X_{t-i})^x}{x!}.$$

***Remarque* 2.7.1**

Comme la distribution conditionnelle de $\{X_t, t \in \mathbb{Z}\}$ dans le modèle INARCH (p) est de Poisson, le modèle (2.48) ne peut pas traiter la surdispersion et la sous-dispersion conditionnelles. Afin de résoudre ce problème, certains auteurs ont proposé de remplacer la distribution de Poisson par d'autres distributions, telles que la distribution de Poisson double (DP) (Heinen, 2003), la distribution Binomiale négative (NB)(Zhu, 2011) et la distribution de Poisson généralisée (GP) (Zhu, 2012).

Définition 2.7.2 (Modèle INARCH (1))

Le modèle INARCH (1) c'est un contrepartie du modèle INAR (1) présenté dans le chapitre précédente et aussi est un cas limite du modèle INGARCH (1, 1) (pour p = 0), donc X_t soit conditionnellement distribué par la loi de Poisson de la manière suivante :

$$X_t|F_{t-1} \sim P(\omega + \alpha_1 X_{t-1}) \qquad t \in \mathbb{Z}. \tag{2.49}$$

Où $\{X_t\}$ est l'observation à l'instant t, $t \in \mathbb{Z}$, $\omega > 0$, $\alpha_1 \geq 0$ et F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_{t-1}\}$.

2.7.1.2 Condition de stationnarité et d'ergodicité

Un processus INARCH (1) est une chaîne de Markov faiblement stationnaire et ergodique alors tous les moments d'un processus INARCH (1) existent (Ferland et al [82]) sous la condition suffisante suivant

$$\alpha_1 < 1. \tag{2.50}$$

Si $p > 1$ la condition suffisante de la stationnarité stricte et d'ergodicité devient $\sum_{i=1}^{p} \alpha_i < 1$.

2.7.1.3 Les moments de premier et second ordre du processus

En particulier, la moyenne et la variance d'un processus INARCH (1) sont données par

$$E[X_t] = \mu_X = \frac{\omega}{1-\alpha}, \tag{2.51}$$

$$V[X_t] = \sigma_X^2 = \frac{\omega}{(1-\alpha)(1-\alpha^2)}, \tag{2.52}$$

et la fonction d'autocovariance satisfait les équations suivantes

$$\gamma_X(k) = \sum_{i=1}^{p} \alpha_i.\gamma_X(|k-1|) + \delta_{k0}.\mu, \qquad k \geq 0,$$

où δ_{ab} désigne le delta de Kronecker [20].

2.7.1.4 Structure d'autocorrélation

La fonction d'autocorrélation d'un INGARCH (p) vaut

$$\rho(k) = \sum_{i=1}^{p} \alpha_i \rho(|k-i|). \tag{2.53}$$

L'équation de la fonction d'autocovariance est identique à celle des équations de Yule-Walker d'un modèle AR (p) et si $(p = 1)$ égale à $\rho(k) = \alpha^k$ comme dans le cas standard AR (1).

2.7.2 Estimation des paramètres du modèle INARCH (p)

Dans cette sous-section, nous citerons les méthodes d'estimation pour les paramètres inconnus du processus INARCH. Supposons que nous ayons une série chronologique $(X_1, X_2, ..., X_N)$ issue d'un processus INARCH (p).

2.7.2.1 Méthode des moindres carrés conditionnels

Dans le cas autorégressif, c'est-à-dire si (q = 0), des estimations des moindres carrés conditionnels (CLS) peuvent également être dérivées facilement. D'après la définition l'espérance conditionnelle $E[X_t|F_{t-1}]$ du modèle (2.48) est simplement égale à $\omega + \sum_{i=1}^{p} \alpha_i X_{t-i}$. Ainsi, les estimations (CLS) des paramètres peuvent être obtenues en minimisant

$$CSS(\omega, \alpha_1, ..., \alpha_p) = \sum_{t=p+1}^{T} (x_t - \omega - \sum_{i=1}^{p} \alpha_i x_{t-i})^2. \tag{2.54}$$

2.7.2.2 Méthode du maximum de vraisemblance

Ferland et al [82] proposent une approche de maximum de vraisemblance conditionnelle (ML) à des valeurs pré-échantillons. Ce type de fonction de vraisemblance est facile à calculer, puisque le processus a une distribution de Poisson conditionnelle. Dans le cas autorégressif, par exemple, on a

$$L(\omega, \alpha_1, ..., \alpha_p) = P(X_T = x_T, ...|X_p = x_p, ..., X_1 = x_1) = \prod_{t=p+1}^{T} e^{-\omega - \sum_{i=1}^{p} \alpha_i x_{t-i}} . (\omega + \sum_{i=1}^{p} \alpha_i x_{t-i})^{x_t} / x_t!. \tag{2.55}$$

Les estimations ML sont obtenues en maximisant numériquement la fonction de vraisemblance logarithmique $\ln L(\omega, \alpha_1, ..., \alpha_p)$. Les erreurs types asymptotiques peuvent être calculées à partir des informations de Fisher observées [82].

2.7.2.3 Méthode des moments

Soit $(X_1, ..., X_n)$ une série chronologique issue d'un processus INARCH (1) stationnaire. Alors $E[X_t] = \frac{\omega}{1-\alpha}$ et $\rho(1) = \alpha$, la méthode des moments implique pour estimer α et ω elle est défini comme suit .

$$\hat{\alpha_{MM}} = \frac{\sum_{t=2}^{T}(X_t - \bar{X_T})(X_{t-1} - \bar{X_T})}{\sum_{t=1}^{T}(X_t - \bar{X_T})^2}, \qquad \hat{\omega_{MM}} = \bar{X_T}.(1 - \hat{\alpha_{MM}}). \tag{2.56}$$

Où $\bar{X_T} = \frac{1}{T}\sum_{t=1}^{T} X_t$ désigne la moyenne empirique de $(X_1, ..., X_n)$.

2.8 Modèle DINARCH (p)

Le modèle INGARCH standard avec sa distribution conditionnelle de Poisson présente une surdispersion non conditionnelle (Ceux-ci indiquent que la variance conditionnelle a un rapport constant avec la moyenne conditionnelle), mais le degré de surdispersion est déterminé par la structure d'autocorrélation.
Pour surmonter cette limitation, Xu et al [42] ont un nouveau modèle INARCH dispersé, appelé DINARCH .

2.8.1 Structure probabiliste du modèle DINARCH (p)

2.8.1.1 Définitions

Définition 2.8.1 (Modèle DINARCH (p))

Un processus stochastique à valeurs entières $\{X_t, t \in \mathbb{Z}\}$ est dit satisfaire un modèle INARCH dispersés **(DINARCH (p))** d'ordre p, et il est défini comme suit

$$\begin{cases} E[X_t|F_{t-1}] = \lambda_t, \\ V[X_t|F_{t-1}] = a\lambda_t \end{cases} \tag{2.57}$$

où λ_t satisfait la deuxième expression du modèle (2.57) et a ($a > 0$ est lié à F_{t-1}) est supposé constant.
Le modèle (2.57) indique une surdispersion et une sous-dispersion conditionnelles (lorsque a ¡ 1 et a ¿ 1), et quand (a = 1) le modèle (2.57) est identique au modèle (2.48). Lorsque p = 1, notre modèle est également un modèle AR (1) linéaire conditionnel non gaussien analysé par Grunwald et al [37]. Si a n'est pas constant, le modèle (2.57) comprend également le modèle NB-INGARCH (proposé par Zhu [104]).

Pour le modèlé (2.57), $\{\varepsilon_t\} = \{X_t - \lambda_t\}$ et $Y_t = X_t - \mu$ alors

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t, \tag{2.58}$$

avec $E[\varepsilon_t] = 0, V[\varepsilon_t] = a\mu$ et $Cov(\varepsilon_t, Y_{t-k}) = 0$ pour $k > 0$ et $K(1) > 0$.

***Remarque* 2.8.1**

Dans la littérature plusieure modèles qui appartient à la famille INARCH dispersé ont été utilisés : modèle (NB-DINARCH) basé sur une distribution $NB(r_t, Q)$ généralement utilisée pour décrire les données de comptage surdispersé qui peut être vu comme un mélange continu d'une distribution de Poisson tel que le taux de Poisson supposé être distribué Gamma, un autre modèle (DP-DINARCH) basé sur une distribution DP (Double Poisson) a été introduite par Efron [28] et aussi un modèle (GP-DINARCH) basé sur une distribution $GP(\lambda^*, k)$ est proposée par Consul et Jain [22].

2.8.1.2 Condition de stationnarité et d'ergodicité

Les conditions de stationnarité et d'ergodicité strictes d'un processus DINARCH (p) sont les mêmes conditions proposées par Ferland et al [82] pour la stationnarité stricte d'un

processus INGARGH (p, q).
Le processus $\{X_t\}$ une chaîne de Markov avec une probabilité

$$P(X_t|X_{t-1}, X_{t-2}...X_{t-p}) = P(X_t|X_{t-1}).$$

Selon le théorème (4.3.3) de Ross [83], $\{X_t, t \in \mathbb{Z}\}$ est irréductible et apériodique, alors le processus $\{X_t, t \in \mathbb{Z}\}$ défini par le modèle (2.45) est ergodique.

***Remarque* 2.8.2**

La propriété d'ergodicité maintient qu'il existe une unique distribution stationnaire qui est la distribution limite d'un processus DINARCH (p).

2.8.1.3 Les moments de premier et second ordre du processus

Le modèle DINARCH (1) est le modèle le plus simple de la famille DINARCH (p) Selon la condition (2.15) lorsque $0 < \alpha_1 < 1$, le processus DINARCH (1) est faiblement stationnaire alors l'espérance et la variance de X_t sont

$$E[X_t] = \frac{\omega}{1-\alpha_1}. \tag{2.59}$$

$$V[X_t] = \frac{a\omega}{(1-\alpha_1)(1-\alpha_1^2)}. \tag{2.60}$$

a permet de contrôler le degré de dispersion indépendamment de α (Xu et al [42]).

2.8.2 Estimation des paramètres du modèle DINARCH (P)

Xu et al [42] a proposé trois méthodes pour l'estimation des paramètres. Ces méthodes sont comparées dans une étude de simulation pour la situation conditionnelle surdispersée avec une série temporelle pure à valeur entière d'ordre un. Une conclusion a été faite dit que lorsque la distribution conditionnelle de $(X_t|F_{t-1})$ est connue, les paramètres peuvent être estimés selon l'estimation du maximum de vraisemblance (MLE). Sinon, si la distribution conditionnelle est inconnue, nous adoptons deux autres méthodes. L'une est l'estimation conditionnelle des moindres carrés (WLSE) et l'autre est l'estimation de Yule-Walker (YW). qui est généralement utilisée pour estimer les paramètres des modèles ARMA.

2.9 Autres type de modèle INGARCH

2.9.1 Modèle DP-INGARCH (p, q)

Les séries chronologiques de comptage présentent souvent une surdispersion, c'est-à-dire une variance supérieure à la moyenne, mais le phénomène inverse peut être rencontré. Les modèles INGARCH de Poisson et binomial négatif ne peuvent pas prendre en compte la sous-dispersion.
Pour résoudre ce problème, Heinen [44] propose un modèle basé sur la distribution à double Poisson [28].

2.9.1.1 Définition (Modèle DP-INGARCH (p, q))

Un prossecus $\{X_t, t \in \mathbb{Z}\}$ est un modèle INGARCH à double Poisson (DP-INGARCH (p, q)) si sa distribution conditionnelle est une distribution DP [(voir Efron [28])].

$$\begin{cases} X_t|F_{t-1} \sim DP(\lambda_t, \gamma) & t \in \mathbb{Z}, \\ \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2.61}$$

où $0 < \gamma, \omega > 0, r > 0, \alpha_i > 0, \beta_j > 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$ F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_t\}$.

2.9.1.2 Quellque propriété du modèle DP-INGARCH

Pour (p, q) = (1, 1) les conditions de stationnarité et d'ergodicité strictes d'un processus DP-INGARCH (1, 1) sont les mêmes conditions proposées par Ferland et al [82] pour la stationnarité stricte d'un processus INGARGH (1, 1), tout ça implique l'existence d'une solution stationnaire (X_t) telle que les moments d'un DP-INGARCH (1, 1) sont tous finis, dans ce cas

$$E[X_t] = \mu_X = \frac{\omega}{1 - (\alpha_1 + \beta_1)}, \tag{2.62}$$

$$V[X_t] = \sigma_X^2 = \frac{(1 - (\alpha_1 + \beta_1)^2 + \alpha_1^2)}{1 - (\alpha_1 + \beta_1)^2} \mu_X \gamma. \tag{2.63}$$

2.9.2 Modèle COM-INGARCH (p, q)

Zhu [103] a proposé le modèle INGARCH de Poisson de Conway-Maxwell (COM) comme alternative aux modèles GP-INGARCH et DP-INGARCH, et a montré que ces modèles sont souvent plus performants que les autres concurrents, en particulier le modèle

DP-INGARCH.
La distribution COM-Poisson est flexible dans la modélisation d'un large éventail de surdispersion et de sous-dispersion avec seulement deux paramètres, tout en possédant des propriétés qui le rendent méthodologiquement utile dans la pratique.

2.9.2.1 Définition (Modèle COM-INGARCH (p, q))

Un prossecus $\{X_t, t \in \mathbb{Z}\}$ est un modèle COM-Poisson INGARCH (p, q) si sa distribution conditionnelle est une distribution COM-Poisson [(voir Efron [103])].

$$\begin{cases} X_t|F_{t-1} \sim COM - Poisson(\lambda_t, v) & t \in \mathbb{Z}, \\ \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2.64}$$

où $v \geq 0, \omega > 0, r > 0, \alpha_i > 0, \beta_j > 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$ F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_t\}$.
Le modèle ci-dessus est noté COM-Poisson INARCH (p) lorsque (q = 0). Clairement, le modèle (2.64) se réduit au modèle (2.11) lorsque $v = 1$.

2.9.2.2 Quellque propriété du modèle COM-INGARCH

Si $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$, alors le modèle (2.64) est approximativement stationnaire. Donc les moments d'un COM-INGARCH (p, q) existes, et ils sont tous finis, dans ce cas

$$E[X_t] \approx \frac{\omega - (1 - \sum_{j=1}^{q} \beta_j) \frac{v-1}{2v}}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}, \tag{2.65}$$

$$V[X_t] \approx \frac{E[\lambda_t]}{v} + V[\lambda_t]. \tag{2.66}$$

2.9.3 Modèle ZIP-INGARCH (p, q)

En raison des caractéristiques de la distribution de Poisson, qui a la même moyenne et la même variance, alors lorsque le modèle présente un inconvénient en ce sens qu'il ne peut pas tenir compte la surdispersion et avec un phénomène de dépassement de zéro qui fait référence à un phénomène dans lequel un plus grand nombre de zéros que prévu dans le modèle sont observés (tel que l'apparition d'une maladie spécifique ou la survenue d'un crime spécifique) en même temps.
Afin de traiter efficacement ce problème Zhu [106] a proposé le ZIP-INGARCH (p, q) [94].

2.9.3.1 Définition (Modèle ZIP-INGARCH (p, q))

Un prossecus $\{X_t, t \in \mathbb{Z}\}$ est un modèle INGARCH de Poisson zero-inflated (ZIP-INGARCH (p, q)) si sa distribution conditionnelle est une distribution ZIP [(voir Johnsonetal [72] section 4.10.3)].

$$\begin{cases} X_t|F_{t-1} \sim ZIP(\lambda_t, \nu) & t \in \mathbb{Z}, \\ \lambda_t = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}, \end{cases} \tag{2.67}$$

où $0 < \nu < 1, \omega > 0, r > 0, \alpha_i > 0, \beta_j > 0$ pour tout $i = 1, 2, ..., p$ et $j = 1, 2, ..., q$ F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_t\}$.

Remarque

Si $\nu = 0$, alors le modèle ZIP-INGARCH se réduit au modèle INGARCH de Poisson considéré dans Ferland et al [82].

2.9.3.2 Quellque propriété du modèle ZIP-INGARCH

La condition nécessaire et suffisante pour que le processus ZIP-INGARCH (p, q) soit strictement stationnaire est

$$1 - \sum_{i=1}^{q}((1-\nu)\alpha_i + \beta_j)z^{-i} + \sum_{i=1}^{q}(1-\nu)\alpha_j z^{-i} = 0. \tag{2.68}$$

2.9.4 Modèle INGARCH log lineaire

Bien que les modèles précédents semblent offrir un cadre adéquat pour la modélisation de données de comptage dépendantes, il y a au moins deux inconvénients liés à ses applications.

Par exemple en considérant le modèle INGARCH poissonienne, du fait que $0 < \alpha_1 + \beta_1 < 1$ et $cov(X_t, X_{t+h}) > 0$ signifie que ce modèle a été conçu principalement pour la modélisation des corrélations positives.

Donc, il ne peut être utilisé lorsque la structure de dépendance admet des corrélations négatives. Pour s'attaquer à ces problèmes, Fokianos et Tjøstheim [33] ont proposé un modèle INGARCH log-linéaire.

2.9.4.1 Définition (INGARCH log-linéaire)

Un prossecus $\{X_t, t \in \mathbb{Z}\}$ est un modèle INGARCH log-linéaire si la distribution conditionnelle de X_t compte tenu de ses valeurs passées est poissonienne avec comme paramètre

d'intensité $\lambda_t = e^{vt}$ où pour un modèle log-linéaire poissonien du premier ordre

$$\begin{cases} X_t|F_{t-1} \sim Poisson(\lambda_t) & t \in \mathbb{Z}, \\ v_t = \omega + \alpha_1 \log(X_{t-1}+1) + \beta_1 v_{t-1}, \end{cases} \tag{2.69}$$

où $\omega, \alpha_1, \beta_1 \in \mathbb{R}$ et F_{t-1} représente la σ-algèbre engendrée par $\{X_0, X_1, ..., X_t\}$.

2.9.4.2 Quellque propriété du modèle INGARCH log-linéaire

Fokianos et Tjøstheim [33] ont montré que sous la condition de stabilité suivante

$$\begin{cases} |\alpha_1 + \beta_1| < 1 & si \quad \alpha_1 > 0, \\ |\beta_1||\alpha_1 + \beta_1| < 1 & sinon, \end{cases} \tag{2.70}$$

le processus $\{X_t, t \in \mathbb{Z}\}$ donné par (2.69) est strictement stationnaire et ergodique.

Conclusion

La catégorie des séries temporelles à valeurs entières basées sur la régression discrète est une catégorie assez grande qui se spécialise dans l'analyse statistique de certaines données discrètes qui montrent une variance supérieure à la moyenne de l'échantillon. Ce phénomène connu sous le nom de la surdispersion a été largement étudié et varié dans la littérature. Ce chapitre se concentre sur la recherche de quelques modèles de cette catégorie et donne leur différentes définitions et caractéristiques de base telles que les structures de probabilité qui caractérisent les modèles proposés ci-dessus, avec quelques principes d'estimation de chaque modèle mentionnés dans la littérature.

3

Application d'un modèle INGARCH pour la modélisation et la prévision du trafic réseau

3.1 Introduction

Le trafic réseau a énormément augmenté au cours de la dernière décennie en raison de l'avènement des nouvelles technologies, industries et applications...etc et afin d'éviter la congestion et la saturation du réseau à court ou long terme, il existe de nombreuses façons d'évaluer les performances du trafic car de nombreux travaux de prévision du trafic ont été réalisés dans lesquels de nombreux modèles ont été utilisés pour vérifier leur adéquation à la modélisation du trafic dans la littérature.

Dans cette section, une présentation d'un modèle INGARCH comme un modèle prédictif du trafic réseau a été fait où l'estimation de ses paramètres avec un algorithme d'ajustement classique (CMLE), tandis que les paramètres du processus de Poisson sont prédits à pas de temps futurs sur la base d'un algorithme de prédiction sur un ensemble de données fourni par le Center for Applied Internet Data Analysis (CAIDA).

3.2 Travaux connexes

L'analyse du trafic réseau est devenue de plus en plus vitale et importante de nos jours pour surveiller le trafic réseau et une raison plus pratique pour les prévisions à long terme qui peuvent être envisagée au niveau de chaque lien d'un réseau et d'avoir une idée plus claire sur les éventuels changements et améliorations nécessaires dans le futur pour l'équipement et les liens du réseau et nous pouvons aussi conclure que les prévisions sur les variables qui caractérisent les paramètres du réseau tel que le débit, le délai, le taux de perte et le nombre de flux permetent de prévoir la performance du réseau [101].
Le réseau est analysé à différents niveaux à savoir, au niveau du paquet, au niveau du flux et au niveau réseau pour la gestion de la sécurité, la figure (3.1) montre trois phases principales d'analyse du trafic réseau.

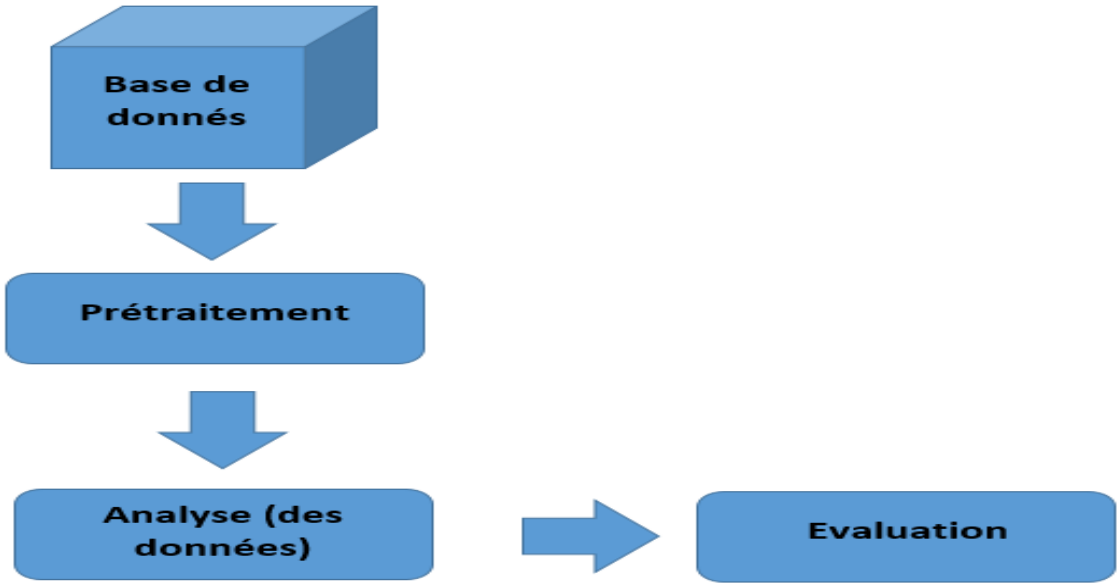

FIGURE 3.1 – La structure générique d'analyse du trafic réseau.

Pour cela il existe de nombreuses études dans la littérature concernant les modèles et les méthodes de prévision de trafic réseau. La prévision est basée sur les modèles des séries temporelles tels que la moyenne mobile autorégressive (ARMA) [99] a été utilisé comme un modèle de trafic de réseau pour la détection des intrusions et des attaques de réseau, la moyenne mobile intégrée autorégressive (ARIMA) [99], la moyenne mobile intégrée fractionnée (FARIMA) [25] qui a permis d'améliorer l'efficacité et la précision des prédictions et un modèle Autorégressif Conditionnellement Hétéroscédastiques Généralisé (GARCH) [49] a été envisagé pour estimer la matrice de trafic du protocole Internet (IP) à grande échelle qui décrit le volume du trafic entre les paires source-destination dans le réseau IP, les modèles hybrides ont été envisagés dans plusieurs études telles que (AR/GARCH) et (ARIMA/ GARCH) ont été proposés pour prédire le trafic Internet et le flux du trafic dans [56] et [18], respectivement, les techniques d'apprentissage automatique (ML) [87] et d'autres théories telles que la chaîne de Markov cachée et la méthode exponentielle [102] et récemment des réseaux neuronaux tels que Les réseaux de neurones récurrents (LSTM RNN) capable d'identifier des modèles récurrents dans diverses mesures a été mis en oeuvre pour prédire le trafic dans les réseaux cellulaires et un mécanisme qui prédit les changements de charge de trafic dans le réseau central de la 5G a été proposé via un LSTM RNN (Recurrent Neural Network) et un DNN (Deep Neural Networks) sur un ensemble de données réelles d'arrivée de trafic dans un réseau mobile [48] et LSTM CNN (Convolutional Neural Network) tandis que ces derniers ont été envisagés pour prédire les changements à court terme dans le trafic traversant un réseau de centre de données [2].

3.3 Ajustement du modèle INGARCH et les données mesurées.

Afin d'adapter un modèle INGARCH approprié au trafic réseau, deux ordres p et q doivent être estimés. La valeur estimée peut être obtenue en testant la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF) [93]. Cependant, comme le but de cet exemple est de tester l'adaptabilité du modèle INGARCH au trafic réseau, alors cette recherche ne considère pas l'estimation de l'ordre optimal. Généralement, l'occurrence d'un événement tend à dépendre davantage des événements récents que des événements survenus il y a longtemps, et la configuration est simple et le temps de calcul est moindre si les ordres sont petits.
Par conséquent, "1" est souvent utilisé comme p et q (p = q = 1) dans le modèle ARIMA/GARCH pour la modélisation du trafic réseau [18].
Sur cette base, les ordres sont fixé ici à 1. Ensuite, une application importante de l'analyse des séries chronologiques la prévision. L'objectif de la prévision est d'estimer les valeurs futures de X_t en tant que combinaison linéaire d'un sous-ensemble d'observations précédentes. La prévision peut être soit en une étape, où X_n est prévu, soit en plusieurs étapes où X_{n+k} $k > 0$. Il existe trois types de prévisions : les prévisions ponctuelles, les prévisions à intervalles et les prévisions probabilistes. Ce chapitre se concentre sur l'utilisation et l'analyse des prévisions probabilistes, par la définition (2.3.2), on obtient $\{\hat{\lambda}_t(n) : n \geq 1\}$ tel que

$$\hat{\lambda}_t(1) = \omega + \alpha\lambda_t + \beta X_t, \qquad si \;\; n = 1, \tag{3.1}$$

$$\hat{\lambda}_t(n) = \sum_{k=0}^{n-2} \omega(\alpha+\beta)^k + (\alpha+\beta)^{n-1}\hat{\lambda}_t(1), \qquad si \;\; n \geq 2. \tag{3.2}$$

3.3.1 Ensemble de donneés

Kim [55] a réalisé une expérience sur un ensemble de données en l'utilisant un langage de programmation (Python 3.6) et (Tensorflow v.1.7.0) sur un système (Intel Core i7, 16 GB RAM).
Pour l'analyse l'auteur a utilisé un ensemble de données de traces passives de (CAIDA) en 2016 contient des traces de trafic passives anonymisées du moniteur equinix-chicago de CAIDA (dir A, dir B) collectées à partir de moniteurs haute vitesse sur une liaison dorsale commerciale [46]. Il s'agit de quatre données des dates différentes 21/01/2016, 18/02/2016, 17/03/2016 et 06/04/2016.

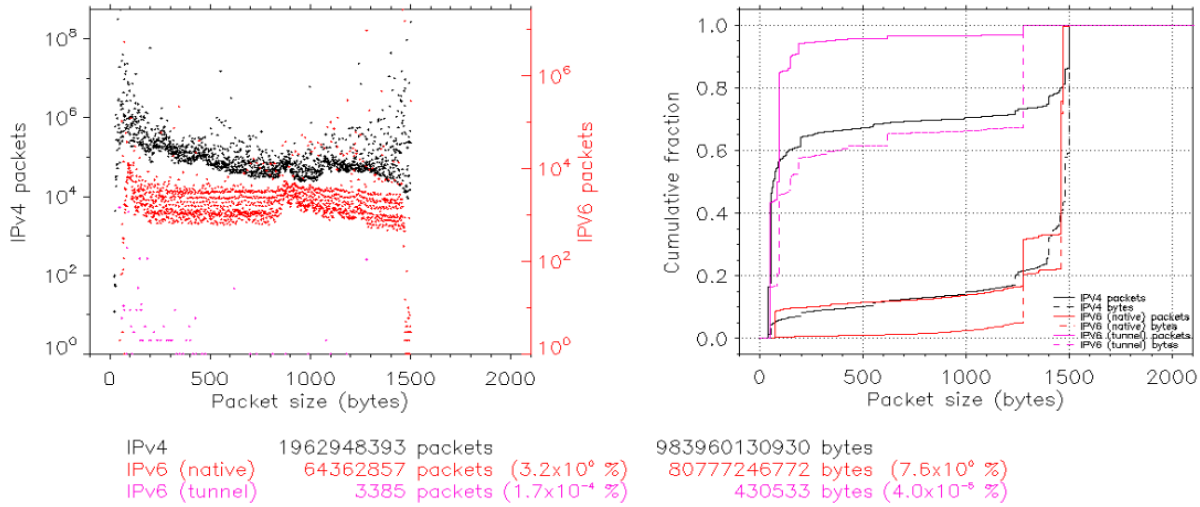

FIGURE 3.2 – Fonction de distribution de taille des paquets Equinix-chicago.dira.21/01/2016-130000.UTC

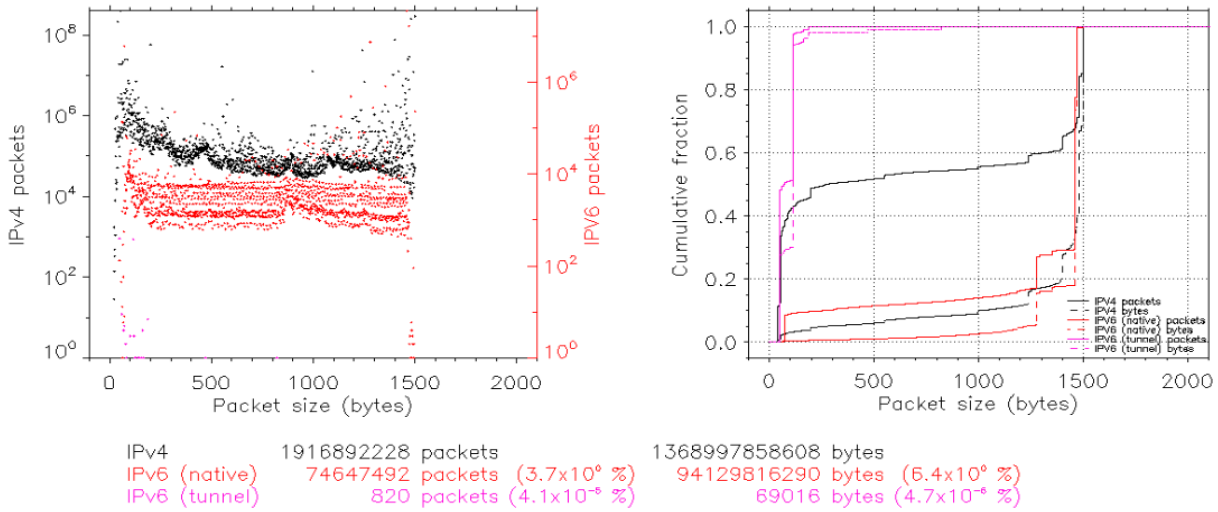

FIGURE 3.3 – Fonction de distribution de taille des paquets Equinix-chicago.dira.18/02/2016-130000.UTC

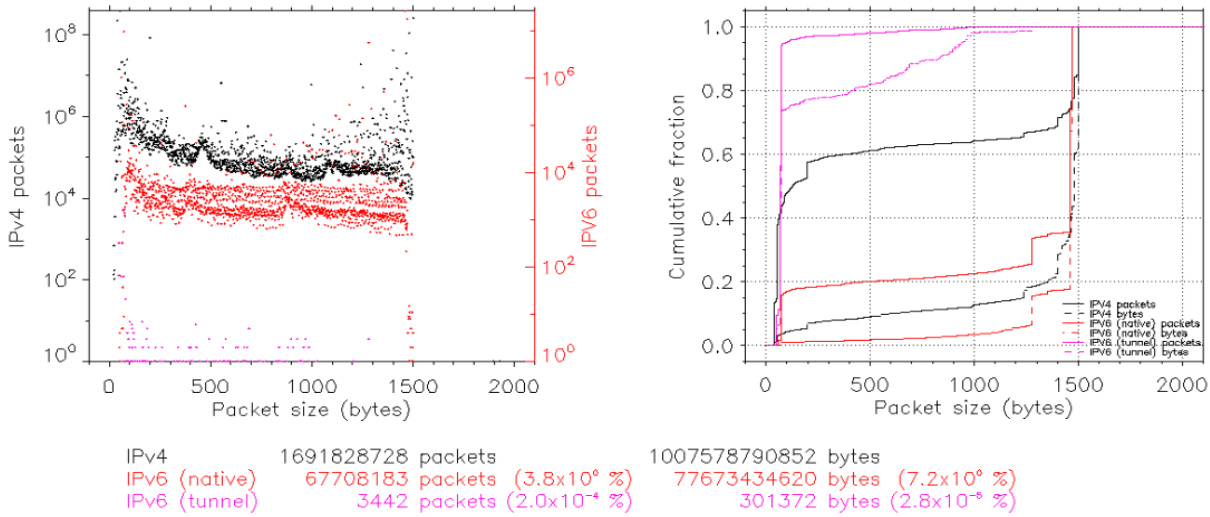

FIGURE 3.4 – Fonction de distribution de taille des paquets Equinix-chicago.dira.17/03/2016-130000.UTC

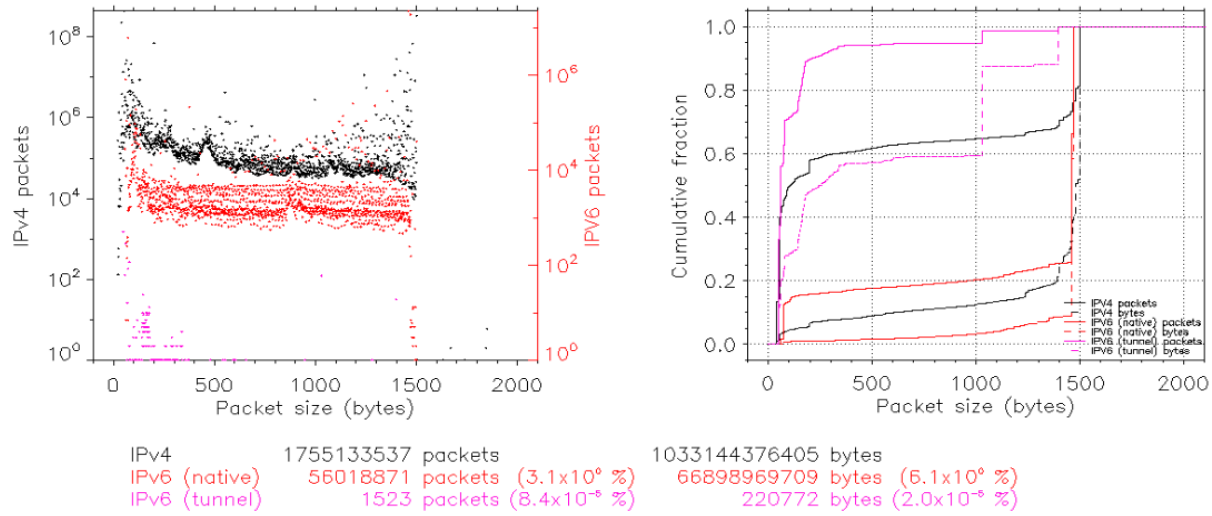

FIGURE 3.5 – Fonction de distribution de taille des paquets Equinix-chicago.dira.06/04/2016-130000.UTC

Chaque date contient une heure de trafic (130000 UTC) et elle est divisée en 64 sous-ensembles séquentiellement par heure. Par conséquent, la période de prévision est d'environ 1 minute pour toute prédiction avancée, un total de 256 sous-groupes est pris en compte. Le nombre d'octets de paquets IPv4 dans le "dir A" est extrait pour chaque date et il est utilisé dans l'expérience. Les performances sont comparées avec deux modèles de séries temporelles (ARIMA, GARCH) et un modèle NN (LSTM est un modèle qui prend en compte le problème du gradient de fuite dans le RNN [80]).
Ces données sont utiles pour la recherche sur les caractéristiques du trafic Internet, notamment la panne des applications, les événements de sécurité, la distribution géographique et topologique, le volume et la durée des flux.

3.3.2 Méthodologie de prévision

Soit les notations suivant :

Notation	**Description**.
X_t	Données brutes
Y_t	Données transformés
F_t	Historique jusqu'a t
λ_t	La moyen conditionnelle $E[X_t\|F_{t-1}]$
$\hat{\lambda}_t$	Prévision de $\hat{\lambda}_t$
p, q	Ordres du modèle INGARCH
$\omega, \alpha_i, \beta_j$	Les paramètres de $\hat{\lambda}_t$
Θ	$\{\omega, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q\}$
$L(\theta)$	La fonction de vraisemblance conditionelle
$\log L(\theta)$	La fonction logarithmique de vraisemblance conditionelle

TABLE 3.1 – Résumé des notations

$\{X_t\}_{t=1}^N$ un ensemble de données donné, où N est le nombre de données dans l'ensemble de données. Pour obtenir une prévision à n étapes, l'estimateur du maximum des vraisemblance conditionnelle de θ est obtenu par l'algorithme de prévision suivant.

A l'étape (1), $\{Y_t\}_{t=1}^N$ est obtenue par :

$$Y_t = \frac{a(X_t - \mu_X)}{\sigma_X} + a^2. \tag{3.3}$$

où $\mu_X = E(X_t)$, $\sigma_X = \sqrt{V(X_t)}$ et a est une constante à déterminer pour la transformation attribue de $\{X_t\}_{t=1}^N$ à $\{Y_t\}_{t=1}^N$ où $E(X_t) = V(X_t) = a$.
Dans l'étape (2.1), $\{Y_t\}_{t=1}^{N_1}$ et $\{Y_t\}_{t=N_1+1}^N$ sont utilisés pour l'apprentissage et le test, respectivement.

Algoritheme de prévision
1.Traiter les données : Transformer les données de trafic $\{X_t\}_{t=1}^N$ en $\{Y_t\}_{t=1}^N$ qui a une distribution de poisson.
2.La formation du processus :
2.1 Divise $\{Y_t\}_{t=1}^N$ en deux sous-ensemble disjoints $\{Y_t\}_{t=1}^{N_1}$ et $\{Y_t\}_{t=N_1+1}^N$, où $N_1 = N \times r$ pour un ration d'apprentissage $r \in (0,1)$ donné.
2.2 En utulisant $\{Y_t\}_{t=1}^{N_1}$ pour trouve $\hat{\theta}$ par $\log L(\theta) = \sum_{i=1}^t \{X_i \log(\hat{\lambda}_i) - \hat{\lambda}_i - \log(X_i!)\}$ avec $\hat{\theta} = \underset{\theta\in\Theta}{argmax} \log L(\theta)$.
3.Le processus de prévision :
3.1 En utilisant $\hat{\theta}$ et (3.2) pour trouver $\{\hat{Y_{N_1(n)}}\}_{n=1}^{N-N_1}$ les données predits à n pas d'avance.
3.2 Prendre la transformation inverse de $\{\hat{Y_{N_1(n)}}\}_{n=1}^{N-N_1}$ et obtenire $\{\hat{X_{N_1(n)}}\}_{n=1}^{N-N_1}$ puis calculer les mesures de performance pour $\{\hat{X_{N_1(n)}}\}_{n=1}^{N-N_1}$ et $\{X_t(n)\}_{t=N_1+1}^N$.
4.Mise à jour : Augmenter N_1 un par un jusqua N et effectuer les etapes $2-3$.

TABLE 3.2 – L'algorithme de la prévision

3.3.3 Mesures de performances

Pour mesurer la précision des prévisions, les mesures suivantes sont considérées : l'erreur absolue moyenne normalisée (NMAE), l'erreur carrée moyenne normalisée (NMSE) et le rapport signal/erreur de prédiction (PSER). Ils sont définies par

$$NMAE = \frac{1}{N-N_1} \sum_{j=N_1+1}^{N} \frac{X_j - \hat{X}_j}{X_j}, \tag{3.4}$$

$$NMSE = \frac{MSE}{M_{test}}, \tag{3.5}$$

$$PSER = 10 \log_{10}(\frac{SM_{test}}{MSE}). \tag{3.6}$$

Où MSE, M_{test} et SM_{test} sont donnés par

$$MSE = \sum_{j=N_1+1}^{N} \frac{(X_j - \hat{X}_j)^2}{N-N_1},$$

$$M_{test} = \sum_{j=N_1+1}^{N} \frac{X_j}{N - N_1},$$

et

$$SM_{test} = \sum_{j=N_1+1}^{N} \frac{X_j^2}{N - N_1}.$$

3.4 Résultats

Afin de valider l'efficacité du modèle, des traces de trafic passif anonymisées fournies par le Center for Applied Internet Data Analysis sont utilisées dans l'expérience est une étude comparative a été faite avec d'autres modèles tels que ARIMA, GARCH et LSTM et une analyse de cette étude comparative est donnée dans les sections suivantes avec l'utilisation des données du trafic il a été observé que INGARCH a les meilleurs performances.

3.4.1 Traitement de données

Pour obtenir $\{Y_t\}_{t=1}^{N}$ pour INGARCH, la valeur de (a) dans (3.3) est fixé à 20, qui est choisi pour séparer les données en entiers donnant $\mu_{Y_t} = \sigma_{Y_t}^2 = 400$. On a remarqué que $a \leq 20$ ne garantit pas tous les entiers distincts de $\{Y_t\}_{t=1}^{N}$ et $a \geq 20$ n'est pas nécessaire car la séparation des données dans l'ensemble de données brutes en entiers distincts est suffisant.
Les paramètres ω, α et β sont trouvés d'une manière extensive en prenant des valeurs dans (0.5,1.5), [0,1), et [0,1), respectivement.
Les intervalles de α et β sont sélectionnés en tenant compte de la stationnarité des données, tandis que la pertinence de l'intervalle de ω est vérifiée par l'équation suivante (Proposition 7 dans [64]).
la moyenne de INGARCH (1,1) :

$$E[Y_t] = \mu_{Y_t} = \frac{\omega}{1 - (\alpha_1 + \beta_1)}. \tag{3.7}$$

Pour la comparaison, avec l'utilisation des processus stationnaire comme : les modèles ARIMA (p, d, q) et GARCH (p, d, q) avec (d = 1) car les données brutes échouent le test de stationnarité.
Les ordres p et q sont variés dans l'expérience et nous avons obtenu que la performance de l'ordre (p, q) = (5, 1) est meilleure que celle de tous les autres ordres en termes de mesures considérées.
La méthode CMLE est appliquée au GARCH.

Pour les LSTM,l'ensemble de données est prétraité par Min-Max, qui est défini par la formule suivante

$$X_t = \frac{X_t - \underbrace{\min(X_t)}_{1\leq t\leq N}}{\underbrace{\max(X_t)}_{1\leq t\leq N} - \underbrace{\min(X_t)}_{1\leq t\leq N}}. \tag{3.8}$$

Le réseau de LSTM RNN se compose des trois couches principales : une couche d'entrée avec une entrée, des couches cachées avec plusieurs blocs LSTM et une couche de sortie qui fait une prédiction d'une seule valeur .

Dans l'exemple, il y a cinq couches cachées et douze blocs sont considérés et on obtient que les performances avec trois couches cachées et douze blocs.

Dans ce qui suit, LSTM (m_1, m_2) implique LSTM avec m_1 couches cachées et m_2 blocs dans chaque couche. LSTM est appelé **LSTM DNN** si $m_1 \geq 3$, sinon c'est un **LSTM RNN**.

3.4.2 Évaluation les performances

Dans cette sous-section, les performances et la complexité du modèle ont été évaluées à l'aide à des résultats ci-dessous. Pour chaque ensemble de données une étude comparative à été fait avec les modèle existant ARIMA, GARCH, LSTM et le modèle proposé INGARCH.

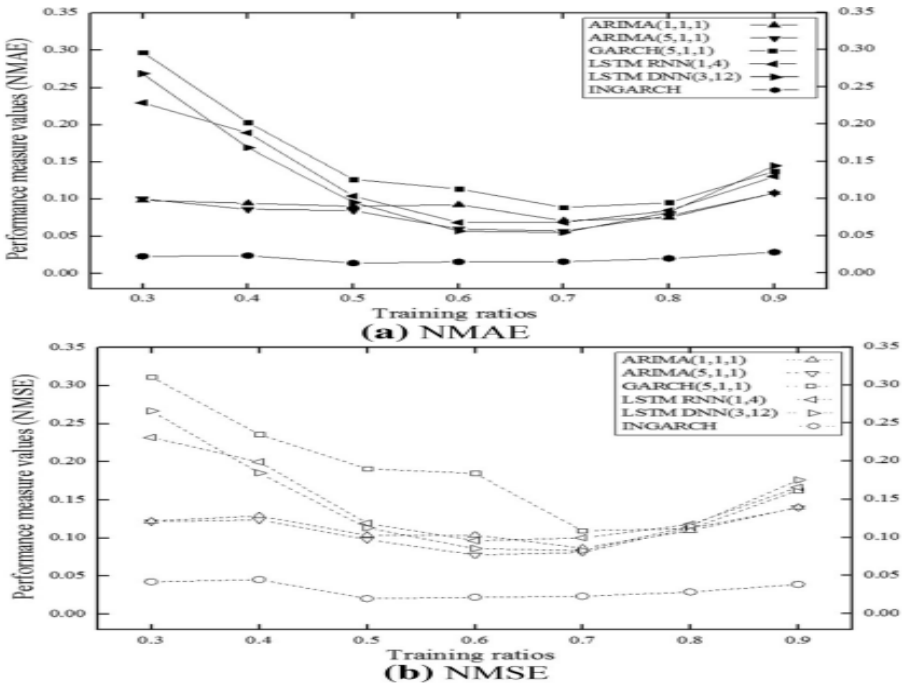

FIGURE 3.6 – Comparaison des mesures de performance pour différentes ratios (prévision à un pas d'avance)

La figure (3.6) compare les mesures de performance pour différentes ratio (r). Elle montre que les performances d'ARIMA et de LSTM RNN (1,4) sont les meilleures lorsque (r = 0,7), tandis que celles d'INGACH et de LSTM DNN (3,12) sont les meilleures lorsque (r = 0,6).
Pour GARCH (5,1,1), la NMAE est meilleure lorsque (r = 0,7) et NMSE est meilleure lorsque (r = 0,6) .
Selon la figure, les résultats suivants sont obtenus lorsque (r = 0,7)

$$\begin{cases} I < LD(3,12) < A(5,1,1) < LR(1,4) < A(1,1,1) < G(5,1,1) & pour \quad NMAE, \\ I < A(5,1,1) < LD(3,12) < A(1,1,1) < LR(1,4) < G(5,1,1) & pour \quad NMSE. \end{cases} \tag{3.9}$$

Où I, G, A, LR, et LD représentent INGARCH, GARCH, ARIMA, LSTM RNN, et LSTM DNN respectivement.

On remarque que le modèle INGARCH dépend pas du ratio, ce qui semble utilisable pour

réaliser l'algorithme de prédiction mentionné ci-dessus. Sur cette base, le ratio d'apprentissage de (r = 0,7) est utilisé de manière fondamentale et les deux ratios d'apprentissage, 0,1 et 0,4, sont utilisés en plus dans certaines résultats.
On observe que le θ actualisé pour les prévisions à un pas en avant fluctuent légèrement puis convergent vers (0.5556, 0.0556, 0.943) pour pour les données de tous les jours et pour les différents valeurs de r.
En d'autres termes, le modèle du traffic pour les données est obtenu par

$$\hat{\lambda}_{N_1}(1) = 0.5556 + 0.556\lambda_{N_1} + 0.943X_{N_1}. \tag{3.10}$$

D'aprés (3.7) et (3.9), on obtient

$$E[Y_t] \approx 396,86.$$

On peut supposer que la prévision est plus précise si n le nombre de pas futurs dans la prédiction est petit. Les résultats suivants vérifient cette hypothèse.

***Remarque* 3.2.1**

Dans les figures ((3.7), (3.8), (3.10)), l'axe des x et l'axe des y représentent respectivement le temps (unité : minutes) et le nombre d'octets (unité : 1010), et les lignes de couleur noire et autre que noire représentent respectivement les données brutes et les prévisions.

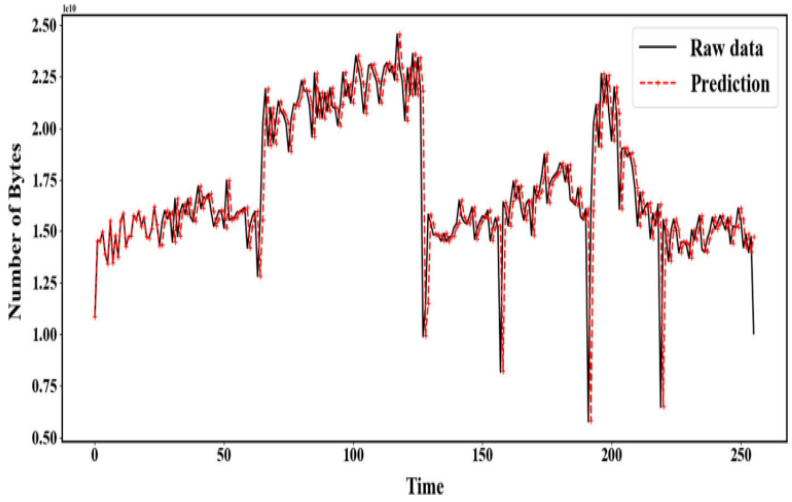

FIGURE 3.7 – Données brutes et prévision à 1 étape pour (r = 0.1) d'INGARCH.

La figure (3.7) montre le total des octets de paquets IPv4 du traffic des quatre heures des quatre dates et sa prévision à un pas en avant dans l'ordre temporel lorsque r est égal à

0,1 (les performances d'INGARCH). Les données des quatre dates sont utilisées pour voir comment le traffic des autres jours affectent (une prévisions pour les autres jours).

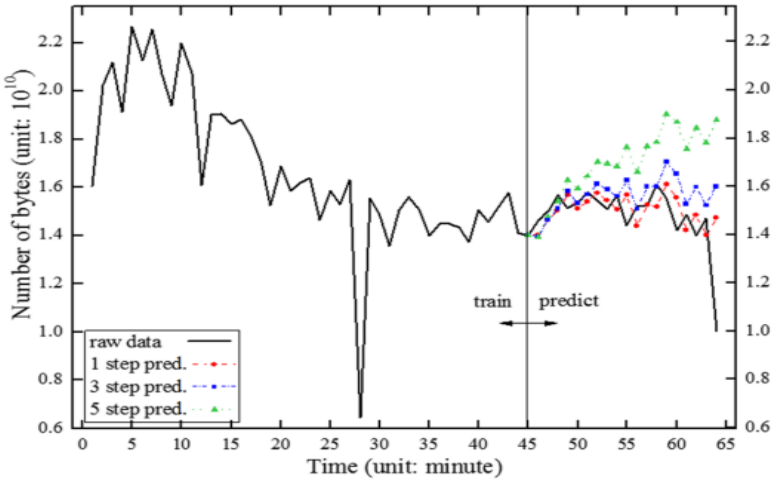

FIGURE 3.8 – Comparaison des prévisions à n étapes pour (r = 0,7) d'INGARCH.

La figure (3.8) compare les prévisions à 1, 3 et 5 pas en avant avec des données de (06/04/2016) uniquement lorsque (r = 0,7) d'INGARCH. Elle montre que la Prévision à 1 étape est la meilleure, comme prévu .

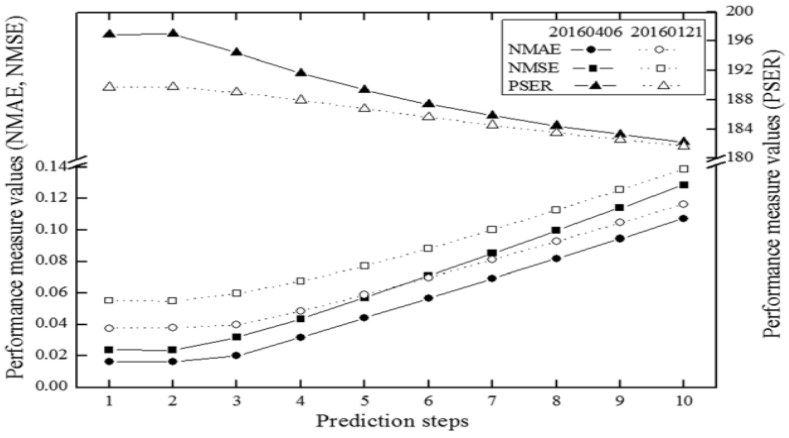

FIGURE 3.9 – Comparaison des mesures de performance d'INGARCH avec des prévisions jusqu'à 10 étapes d'avance pour un ration (r = 0,7).

La figure (3.9) compare les mesures de performance d'INGARCH avec des données de (06/04/2016) et (21/01/2016) jusqu'aux prévisions à 10 étapes lorsque (r = 0,7). Quand le pas augmente, le NMAE et le NMSE augmentent, tandis que le PSER diminue comme prévu.

Données	Etape	NMAE(0.1/0.7)	NMSE(0.1/0.7)	SER(0.1/0.7)
Tout les dates	1	0.0191/0.0239	0.0329/0.0423	206.16/198.65
	3	0.1582/0.0630	0.1858/0.0754	191.13/193.63
06/04/2016	1	0.0235/0.0157	0.0407/0.0237	197.37/196.82
	3	0.0508/0.0193	0.0626/0.0316	193.63/194.41
	5	0.1261/0.0432	0.1444/0.0568	186.37/189.31
01/21/2016	1	0.0375/0.0365	0.0490/0.0550	195.47/189.70
	3	0.0542/0.0383	0.0687/0.0596	192.55/189.01
	5	0.1230/0.0569	0.1464/0.0770	185.97/186.79

TABLE 3.3 – Résumé des valeurs des mesures de performance d'INGARCH obtenues lorsque (r = 0,1) et (r = 0,7)

Le tableau (3.3) résume les valeurs des mesures de performance d'INGARCH lorsque (r = 0,1) et (r = 0,7). Il montre que la prévision à une étape est plus précise lorsque r égale à 0,1 quand les données de toutes les dates sont utilisées, tandis que lorsque le nombre d'étapes augmente et r égale à 0,7 la précision revient à les données d'une seule date. Cela s'explique par le fait que le traffic est plus susceptible d'être corrélé au traffic du même jour plutôt qu'au traffic d'un autre jour.

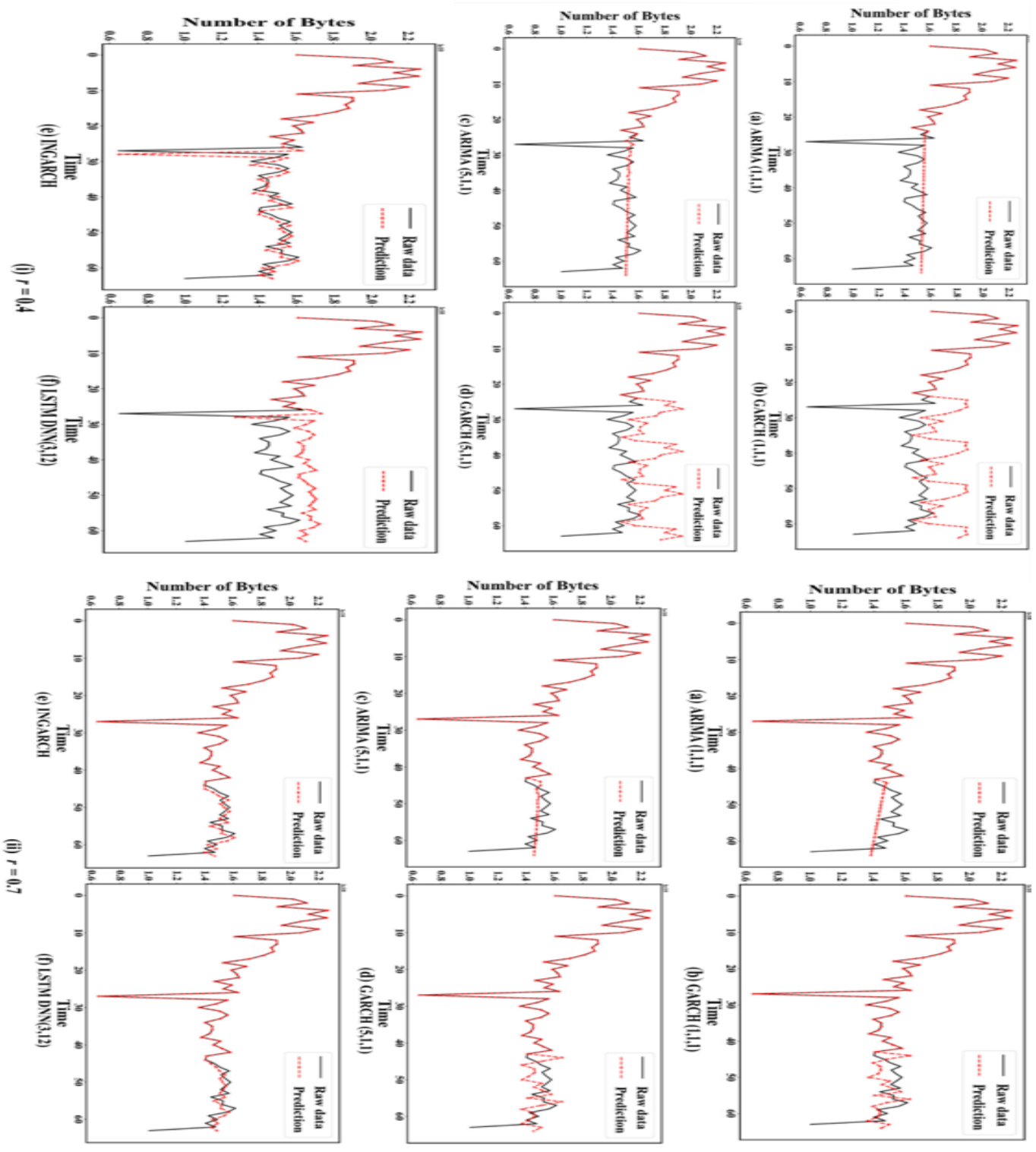

FIGURE 3.10 – Comparaison de quatre modèles pour un ration (i) r = 0,4 et (ii) r = 0,7 .

La figure (3.10) compare les prévisions à 1 étape pour quatre différents modèles ont utilisont les données de (06/04/2016) avec deux valeurs différentes de r, (i) r =0,4 et (ii) r = 0,7.
Les performances de deux ordres différents de (p) pour un modèle ARIMA et GARCH sont comparés, [(a) & (c)] et [(b) & (d)] et les performances d'un modèle INGARCH et d 'un modèle LSTM DNN avec 3 couches cachées et 12 blocs dans chaque couche [(e) & (f)].
Pour évaluer la précision de la prévision du modèle INGARCH, nous avons réalisé une étude comparative avec les modèles existants ARIMA, GARCH et LSTM DNN. Comme le montre le tableau (3.4) qui présente les mesures de performance des quatre modèles présentés dans la figure (3.10).

Modèles	NMAE(0.4/0.7)	NMSE(0.1/0.7)
ARIMA(1,1,1)	0.0944/0.0712	0.1293/0.0867
ARIMA(5,1,1)	0.0862/0.0573	0.1238/0.0814
GARCH(1,1,1)	0.2032/0.0865	0.2329/0.1062
GARCH(5,1,1)	0.2031/0.0888	0.2359/0.1095
LSTM RNN(1,4)	0.1891/0.0691	0.1998/0.1003
LSTM DNN(3,12)	0.1691/0.0555	0.1851/0.0833
INGARCH	0.0243/0.0161	0.0453/0.0237

TABLE 3.4 – Comparaison les mesure de performance (NMAE & NMSE) pour les quatre modèles avec un ration (r = 0,4) et (r = 0,7).

Sur la base de la figure (3.10) et du tableau (3.4), les NMAE et NMSE des modèles sont obtenus comme suit

$$\begin{cases} I < A(5,1,1) < A(1,1,1) < LD(3,12) < LR(1,4) < G(5,1,1) < G(1,1,1) & pour \ \ NMAE(0,4), \\ I < LD(3,12) < A(5,1,1) < LR(1,4) < A(1,1,1) < G(1,1,1) < G(5,1,1) & pour \ \ NMAE(0,7), \\ I < A(5,1,1) < LD(3,12) < A(1,1,1) < LR(1,4) < G(1,1,1) < G(5,1,1) & pour \ \ NMSE(0,7), \\ I < A(5,1,1) < A(1,1,1) < LD(3,12) < LR(1,4) < G(1,1,1) < G(5,1,1) & pour \ \ NMSE(0,4). \end{cases} \tag{3.11}$$

Où I, G, A, LR, et LD représentent INGARCH, GARCH, ARIMA, LSTM RNN, et LSTM DNN respectivement.
Selon (3.11), INGARCH semble être le meilleur modèle parmi les quatre modèles pour les mesures étudie et sur la base de cette comparaison les valeurs obtenues sont considérées comme fiables. Par conséquent, le modèle INGARCH est approprié pour le modèle de prévision du trafic.

3.4.3 Conclusion

Dans notre exemple, le modèle INGARCH introduit est un modèle de série chronologique non linéaire, qui est capable de capturer les caractéristiques du trafic de réseau et la propriété de dépendance à longue distance car l'arrivée des paquets dans les topologies de réseau telles que IoT, WLAN et VANET avec taux de connexion élevés apparaît toujours comme un processus de Poisson, alors pour une modélisation le modèle INGARCH semble être un modèle adéquat pour la prévision du trafic de réseau que les séries chronologiques linéaires traditionnelles n'ont pas réussi à prendre en compte.
Les performances du modèle proposé sont comparées à celles de trois modèles différents : moyenne mobile intégrée autorégressive, GARCH, et réseau neuronal récurrent à mémoire à long terme, Le résultat obtenu est que le modèle INGARCH est plus facile à mettre en oeuvre que le modèle de réseau neuronal (NN) et qu'il peut mieux capturer les caractéristiques du trafic réseau que les autres modèles statistiques.

Conclusion générale

Dans ce travail nous avons présenté une synthèse bibliographique sur les modèles de séries chronologiques à valeurs entières où ces modèles sont utilisés dans le cadre de la résolution des problèmes de comptage. Nous avons scindé ce mémoire en trois chapitres, les deux premiers chapitres concernent la présentation des de classes des modèles de séries temporelles à valeurs entières tandis que le dernier chapitre porte un exemple d'application d'un modèle INGARCH pour la modélisation et la prévision sur du trafic réseau internet qui a été étudié récemment par Kim [55].

Concernant les deux catégories de modèles de séries chronologiques à valeurs entières cités dans ce mémoire, à savoir : les modèles basés sur l'opérateur d'amincissement et les modèles basés sur la régression discrète. Dans un premier temps, nous avons présenté les structures probabilistes pour certains modèles cités dans la littérature, puis dans un second temps nous avons présenté les principales méthodes d'estimations pour chaque modèle.
Le troisième chapitre a pour but de présenter une application d'un modèle INGARCH sur des données réelles (Trafic reseau internet) [55].

Enfin, Les modèles décrites dans les deux catégories sont devenue très populaires ces dernières années pour la modélisation des processus de comptage stationnaires. Mais un grand nombre d'autres modèles de séries chronologiques de comptage ont également été proposés dans la littérature comme les modèles de régression conditionnelle, les modèles de markov cachés, les modèles d'espace d'état et Les nouveaux modèles ARMA discrets (NDARMA) [19].

Bibliographie

[1] Ahmed, A. *Contribution à l'économétrie dans des séries tomporelles à valeurs entières*. Université Charle de Gaulls Lille 3, 2016.

[2] Alberto. Mozo, B. O., and Gomez-Canaval, S. *Forecasting short-term data center network traffic load with convolutional neural networks*. PLoS ONE, 13(2), e0191939., 2018.

[3] Alzaid, A., and Al-Osh, M. *First ordre integer autoregressive (INAR (1)) process*. Journal of Time Series Analysis,8(3), 261-275, 1987.

[4] Alzaid, A., and Al-Osh, M. *First-order integer-valued autoregressive process : distributional and regression properties*. Statistica Neerlandica 42(1), 53-61, 1988.

[5] Alzaid, A., and Al-Osh, M. *Integer-Valued Moving Average (INMA) Process*. Statistical Papers.29, 281-300, 1988.

[6] Alzaid, A., and Al-Osh, M. *An integer valued autoregressive structure (INAR(p)) process*. Journal Appl probab.27(2), 314-324, 1990.

[7] Amenan, C. *Les modèles VAR (p)*. livre p. 58-76, 2019.

[8] Arthur. P. Dempster, N. M. L., and Rubin, D. B. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. J. R. Statist. Soc.B 39, 1-38, 1977.

[9] Bendjeddou, S. *Inférence du quasi-maximum de vraisemblance de modèles de séries chronologique à valeurs entières*. Thèse doctorat, Université de Houari Boumediene (USTHB), 2018.

[10] Bentarzi, M., and Aries, M. *On some periodic INARMA (p,q) models.Communications in Statistics-Simulation and Computation*. 2020.

[11] Bentarzi, M., and Bentarzi, O. *Periodic Integer-Valued GARCH Model*. Communications in Statistics - Simulation and Computation. Vol. 46,1167-1188, 2017.

[12] Bentarzi, O. *Sur des modèles de séries chronologiques périodiques à valeur entières : structure,estimation et application.* Thèse doctorat, Université de Houari Boumediene (USTHB), 2018.

[13] Boris, A., and Christian, H. *Parameter estimation and diagnostic tests for INMA (1) processes.* 2019.

[14] Bourguignon, M. *poisson géometrique INAR (1) process for modeling cont time serie with overdispresion.* Statistica Neerlandica,70,176-192, year=2016.

[15] Brännäs, K., and Hall, A. *Estimation dans les modèles de moyenne mobile à valeur entière.* Applied Stochastic Models in Business and Industry.17, 277-291, 2001.

[16] Brännäs, K., and Quoreshi, A. *Integer-value moving average modeling of the transactions in stocks.* AppliedFinancial Economics 20, 1429-440, 2010.

[17] Chahine, J. *Une généralisation de la loi binomiale négative.* Revue de statistique appliquée, tome 13, p. 33-43, 1965.

[18] Chenyi. Chen, J. Hu, Q. M., and Zhang, Y. *Short-time traffic flow prediction with ARIMA-GARCH Model.* In Proceedings of IEEE IV, 607-612, 2011.

[19] Christian, H., and Weiss, A. *An introduction to discrete-valued time series.* 2018.

[20] Christian, W. *Modelling time series of counts with overdispersion.* Article, Stat Methods Appl 18,507-519, 2009.

[21] Christou, V., and Fokianos, K. *Quasi-likelihood inference for negative binomial time series models.* Journal of Time Series Analysis, 35, 55-78, 2014.

[22] Consul, P., and Jain, G. *A generalization of the Poisson distribution.* Technometrics 15 (4), 791-799, 1973.

[23] Cox, D., and Gudmundsson., G. *Statistical analysis of time series : some recent developments.* Scandinavian Journal of Statistics, 8, 93-115, 1981.

[24] Davis, R., and Liu, H. *Theory and inference for a class of observation-driven models with application to time series of counts.* Statistica Sinica, 26, 1673-1707, 2016.

[25] Dingding . Zhou, S. C., and Dong, S. *Network traffic prediction based on ARFIMA model.* International Journal of Computer Science Issues, 9(6, 3), 106-111, 2012.

[26] Donovan, T. *Short term forecasting : Introduction to the Box-Jenkinz Approach Wiley.* 1983.

[27] Du, J., and li, Y. *The integer valued autoregressive modele (INAR (p)) process.* Journal Time Ser129-142., 1991.

[28] Efron, B. *Double exponential families and their use in generalized linear regression.* Journal of the American Statistical Association 81,709-721, 1986.

[29] Elsaied, H., and Fried, R. *On robust estimation of negative binomial INARCH models.* Article, 79 :137-158, 2021.

[30] Fan, J., and Li, R. *Variable selection via nonconvace penalized likelihood and its oracle properties.* Journal of the American Statistical Association, 96,1348-1360, 2001.

[31] Feike. Drost, R. v. d. A., and Werker, B. *Note on integer-valued bilinear time series models.* Statistics Probability Letters, 78, 992-96, 2008.

[32] Fieke. Drost, R. Akker, B., and Werker. *Efficient estimation of auto-regression parameters and innovation distributions for semiparametric integer-valued AR(p) models.* Journal of the Royal Statistical Society : Series B (Statistical Methodology), 71, 467-485, 2009.

[33] Fokianos, K., and Tjøstheim, D. *Log-linear Poisson autoregression.* Journal of Multivariate Analysis 102,563-578, 2011.

[34] Franke, J., and Rao, T. S. *Multivariate first-order integer-valued autoregressions.* Technical report, Forschung Universitat Kaiserslautern, 1993.

[35] Franses, P., and Paap, R. *Periodic Time Series Models.* Oxford University Press, 2004.

[36] Freeland, R. K. *Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data.* PhD thesis, University of British Columbia, Canada, 1998.

[37] Gary.K. Grunwald, R.J. Hyndman, L. T., and Tweedie, R. *Nongaussian conditional linear AR(1) models.* Australian and New Zealand Journal of Statistics 42, 479-495, 2000.

[38] Gauthier, G. *Modele de type autorégressif pour les sèrie chronologie à valeur entiers non négatives.* Département de mathématique et informatique, 179 page.

[39] Gladyshev, E. *Periodically and almost-Periodically correlates Random processes with continuous time parameter.* Theory Prob et its App.8(2), 173-177., 1963.

[40] Goldberg, S. *Introduction to Difference Equations.* New York : Wiley, 1958.

[41] Gouriéroux, C. *ARCH Models and Financial Applications.* Springer Series in Statistics. NewYork : Springer, 1997.

[42] Hai.Y. Xu, M. Xie, T. G., and Fu, X. *A model for integer-valued time series with conditiona overdispersion.* Computational Statistics and Data Analysis 56(12), 4229-4242, 2012.

[43] Harry, J. *Likelihood Inference for generalized Integer autoregressive Time models.* 2019.

[44] Heinen, A. *Modelling time series count data : an autoregressive conditional poisson model.* Center for operations research and econometrics (CORE), Catholic University of Louvain, Belgium, 2003.

[45] Heinen, A. *Modelling time series count data : an autoregressive conditional Poisson model, Center for Operations Research and Econometrics.* CORE Discussion Paper No. 2003-63. University of Louvain. Belgium, 2003.

[46] www.caida.org.

[47] Hui. Zhang, D. W., and Sun, L. *Regularized estimation in GINAR(p) process.* Journal of the Korea Statistical Society, 2016.

[48] Imad. Alawe, A. Ksentini, Y. H.-A., and Bertin, P. *Improving traffic forecasting for 5G core network scalability : A machine learning approach.* IEEE Network, 32(6), 42-49, 2018.

[49] Jiang, D., and Hu, G. *GARCH model-based large-scale IP traffic matrix estimation.* IEEE Communications Letters, 13(1), 52-54, 2009.

[50] J.P. Dion, G. G., and Latour, A. *Branching processes with immigration and integer valued autoregressive time serie.* J.Serdia Math , 21(2), 123-136, 1995.

[51] Kachour, M. *Le processus autorégressifs à valeurs entiérautorégressifs à valeurs entiéres d'arrondi d'ordre p RINAR (p).* Journal de statistique, 2009.

[52] Kachour, M. *Une nouvelle classe de modèles autorégressifs à valeurs entiéres.* Thèse doctorat, Université de Rennes1, 2009.

[53] Kachour, M., and Yao, J. *The first order rounded integer valued autoregressiv RINAR (p) proces.* J.Time Ser Anal.4, 417-448, 2009.

[54] Kang. Yao, W. Dehui, Y. K., and Yulin, Z. *A new thinning-based INAR(1) process for underdispersed or overdispersed counts.* Journal of the Korean Statistical Society, 2019.

[55] Kim, M. *Network traffic prediction based on INGARCH model.* Springer Science, part of Springer Nature, 2020.

[56] Kim, S. *Forecasting Internet traffic by using seasonal GARCH models.* Journal of Communications and Networks, 13(6), 621-624., 2011.

[57] Klimko, L., and Nelson, P. *On conditional least squares estimation forstochastic processes.* Annals of Statistics, 6 :629-642, 1978.

[58] Konstantinos. Fokianos, A. A. R., and Tjøstheim, D. *Poisson Autoregression.* Journal of the American Statistical Association, 104 :488, 1430-1439, 2009.

[59] Latour, A. *The multivariate GINAR (p) process.* Adv.in Appl Probab.29(1), 298-248, 1997.

[60] Latour, A. *Existence and stochastic structure of a non negative integer valued autoregressive process.* J.Time Ser Anal.19(1), 439-455, 1998.

[61] Latour, A. *The Multivariate GINAR(p) Process.* Advances in Applied Probability, Vol. 29, No.1 (Mar., 1997), pp. 228-248, 2014.

[62] Lionel, T. *Séries chronologiques à valeurs entières.* Journées MAS et Journée en l'honneur de Jacques Neveu, Talence, France, 2011.

[63] Magda. Monteiro, M. S., and Pereira, I. *Integer valeud autoregressive processes with periodec structure.* J.statistical planning interfence, 2010.

[64] McCullagh, P., and Nelder, J. *Generalized Linear Models.* 2nd Edition. Chapman and Hall, 1989.

[65] McKnezie, E. *Contribution the discussion of lawrance and lewis.* J.R.Statist.Soc.47, 187-185, 1985a.

[66] McKnezie, E. *Somme simple modèld for discret variante time serie.* Water Resources Bulletin, 21, 645-650, 1985b.

[67] McKnezie, E. *Some ARMA models for dependent sequences of poisson counts.* Adv.Appl.Prob.20, 822-835, 1988.

[68] Michael.A. Benjamin, R. R., and Stasinopoulos, D. *Generalized autoregressive moving average model.* Journal of the American Statistical Association 98(1), 214-223, 2003.

[69] Miroslav.M. Ristic, H. B., and Nastic, A. *A new geomtric First order integer valued autoregressiv (NGINAR(1)) process.* J.Statist.plann.Inference, 139(7),2218-2226, 2009.

[70] Naushad. Mamode Khan, Y. Sunecher, V. J. M. R., and Khan, M. H.-M. *Investigating GQL-based inferential approaches for non-stationary BINAR (1) model under different quantum of over- dispersion with application, Computational Statistics*. 2018.

[71] Neal, P., and Rao, T. S. *MCMC for integer-valued ARMA processes*. J. Time Ser. Anal. 2007.28, 92-100, 2004.

[72] Norman.Lloyd . Johnson, A. K., and Kotz, S. *Univariate Discrete Distributions,*. 3rd ed. Wiley, New Jersey, 2005.

[73] Norman.Lloyd . Johnson, S. K., and Balakrishnan, N. *Discrete Multivariate Distributions*. Wiley-Interscience, 1997.

[74] Norman.Lloyd. Johnson, S. K., and Balakrishnan, N. *Discrete Multivariate Distributions*. Wiley-Interscience, 1997.

[75] Ouoreshi, A. S. *A review of INMA Integer valued model class : application and futuer development*. 2020.

[76] Paul. Doukhan, A. L., and Oriachi, D. *A simple Integer valued bilinear time serie model*. Adv.apl.prob, 559-578, 2006.

[77] Pedeli, X., and Karlis, D. *A bivariate Poisson INAR (1) model with application*. Statistical Modelling, 11, 325-349, 2011.

[78] Pedeli. Xanthi, C. A., and Konstantinos, F. *Likelihood estimation for the INAR(p) model by saddlepoint approximation*. Journal of the American Statistical Association, 110(511), 2015.

[79] Puig, P., and Valero, J. *Characterization of count data distributions involving additivity and binomial subsampling*. Bernoulli, 13, 544-55, 2007.

[80] Qinzhzng. Zhuo, Q. Li, H. Y., and Qi, Y. *Long short-termmemory neural network for network traffic prediction*. In Proceedings of ISKE (pp. 1-6), 2017.

[81] Quoreshi. Shahiduzzaman, A. N., and Reaz, U. *A Review of INMA Integer-valued Model Class, Application and Further Development*. 2020.

[82] René. Ferland, A. L., and Oraichi, D. *Integer-valued GARCH process*. Journal of Time Series Analysis, 27, 923-942, 2006.

[83] Ross, S. *Stochastic Processes*. Wiley, New York, 1996.

[84] Rydberg, T., and Shephard, N. *BIN Models for Trade-by-Trade Data. Modelling the Number of Trades in a Fixed Interval of Time*. Technical report 0740, Econometric Society, 2000.

[85] Sadoun, M. *Estimation et Test Localement Asymptotiquement Efficace dans les Modèles de Séries Chronologiques à Valeurs Entières Périodiques : Cas Paramétrique et Semi-Paramétriques.* Thèse de doctorat.Université USTHB, 2020.

[86] Sadoun, M., and Bentarzi, M. *Modèlisation PINAR (p) et prévision du nombre d'admmissions Hospiitaliéres.*

[87] Samira. Chabaa, A. Z., and Antari, J. *Identification and prediction of internet traffic using artificial neural networks.* Journal of Intelligent Learning Systems and Applications, 2, 147-155, 2010.

[88] Scotto. Manuel, H. C., and Sonia, G. *Thinning-based models in the analysis of integer-valued time series.* Statistical Modelling . 1-29, 2015.

[89] Silva, I. M. M. D. *Contributions to the analysis of discrete-valued time series.* PhD thesis, University of Porto, Portugal, 2005.

[90] Steutel, F., and Harn, K. V. *Discrete analogues of self decomposability and stability.* The Annals of probability.7(5),893-899, 1979.

[91] Weib. Christian, H.J.M.F. Martin, M. N., and A.Yuvraj. *INARMA (p) modeling of count time serie.* Article, 285-287, 2019.

[92] Xanthi. Pedeli, A. D., and Fokianos, K. *Likelihood estimation for the INAR(p) model by saddlepoint approximation.* Journal of the American Statistical Association, 2014.

[93] Xu, S., and Zeng, B. *Network traffic prediction model based on auto-regressive moving average.* Journal of Networks, 9(3), 653-659, 2014.

[94] Yoon, J. E., and Hwang, S. Y. *Zero-Inflated INGARCH Using Conditional Poisson and Negative Binomial : Data Application.* The Korean Journal of Applied Statistics 28(3), 583-592, 2015.

[95] Yu, G., and Kim, S. *Parametre change test for periodic integer valeud autoregresive proccess, Communications in Statistics.* theory and methods.49(9), 912-2898., 2020.

[96] Yu, K., and Zou, H. *The combined Poisson INMA(q) models for time series of counts.* J Appl Math. Article, 2015.

[97] Yunwei, C. *Integer valued time serie and renwal processes.* Thèse doctorat, 2009.

[98] Yuvraj. Sunecher, N. M. K., and Jowaheer, V. *Estimating the parameters of a BINMA Poisson model for a nonstationary bivariate time series.* Communication in Statistics : Simulation and Computation, 47 :9,6803-6827, 2017.

[99] Zhan, M. F., and Elbiaze, H. *Analysis and prediction of real network trafic.* Journal of Networks, 4(9), 855-865, 2009.

[100] Zhang, H. *Inference for INAR(p) processes with signed generalized power series thinning operator.* Journal of Statistical Planning and Inference, 140, 667-683, 2010.

[101] Zhani, M. *Prévision du trafic internet - Modèles et applications.* Thése du Doctorat en Informatique, UNIVERSITÉ DU QUÉBEC À MONTRÉAL, 2011.

[102] Zhitang. Chen, J. W., and Geng, Y. *Predicting future traffic using hidden Markov models.* In Proceedings of IEEE ICNP (pp.1-6), 2016.

[103] Zhu, F. *Modeling time series of counts with COM-Poisson INGARCH models.* School of Mathematics, Jilin University, Changchun 130012, China, 2011.

[104] Zhu, F. *A negative binomial integer-valued GARCH model.* Journal of Time Series Analysis, 32, 54-67, 2011.

[105] Zhu, F. *Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models.* Journal of Mathematical Analysis and its Applications, 389, 58-71, 2012.

[106] Zhu, F. *Zero-inflated poisson and negative binomial integer-valued garch models.* Journal of Statistical Planning and Inference, vol. 142, no. 4, pp. 826-839, 2012.

[107] Zucchini, W., and MacDonald, I. *Markov Models for Time Series.* CRC Press, Boca Raton, 2009.

Résumé

Plusieurs problèmes dans de nombreux domaines scientifiques sont résolus à l'aide des modèles de séries chronologiques à valeurs entières avec quelques applications et simulations, car lorsqu'une série ne prend qu'un nombre limité de valeurs entières, elle ne peut être approchée correctement par un modèle de séries chronologiques classique à valeurs réelles.

Notre objectif dans ce travail est de faire une synthèse bibliographique sur les modèles de série chronologique à valeurs entières existants et les classifier selon le classement de (Cox et al, 1981) en deux catégories principales qui jouent un rôle majeur dans la modélisation des données issues d'un phénomène de comptage.

Nous avons donné dans un premier temps, pour chacune de ces deux catégories de modèles de séries temporelles entières, la structure de probabilité de certains modèles cités dans la littérature, puis au second temps nous avons mis l'accent sur quelques approches d'estimation considérées dans la littérature pour ces modèles.

Pour exhiber la bonne performance des modèles de séries chronologiques à valeurs entières nous avons présenté une étude récente de (Kim, 2020) qui révèle les avantages de l'utilisation de ces modèles dans la prévision et l'analyse du trafic réseau, qui est devenue de plus en plus importante à l'heure actuelle et à l'avenir pour la surveillance du trafic réseau.

Mots clès : Séries chronologiques à valeurs entières, processus de comptage, stationnaire, ergodique, structure de probabilité, approches d'estimation, opérateur d'amincissement.

Abstract

Several problems in many scientific fields are solved using integer value time series models with some applications and simulations. Because when a series takes only a limited number of integer values, it cannot be approximated correctly by a classical real valued time series model.

Our objective in this work then is to make a bibliographical synthesis on the existing integer time series models and to classify them according to the classification of (Cox et al, 1981) in two main categories which play a major role in the modeling of data resulting from a counting phenomenon.

We have first given, for each of these two categories of integer time series models, the probability structure of some models cited in the literature, and then we have focused on some estimation approaches considered in the literature for these models.

To demonstrate the good performance of integer-valued time-series models, we presented a recent study conducted by (Kim, 2020) that reveals the advantages of using these models in network traffic prediction and analysis, which are becoming increasingly important nowadays and in the future on network traffic monitoring.

Keywords : Integer time series, counting processes, stationary , ergodic, probability structure, estimation approaches, operator of thinning.