

UNIVERSIDADE ABERTA



**Modelos autorregressivos com valores inteiros não negativos:  
aplicação em séries económicas de Cabo Verde**

**Nataniel Lopes Barros**

Mestrado em Estatística, Matemática e Computação: ramo de  
Estatística Computacional

2016

UNIVERSIDADE ABERTA



**Modelos autorregressivos com valores inteiros não negativos:  
aplicação em séries económicas de Cabo Verde**

**Nataniel Lopes Barros**

Mestrado em Estatística, Matemática e Computação: ramo de Estatística  
Computacional

Dissertação de mestrado orientada pela  
Professora Doutora Maria do Rosário Olaia Duarte Ramos  
e co-orientada pelo Professor Doutor Amílcar Oliveira

2016

# Resumo

No estudo de séries temporais, os processos estocásticos usuais assumem que as distribuições marginais são contínuas e, em geral, não são adequados para modelar séries de contagem, pois as suas características não lineares colocam alguns problemas estatísticos, principalmente na estimação dos parâmetros. Assim, investigou-se metodologias apropriadas de análise e modelação de séries com distribuições marginais discretas. Neste contexto, Al-Osh and Alzaid (1987) e McKenzie (1988) introduziram na literatura a classe dos modelos autorregressivos com valores inteiros não negativos, os processos INAR. Estes modelos têm sido frequentemente tratados em artigos científicos ao longo das últimas décadas, pois a sua importância nas aplicações em diversas áreas do conhecimento tem despertado um grande interesse no seu estudo.

Neste trabalho, após uma breve revisão sobre séries temporais e os métodos clássicos para a sua análise, apresentamos os modelos autorregressivos de valores inteiros não negativos de primeira ordem INAR (1) e a sua extensão para uma ordem  $p$ , as suas propriedades e alguns métodos de estimação dos parâmetros nomeadamente, o método de Yule-Walker, o método de Mínimos Quadrados Condicionais (MQC), o método de Máxima Verosimilhança Condicional (MVC) e o método de Quase Máxima Verosimilhança (QMV). Apresentamos também um critério automático de seleção de ordem para modelos INAR, baseado no Critério de Informação de Akaike Corrigido, AICC, um dos critérios usados para determinar a ordem em modelos autorregressivos, AR. Finalmente, apresenta-se uma aplicação da metodologia dos modelos INAR em dados reais de contagem relativos aos setores dos transportes marítimos e atividades de seguros de Cabo Verde.

**Palavras-chave:** Séries temporais de contagem, modelo INAR, métodos de estimação, séries económicas.



# Abstract

In time series analysis, it is usually assumed that the stochastic processes have continuous marginal distributions and generally, that is not suitable for modeling count series since its characteristics pose some statistical problems, mainly in the estimation of parameters. Thus, it was investigated appropriate methods of analysis and modeling time series with discrete marginal distributions. In this context, Al-OSH and Alzaid (1987) and McKenzie (1988) introduced in the literature the class of non-negative integer values autoregressive models, INAR processes. It has often been treated in many scientific articles over the past decades since, its importance in the application in different areas of knowledge has aroused a great interest in their study.

In this work, after a brief review of the classic time series models, we present the non-negative integer values autoregressive models of first order, INAR(1), and its extension to order  $p$ , INAR( $p$ ), their properties and some methods of estimation of parameters namely, the Yule-Walker method, the Conditional Least Squares method (MQC), the Conditional Maximum Likelihood method (MVC) and finally, the Quasi Maximum Likelihood. We also present an automatic criterion for order selection for INAR models, based on the Akaike Information Criterion Corrected, AICC, one of the criteria used to determine the order in autoregressive models, AR. Finally, real count data on the sectors of maritime transport and insurance activities in Cape Verde are analyzed considering the methodology of INAR models.

**Keywords:** Count time series, INAR model, estimation methods, economic time series.



# Agradecimentos

Agradeço à Professora Doutora Maria do Rosário Olaia Duarte Ramos pelas orientações e todo o auxílio durante a realização deste trabalho e ao Professor Doutor Amílcar Oliveira, pelos conselhos e incentivo.

Agradeço à minha mãe e meus irmãos por acreditarem sempre em mim e pelo apoio incondicional, sem o qual esta dissertação não seria possível. Também sou grato aos meus amigos e conhecidos que de alguma forma me apoiaram nos meus estudos.

Finalmente, agradeço ao Banco de Cabo Verde e ao Instituto Nacional de Estatística de Cabo Verde, por me facultarem os dados utilizados no estudo de caso prático.



# Sumário

<b>Resumo</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Abreviaturas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Séries Temporais</b>	<b>3</b>
2.1 Definição . . . . .	3
2.2 Tendência e sazonalidade . . . . .	4
2.3 Eliminação de tendência na ausência de sazonalidade . . . . .	5
2.3.1 Estimação paramétrica da tendência . . . . .	5
2.3.2 Estimação não paramétrica da tendência . . . . .	6
2.3.3 Método das diferenças iteradas . . . . .	7
2.4 Eliminação da tendência e da sazonalidade . . . . .	9
2.4.1 Método paramétrico . . . . .	9
2.4.2 O método das diferenças . . . . .	10
2.5 Processos Estacionários . . . . .	10
2.6 Processos lineares . . . . .	12
2.6.1 Processos lineares e processos lineares gerais . . . . .	12
2.6.2 Estimação da média e autocovariância . . . . .	14
2.7 Processos ARMA . . . . .	17
2.7.1 Processos MA(q) . . . . .	18
2.7.2 Processos AR(p) - Estacionaridade e casualidade . . . . .	19
2.7.3 Invertibilidade . . . . .	23
2.7.4 Autocovariância dos processos ARMA . . . . .	25
2.8 Teoria preditiva de processos lineares . . . . .	27

2.8.1	Estimação dos parâmetros ARMA . . . . .	27
2.9	Predição . . . . .	31
2.9.1	A previsão linear dos processos ARMA(p,q) . . . . .	31
2.9.2	Algoritmo de Durbin-Levinson . . . . .	32
<b>3</b>	<b>Modelos INAR</b>	<b>35</b>
3.1	Introdução . . . . .	35
3.2	Operador Thinning Binomial . . . . .	36
3.3	Processos INAR(1) . . . . .	38
3.3.1	Propriedades gerais do modelos INAR(1) . . . . .	39
3.3.2	Estimação dos parâmetros . . . . .	41
3.4	Processos INAR(p) . . . . .	46
3.4.1	Introdução . . . . .	46
3.4.2	Métodos de estimação . . . . .	50
3.4.3	Método de quase-verossimilhança utilizando a abordagem de mo- delos lineares generalizados . . . . .	60
3.4.4	Propriedades de correlação . . . . .	61
<b>4</b>	<b>Seleção de ordem nos modelos INAR</b>	<b>65</b>
4.1	Introdução . . . . .	65
4.2	Seleção de ordem autorregressiva . . . . .	66
4.3	Seleção de ordem nos modelos INAR . . . . .	70
<b>5</b>	<b>Aplicação a dados reais</b>	<b>73</b>
5.1	Transporte marítimo em Cabo Verde . . . . .	73
5.2	Atividades de seguros . . . . .	77
	<b>Conclusões</b>	<b>81</b>
	<b>Referências Bibliográficas</b>	<b>83</b>

# Lista de Figuras

3.1	Função de autocorrelação amostral do processo INAR(1), com $\alpha = 0.9$ e $\lambda = 1$ . . . . .	40
5.1	Número de navios estrangeiros de longo curso fundeados . . . . .	74
5.2	Funções de autocorrelação e autocorrelação parcial do número de navios estrangeiros de longo curso fundeados nos portos de Cabo Verde . . . . .	74
5.3	Função de autocorrelação dos erros do modelo 5.1 . . . . .	75
5.4	Número de navios nacionais de cabotagem fundeados nos portos de Cabo Verde . . . . .	76
5.5	Funções de autocorrelação e autocorrelação parcial do número de navios nacionais de cabotagem fundeados nos portos de Cabo Verde . . . . .	76
5.6	Número de sinistros comunicados no mês no ramo de acidentes e doença . . . . .	77
5.7	Funções de autocorrelação e autocorrelação parcial do número de sinistros comunicados no mês - ramo de acidentes e doença . . . . .	78
5.8	Número de sinistros comunicados no mês - ramo transportes automóveis . . . . .	79
5.9	Funções de autocorrelação e autocorrelação parcial do número de sinistros comunicados no mês no ramo de transportes automóveis . . . . .	79



# Lista de Tabelas

5.1	Estimativa dos coeficientes do modelo INAR(2) ajustado sobre o número mensal de navios de cabotagem nacionais fundeados . . . . .	76
5.2	Estimativa dos coeficientes do modelo INAR(5) ajustado sobre o número mensal de sinistros comunicados no ramo acidentes e doenças .	78
5.3	Estimativa dos coeficientes do modelo INAR(5) ajustado sobre o número mensal de sinistros comunicados no ramo transportes automóveis	80



# Lista de Abreviaturas

AICC	Akaike Information Criterion Corrected
AR	AutoRegressive
ARMA	AutoRegressive Moving Average
FPE	Erro de Previsão Final
i.i.d.	Independente e identicamente distribuído
INAR	INteger-valued AutoRegressive model
MA	Moving Average
MQC	Mínimos Quadrados Condicionais
MVC	Máxima Verosimilhança Condicional
QMV	Quase Máxima Verosimilhança
WN	White Noise (Ruído Branco)
YW	Yule-Walker Estimator



# Capítulo 1

## Introdução

Grande parte da literatura sobre séries temporais apresentam modelos do tipo ARMA, assumindo uma distribuição normal ou, pelo menos, uma distribuição contínua dos dados. No entanto, séries de valores discretos inteiros, como as séries de contagem, cujos valores pertencem a conjuntos finitos ou infinitos numeráveis, aparecem frequentemente na prática. Quando os valores destas séries temporais são elevados, podem ser analisados utilizando métodos tradicionais para variáveis aleatórias de valores contínuos. Todavia, quando estes valores são relativamente baixos, não se pode utilizar os métodos usuais pois, a simples multiplicação de uma variável aleatória inteira por um constante real (parâmetro) pode resultar numa variável aleatória não inteira.

Esta situação pode ser ultrapassada substituindo a multiplicação escalar por uma operação aleatória, *operação thinning*, (Steutel e Van Harn, 1979), o correspondente da multiplicação para o contexto de valores inteiros. Vários processos análogos dos modelos tradicionais de análise de séries temporais lineares, para séries de contagem e com base na operação thinning, tem sido propostos.

Em particular, os modelos autorregressivos com valores inteiros Não Negativos (INAR) e os processos AR lineares apresentam estruturas de correlação similares e as condições de estacionaridade são também as mesmas. Os processos INAR constituem uma classe interessante de processos permitindo especificar a estrutura de dependência e escolher a distribuição marginal dentre um conjunto de distribuições discretas. Devido à importância da sua aplicação em diversas áreas de conhecimento, o

interesse no estudo e modelação de séries temporais de valores discretos vem crescendo nas últimas décadas.

Este trabalho tem como objetivo estudar os processos autorregressivos com valores inteiros não negativos, entender o seu comportamento, características e como utilizá-los para a modelação de séries de dados reais. O capítulo 2 apresenta uma revisão sobre as séries temporais e alguns métodos tradicionais utilizados na sua análise, estimação dos parâmetros e previsão. As séries de contagem, o conceito de modelo autorregressivo com valores inteiros não negativos (INAR) e os métodos de estimação dos parâmetros são apresentados no capítulo 3. No capítulo 4 é apresentado um método de escolha de ordem nos processos INAR e finalmente, no capítulo 5, a aplicação dos processos INAR em séries temporais do setor dos transportes marítimos e atividades de seguros de Cabo Verde.

# Capítulo 2

## Séries Temporais

### 2.1 Definição

Uma série temporal ou cronológica é uma sequência de observações ordenadas ao longo do tempo. Assume-se ainda a existência de dependência entre instantes. A definição matemática adequada para o estudo e a previsão de tais sequências consiste em considerá-las como uma realização particular de um Processo Estocástico, um conjunto de variáveis aleatórias  $\{X_t, t \in I\}$ , definidas sobre um espaço de probabilidade, onde  $I$  é um intervalo de tempo que pode ser discreto, (neste caso,  $I = \{1, 2, \dots, T\}$  e  $T$  é o número de observações) ou contínuo (e neste caso nota-se  $\{X_t, t \in I\}$ ).

A análise de séries temporais consiste inicialmente na especificação do modelo, com um número finito de parâmetros em que se deve ter em consideração a delimitação do fenómeno de estudo, a identificação das variáveis, o estudo das correlações entre estas e a definição da finalidade do modelo, a fim de orientar a especificação da forma matemática, a seleção de variáveis e o número de equações.

A seguir, deve-se estimar os parâmetros do modelo, de modo que os desvios entre os valores observados e estimados sejam mínimos, e realizar os testes de ajustamento adequados que permitem verificar a compatibilidade do conjunto de valores observados com uma distribuição específica.

Um dos principais objetivos do estudo de uma série temporal consiste na previsão dos valores futuros  $X_{T+h}$  ( $h = 1, 2, \dots$ ) da série a partir de seus valores observados até ao

instante  $T : X_1, \dots, X_T$ . Esta predição  $\hat{X}_T(h)$  é geralmente diferente do valor real  $X_{T+h}$  no período  $T + h$ . Esta diferença mede-se definindo o erro de predição  $\hat{X}_T(h) - X_{T+h}$  com a ideia de que quanto maior for  $h$ , maior será o erro. O intervalo de predição  $\hat{X}_{T+h}^{(1)}$  e  $\hat{X}_{T+h}^{(2)}$  é suscetível de conter o valor desconhecido de  $X_{T+h}$ . Para tal análise, deve ser eliminada a tendência da série que representa a evolução a médio e longo prazo do fenómeno estudado e age como uma forte correlação entre as variáveis  $X_t$ , mas não exprime nenhuma ligação de carácter explicativo. Corrige-se também a série observada de eventuais variações sazonais resultantes de um comportamento cíclico. Pode-se também desejar estudar a influência que exerce uma série sobre outra. Sejam  $(X_t)$  e  $(Y_t)$  duas séries temporais, examina-se por exemplo, se existem entre elas relações do tipo:

$$Y_t = a_1 X_{t-1} + a_3 X_{t-3}.$$

A seguir, pode-se interessar em saber qual série influencia a outra (casualidade) e conhecer o período e a duração de tal influência (desfasamento temporal).

## 2.2 Tendência e sazonalidade

Nas representações gráficas de séries temporais, pode-se observar em certos casos, um componente determinístico que se apresenta como uma tendência ou ciclo sazonal. Uma série pode também apresentar estes dois componentes ao mesmo tempo.

Uma série temporal  $X_t$  pode ser decomposta em

$$X_t = m_t + S_t + Y_t$$

onde,

$m_t$  é a tendência, e pode ser uma função descrita por um número finito de parâmetros, por exemplo, uma função linear  $a + bt$  ou um polinômio em  $t$ ;

$S_t$  é a sazonalidade e é uma função periódica. Seja  $d$  o período de  $S_t$ , então  $S_{t+d} = S_t$  para todo  $t$ . Supõe-se também que  $S_t$  não contém tendência e escreve-se

$$\sum_{j=1}^d S_{t+j} = 0;$$

$m_t + S_t$  é o componente determinístico do modelo;

$Y_t$  é o componente aleatório do modelo, de média nula mas possuindo geralmente uma estrutura de correlação não nula. É sem dúvida a parte à qual se interessa para a modelação.

## 2.3 Eliminação de tendência na ausência de sazonalidade

A parte determinística do modelo é unicamente composta por uma tendência  $m_t$ , ou seja,

$$X_t = m_t + Y_t.$$

### 2.3.1 Estimação paramétrica da tendência

Este método consiste em estimar a tendência através do método de mínimos quadrados ordinários. Supõe-se que o componente pode ser estimado por uma função linear

$$\hat{m}_t = \hat{a} + \hat{b}t.$$

Para encontrar os estimadores, minimiza-se o erro quadrático cometido quando se substitui  $m_t$  por  $\hat{m}_t$ . Observando a série  $X_1, \dots, X_T$ , deve-se encontrar os coeficientes  $\hat{a}$  e  $\hat{b}$  que minimizam

$$\sum_{t=1}^T (X_t - a - bt)^2.$$

A solução deste problema de minimização é então

$$\begin{cases} \hat{a} = \frac{4T+2}{T-1} \bar{X} - \frac{6}{T(T-1)} \sum_{t=1}^T tX_t \\ \hat{b} = \frac{6}{T(T-1)} \left( \frac{2}{T(T+1)} \sum_{t=1}^T tX_t - \bar{X} \right) \end{cases}$$

#### Demonstração:

Derivando em relação a  $a$  e  $b$ , após algumas simplificações e igualando a zero, resulta o sistema:

$$(i) \quad \sum_{t=1}^T X_t - aT - b \sum_{t=1}^T t = 0;$$

$$(ii) \quad \sum_{t=1}^T tX_t - a \sum_{t=1}^T t - b \sum_{t=1}^T t^2 = 0.$$

Utilizando as fórmulas usuais:

$$1. \sum_{t=1}^T t = \frac{T(T+1)}{2}$$

$$2. \sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$$

e dividindo (i) e (ii) por  $T$ , tem-se:

$$(i') \quad \bar{X} - a - \frac{(T+1)b}{2} = 0$$

$$(ii') \quad \frac{1}{T} \sum_{t=1}^T tX_t - \frac{(T+1)a}{2} - \frac{(T+1)(2T+1)b}{6} = 0$$

onde,

$\bar{X}$  é a média empírica de  $\{X_t\}_{t=1}^T$ . Pode-se verificar que de fato,  $a$  e  $b$  são mínimos da função.

Todavia, a modelização linear da tendência pode ser muito simplificadora. Em certos casos, pode-se estimar a tendência quadrática pelo mesmo método:  $\hat{m}_t = \hat{a} + \hat{b}t + \hat{c}t^2$ .

### 2.3.2 Estimação não paramétrica da tendência

Em certas ocasiões, o polinómio utilizado para estimar  $m_t$  pode não ser nem linear nem quadrático, o que dificulta encontrar o grau do polinómio de ajustamento para a tendência  $m_t$ , tornando assim impossível a utilização do método dos mínimos quadrados. Todavia, pode-se utilizar um polinómio de grau elevado o que torna o cálculo fastidioso. Neste caso, recorre-se à teoria de estimação não paramétrica da tendência que não supõe que esta seja polinomial *a priori*, mas que  $m_t$  seja linear num intervalo  $[t - q, t + q]$ . Um bom estimador da tendência seria dado por

$$m_t = \frac{1}{2q+1} \sum_{k=-q}^q X_{t-k}.$$

Assim, pode-se calcular  $\hat{m}_t$  para cada valor de  $t$  calculando esta média sobre as  $2q$  observações à volta de  $t$ . Para evitar problemas nos extremos da série (quando  $t \leq q$

ou  $t > T - q$ ), considera-se:

$$\begin{cases} X_t = X_1, & \text{se } t \leq 1 \\ X_t = X_T, & \text{se } t \geq T \end{cases}$$

### 2.3.3 Método das diferenças iteradas

Pode-se também eliminar a tendência sem ter de estimá-la. Definimos o operador de retardo como sendo a função linear  $B$  que, a todo  $X_t$  faz corresponder a observação precedente  $X_{t-1}$ , tal que

$$B(X_t) = X_{t-1}.$$

Este operador pode ser utilizado recursivamente, para obter a penúltima observação antes de  $t$ ,

$$\begin{aligned} B^2(X_t) &= B(B(X_t)) \\ &= B(X_{t-1}) \\ &= X_{t-2} \end{aligned}$$

e pode-se escrever, por indução, que

$$B^n(X_t) = X_{t-n}.$$

Introduzimos também o operador de diferença  $\nabla$  cuja aplicação a  $X_t$  fornece a diferença entre  $X_t$  e o valor precedente  $X_{t-1}$  da série,

$$\nabla X_t = X_t - X_{t-1}.$$

Em virtude da definição precedente, temos diretamente:

$$\begin{aligned} \nabla X_t &= X_t - BX_t \\ &= (1 - B)X_t. \end{aligned}$$

Assim como o operador  $B$ , pode-se iterar  $\nabla$  para obter

$$\begin{aligned}\nabla^2 X_t &= \nabla \nabla X_t \\ &= (1 - B)[(1 - B)X_t] \\ &= (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}.\end{aligned}$$

De um modo geral, o operador de diferença de ordem  $d$  resulta, do valor  $X_t$  da série observada, na diferença de  $X_t$  com o valor da série no período  $t - d$ ,

$$\nabla_d X_t = X_t - X_{t-d}$$

e demonstra-se que

$$\nabla_d X_t = (1 - B^d)X_t.$$

Supõe-se que a tendência seja da forma de um polinómio de ordem  $k$ , descrevendo-se como

$$m_t = \sum_{j=0}^k a_j t^j.$$

Se se aplica a  $m_t$  o operador de diferença  $\nabla$  (de ordem 1), obtém-se

$$\begin{aligned}\nabla m_t &= m_t - m_{t-1} \\ &= \sum_{j=0}^k a_j t^j - \sum_{j=0}^k a_j (t-1)^j\end{aligned}$$

e, desenvolvendo  $(t-1)^j$ , obtém-se no segundo membro da equação,

$$\sum_{j=0}^k a_j (t-1)^j = \sum_{j=0}^k a_j t^j + \sum_{j=1}^k b_{j-1} t^{j-1}.$$

Para os coeficiente  $b_{j-1}$ ,  $\nabla m_t$  torna-se

$$\begin{aligned}\nabla m_t &= \sum_{j=1}^k b_{j-1} t^{j-1} \\ &= \sum_{i=0}^{k-1} b_i t^i \quad \text{com } i = j - 1,\end{aligned}$$

e vê-se que o operador  $\nabla$  diminuiu a ordem do polinómio  $m_t$ :  $m_t$  é de ordem  $k$  enquanto que  $\nabla m_t$  é de ordem  $k - 1$ . Assim, se se aplica o operador  $\nabla$   $k$  vezes ao polinómio  $m_t$ , obtém-se

$$\nabla^k m_t = k! a_k = \text{constante independente de } t,$$

e assim aplicado a  $X_t = m_t + Y_t$ , o operador  $\nabla$  resulta

$$\nabla^k X_t = \text{constante} + \nabla^k Y_t.$$

Na prática, ignora-se o grau do polinómio de  $m_t$  e por isso, aplica-se o operador  $\nabla$  o número de vezes suficiente para obter uma série  $\nabla^k X_t$  cuja média seja nula (e então tem-se apenas o componente aleatório).

## 2.4 Eliminação da tendência e da sazonalidade

### 2.4.1 Método paramétrico

Seja uma série, que além da tendência  $m_t$ , apresenta um componente sazonal  $s_t$ ,

$$X_t = m_t + S_t + Y_t.$$

Supõe-se que se observa uma série de dados mensais com o período  $d = 12$ . Neste caso,  $\{X_t\}_{t=1, \dots, T} = \{X_{jk}\}$ ,  $j = 1, \dots, n$  anos e  $k = 1, \dots, 12$  meses, com a relação

$$X_{jk} = X_{k+12(j-1)}.$$

Um estimador natural da tendência para o ano  $j$  é então

$$\hat{m}_j = \frac{1}{12} \sum_{k=1}^{12} X_{jk}, \quad j = 1, \dots, n.$$

Os resíduos  $X_j - \hat{m}_j$  incluem o componente sazonal. Um estimador deste componente  $S_k$  é dado pela média empírica sobre o mesmo mês de cada ano dos resíduos,

$$\hat{S}_k = \frac{1}{n} \sum_{j=1}^k (X_{jk} - \hat{m}_{jk}), \quad n = \text{anos}, \quad k = 1, \dots, 12.$$

Obtém-se assim, após a eliminação da tendência e da sazonalidade, a série

$$X_{jk} - \hat{m}_j - \hat{S}_k$$

ou seja, o componente aleatório da série que se utiliza para a análise.

### 2.4.2 O método das diferenças

O método das diferenças é utilizado para eliminar a tendência e a sazonalidade. Aplica-se o operador de diferença de ordem  $d$ ,  $\nabla_d$  onde  $d$  é o período de  $s_t$  na série  $X_t$ ,

$$\begin{aligned} \nabla_d X_t &= X_t - X_{t-d} \\ &= m_t + S_t + Y_t - m_{t-d} - S_{t-d} - Y_{t-d} \\ &= (m_t - m_{t-d}) + (S_t - S_{t-d}) + (Y_t - Y_{t-d}) \end{aligned}$$

onde, por definição,  $S_t - S_{t-d}$  é nulo.  $\nabla_d X_t$  é assim uma série temporal dessazonalizada, mas apresenta uma tendência  $\nabla_d m_t$ . Os métodos de eliminação da tendência na ausência de sazonalidade se aplicam.

## 2.5 Processos Estacionários

Seja uma série de variáveis aleatórias  $\{X_t\}_{t=1,2,\dots}$ . Diz-se que esta série é estacionária em média quando ela se desenvolve no tempo, aleatoriamente, ao redor de uma

média constante ou seja, a média não muda no decorrer do tempo.

$$EX_t = EX_0, \quad \forall t = 0, 1, 2, \dots$$

Esta série é também estacionária em variância quando

$$Var(X_t) = Var(X_0), \quad \forall t = 0, 1, 2, \dots$$

**Definição.**  $\{X_t\}_{t=0,1,2,\dots}$  é uma série estacionária em sentido lato ou de segunda ordem quando se verifica simultaneamente as três condições seguintes:

- $EX_t = \mu < \infty, \quad \forall t \in \mathbb{N}$  - Estacionaridade em média
- $EX_t^2 < \infty, \quad \forall t \in \mathbb{N}$  - Variância finita
- $Cov(X_s, X_{s+t}) = Cov(X_{s-1}, X_{s-1+t}) = \dots = Cov(X_0, X_t) \quad \forall s, t \in \mathbb{N}$  - Invariância da estrutura de covariância.

A partir desta propriedade, pode-se introduzir a sequência

$$r_h^X = cov(X_s, X_{s+h}),$$

que é independente de  $s$ . É a sequência das autocovariâncias de  $\{X_t\}$  e é bem definida apenas se  $\{X_t\}$  é estacionária de segunda ordem ou seja, respeita a terceira condição. Pode-se normalizar esta série e definir a sequência das autocorrelações por

$$\rho_h^X = \frac{r_h^X}{r_0^X},$$

com as seguintes propriedades, válidas por toda série de variáveis aleatórias estacionárias de segunda ordem :

- $r_0 = Var(X_t) = \text{Constante}$
- Simetria:  $r_h = r_{-h} \quad \forall h$
- $|r_h| < r_0 \quad \forall h$

e, para a série das autocorrelações,

- $\rho_0 = 1$
- Simetria:  $\rho_h = \rho_{-h} \quad \forall h$
- $|\rho_h| \leq 1 \quad \forall h.$

Pode-se também definir um conceito de estacionaridade a partir das leis de probabilidade conjunta das variáveis aleatórias da sequência  $X_t$ .

**Definição.** A série  $\{X_t\}$  é estacionária no sentido estrito se

$$\mathcal{L}(X_{t+h_1}, X_{t+h_2}, \dots, X_{t+h_k}) = \mathcal{L}(X_{h_1}, X_{h_2}, \dots, X_{h_k}) \quad \forall t, \forall (h_1, \dots, h_k), \forall k$$

Este conceito de estacionaridade é mais exigente do que a estacionaridade no sentido lato, como indica o lema seguinte.

**Lema 1:** Se  $\{X_t\}$  é estacionária em sentido estrito, e  $EX_t^2 < \infty$ , então  $(X_t)$  é estacionária em sentido lato. O recíproco geralmente é falso. Todavia, se  $(X_t)$  é gaussiano, então  $\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = N_n(\mu_n, \Sigma_n)$  e os conceitos de estacionaridade em sentido estrito e lato se coincidem.

### Ruído branco, o exemplo mais simples de processo estacionário

A série de variáveis aleatórias  $\{\varepsilon_t\}$  é um ruído branco em sentido lato (e respectivamente estrito) se:

- $E\varepsilon_t = 0, \quad \forall t \in \mathbb{Z}$
- $E\varepsilon_t^2 = \sigma^2$  é constante e estritamente positiva.
- $cov(\varepsilon_s, \varepsilon_t) = 0$  se  $t \neq s$  (respectivamente os  $\varepsilon_t$  são independentes e identicamente distribuídos (i.i.d.))

## 2.6 Processos lineares

### 2.6.1 Processos lineares e processos lineares gerais

Um processo linear é um processo estocástico  $X_t$  formado por uma combinação linear de ruídos brancos em sentido estrito. A classe de processos lineares gerais são

constituídos de combinações lineares de ruídos brancos em sentido lato.

**Definição.**  $\{X_t\}_{t \in \mathbb{Z}}$  é um processo linear (respetivamente um processo linear geral) de média  $\mu$  se pode ser escrito da forma

$$X_t = \mu + \sum_{k=-\infty}^{\infty} b_k \varepsilon_{t-k},$$

onde  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é um ruído branco em sentido estrito (respetivamente lato) com variância  $\sigma_\varepsilon^2$ , e onde a série dos coeficientes  $b_k$  é suposta tal que

$$\sum_{k=-\infty}^{\infty} b_k^2 < \infty.$$

Consideramos uma série de números  $m_n \rightarrow \infty$  e formemos o processo truncado

$$\mu + \sum_{k=-m_n}^{m_n} b_k \varepsilon_{t-k},$$

que é bem definido pois o número dos termos da soma é finito. A convergência significa que a média quadrática

$$E \left( \mu + \sum_{k=-m_n}^{m_n} b_k \varepsilon_{t-k} - X_t \right)^2 \leq \sum_{|k| \geq m_n} b_k^2 E(\varepsilon_{t-k}^2) = \sigma_\varepsilon^2 \sum_{|k| \geq m_n} b_k^2$$

tende para 0 quando  $n \rightarrow \infty$  (e então quando  $m_n \rightarrow \infty$ ).

**Proposição.** Se  $\{X_t\}$  é um processo linear geral, como  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , então  $\{X_t\}$  é estacionário e temos:

$$\begin{cases} r_0^X = \text{Var } X_t = \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} b_k^2 \\ r_s^X = \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} b_k b_{k+s} \quad \forall s \in \mathbb{Z} \end{cases}$$

### Demonstração

Sabe-se que para todo  $t \in \mathbb{Z}$ ,  $E X_t = \mu$ . Para a verificação das outras duas condições,

calcula-se a covariância entre  $X_t$  e  $X_{t+s}$ ,

$$\begin{aligned}
 \text{Cov}(X_t, X_{t+s}) &= E(X_t, X_{t+s}) - E(X_t)E(X_{t+s}) \\
 &= E\left(\mu^2 + \mu \sum_{l=-\infty}^{\infty} b_l \varepsilon_{t+s-l} + \mu \sum_{k=-\infty}^{\infty} b_k \varepsilon_{t-k} + \sum_{k,l=-\infty}^{\infty} b_k b_l \varepsilon_{t-k} \varepsilon_{t+s-l}\right) - \mu^2 \\
 &= \mu \sum_{l=-\infty}^{\infty} b_l E(\varepsilon_{t+s-l}) + \mu \sum_{k=-\infty}^{\infty} b_k E(\varepsilon_{t-k}) + \sum_{k,l=-\infty}^{\infty} b_k b_l E(\varepsilon_{t-k} \varepsilon_{t+s-l})
 \end{aligned}$$

onde a passagem da média (esperança) na soma finita é justificada pelo teorema de convergência monótona e o facto de que  $E X_t = \mu < \infty$  para todo  $t$ . Como  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , temos:

$$\begin{aligned}
 \text{Cov}(X_t, X_{t+s}) &= \sum_{k,l=-\infty}^{\infty} b_k b_l \text{Cov}(\varepsilon_{t-k}, \varepsilon_{t+s-l}) \\
 &= \sigma_\varepsilon^2 \sum_{k,l=-\infty}^{\infty} b_k b_l \delta_{t-k, t+s-l} \\
 &= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} b_k b_{k+s}.
 \end{aligned}$$

Portanto  $\text{Cov}(X_t, X_{t+s})$  não depende de  $t$  e temos os resultados anunciados.

## 2.6.2 Estimação da média e autocovariância

Sejam  $X_1, \dots, X_t$  uma série de  $T$  observações produzidas por um processo estacionário  $X_t$  de média  $\mu$  e de função de autocovariância  $r_h^X$ . Nesta secção estimamos as duas quantidades desconhecidas  $\mu$  e  $r_h^X$ .

### 2.6.2.1 Estimação da média

Para estimar a média (teórica)  $\mu$ , um estimador natural é a média empírica é

$$\hat{\mu} = \bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t.$$

**Proposição.** Se  $X_t$  é um processo estacionário e que  $\bar{X}_t$  designa a média para as  $T$  observações da série, então:

- $E(\bar{X}_T) = \mu$
- $Var(\bar{X}_T) = \frac{1}{T} \sum_{|h| < T-1} \left(1 - \frac{|h|}{T}\right) r_h^X$
- Se os  $X_t$  são *i.i.d.*, então  $Var(\bar{X}_T) \rightarrow 0$  quando  $T \rightarrow \infty$ .

Por exemplo, se se sabe que o processo linear  $\{X_t\}$  é um processo gaussiano, então

$$\sqrt{T}(\bar{X}_T - \mu) \stackrel{d}{\sim} N(0, T \times Var(\bar{X}_T)).$$

Supõe-se que se quer efetuar um teste bilateral sobre a média

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

a um nível  $\alpha$  fixo. Rejeita-se  $H_0$  se a estatística

$$\frac{\bar{X}_T - \mu}{\sqrt{Var(\bar{X}_T)}} > Z_{\frac{\alpha}{2}} \text{ ou}$$

$$\frac{\bar{X}_T - \mu}{\sqrt{Var(\bar{X}_T)}} < -Z_{\frac{\alpha}{2}}$$

onde  $Z_{\frac{\alpha}{2}}$  é o quantil de uma variável aleatória normal centrada reduzida.

O intervalo de confiança ao nível  $1 - \alpha$  é

$$\left[ \bar{X}_T - Z_{\frac{\alpha}{2}} \sqrt{Var(\bar{X}_T)}; \bar{X}_T + Z_{\frac{\alpha}{2}} \sqrt{Var(\bar{X}_T)} \right].$$

Quando se desconhece a lei de  $X_T$ , utiliza-se a distribuição assintótica  $\sqrt{T}(\bar{X}_T - \mu)$  dada por

$$\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{L} N(0, \sum_{h \in \mathbb{Z}} r_h^X),$$

quando  $T \rightarrow \infty$ .

### 2.6.2.2 Estimação da função de autocovariância

Para construir um estimador da função de autocovariância teórica  $r_h^X$ , sabe-se que, se  $(X_1, Y_1), \dots, (X_T, Y_T)$  são observações bivariadas *i.i.d.* de variância finita, o estimador

de covariância entre  $X$  e  $Y$  é dado por

$$\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)(Y_t - \bar{Y}_T).$$

Então estima-se  $r_h^X$  pela função de autocovariância empírica

$$\hat{r}_h^X = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T),$$

definida para  $0 \leq h \leq T - 1$ . Para os valores negativos de  $h$ , utiliza-se a simetria da autocovariância teórica  $r_h^X$ , e define-se  $\hat{r}_h^X(h) = \hat{r}_h^X(-h)$  para  $h < 0$ . Este estimador tem a vantagem de ser um estimador sem viés da autocovariância  $r_h^X$  teórica, ou seja, a sua média (esperança) é exatamente igual a  $r_h^X$ . Todavia, tem a inconveniência de não verificar a propriedade da série das autocovariâncias teóricas  $\{r_h^X\}_{h \geq 0}$  ser não negativa o que é embaraçoso, por exemplo, quando se quer utilizar este estimador nas equações de Yule-Walker.

Se  $X_T$  é um processo linear e sob certas condições, as propriedades de  $\hat{r}_h^X$  são do mesmo tipo que para o estimador da média:  $\hat{r}_h^X$  é um estimador consistente e assintoticamente não enviesado. Portanto,

$$E \hat{r}_h^X \rightarrow r_h^X \quad \text{quando } T \rightarrow \infty.$$

O mesmo tipo de estimador pode ser utilizado para estimar a função de autocorrelação. Neste caso utilizamos

$$\hat{\rho}_h^X = \frac{\hat{r}_h^X}{\hat{r}_0^X},$$

que se chama função de autocorrelação empírica.

A proposição seguinte apresenta o teorema de limite central para  $\hat{\rho}_h^X$  que será muito útil para construir testes sobre a série das autocorrelações empíricas.

**Proposição.** Se  $\{X_t\}$  é um processo linear com  $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$  e cujas duas condições seguintes são respeitadas:

- $\sum_{j \in \mathbb{Z}} |b_j| < \infty$

- $E\varepsilon_j^4 < \infty$

então, para todo  $h \geq 1$  fixo, tem-se, quando  $T \rightarrow \infty$ ;

$$\sqrt{T}(\hat{\rho}_h^X - \rho_h^X) \xrightarrow{L} N(0, \Sigma)$$

onde  $\Sigma$  é a matriz de covariância assintótica, cujo elemento  $(i, j)$  é dado pela fórmula de Bartlett

$$\begin{aligned} \Sigma_{ij} &= \sum_{k \in \mathbb{Z}} \left( \rho_{k+i}^X \rho_{k+j}^X + \rho_{k-i}^X \rho_{k+j}^X + 2\rho_i^X \rho_j^X (\rho_k^X)^2 - 2\rho_i^X \rho_k^X \rho_{k+j}^X - 2\rho_j^X \rho_k^X \rho_{k+i}^X \right) \\ &= \sum_{k \geq 1} \left( \rho_{k+i}^X \rho_{k-i}^X - 2\rho_i^X \rho_k^X \right) \left( \rho_{k+j}^X \rho_{k-j}^X - 2\rho_j^X \rho_k^X \right) \end{aligned}$$

## 2.7 Processos ARMA

Os processos ARMA são lineares e sob certas condições, estacionários. A sua linearidade fornecerá uma teoria simples de previsão.

**Definição.** Um processo linear estacionário  $\{X_t\}$  chama-se  $ARMA(p, q)$ ,  $p \geq 0$ ,  $q \geq 0$  se existem constantes  $a_1, \dots, a_p$  ( $a_p \neq 0$ ) e  $\theta_1, \dots, \theta_q$  ( $\theta_q \neq 0$ ) e um processo  $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$  tais que

$$X_t - \sum_{k=1}^p a_k X_{t-k} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}.$$

O processo  $\{\varepsilon_t\}$  chama-se processo das inovações ou processo de ruído branco (WN-white noise).

**Proposição.** Para uma determinada sequência de autocovariâncias  $\{r_h\}$  e para cada número inteiro  $K > 0$ , existe um processo  $ARMA\{X_t\}$  tal que as suas  $K$  primeiras autocovariâncias  $r_h^X$  coincidem com  $r_h$  :  $r_h^X = r_h$  para  $h = 0, \dots, K$ .

### Exemplos

Se  $p = q = 0$ , o processo torna-se

$$X_t = \varepsilon_t \sim WN(0, \sigma_\varepsilon^2),$$

e portanto, um processo  $ARMA(0, 0)$  é um ruído branco no sentido lato.

As duas grandes categorias de processos ARMA que estudaremos são aquelas em que um dos dois parâmetros  $p$  ou  $q$  é nulo e o outro não nulo.

- Caso em que  $p = 0$  e  $q > 0$ , os processos  $ARMA(0, q)$  escritos  $MA(q)$ , da forma

$$X_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

e chamam-se processos de média móvel de ordem  $q$ .

- Caso em que  $p > 0$  e  $q = 0$ , os processos  $ARMA(p, 0)$ , escritos simplesmente  $AR(p)$ , da forma

$$X_t = \varepsilon_t + \sum_{k=1}^p a_k X_{t-k}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

e chamam-se autorregressivos de ordem  $p$ .

Na definição dos processos ARMA, indicamos que estes processos são estacionários. Esta precisão é necessária porque um processo definido tal como acima pode não ser estacionário. A seguir, examinamos as condições sobre os coeficientes  $a_1, \dots, a_p$  e  $\theta_1, \dots, \theta_q$  para que o processo definido seja estacionário.

### 2.7.1 Processos $MA(q)$

Estes processos aparecem naturalmente como casos particulares de um processo linear geral no qual, o número de coeficientes não nulos é finito, e então a série infinita torna-se finita. Seja um processo da forma

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2),$$

para  $q > 0$  fixo e por todo  $t \in \mathbb{Z}$ . A primeira pergunta que fazemos é se este processo é estacionário. Sabe-se que  $EX_t = 0$  e, em termos de estrutura de segunda ordem, obtemos:

$$r_h^X = \begin{cases} \sigma_\varepsilon^2 \sum_{k=0}^{q-h} \theta_k \theta_{k+h} & \text{se } 0 \leq h \leq q \\ 0 & \text{se } h > q \end{cases}$$

com  $\theta_0 = 1$ . Consequentemente todo processo de média móvel é automaticamente um processo estacionário.

A função de autocorrelação de um processo MA(q) resulta automaticamente:

$$\rho_h^X = \begin{cases} \frac{\theta_h + \theta_1\theta_{h+1} + \dots + \theta_{q-h}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{se } 1 \leq h \leq q \\ 0 & \text{caso contrário} \end{cases}$$

e em particular,  $\rho_q^X = \frac{\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}$  anula apenas se  $\theta_q = 0$ , o que significa que o processo não é de ordem  $q$ . Se  $h > q$ , a função de autocorrelação de um processo MA(q) se anula. Esta observação é útil para a modelação se, a partir dos dados  $X_1, \dots, X_T$ , a função de autocorrelação empírica não é significativamente diferente de zero além de um certo número  $q_0$ , deve-se então escolher ajustar um modelo MA( $q_0$ ) aos dados observados.

## 2.7.2 Processos AR(p) - Estacionaridade e casualidade

Esses processos formam uma classe flexível de modelos para muitos fenómenos estudados. Definem-se implicitamente pela relação

$$X_t = \varepsilon_t + \sum_{k=1}^p a_k X_{t-k}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2). \quad (2.1)$$

Este processo é estacionário? Qual é sua forma explícita?

Consideremos o processo AR(1). Temos:

$$X_t = \varepsilon_t + aX_{t-1}$$

e em iterando esta equação recursiva, temos para todo  $K \in \mathbb{N}_0$

$$X_t = a^K X_{t-K} + \sum_{k=1}^K a^{k-1} \varepsilon_{t-k+1}.$$

Suponhamos primeiramente que  $|a| < 1$ . Neste caso temos:

$$\begin{aligned} E \left[ X_t - \sum_{k=1}^K a^{k-1} \varepsilon_{t-k+1} \right]^2 &= a^{2K} E [X_{t-K}]^2 \\ &= \text{constante } a^{2K} \end{aligned}$$

se  $X_t$  é estacionário, portanto, de variância constante. Como supomos que  $|a| < 1$ , temos  $a^{2K} \rightarrow 0$  quando  $K \rightarrow \infty$  e portanto

$$E \left[ X_t - \sum_{k=1}^K a^{k-1} \varepsilon_{t-k+1} \right]^2 \rightarrow 0$$

quando  $K \rightarrow \infty$ , o que permite escrever

$$X_t = \sum_{k=1}^{\infty} a^{k-1} \varepsilon_{t-k+1} = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}. \quad (2.2)$$

Pelo teorema de convergência monótona, vemos que  $EX_t = 0$ . Para a função de autocovariância, o mesmo argumento permite escrever:

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \lim_{K \rightarrow \infty} \left( \sum_{k=1}^K a^{k-1} \varepsilon_{t-k+1}, \sum_{k=1}^K a^{k-1} \varepsilon_{t+h-k+1} \right) \\ &= \lim_{K \rightarrow \infty} \left( \sigma_\varepsilon^2 a^{|h|} \sum_{k=1}^K a^{2k} \right) \\ &= a^{|h|} \sigma_\varepsilon^2 \sum_{k=0}^{\infty} a^{2k} = \sigma_\varepsilon^2 \frac{a^{|h|}}{1-a^2} \end{aligned}$$

pela fórmula de uma série geométrica que converge porque  $|a| < 1$ . Se  $|a| > 1$ , a série não converge. Pode-se todavia ultrapassar esta dificuldade em reescrevendo o processo AR(1) como:

$$X_t = -\frac{1}{a} \varepsilon_{t+1} = \frac{1}{a} X_{t+1}$$

Pelos mesmos argumentos relativos ao caso  $|a| < 1$ , pode-se mostrar que

$$X_t = - \sum_{k=1}^{\infty} a^{-k} \varepsilon_{t+k}$$

é solução da equação implícita e que a série converge. Todavia, esta solução liga

$\{X_t\}$  com os valores futuros das inovações  $\{\varepsilon_{t+k}\}$ . Esta solução não é natural pois o processo está correlacionado com realizações ainda não observadas dos processos  $\varepsilon_t$ . Diz-se então que esta solução não é casual.

**Definição.** Um processo AR(p) é casual quando existe uma série de números  $\{\alpha_k\}$  tal que  $\sum_{k \in \mathbb{Z}} |\alpha_k| < \infty$  e

$$X_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k}.$$

Portanto, todo processo MA é por definição casual.

Nota-se que uma condição necessária e suficiente para que um processo AR(1) seja estacionário e casual é:

$$|a| < 1$$

e neste caso, os coeficientes  $\alpha_k$  de sua representação casual  $MA(\infty)$  são dados por  $\alpha_k = a^k$ .

Observamos igualmente que a situação  $|a| = 1$  não fornece solução estacionária o que se explica pela forma da autocovariância.

A generalização deste resultado em casos onde o modelo autorregressivo é de ordem  $p > 1$  não é direta. Consideremos o modelo 2.1 no qual a introdução do operador de retardo  $B$  permite reescrevê-lo sob a forma:

$$X_t = \varepsilon_t + \sum_{k=1}^p a_k (B^k X_t)$$

ou ainda:

$$\left(1 - \sum_{k=1}^p a_k B^k\right) X_t = \varepsilon_t.$$

ou seja, forma-se o operador  $A(B) = 1 - \sum_{k=1}^p a_k B^k$  que é aplicado a  $X_t$ . Temos então

$$A(B)X_t = \varepsilon_t. \quad (2.3)$$

O resultado seguinte fornece as condições necessárias e suficientes de estacionariedade e de casualidade dos modelos autorregressivos sobre o comportamento do operador  $A$ .

**Proposição.** O processo autorregressivo 2.3 é casual e estacionário se e apenas se o polinômio em  $z$  :  $A(z) = 1 - \sum_{k=1}^p a_k z^k$  é tal que:

$$A(z) \neq 0 \text{ para todo } z \in \mathbb{C} \text{ tal que } |z| \leq 1. \quad (2.4)$$

Os coeficientes  $\alpha_k$  aparecem na representação casual  $MA(\infty)$  ou seja, são determinados por:

$$1 - \sum_{j=1}^{\infty} \alpha_j z^j = \frac{1}{A(z)}$$

se  $|z| \leq 1$ .

Por motivação destas condições, vejamos ainda o exemplo de um processo AR(1). Neste caso, o estudo da raiz do polinômio é particularmente simples:  $A(z) = 1 - az$  tem como raiz  $z = \frac{1}{a}$ . Então, a condição  $|a| < 1$  para a estacionaridade e casualidade de um processo AR(1) coincide com a condição  $|z| > 1$  enunciada acima:  $A(z) = 1 - az \neq 0$  para todo  $z \in \mathbb{C}$  tal que  $|z| \leq 1$ . Para os processos AR(p) com  $p > 1$ , as considerações seguintes permitem estabelecer a condição 2.4. Seja o caso  $p = 2$ .

As raízes  $z_1, z_2 \neq 0$  do polinômio de um AR(2)  $A(z) = 1 - a_1 z - a_2 z^2 = a_2(z - z_1)(z - z_2)$ , são funções dos coeficientes  $a_1, a_2$ . É mais fácil fornecer a condição de estacionaridade em função destas raízes  $z_1, z_2$ , pois a equação de um AR(2)  $A(B)X_t = \varepsilon_t$  se transforma como segue,

$$A(B)X_t = z_1 z_2 \left(1 - \frac{B}{z_1}\right) \left(1 - \frac{B}{z_2}\right) X_t = \varepsilon_t,$$

o que resulta

$$\left(1 - \frac{B}{z_1}\right) \left(1 - \frac{B}{z_2}\right) X_t = \frac{\varepsilon_t}{z_1 z_2} =: \eta_t \sim WN(0, \sigma_\eta^2).$$

Portanto, o processo definido por  $Y_t := \left(1 - \frac{B}{z_1}\right) X_t$  é um processo AR(1) com o parâmetro  $z_1^{-1}$ ,

$$\left(1 - \frac{B}{z_1}\right) Y_t = \eta_t,$$

que é estacionário e casual se  $|z_1^{-1}| < 1 \Leftrightarrow |z_1| > 1$ . De seguida, o processo  $X_t$  definido por

$$\left(1 - \frac{B}{z_2}\right) X_t = Y_t$$

é um AR(1) com parâmetro  $z_2^{-1}$ ; existe uma solução estacionária e casual para  $X_t$  se

$|z_2^{-1}| < 1 \Leftrightarrow |z_2| > 1$ . Então, um AR(2) é estacionário e casual se as raízes  $z_1, z_2$  de seu polinômio  $A(z)$  verificam:  $z_1, z_2 \in \{z : |z| > 1\}$ .

O mesmo raciocínio se aplica aos processos AR( $p$ ),  $p > 2$ .

Nota-se que a proximidade das raízes à região limite de não estacionaridade  $\{z : |z| \leq 1\}$  influencia como simular uma realização de um processo AR( $p$ ) estacionário.

Então examina-se a existência de uma solução estacionária da equação AR(1) pelas equações diferenciais: uma solução  $X_t = \xi_t + \eta_t$  se escreve como uma soma de  $\xi_t$  sendo solução da equação homogênea  $\xi_t - \xi_{t-1} = 0$  e uma solução especial  $\eta_t$  da equação heterogênea  $\eta_t - a\eta_{t-1} = \varepsilon_t$ .

A solução da equação homogênea  $\xi_t - \xi_{t-1} = 0$  é  $\xi_t = a^t \xi_0$  (onde  $\xi_0$  é um valor inicial qualquer) o que não é estacionário, mas tende para zero,  $\xi_t \rightarrow 0, |t| \rightarrow \infty$  se  $|a| < 1$ . Para encontrar a solução da equação especial, seja  $\eta_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k}$  um processo linear (com  $b_k = a^k \Rightarrow \sum_{k=0}^{\infty} |b_k| < \infty$  pois  $|a| < 1$  é estacionário). Temos

$$\eta_t - a\eta_{t-1} = \sum_{k \geq 0} a^k \varepsilon_{t-k} - \sum_{k \geq 0} a^{k+1} \varepsilon_{t-1-k} = \varepsilon_t,$$

então  $\eta_t$  é efetivamente solução da equação especial.

Portanto, a soma de  $\xi_t$  e  $\eta_t$ , isto é,  $X_t = \xi_t + \eta_t = a^t \xi_0 + \eta_t \approx \eta_t$ , se  $t$  é suficientemente grande. Então, a partir de um valor  $t_0$ , para  $t > t_0$ ,  $X_t$  se comporta como um processo estacionário.

### 2.7.3 Invertibilidade

Consideremos o processo MA(1)

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

que se assemelha formalmente ao processo AR(1) trocando os  $\varepsilon_t$  por  $X_t$ . Se aplicarmos as mesmas técnicas que na seção precedente, mas mudando os  $\varepsilon_j$  por  $X_j$ ,

então a relação acima implica, sob as condições sobre  $\theta$  discutidas a seguir, que:

$$\varepsilon_t = - \sum_{k=1}^{\infty} \theta^{-k} X_{t+k}.$$

Pelas mesmas razões que na secção precedente, esta expressão não apresenta nenhuma interpretação física razoável pois significa que as inovações  $\varepsilon_t$  no tempo  $t$  dependem dos valores futuros da série  $\{X_{t+k}\}_{k>1}$ . Diz-se neste caso que o processo não é invertível: a partir do processo MA(1) apresentado anteriormente, as inovações não podem ser representadas em função dos valores passados do processo.

A invertibilidade do processo MA(1) depende do valor de  $\theta$ . Pelos mesmos cálculos que na secção precedente, mas trocando as inovações  $\varepsilon_t$  com os valores  $X_t$  do processo, vemos que o processo MA(1) é invertível se e apenas se  $|\theta| < 1$ , lembrando que um processo de médias móveis é sempre um processo estacionário.

**Definição.** Se  $X_t$  é um processo MA( $q$ ) como definido acima, diz-se que  $X_t$  é invertível quando existe uma sequência  $\phi_j$  com  $\sum_j |\phi_j| < \infty$  tal que

$$\varepsilon_t = \sum_{j=0}^{\infty} \phi_j X_{t-j}.$$

Por definição, vemos diretamente que todo processo autorregressivo é automaticamente invertível. A equação precedente é por vezes qualificada de "representação  $AR(\infty)$ " do processo e constitui uma das principais motivações para o seu estudo: um processo linear que é invertível permite uma tal representação  $AR(\infty)$  com os coeficientes  $\phi_j$  rapidamente crescentes. Então existe uma boa aproximação por um processo AR( $p$ ), com uma ordem  $p$  relativamente pequena. Isto permite tratar o processo linear no quadro simples e eficaz de estimação e predição dos processos autorregressivos de ordem  $p$ . O resultado seguinte apresenta as condições necessárias e suficientes de invertibilidade dos modelos de médias móveis.

**Proposição.** O processo de médias móveis é invertível se e apenas se o polinômio em  $z$  :  $\Theta(z) = \sum_{j=0}^q \theta_j z^j$  é tal que:

$$\Theta(z) \neq 0 \text{ para todo } z \in \mathbb{C} \text{ tal que } |z| \leq 1$$

Os coeficientes  $\theta_j$  aparecendo na definição do processo MA(q) são determinados por:

$$\sum_{j=0}^{\infty} \theta_j z^j = \frac{1}{\Theta(z)}$$

se  $|z| \leq 1$ .

De um modo mais geral, temos o mesmo resultado para a classe dos processos lineares.

**Proposição.** Seja  $X_t = \sum_{k \geq 0} b_k \varepsilon_{t-k}$  um processo linear (geral) casual onde  $\sum_k |b_k| < \infty$  se  $B(z) = \sum_k b_k z^k \neq 0 \forall |z| \leq 1$ . Então existe uma representação AR( $\infty$ )

$$\forall t \quad X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t,$$

onde  $A(z) = 1 - \sum_{k \geq 1} a_k z^k = \frac{1}{B(z)}, |z| \leq 1$ .

### 2.7.4 Autocovariância dos processos ARMA

Os resultados propostos pelos processos AR(p) ou MA(q) podem ser resumidos considerando um processo geral ARMA(p,q).

**Proposição.** Seja  $X_t$  um processo ARMA(p,q). Suponhamos que os polinômios  $A(Z) = 1 - a_1 Z - \dots - a_p Z^p$  e  $\Theta(Z) = 1 + \theta_1 Z + \dots + \theta_q Z^q$  não tem raiz comum.

1.  $X_t$  é estacionário e casual se e apenas se  $A(Z) \neq 0$  para todo  $Z \in \mathbb{C}$  tal que  $|Z| \leq 1$ .

Os coeficientes  $\alpha_k$  são determinados por:

$$\sum_{j=0}^{\infty} \alpha_j Z^j = \frac{\Theta(Z)}{A(Z)} \quad \text{se } |Z| \leq 1.$$

2.  $X_t$  é invertível se e apenas se  $\Theta(Z) \neq 0$  para todo  $Z \in \mathbb{C}$  tal que  $|Z| \leq 1$ . Os coeficientes  $\phi_j$  são determinados por:

$$\sum_{j=0}^{\infty} \phi_j Z^j = \frac{A(Z)}{\Theta(Z)} \quad \text{se } |Z| \leq 1.$$

Calculamos igualmente a função de autocorrelação de um processo MA(q). Se  $X_t$  é

da forma MA(q), então:

$$\rho_h^X = \begin{cases} \frac{\theta_h + \theta_1\theta_{h+1} + \dots + \theta_{q-h}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{se } 1 \leq h \leq q \\ 0 & \text{caso contrário} \end{cases}.$$

Esta fórmula nos permite encontrar a função de autocovariância de todo processo ARMA casual, pois estes processos podem ser postos sob a forma AR(p), e temos:

$$r_h^X = \sigma_\varepsilon^2 \sum_{k=0}^{\infty} \alpha_k \alpha_{k+h}.$$

Para encontrar os coeficientes  $\alpha_k$ , utiliza-se a última proposição que liga estes coeficientes aos polinômios  $A(z)$  e  $\Theta(z)$ :

$$\sum_{j=0}^{\infty} \alpha_j z^j = \frac{\Theta(z)}{A(z)} \quad \text{para } |z| \leq 1.$$

Pode-se reescrever esta equação sob a forma

$$A(z) \sum_j \alpha_j z^j = \Theta(z)$$

e em identificando os coeficientes dos polinômios tendo o mesmo grau, obtém-se:

$$\alpha_j - \sum_{0 < k \leq j} a_k \alpha_{j-k} = \theta_j \quad \text{para } 0 \leq j < \max(p, q + 1)$$

e,

$$\alpha_j - \sum_{0 < k \leq p} a_k \alpha_{j-k} = 0 \quad \text{para } j \geq \max(p, q + 1).$$

Estas equações podem ser recursivamente resolvidas para  $\alpha_0, \alpha_1, \alpha_2, \dots$ :

- $\alpha_0 = \theta_0 = 1$
- $\alpha_1 = \theta_1 + a_1 \alpha_0 = \theta_1 + a_1$
- $\alpha_2 = \theta_2 + \alpha_0 a_2 + \alpha_1 a_1 = \theta_2 + a_2 + \theta_1 a_1 + a_1^2 \dots$

## 2.8 Teoria preditiva de processos lineares

Nesta secção, supomos que a ordem do modelo foi escolhida. Sabemos que os dados provêm, digamos, de um processo ARMA( $p, q$ ) e perguntamos como estimar os  $p + q$  parâmetros,  $a_1, \dots, a_p, \theta_1, \dots, \theta_q$ , e como construir uma teoria preditiva para este modelo.

### 2.8.1 Estimação dos parâmetros ARMA

#### 2.8.1.1 Estimação dos parâmetros autorregressivos: as equações de Yule-Walker

Consideremos um processo autorregressivo de ordem  $p$ , casual e de média nula:

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2).$$

Multiplicando cada membro desta equação por  $X_{t-h}$  ( $h = 0, \dots, p$ ) e tomando a esperança de cada lado, obtém-se, para o primeiro membro, a função de autocovariância

$$E(X_t X_{t-h}) = r_h^X$$

e para o segundo membro

$$\sum_{k=1}^p a_k E(X_{t-k} X_{t-h}) + E(\varepsilon_t X_{t-h}) = \begin{cases} \sum_{k=1}^p a_k r_{k-h}^X & \text{se } h = 1, \dots, p \\ \sum_{k=1}^p a_k r_{k-h}^X + \sigma_\varepsilon^2 & \text{se } h = 0 \end{cases}$$

pois, pela hipótese de casualidade do processo, mostra-se que  $E(\varepsilon_t X_{t-h})$  se anula exceto quando  $h = 0$ . Igualando as duas últimas expressões, temos, para  $h > 0$  a seguinte relação:

$$r_h^X = \begin{cases} \sum_{k=1}^p r_{k-h}^X a_k & \text{se } h = 1, \dots, p \\ \sum_{k=1}^p r_{k-h}^X a_k + \sigma_\varepsilon^2 & \text{se } h = 0 \end{cases}$$

Este sistema pode-se escrever sob a forma matricial: seja  $r_p^X$  o vetor coluna de comprimento  $p$  formado por  $(r_1^X, \dots, r_p^X)'$ , seja  $a_p$  o vetor coluna  $(a_1, \dots, a_p)'$  e formando a

matriz  $p \times p$  cujo elemento  $(i, j)$  vale  $r_{i-j}^X$ , ou seja

$$R_p^X = \begin{pmatrix} r_0^X & r_1^X & \cdots & r_p^X \\ r_1^X & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1^X \\ r_p^X & \cdots & r_1^X & r_0^X \end{pmatrix}.$$

então o sistema pode se reescrever matricialmente e temos as equações

$$\begin{cases} r_p^X = R_p^X a_p \\ r_0^X = a_p' r_p^X + \sigma_\varepsilon^2 \end{cases}$$

denominadas de equações de Yule-Walker. Ligam os parâmetros  $a_p$  do processo autorregressivo à função de autocovariância teórica que define  $r_p^X$  e  $R_p^X$ .

Para estimar os parâmetros  $a_1, \dots, a_p$ , pode-se servir das equações de Yule-Walker utilizando a função de autocovariância empírica  $\hat{r}_p^X$  como um estimador de  $r_p^X$ . As relações entre  $a_p, r_p^X$  e  $R_p$  permitem encontrar o estimador  $\hat{a}_p$  de  $a_p$  e um estimador  $\hat{\sigma}_\varepsilon$  de  $\sigma_\varepsilon$  através do sistema de Yule-Walker seguinte, a resolver em  $\hat{a}_p$  e  $\hat{\sigma}_\varepsilon$

$$\begin{cases} \hat{r}_p^X = \hat{R}_p^X \hat{a}_p \\ \hat{r}_0^X = \hat{a}_p' \hat{r}_p^X + \hat{\sigma}_\varepsilon \end{cases}$$

Para estimar  $a_p$ , resolve-se o sistema anterior, invertendo a matriz  $R_p^X$ :

$$\hat{a}_p = (\hat{R}_p^X)^{-1} \hat{r}_p^X$$

e depois calculando  $\hat{\sigma}_\varepsilon = \hat{r}_0^X - \hat{a}_p' \hat{r}_p^X$ .

### Propriedades

O processo AR(p) definido por:

$$Y_t - \hat{a}_1 Y_{t-1} - \cdots - \hat{a}_p Y_{t-p} = \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$$

é estacionário e casual. É o modelo ajustado às observações  $X_1, \dots, X_T$ . A função de autocovariância do modelo ajustado nota-se  $r_h^Y$  e obedece as equações de Yule-

Walker:

$$\begin{cases} r_p^Y = R_p^Y \hat{a}_p, & h = 1, \dots, p \\ r_0^Y = R_p^Y \hat{a}_p + \hat{\sigma}_\varepsilon, & h = 0 \end{cases}$$

E, nota-se que

$$r_p^Y = \hat{r}_p^X, \quad h = 0, \dots, p$$

e portanto a função de autocovariância do modelo ajustado coincide com a função de autocovariância empírica.

**Proposição.** Se  $X_t$  é um processo estacionário e casual com  $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$  e se  $\hat{a}_p$  é o estimador de Yule-Walker de  $a_p$ , obtido a partir de  $T$  observações do processo autorregressivo, então, quando  $T \rightarrow \infty$ :

$$\sqrt{T}(\hat{a}_p - a_p) \rightarrow N_p(0, \sigma_\varepsilon^2 R_p^{-1})$$

onde  $R_p$  é a matriz  $p \times p$  de covariância dada pela matriz  $R_p^X$ . Esta matriz é invertível se  $r_0^X > 0$ . Ainda, para  $T \rightarrow \infty$ :

$$\hat{\sigma}_\varepsilon \xrightarrow{P} \sigma_\varepsilon^2.$$

### 2.8.1.2 Estimação pela máxima verosimilhança

O método de Yule-Walker não permite estimar de forma eficaz os parâmetros de um modelo com uma parte MA porque conduz a um sistema de equações não lineares. Nesta secção aplicamos como método alternativo, o método geral de *máxima verosimilhança* para estimar os parâmetros de um modelo ARMA.

Seja o processo casual ARMA(p,q) de média nula

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \theta_0 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q}, \quad \{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

onde  $\theta_0 = 1$  no qual supõe-se que

$$1 - a_1 z - \dots - a_p z^p \neq 0 \quad \text{para } |z| \leq 1$$

e

$$1 + \theta_1 z + \dots + \theta_q z^q \neq 0 \quad \text{para } |z| \leq 1.$$

O objetivo é estimar os vetores coluna  $a_p$  e  $\theta_q$ . Suponhamos que conhecemos a distribuição de  $\{X_t\}$  e que é gaussiano ou seja, que todas as funções de distribuição de  $\{X_t\}$  são multivariadas normais. Esta hipótese nos permite escrever a função de verosimilhança do modelo que é a probabilidade de observar  $X_T = (X_1, \dots, X_T)'$  sob a hipótese gaussiana:

$$\mathcal{L}(R_T^X) = (2\pi)^{-\frac{T}{2}} | \det R_T^X |^{-\frac{1}{2}} e^{-\frac{1}{2} X_T' (R_T^X)^{-1} X_T}$$

onde  $R_T^X$  é a matriz de covariância cujo elemento  $(i, j)$  é dado por  $E(X_i X_j) = r_{|i-j|}^X$ . Esta matriz é constituída pela função de autocovariância  $r_h^X$ .

Esta função de autocovariância pode ser exprimida em função dos parâmetros  $a_p, \theta_q, \sigma_\varepsilon^2$ . Então:

$$R_T^X = R_T^X(a_p, \theta_q, \sigma_\varepsilon^2)$$

e a função de verosimilhança do processo gaussiano se escreve:

$$\mathcal{L}(a_p, \theta_q, \sigma_\varepsilon^2) = \mathcal{L}(R_T^X(a_p, \theta_q, \sigma_\varepsilon^2)).$$

Se  $X_t$  não é um processo gaussiano, não se conhece as funções de distribuição de  $X_t$  e a função de verosimilhança não se encontra disponível. Todavia, mesmo nestes casos, utiliza-se a função de verosimilhança acima como uma medida de qualidade do ajustamento da matriz de covariância aos dados e considera-se esta função para encontrar os estimadores de  $a_p, \theta_q$  e  $\sigma_\varepsilon^2$ .

Estes estimadores são dados pela maximização de  $\mathcal{L}(a_p, \theta_q, \sigma_\varepsilon^2)$ . Pode-se mostrar que, se o processo ARMA(p,q) é casual, invertível e com  $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ , então o processo

$$Y_t = \sum_{i=1}^p \hat{a}_i X_{t-i} + \varepsilon_t + \sum_{j=1}^q \hat{\theta}_j \varepsilon_{t-j}, \quad \{\varepsilon_t\} \sim i.i.d.(0, \sigma_\varepsilon^2)$$

é também estacionário, casual e invertível.

**Proposição.** Se  $X_t$  é um processo casual e invertível, com  $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$  e se  $\beta = (a_p, \theta_q)$ , o estimador de máxima verosimilhança  $\hat{\beta} = (\hat{a}_p, \hat{\theta}_q)$  é assintoticamente normal

$$\sqrt{T}(\hat{\beta} - \beta) \rightarrow N(0, V(\beta))$$

onde a matriz variância-covariância  $V(\beta)$  é uma função do parâmetro  $\beta$  que coincide, no caso de um processo AR de ordem  $p$ , com a matriz da variância assintótica dos estimadores de Yule-Walker  $\sigma_\varepsilon^2 R_p^{-1}$ .

## 2.9 Predição

Para tratar o problema de predição de valores futuros  $X_{T+1}, X_{T+2}, \dots$  de uma série temporal observada até o tempo  $T$ , supõe-se que as observações provêm de um processo estacionário ARMA( $p, q$ ) casual, de média nula e cuja ordem  $(p, q)$  é conhecida bem como os seus coeficientes.

### 2.9.1 A previsão linear dos processos ARMA( $p, q$ )

Se se observa as realizações  $X_1, \dots, X_T$  da série  $\{X_t\}$ , procura-se prever o valor  $X_{T+h}$  ( $h$  é um número fixo) a partir de uma combinação linear das observações. Notando  $\hat{X}_{T+h}$  o valor previsto em  $t+h$  de  $\{X_t\}$ , estudaremos as previsões que se escrevem sob a forma:

$$\hat{X}_{T+h} = k_0 + k_1 X_T + \dots + k_T X_1$$

Os coeficientes  $k_0, \dots, k_T$  desta expressão são escolhidos de tal forma que a predição  $\hat{X}_{T+h}$  seja o mais próximo possível da real realização  $X_{T+h}$  do processo em  $T+h$ .

O erro de predição a minimizar é dado pelo erro quadrático médio

$$S(K) = E (X_{T+h} - k_0 - k_1 X_T - \dots - k_T X_1)^2$$

Os coeficientes  $K$  formam a solução de um sistema de  $T$  equações correspondentes a cada derivada parcial de  $S(K)$ . Este sistema pode-se escrever da forma seguinte:

$$E \left[ X_{T+h} - k_0 - \sum_{i=1}^T k_i X_{T+1-i} \right] = 0$$

$$E \left[ \left( X_{T+h} - k_0 - \sum_{i=1}^T k_i X_{T+1-i} \right) X_{T+1-j} \right] = 0, \quad j = 1, \dots, T$$

Da primeira equação resulta que  $k_0 = 0$ , o que corresponde à hipótese de que o processo apresenta média nula. Quanto às  $T-1$  equações seguintes, elas podem se

reescrever (com  $k_0 = 0$ ):

$$r_{h-1+j}^X - \sum_{i+1}^T k_i r_{|i-j|}^X = 0, \quad j = 1, \dots, T$$

ou ainda em notação matricial:

$$R_T^X K_T = r_T^X(h)$$

onde  $K_T = (k_1, \dots, k_T)'$ ,  $r_T^X(h) = (r_h^X, \dots, r_{h+T-1}^X)$  e  $R_T^X$  é a matriz  $T \times T$  de covariância cujo elemento  $(i, j)$  é:

$$(R_T^X)_{ij} = r_{|i-j|}^X \quad i, j = 1, \dots, T.$$

Conhecendo a função de autocovariância do processo, a predição  $\hat{X}_{T+h}$  é dada pela relação acima, com os coeficientes  $K_T$  tais que:

$$K_T = (R_T^X)^{-1} r_T^X(h)$$

onde este inverso da matriz  $R_T^X$  existe se  $r_0^X > 0$  e  $r_h^X \rightarrow 0$  para  $|h| \rightarrow \infty$ .

### Propriedades de $\hat{X}_{T+h}$

- O erro de predição é igual a:

$$E(X_{T+h} - \hat{X}_{T+h})^2 = r_0^X - K_T' r_T^X(h)$$

- $E(X_{T+h} - \hat{X}_{T+h}) = 0$
- $E[(X_{T+h} - \hat{X}_{T+h})X_j] = 0, \quad j = 1, \dots, T$

## 2.9.2 Algoritmo de Durbin-Levinson

A ideia do algoritmo de Durbin-Levinson é que, para prever o valor da série no período  $T + h$ , pode-se utilizar a previsão da série para  $T + h - 1$ . No intervalo de predição, este algoritmo propõe o esquema de predição seguinte: a partir das  $T$  observações, avalia-se  $\hat{X}_{T+1}$ . A partir das  $T$  observações às quais juntamos  $\hat{X}_{T+1}$ , avalia-se  $\hat{X}_{T+2}$  e assim sucessivamente até  $\hat{X}_{T+h}$  para o  $h$  desejado. Para a primeira etapa deste procedimento, encontra-se, notando  $\Phi_{T_i} = K_i$

$$\hat{X}_{T+1} = \Phi_{T1} X_T + \dots + \Phi_{TT} X_1.$$

O erro quadrático médio de predição é igual a:

$$\begin{aligned} v_T &= E(X_{T+1} - \hat{X}_{T+1})^2 \\ &= r_0^X - \Phi_T' r_T^X \end{aligned}$$

onde  $\Phi_T = (\Phi_{T1}, \dots, \Phi_{TT})'$ .

O algoritmo de Durbin-Levinson permite exprimir os coeficientes  $\Phi_T$  em função dos coeficientes  $\Phi_{T-i}$ . Assim, para prever  $\hat{X}_{T+1}$ , utiliza-se os coeficientes  $\Phi_T$  encontrados por este algoritmo e para encontrar  $\hat{X}_{T+2}$ , utiliza-se os coeficientes  $\Phi_{T+1}$  calculados a partir de  $\Phi_T$  e assim sucessivamente.

**Proposição.** Se  $X_t$  é um processo estacionário, de média zero e de função de autocovariância  $r_h^X$  tal que  $r_0^X > 0$  e  $r_h^X \rightarrow 0$  quando  $h \rightarrow \infty$ , então os coeficientes  $\Phi_{Tj}$  e os erros quadráticos médios de predição  $v_T$  satisfazem o algoritmo seguinte:

**Etapa 1:**  $\Phi_{11} = \frac{r_1^X}{r_0^X}, \quad v_0 = r_0^X$

**Etapa T:**

- $\Phi_{TT} = \left[ r_T^X - \sum_{j=1}^{T-1} \Phi_{T-1,j} r_{T-j}^X \right] v_{T-1}^{-1}$
- $\begin{pmatrix} \Phi_{T1} \\ \vdots \\ \Phi_{T,T-1} \end{pmatrix} = \begin{pmatrix} \Phi_{T-1,1} \\ \vdots \\ \Phi_{T-1,T-1} \end{pmatrix} - \Phi_{TT} \begin{pmatrix} \Phi_{T-1,T-1} \\ \vdots \\ \Phi_{T-1,1} \end{pmatrix}$
- $v_T = v_{T-1} (1 - \Phi_{TT}^2)$ .

Substituindo as autocovariâncias  $r_h^X$  pelas funções de autocorrelação empíricas  $\hat{\rho}_h^X$ , encontra-se a realização, o algoritmo de Durbin-Levinson empírico com os coeficientes "empíricos"  $\hat{\Phi}_T = (\hat{\Phi}_{T1}, \dots, \hat{\Phi}_{TT})$  e um estimador do erro de predição  $\hat{v}_T$ .



# Capítulo 3

## Modelos INAR

### 3.1 Introdução

Os modelos para dados de contagem tem vindo a conquistar, ao longo das últimas décadas, grande espaço na literatura e o interesse na modelação de processos estacionários discretos com distribuições marginais discretas tem crescido de forma significativa, tendo em conta a sua importância em aplicações em diversas áreas de conhecimento. Os dados de contagem do número de ocorrências de um evento, num determinado período de tempo, são inteiros não negativos, tais como:

- Número de transações num intervalo de tempo;
- Número mensal de ocorrência de um determinado acidente;
- Número mensal de crimes cometidos num determinado meio.

No desenvolvimento de tais modelos, deu-se uma atenção especial aos processos autorregressivos com valores inteiros não negativos, introduzidos por McKenzie (1985) e Al-Osh e Alzaid (1987).

O modelo INAR (INteger Valued AutoRegressive process) é um processo de contagem que considera variáveis inteiras não negativas e especialmente, as séries temporais constituídas por dados de contagens. O primeiro modelo INAR proposto por Al-Osh e Alzaid (1987) [3], no âmbito do estudo de tais séries, considera apenas a observação do período imediatamente anterior, à semelhança dos modelos autorregressivos de ordem 1. Tendo em conta que apenas a última observação pode não

apreender toda informação de uma série temporal e, com a necessidade de modelos mais robustos quanto à estrutura do histórico, Alzaid e Al-Osh (1990) [4] e Jin-Guan e Li-Yuan (1991) [10] investigaram modelos com maior complexidade, analogamente aos modelos AR(p), definindo o modelo genérico INAR(p) de ordem  $p$ .

A popularidade desta classe de modelos resulta, principalmente, da possibilidade de interpretações práticas variadas (são inspirados do conceito de população dinâmica, processo de branching, ...) e do fato da sua estrutura de correlação ser idêntica à de um processo AR estacionário (com coeficientes de regressão positivos). Todavia, apresenta algumas limitações: por definição, os coeficientes de regressão são positivos, o que exclui a modelagem de séries com autocorrelações negativas; por construção, não permite a análise de séries temporais discretas com valores negativos.

O modelo INAR se escreve como segue [16]:

$$X_t = C_{t-1} + \varepsilon_t$$

em que:

$C_{t-1} = \alpha \circ X_{t-1}$  é o componente que apreende toda informação contida no histórico, sendo uma função da soma de operadores *thinning binomial*, calculados com fontes em diferentes pontos no tempo até  $t-1$ , inclusive. Cada operador thinning binomial possui uma probabilidade associada, que é um parâmetro a ser estimado;

$\varepsilon_t$  é uma variável aleatória discreta não negativa com média  $\mu_\varepsilon$  e variância  $\sigma_\varepsilon^2$ , independente de  $C_{t-1}$ . Representa a inovação em  $t$ , uma parte não correlacionada com o passado. Mckenzie (1988) e Al-Osh e Alzaid(1987) [3] propuseram modelos com a inovação seguindo a distribuição de Poisson mas as distribuições geométrica ou binomial negativa também podem ser consideradas como sugerido por Mckenzie (1986). Este modelo INAR pode ser diferenciado pela distribuição marginal ou pela memória temporal considerada.

## 3.2 Operador Thinning Binomial

Tendo em conta que, quando os valores da série de contagem em análise são baixos, a multiplicação de uma variável aleatória por um constante real (parâmetro) pode

resultar numa variável aleatória não inteira, precisamos definir o operador aleatório *thinning binomial*, que substitui a multiplicação escalar, no contexto de valores inteiros.

Seja  $X$  uma variável aleatória de valor inteiro não negativo e  $p$  um número real tal que  $p \in [0, 1]$ . O operador thinning binomial, denotado por " $\circ$ ", é definido por Steutel e Van Harn (1979) [28] como

$$p \circ X = \sum_{j=1}^X z_j,$$

em que  $\{z_j\}_{j=1}^X$  é uma sequência de variáveis aleatórias de Bernouilli, independentes e identicamente distribuídas (*i.i.d.*), independentes de  $X$ , tal que  $P(z_j = 1) = 1 - P(z_j = 0) = p$ .

**Propriedades:** Sejam  $X$  e  $Y$  variáveis aleatórias de valores inteiros não negativos, e  $p$  e  $q$  números reais tais que  $p \in [0, 1]$  e  $q \in [0, 1]$ . O operador thinning binomial definido apresenta as seguintes propriedades:

- $0 \circ X = 0$
- $1 \circ X = X$
- $p \circ (q \circ X) = (pq) \circ X$
- $p \circ (X + Y) = p \circ X + p \circ Y$
- $E(p \circ X | X) = pX$
- $E(p \circ X) = pE(X)$
- $Var(p \circ X | X) = p(1 - p)X$
- $Var(p \circ X) = p^2Var(X) + p(1 - p)E(X)$
- $Cov(p \circ X, q \circ Y) = pqCov(X, Y)$ , se as variáveis  $z_i$  incluídas em  $p \circ X$  são independentes das variáveis  $z_j$  incluídas em  $q \circ Y$ .

### 3.3 Processos INAR(1)

Diz-se processo INAR(1) um processo estocástico discreto, de valor inteiro não negativo,  $\{X_t\}$ ,  $t \in \mathbb{Z}$  que satisfaz a seguinte equação,

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t$$

em que  $\alpha \in [0, 1]$ ,  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é uma sequência de variáveis *i.i.d.* com valores inteiros não negativos e,  $E(\varepsilon_t) = \mu$  e  $Var(\varepsilon_t) = \sigma^2$ . Se  $\alpha = 1$ , o processo é um passeio aleatório e se  $\alpha = 0$ , o processo será a sequência  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  e

$$\alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} Y_i$$

onde os  $Y_i$  são variáveis aleatórias de Bernoulli, com  $P(Y_i = 1) = 1 - P(Y_i = 0) = \alpha$  e independentes de  $X_t$ .

Segundo Alzaid e Al-Osh (1988), um processo de ramificação Galton-Watson com imigração pode ser representado pelo processo INAR(1) definido pela equação acima. Nota-se que a realização de  $\{X_t\}$ ,  $t \in \mathbb{Z}$  tem dois componentes aleatórios: o total de elementos sobreviventes no período  $t - 1$ ,  $\alpha \circ X_{t-1}$ , cada um com probabilidade de sobrevivência  $\alpha$ , representando o operador thinning binomial de ordem um (1) a dependência da variável no último instante, e  $\varepsilon_t$  é uma variável aleatória discreta *i.i.d.* com distribuição de Poisson ( $\lambda$ ), que representa a inovação ou os elementos que entram no processo no intervalo  $(t - 1, t)$ , chamados de elementos de entrada.

A série em análise é caracterizada, em relação ao tipo de estrutura de dependência que ela possui, pela autocorrelação e correlação parcial. Cada processo possui um comportamento idiossincrático da autocorrelação parcial, e essas características são identificadas através da comparação das autocorrelações teóricas e amostrais. A autocorrelação de ordem  $k$  de um processo INAR(1) é dada por  $\rho(k) = \alpha^k$ .

A distribuição marginal do processo INAR(1) se escreve:

$$X_t \stackrel{d}{=} \alpha^k \circ X_{t-k} + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t-j}$$

Usando a definição de um modelo INAR(1) e substituindo recursivamente em  $X_{t-i}$ , com  $i = 1, 2, 3, \dots$ , temos:

$$\begin{aligned} X_{t-1} &= \alpha \circ X_{t-2} + \varepsilon_{t-1} \\ X_{t-2} &= \alpha \circ X_{t-3} + \varepsilon_{t-2} \\ &\vdots \end{aligned}$$

Usando a propriedade 3 do operador thinning binomial, temos:

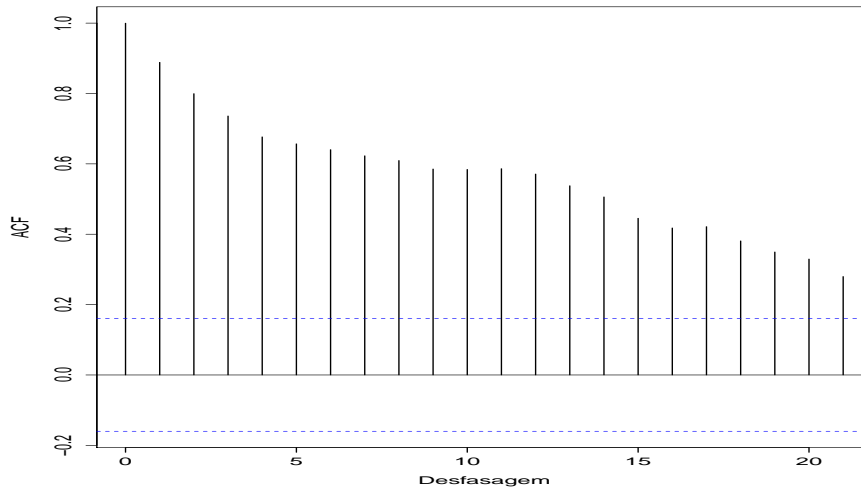
$$\begin{aligned} X_t &\stackrel{d}{=} \alpha \circ (\alpha \circ X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &\stackrel{d}{=} \alpha^2 \circ X_{t-2} + \alpha \circ \varepsilon_{t-1} + \varepsilon_t \\ &\stackrel{d}{=} \alpha^2 \circ (\alpha \circ X_{t-3} + \varepsilon_{t-2}) + \alpha \circ \varepsilon_{t-1} + \varepsilon_t \\ &\stackrel{d}{=} \alpha^3 \circ X_{t-3} + \alpha^2 \circ \varepsilon_{t-2} + \alpha \circ \varepsilon_{t-1} + \varepsilon_t \\ &\stackrel{d}{=} \alpha^3 \circ (\alpha \circ X_{t-4} + \varepsilon_{t-3}) + \alpha^2 \circ \varepsilon_{t-2} + \alpha \circ \varepsilon_{t-1} + \varepsilon_t \\ &\stackrel{d}{=} \alpha^4 \circ X_{t-4} + \alpha^3 \circ \varepsilon_{t-3} + \alpha^2 \circ \varepsilon_{t-2} + \alpha \circ \varepsilon_{t-1} + \varepsilon_t \\ &\stackrel{d}{=} \dots \\ &\stackrel{d}{=} \alpha^k \circ X_{t-k} + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t-j} \end{aligned}$$

### 3.3.1 Propriedades gerais do modelos INAR(1)

Para a modelação de séries temporais, estabelece-se geralmente hipóteses sobre a estrutura de correlação do processo estocástico em estudo. A principal delas, a da estacionariedade em sentido lato, estabelece que os momentos de primeira e segunda ordens do processo não variam sob translações do tempo. Seja  $X_t$ ,  $t \in \mathbb{Z}$ , um processo estocástico com esperança  $\mu_t = E(X_t)$  e autocovariância  $\gamma^t(k) = Cov(X_t, X_{t-k})$  para a defasagem  $k$ . O processo  $X_t$  é dito estacionário se  $\mu_t = \mu$  e  $\gamma^t(k) = \gamma(k)$  são constantes com respeito a  $t$ ,  $\forall k \in \mathbb{Z}$ . Du & Li (1991) [10] e Latour (1998) [14] mostraram que, se  $p < 1$ , o processo INAR(1) definido é estacionário.

Na figura seguinte é apresentada a função de autocorrelação amostral, de um processo INAR(1) simulado, com parâmetros  $\alpha = 0.9$  e  $\lambda = 1$ , que como nos modelos AR(1) decai exponencialmente com o desfasamento  $k$ . Todavia, ao contrário dos

modelos AR(1), assume somente valores positivos. Nos modelos de Box e Jenkins,



**Figura 3.1:** Função de autocorrelação amostral do processo  $INAR(1)$ , com  $\alpha = 0.9$  e  $\lambda = 1$

quando satisfeitas as condições de estacionaridade e invertibilidade, podemos representar o modelo AR(1) através de um modelo  $MA(\infty)$ . De modo análogo, a distribuição marginal de  $X_t$  pode ser escrita em função da distribuição das inovações  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  e é dada por Al-Osh e Alzaid (1987) [3].

$$X_t \stackrel{d}{=} \sum_{j=0}^{\infty} \alpha^j \circ \varepsilon_{t-j}, \quad \text{para } \alpha \in [0, 1]$$

Como observado acima, a dependência de  $X_t$  na sequência das inovações decai exponencialmente com a desfasagem de tempo. Supondo que  $E(\varepsilon_t) = \mu_\varepsilon < \infty$ ,  $Var(\varepsilon_t) = \sigma_\varepsilon^2$ , então o valor esperado de  $X_t$  é dado por:

$$\begin{aligned} E(X_t) &= \alpha \circ E(X_{t-1}) + \mu_\varepsilon \\ &= \alpha^t E(X_0) + \mu_\varepsilon \sum_{j=0}^{t-1} \alpha^j. \end{aligned}$$

Por sua vez, a variância é dada por:

$$\begin{aligned} Var(X_t) &= \alpha^2 Var(X_{t-1}) + \alpha(1 - \alpha)E(X_{t-1}) + \sigma^2 \\ &= \alpha^{2t} Var(X_0) + (1 - \alpha) \sum_{j=1}^t \alpha^{2(j-1)} E(X_{t-j}) + \sigma^2 \sum_{j=1}^t \alpha^{2(j-1)} \end{aligned}$$

A estacionaridade de segunda ordem é obtida a partir do valor inicial com:

- $E(X_0) = \frac{\mu}{1-\alpha}$
- $Var(X_0) = \frac{\alpha\mu + \sigma^2}{1-\alpha^2}$

Neste caso, para  $k$  inteiro não negativo, temos as funções de autocovariância e autocorrelação fornecidas respetivamente por:

$$\begin{aligned}
 \gamma(k) &= Cov(X_t, X_{t-k}) \\
 &= Cov(X_{t-k}, \alpha^k X_{t-k}) + Cov\left(X_{t-k}, \sum_{j=0}^{k-1} \alpha^j \varepsilon_{t-j}\right) \\
 &= \alpha^k Var(X_{t-k}) + \sum_{j=0}^{k-1} \alpha^j Cov(X_{t-k}, \varepsilon_{t-j}) \\
 &= \alpha^k \gamma(0)
 \end{aligned}$$

e

$$\begin{aligned}
 \rho(k) &= \frac{\gamma(k)}{\gamma(0)} \\
 &= \alpha^k; \quad \rho(k) \in [0, 1]
 \end{aligned}$$

Pode-se ver que, se  $\{\varepsilon_t\}_{t \in \mathbb{Z}} \sim P(\lambda)$ , a média e a variância do processo são dadas, respetivamente por

$$\begin{aligned}
 E(X_t) &= \frac{\lambda}{1-\alpha} \\
 Var(X_t) &= \frac{\lambda}{1-\alpha}
 \end{aligned}$$

Para a identificação e a estimação, no contexto de processos estacionários, destacam-se como principais ferramentas as funções de autocorrelação e de autocorrelação parcial.

### 3.3.2 Estimação dos parâmetros

Os estimadores discutidos nesta secção têm sido, em geral, propostos no contexto do processo INAR(1) com distribuição marginal de Poisson.

O problema relacionado à estimação de parâmetros no processo INAR(1) é mais complicado do que para o AR(1), devido ao fato de que a distribuição condicional de  $X_t$

dados  $X_{t-1}$  no processo INAR(1) é a convolução da distribuição de  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  e de uma binomial de parâmetros  $X_{t-1}$  e  $p$ .

### 3.3.2.1 Estimador de Yule-Walker

Seja uma amostra  $X_1, \dots, X_n$  de um processo estacionário  $\{X_t\}_{t \in \mathbb{Z}}$ . A sua função de autocorrelação amostral é dada por

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2},$$

em que  $\hat{\gamma}(k)$  representa a função de autocovariância amostral e  $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ , a média aritmética da série.

O estimador de Yule-Walker para o parâmetro  $\alpha$  é dado por

$$\hat{\alpha} = \hat{\alpha}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2},$$

uma vez que  $\rho(k) = \alpha^k$ .

Com efeito:

$$\varepsilon_t = X_t - \alpha \circ X_{t-1}$$

$$E(\varepsilon_t) = E(X_t) - \alpha E(X_{t-1})$$

Então, um estimador para  $\lambda$  é:

$$\begin{aligned} E(\hat{\varepsilon}_t) &= \sum_{t=1}^n \frac{X_t}{n} - \hat{\alpha} \sum_{t=1}^n \frac{X_{t-1}}{n} \\ &= \frac{1}{n} \left( \sum_{t=1}^n X_t - \hat{\alpha} \sum_{t=1}^n X_{t-1} \right) \\ &= \frac{1}{n} \sum_{t=1}^n (X_t - \hat{\alpha} X_{t-1}) \end{aligned}$$

pois  $\varepsilon_t$  possui uma distribuição de Poisson de parâmetro  $\lambda$ .

### 3.3.2.2 Estimador de Mínimos Quadrados Condicionais

Um método de mínimos quadrados condicionais para processos estocásticos foi desenvolvido por Klimko e Nelson (1978) [11], e aplicaram-no a um processo ramificado com imigração. Este método consiste em minimizar a soma dos quadrados das distâncias de cada observação,  $X_t$ , em relação ao valor esperado condicional dadas as observações anteriores. Seja  $X_t$  um processo estocástico cuja distribuição depende do parâmetro  $\theta = (\theta_1, \dots, \theta_p)$ , e seja  $F_n = \sigma(X_0, \dots, X_n)$ . Então, se  $\{X_n\}_{n \in \mathbb{Z}}$  é um processo INAR(1) com  $\{\varepsilon_n\} \sim P_\circ(\lambda)$ , a esperança condicional de  $X_n$  dado  $F_{n-1}$  é dada por

$$E[X_t | F_{t-1}] = \alpha X_{t-1} + \lambda = g(\theta, X_{t-1})$$

em que  $\theta = (\alpha, \lambda)^T$  representa o vetor de parâmetros a ser estimado. Dada uma amostra de tamanho  $n$  e, seja  $E(X_n | F_{n-1}) = Q_n(\theta)$ , o estimador de mínimos quadrados condicionais de  $\theta$ ,  $\hat{\theta}$ , é

$$Q_n(\hat{\theta}) = \min_{\theta} \{Q_n(\theta)\} = \min_{\theta} \left\{ \sum_{t=2}^n (X_t - \alpha X_{t-1} - \lambda) \right\}.$$

Derivando  $Q_n(\theta)$  em relação a  $\alpha$  e a  $\lambda$  e, igualando as duas equações a zero, obtemos:

$$\hat{\alpha} = \frac{\sum_{t=2}^n X_t X_{t-1} - \frac{1}{n-1} \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{\sum_{t=2}^n X_t^2 - \frac{1}{n-1} \left( \sum_{t=2}^n X_{t-1} \right)^2}$$

e

$$\hat{\lambda} = \frac{1}{n-1} \left( \sum_{t=2}^n X_t - \hat{\alpha} \sum_{t=2}^n X_{t-1} \right).$$

Pode-se verificar que as funções  $g$ ,  $\frac{\partial g}{\partial \alpha}$ ,  $\frac{\partial g}{\partial \lambda}$  e  $\frac{\partial^2 g}{\partial \alpha \partial \lambda}$  satisfazem todas as condições de regularidade do teorema de Klimko e Nelson (1978) [11], que são dadas por:

- $\lim_{n \rightarrow \infty} \sup_{\delta \rightarrow 0} \left( \frac{|T_n(\theta)_{ij}|}{n\delta} \right) < \infty, \quad i \leq 2, j \leq 2,$   
em que  $T_n(\theta) = \left( \frac{\partial^2 Q_n(\theta)}{\partial \theta^2} - v_n \right)$  e  $v_n = \left( \frac{\partial^2 Q_n(\theta)}{\partial \theta_i \partial \theta_j} \right),$
- $(2n)^{-1} v_n \rightarrow \frac{\partial^2 Q_n(\theta)}{\partial \theta^2}$
- $n^{-1} \frac{\partial Q_n(\theta)}{\partial \theta_i} \rightarrow 0, \quad i \leq 2$

e conseqüentemente, estes estimadores são consistentes. Se  $E(|\varepsilon_t|^\beta) < \infty$ , que é o caso de uma distribuição de Poisson, pelo teorema de Klimko e Nelson,  $(\hat{\alpha}, \hat{\lambda})^T$  possuem assintoticamente distribuição normal.

**Teorema.** Se  $\{\varepsilon_t\}_{t \in \mathbb{Z}} \sim P_o(\lambda)$  e o processo  $\{X_t\}_{t \in \mathbb{Z}}$  é estacionário, a distribuição assintótica do estimador de mínimos quadrados condicional de  $\theta$  é dada por

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V^{-1}WV^{-1}),$$

em que  $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})^T$  e  $V$  é uma matriz  $2 \times 2$  com elementos

$$V_{ij} = E \left( \frac{\partial g(\theta, x_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta, x_{t-1})}{\partial \theta_j} \right), \quad i, j = 1, 2,$$

e os elementos da matriz  $W$  são da forma

$$W_{ij} = E \left( \mu_t^2(\theta) \frac{\partial g(\theta, x_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta, x_{t-1})}{\partial \theta_j} \right), \quad i, j = 1, 2$$

com  $\mu_t(\theta) = x_t - g(\theta, x_{t-1})$  e a notação  $\xrightarrow{d}$  indica convergência em distribuição.

Na sua forma matricial escreve-se,

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\alpha(1-\alpha)^2}{\lambda} + (1-\alpha)^2 & -(1-\alpha)\lambda \\ -(1+\alpha)\lambda & \lambda + \frac{1+\alpha}{1-\alpha}\lambda^2 \end{pmatrix} \right).$$

Observa-se, para  $n$  grande, que o estimador de MQC de  $\theta = (\alpha, \lambda)^T$  e o estimador de Yule Walker para  $\theta = (\alpha, \lambda)^T$  são equivalentes.

### 3.3.2.3 Estimador de Máxima Verosimilhança Condicional

Duas alternativas para a obtenção das estimativas dos parâmetros do modelo INAR(1), através da maximização da função de verosimilhança, foram propostos por Al-Osh e Alzaid (1987) [3]: estimadores condicionais e não condicionais.

Para o processo INAR(1) com  $\{\varepsilon_t\}_{t \in \mathbb{Z}} \rightarrow P_o(\lambda)$ , a distribuição condicional de  $X_t$  dado  $X_{t-1}$ , denotada por  $f(X_t | X_{t-1})$ , é a convolução de uma distribuição binomial resultante do operador thinning binomial,  $\alpha \circ X_{t-1}$  e de uma distribuição de Poisson resultante do processo de chegada,  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ .

Assim, se  $*$  denota a convolução e

$$f_1(i) = \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{(X_{t-1})-i}, \quad f_2(i) = \frac{\lambda^i}{i!} e^{-\lambda},$$

então

$$f(X_t | X_{t-1}) = f_1 * f_2 = \sum_{i=0}^{\infty} f_1(i) f_2(x_{t-1}) = e^{-\lambda} \sum_{i=0}^{\min(X_t, X_{t-1})} \frac{\lambda^{(X_t)-i}}{((X_t)-i)!} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{(X_{t-1})-i}$$

Sendo  $X = (X_0, \dots, X_n)$  uma amostra de  $n + 1$  observações, a função de verosimilhança é dada por

$$L(\alpha, \lambda; X) = \frac{\left[\frac{\lambda}{1-\alpha}\right]^{X_0}}{X_0!} \exp\left(\frac{\lambda}{1-\alpha}\right) \prod_{t=1}^n f(X_t | X_{t-1}),$$

e a função de verosimilhança condicional do processo INAR(1), dado  $X_0$  é definida como

$$L(\alpha, \lambda; X | X_0) = \prod_{t=1}^n f(X_t | X_{t-1})$$

A função log-verosimilhança condicional é da forma

$$l(\alpha, \lambda; X | X_0) = \log\left(\prod_{t=1}^n f(X_t | X_{t-1})\right) = \sum_{t=1}^n \log(f(X_t | X_{t-1})).$$

A função escore condicional é dada por:

$$\frac{\partial l(\alpha, \lambda; X | X_0)}{\partial \alpha} = \sum_{t=1}^n \frac{X_{t-1} [f(X_{t-1} | X_{t-1} - 1) - f(X_t | X_{t-1})]}{1 - \alpha} \frac{1}{f(X_t | X_{t-1})}$$

$$\frac{\partial l(\alpha, \lambda; X | X_0)}{\partial \lambda} = \sum_{t=1}^n \frac{[f(X_t - 1 | X_{t-1}) - f(X_t | X_{t-1})]}{f(X_t | X_{t-1})}$$

O estimador de máxima verosimilhança condicional de  $\theta = (\alpha, \lambda)^T$  é o valor  $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})^T$  que maximiza a função de verosimilhança  $L(\alpha, \lambda; X | X_0)$ . O valor de  $\theta$  que maximiza a função de verosimilhança condicional  $L(\alpha, \lambda; X | X_0)$  também maximiza  $l(\alpha, \lambda; X | X_0)$ . Assim os estimadores de máxima verosimilhança condicional de  $\alpha$  e  $\lambda$  são obtidos através da maximização do logaritmo da função de verosimilhança con-

dicional recorrendo em geral a métodos numéricos, uma vez que as equações de log-verosimilhança conduzem-nos a um sistema de equações não lineares.

A distribuição assintótica do estimador de Máxima Verosimilhança Condicional para os parâmetros do modelo INAR(1) é estabelecida no seguinte teorema.

**Teorema:** A distribuição assintótica dos estimadores de máxima verosimilhança condicional de  $\alpha$  e  $\lambda$  é dada por

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} N(0, K^{-1})$$

em que  $K$  é a matriz de informação de Fisher, isto é,  $-K$  é a esperança das segundas derivadas do logaritmo da função de verosimilhança condicional.

## 3.4 Processos INAR(p)

### 3.4.1 Introdução

As realizações  $\{X_t\}$  de alguns processos de contagem não dependem apenas do seu valor anterior imediato  $X_{t-1}$  mas também de realizações prévias do processo  $\{X_{t-i}\}_{i=2}^p$ , para algum constante  $p$ . Tendo em consideração estas realizações prévias, define-se assim, a classe dos modelos INAR(p) que considera os  $p$  últimos valores como fatores que permitem determinar o valor da série no período  $t$ . É a extensão do processo INAR(1) a um processo autorregressivo com valores inteiros de ordem  $p$ , INAR(p).

De uma forma direta pode-se estender o processo INAR(1) considerando a forma do processo usual AR(p), com substituição da multiplicação escalar,  $\alpha_i X_{t-i}$ , pela operação ' $\alpha_i \circ X_{t-i}$ ' para  $i = 1, 2, \dots, p$ . Todavia, com cada substituição, são necessárias algumas hipóteses suplementares sobre o modelo, permitindo assim guardar a estrutura de dependência do processo e conseqüentemente, a boa definição do modelo. O processo  $\{X_t\}$  é considerado um modelo INAR(p) se admite a representação seguinte:

$$X_t = \sum_{i=1}^p \alpha_i \circ X_{t-i} + \varepsilon_t \quad \text{para } t = 0, \pm 1, \dots, \quad (3.1)$$

onde  $\{\varepsilon_t\}$ , designado processo de inovação, é uma sequência de variáveis aleatórias *i.i.d.* de valores inteiros não negativos, com distribuição de Poisson ( $\lambda$ ) e com média  $\mu_\varepsilon$  e de variância finita  $\sigma_\varepsilon^2$ ;

$\sum_{i=1}^p \alpha_i \circ X_{t-i}$  é uma combinação linear de operadores thinning binomial em distintas ordens de defasagens. Cada ordem de defasagem  $k$  possui uma probabilidade  $\alpha_k$  associada,  $k = 1, \dots, p$ ,

$$\alpha_i \circ X_{t-i} = \sum_{j=1}^{t-i} Y_{i,j}, \quad i = 1, \dots, p,$$

onde  $\{Y_{i,j}\}$ , a série de contagem, é um conjunto de variáveis aleatórias *i.i.d.* com valores inteiros não negativos tal que  $E[Y_{i,j}] = \alpha_i$ ,  $Var[Y_{i,j}] = \alpha_i^2$ ,  $E[Y_{i,j}^3] = \gamma_i$  e  $E[Y_{i,j}^4] = k_i$ . Todas as séries de contagem são consideradas independentes de  $\{\varepsilon_t\}$  e  $0 \leq \alpha_i < 1$ ,  $i = 1, \dots, p-1$ , e  $0 < \alpha_p < 1$ .

Alternativamente, um processo INAR(p) pode ser representado como um processo AR(1)  $p$ -dimensional. Assim, usando o vetor do operador thinning, definido como um vetor aleatório cujo  $i$ -ésimo elemento é dado por

$$[A \circ X]_i = \sum_{j=1}^p a_{ij} \circ X_j, \quad i = 1, \dots, p,$$

onde  $X = [X_1, \dots, X_p]^T$ , é um vetor aleatório,  $A$  é uma matriz  $p \times p$  com entradas  $a_{ij}$  satisfazendo  $0 \leq a_{ij} \leq 1$  para  $i, j = 1, \dots, p$ , e as séries de contagem de todo  $a_{ij} \circ X_j$ ,  $i, j = 1, \dots, p$ , são independentes. O processo INAR(p) definido acima pode ser escrito como

$$X_t = A \circ X_{t-1} + W_t,$$

$$X_t = HX_t \tag{3.2}$$

onde  $H = [1 \ 0 \ \dots \ 0]$ ,  $X_t = [X_t \ X_{t-1} \ \dots \ X_{t-p+1}]^T$ ,  $W_t = [\varepsilon_t \ 0 \ \dots \ 0]^T$ , para  $\{\varepsilon_t\}$  uma sequência de variáveis aleatórias *i.i.d.* com  $E[\varepsilon_t] = \mu_\varepsilon$ ,  $Var[\varepsilon_t] = \sigma_\varepsilon^2$ ,  $E[\varepsilon_t^3] =$

$\gamma_\varepsilon$  e  $E[\varepsilon_t^4] = k_\varepsilon < \infty$  e

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Além disso, o modelo 3.1 pode se apresentar como

$$X_t \stackrel{d}{=} A^k \circ X_{t-k} + \sum_{j=0}^{k-1} A^j \circ W_{t-j},$$

onde  $\stackrel{d}{=}$  significa igualdade na distribuição. Então as equações 3.2 podem se escrever como:

$$X_t \stackrel{d}{=} \sum_{j=0}^{\infty} A^j \circ W_{t-j},$$

$$X_t \stackrel{d}{=} HX_t,$$

onde  $H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ , desde que o raio espectral da matriz é menos do que um, isto é,  $\rho(A) < 1$ .

Na literatura foram propostos muitos métodos de estimação do modelo INAR(p), nomeadamente, os estimadores de Yule-Walker, Mínimos Quadrados Condicionais, o método Generalizado dos Momentos, o de Máxima Verossimilhança Condicional e, no domínio de frequência, o Critério de Whittle. Métodos baseados em estatísticas de ordem superior (momentos e cumulantes) também foram considerados (I. Silva, 2005) e as propriedades amostrais destes estimadores foram estudadas empiricamente pelos métodos de Monte Carlo (I. Silva, 2005), enquanto pouco se sabe sobre as propriedades assintóticas pois que estas dependem dos momentos e cumulantes de ordem superior que são difíceis de obter.

A distribuição assintótica de autocovariância da amostra dos processos INAR(p) é obtida e é demonstrado que o estimador de Yule-Walker é assintoticamente normalmente distribuído, não enviesado e consistente. Este resultado generaliza o trabalho de Park e Oh (1997), que derivaram a distribuição assintótica do estimador de Yule-

Walker para uma parametrização alternativa do processo INAR(1) de Poisson com operação thinning binomial. [23]

O processo INAR(p) pode ser visto como uma extensão (ou tipo especial) de um processo de "ramificação" com a imigração. Para isto, considera uma população humana ou biológica na qual a fêmea pode reproduzir no máximo uma prole feminina durante o seu período reprodutivo, que é dividido em  $p$  períodos não sobrepostos.

A probabilidade que cada fêmea tenha sua prole durante o seu  $i$ -ésimo período de reprodução é  $\alpha_i$ . Agora isto é claro que o tamanho da  $n$ -ésima geração  $\{X_n\}$  é apenas o total das proles das últimas  $p$  gerações somando o processo de imigrantes  $\{\varepsilon_n\}$  que entraram no sistema durante o período de intervalo  $(n-1, n)$ . Isto pode ser observado que para  $\sum_{i=1}^p \alpha_i \geq 1$ , o tamanho da geração  $\{X_n\}$  irá crescer monotonicamente e temos um processo explosivo. A interpretação prévia é merecidamente uma extensão do processo de ramificação INAR(1) no qual apenas a última geração pode reproduzir.

Pode-se notar que a forma do processo INAR(p) aparenta ser um processo autor-regressivo com valores de ordem  $p$ , contudo a dependência ao longo do tempo da operação 'o' faz este processo ser diferente do processo AR(p). Para explicar este ponto, consideremos o caso  $p = 2$ , por simplicidade, isto é:

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \varepsilon_t$$

No processo AR(2),  $X_t$  é obtido por uma multiplicação direta dos constantes  $\alpha_1$  e  $\alpha_2$  a  $X_{t-1}$  e  $X_{t-2}$ , respetivamente, no período  $t$  e independente de toda estrutura estocástica prévia. Isso não é o caso para o modelo INAR(2). Para este processo, as variáveis aleatórias  $\alpha_1 \circ X_{t-2}$  e  $\alpha_2 \circ X_{t-2}$  que são elementos de  $X_{t-1}$  e  $X_t$  respetivamente, são intimamente ligados de uma forma muito mais poderosa do que no caso do processo AR(2) onde apenas a presença de  $X_{t-2}$  pode os ligar. Aqui as operações " $\alpha_1 \circ$ " e " $\alpha_2 \circ$ " sobre  $X_{t-2}$  são dependentes mas parecem ser realizadas em períodos diferentes. A estrutura é talvez clara quando se estima o processo. No período  $t$ ,  $X_t$  é observado e temos disponíveis  $X_{t-1}$  e  $V_{t-1} = \alpha_2 \circ X_{t-1}$ , tendo usado  $\alpha_1 \circ X_{t-1}$  para calcular  $X_t$ . O primeiro passo é formar  $U_t = \alpha_1 \circ X_t$  e  $V_t = \alpha_2 \circ X_t$  e então  $X_{t+1} = U_t + V_{t-1} + \varepsilon_{t+1}$ , e  $V_t$  fica disponível para ser utilizado no cálculo de  $X_{t+2}$ .

A estrutura de dependência mútua entre os componentes de  $X_t$  i.e.  $\alpha_i \circ X_t$ ,  $i = 1, 2, \dots, p$  aparecendo em diferentes períodos induzem uma estrutura de média móvel no processo. Portanto, o comportamento da função de autocorrelação do processo INAR(p) imita o processo ARMA(p,p-1).

### 3.4.2 Métodos de estimação

A estimação dos parâmetros do modelo INAR possui maior complexidade comparada à estimação dos parâmetros do modelo autorregressivo AR. Isto se dá porque a distribuição de  $X_t$ , condicionada à informação observada até ao instante anterior  $t - 1$  ( $X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots$ ) é a convolução da distribuição da inovação  $\varepsilon_t$  e das somas das binomiais como parâmetros  $(x_{t-i}, \alpha_i)$ ,  $i = 1, 2, \dots$ .

Seguindo o que foi desenvolvido por Al-Osh e Alzaid (1987), foram propostos por Al-Osh e Alzaid (1990) e Jin-Guan e Li-Yuan (1991), os estimadores para o modelo INAR(p); esta secção apresenta os métodos de estimação dos processos INAR com distribuição marginal dada por Poisson. São apresentados quatro estimadores.

#### 3.4.2.1 Estimador de Yule-Walker

Os estimadores de Yule-Walker de  $\alpha_i$ ,  $i = 1, 2, \dots, p$  e  $\lambda$  de um modelo INAR(p) são obtidos de modo equivalente aos estimadores do modelo AR(p). É possível obter uma expressão analítica para cada  $\alpha_i$  em função das autocorrelações. Tomando  $p$  e usando a autocorrelação amostral chega-se a um estimador para  $\alpha$ .

$$\begin{aligned} \hat{\alpha} &= \frac{\hat{\rho}(1)}{\hat{\rho}(0)} \\ &= \frac{\sum_{t=1}^{n-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=0}^n (X_t - \bar{X})^2} \end{aligned}$$

em que  $\bar{X}$  é a média amostral da série  $X_t$ . É razoável pensar que o estimador de  $\lambda$ , média de distribuição de Poisson, é dado como a média da sequência  $\hat{I}_t$ , onde

$$\hat{I}_t = X_t - \hat{\alpha}X_{t-1}, \text{ para } t = 2, 3, \dots, N,$$

$$\hat{\lambda} = \sum_{t=2}^N \frac{\hat{I}_t}{N-1}.$$

Generalizando o modelo INAR e considerando o referido modelo, pode-se estimar os parâmetros  $\alpha_i$  ( $i = 1, \dots, p$ ) através da estrutura de autocorrelação, da mesma forma que para o modelo INAR(1), utilizando as autocorrelações amostrais. A estrutura recursiva das autocorrelações pode ser expressa em forma de matriz, tendo a expressão

$$\rho_k = \sum_{i=1}^{\min(p,k)} \alpha_i \rho_{k-i}$$

apresentada como  $\Gamma\alpha = \rho$  em que

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho_1 & 1 & 0 & \cdots & \vdots \\ \rho_2 & \rho_1 & 1 & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \quad e \quad \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}.$$

Novamente, substituindo  $\rho_i$  por  $\hat{\rho}_i$  temos os estimadores de Yule-Walker dados por:

$$\hat{\Gamma}\hat{\alpha} = \hat{\rho}$$

Os estimadores de  $\lambda$  e de  $\sigma^2$  são respetivamente,

$$\hat{\lambda} = \sum_{t=p+1}^N \frac{\hat{I}_t}{N-p}$$

e

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{t=p+1}^N (\hat{I}_t - \bar{I})^2,$$

considerando agora  $\hat{I}_t = X_t - \sum_{j=1}^p \hat{\alpha}_j X_{t-j}$  e  $\bar{I} = \frac{1}{N-p} \sum_{t=p+1}^N \hat{I}_t$ .

Do mesmo modo, o estimador de Yule-Walker  $\hat{\alpha}$  de  $\alpha$  para a realização  $X_1, \dots, X_N$  de um processo INAR(p), pode ser obtido pela resolução do sistema de equações

lineares seguinte

$$\hat{R}_{p-1}\hat{\alpha} = \hat{r}_p \Leftrightarrow \begin{pmatrix} \hat{R}(0) & \hat{R}(1) & \hat{R}(2) & \cdots & \hat{R}(p-1) \\ \hat{R}(1) & \hat{R}(0) & \hat{R}(1) & \cdots & \hat{R}(p-2) \\ \hat{R}(2) & \hat{R}(1) & \hat{R}(0) & \cdots & \hat{R}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{R}(p-1) & \hat{R}(p-2) & \hat{R}(p-3) & \cdots & \hat{R}(0) \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \vdots \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} = \begin{pmatrix} \hat{r}(1) \\ \hat{r}(2) \\ \hat{r}(3) \\ \vdots \\ \hat{r}(p) \end{pmatrix}$$

com  $\hat{R}(k) = \frac{1}{n} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})$ ,  $k \in \mathbb{Z}$ , e  $\bar{X} = \frac{1}{n} \sum_{t=1}^N X_t$ .

Os estimadores para  $\mu_\varepsilon$  e  $\sigma_\varepsilon^2$  são respetivamente,

$$\hat{\mu}_\varepsilon = \bar{X} \left( 1 - \sum_{i=1}^p \hat{\alpha}_i \right), \quad \hat{\sigma}_\varepsilon^2 = \hat{V}_p - \bar{X} \sum_{i=1}^p \hat{\sigma}_i^2,$$

onde

$$\hat{V}_p = \hat{R}_p(0) - \sum_{i=1}^p \hat{\alpha}_i \hat{R}_p(i)$$

e  $\hat{\sigma}_i^2$  é um estimador da variância das séries de contagens para o  $i$ -ésimo operador thinning,  $\alpha_i \circ X_{t-i}$ ,  $i = 1, \dots, p$ . A estimação de  $\hat{\sigma}_i^2$  depende da distribuição das séries de contagens. Por exemplo, no caso da operação thinning binomial (quando as séries de contagem tem distribuição de Bernoulli),  $\hat{\sigma}_i^2 = \hat{\alpha}_i (1 - \hat{\alpha}_i)$ , para  $i = 1, \dots, p$ .

Para obter a distribuição assintótica do estimador de Yule-Walker, as propriedades assintóticas da matriz de covariância é necessária.

### 3.4.2.2 Distribuição assintótica da função de autocovariância

No teorema seguinte, a normalidade assintótica multivariada da média amostral,  $\bar{X}$  e da função de autocovariância amostral,  $\hat{R}(k)$  é estabelecida.

Consideremos a função auxiliar

$$R^*(k) = \frac{1}{N-k} \sum_{t=1}^{N-k} (X_t - \mu_X)(X_{t+k} - \mu_X), \quad k \in \mathbb{N}_0$$

obtendo suas propriedades assintóticas e provando que  $\hat{R}(\cdot)$  tem as mesmas propriedades assintóticas que a função auxiliar  $R^*(\cdot)$ .

**Teorema:** Se  $\{X_t\}$  é um processo INAR com representação como a definida acima, onde  $\sum_{j=0}^{\infty} |A^j| < \infty$ , e se  $R^*(\cdot)$  é a função de autocovariância de  $\{X_t\}$ , então para qualquer número inteiro não negativo  $h$ ,

$$\begin{pmatrix} \bar{X} \\ \hat{R}(0) \\ \vdots \\ \hat{R}(h) \end{pmatrix}$$

é um

$$\left( \begin{pmatrix} \mu_X \\ R(0) \\ \vdots \\ R(h) \end{pmatrix}, N^{-1}V_R \right)$$

onde

$$V_R = \begin{pmatrix} [V_{11}]_{1 \times 1} & [V_{12}]_{1 \times (h+1)} \\ [V_{21}]_{1 \times (1+h)}^T & [V_{22}]_{(h+1) \times (h+1)} \end{pmatrix}$$

é uma matriz  $(h+1) \times (h+2)$  para qualquer número inteiro não negativo  $h$ , tal que

$$\begin{aligned} V_{11} &= \lim_{N \rightarrow \infty} N \text{Var}(\bar{X}) \\ &= \sum_{h=-\infty}^{\infty} R(h), \end{aligned}$$

$$\begin{aligned} [V_{12}]_{k+1} &= \lim_{N \rightarrow \infty} N \text{Cov}(\bar{X}, R^*(k)) \\ &= \sum_{h=-\infty}^{\infty} C_X(h, k+h), \quad k = 0, 1, \dots, h \end{aligned}$$

$$\begin{aligned} [V_{22}]_{k+1, j+1} &= \lim_{N \rightarrow \infty} N \text{Cov}(R^*(k), R^*(j)) \\ &= \sum_{h=-\infty}^{\infty} R(h)R(h+j-k) + R(h+j)R(h-k) \\ &\quad + C_Y(h, k, j+h), \quad k, j = 0, 1, \dots, h \end{aligned}$$

onde  $C_X(\cdot, \cdot)$  é o cumulante de terceira ordem de  $X_t$  e  $C_Y(\cdot, \cdot, \cdot)$  é o cumulante de quarta ordem do processo  $Y_t = X_t - \mu_X$ .

### 3.4.2.3 Distribuição assintótica do estimador de Yule-Walker

Nesta secção, as propriedades assintóticas da matriz de covariância amostral são usadas para obter a distribuição assintótica do estimador de Yule-Walker. Daí, considerar a matriz de autocovariância amostral  $p \times p$  de Toeplitz,  $\hat{R}_{p-1}$ , e o vetor da função de autocovariância amostral com  $p$  elementos,  $\hat{r}_p$ . Seja

$$\hat{V}_{Rr} = \begin{bmatrix} \text{vec}(\hat{R}_{p-1}) \\ \hat{r}_p \end{bmatrix}$$

um vetor com  $p(p+1)$  elementos, onde o operador  $\text{vec}$  é o operador coluna de empilhamento de uma matriz, isto é,

$$\text{vec}(A) = \begin{bmatrix} A_{.1} \\ A_{.2} \\ \vdots \\ A_{.q} \end{bmatrix}$$

é um vetor  $pq \times 1$  onde  $A_{.i}$  é a  $i$ -ésima coluna da matriz  $A$  de tamanho  $p \times q$ . Nota-se que para  $p = 1$ ,

$$\hat{V}_{Rr} = \left( \hat{R}(0) \quad \hat{R}(1) \right)^T$$

e para  $p > 1$ , o  $i$ -ésimo elemento de  $\hat{V}_{Rr}$  pode se escrever como

$$\hat{V}_{Rr}(i) = \begin{cases} \hat{R}(|(i-1) \bmod p - [(i-1)/p]|), & \text{se } i \leq p^2, \\ \hat{R}(|i \bmod p|), & \text{se } i > p^2, \end{cases}$$

onde  $[a]$  representa a parte inteira de  $a \in \mathbb{R}$ .

Então usando o resultado precedente, encontra-se que

$$\hat{V}_{Rr} = \begin{bmatrix} \text{vec}(\hat{R}_{p-1}) \\ \hat{r}_p \end{bmatrix} \acute{e} AN \left( \begin{bmatrix} \text{vec}(R_{p-1}) \\ r_p \end{bmatrix}, \frac{1}{N} \sum R_r \right),$$

onde  $\sum R_r$  é a matriz de covariância  $p(p+1) \times p(p+1)$  de  $\hat{V}_{R_r}$ ,

$$\sum R_r(i, j) = \text{Cov}(\hat{V}_{R_r}(i), \hat{V}_{R_r}(j)), \quad (3.3)$$

para  $i, j = 1, 2, \dots, p(p+1)$ .

Seja,

$$\begin{aligned} D^T &= - \left[ R(1)I_p \quad \dots \quad R(p)I_p \right] \left( (R_{p-1}^{-1})^T \otimes R_{p-1}^{-1} \right) \begin{bmatrix} I_{p^2} & 0_{p^2 \times p} \\ 0_{p \times p^2} & R_{p-1}^{-1} \end{bmatrix} \\ &+ \begin{bmatrix} 0_{p \times p^2} & R_{p-1}^{-1} \end{bmatrix} \end{aligned} \quad (3.4)$$

onde  $I_n$  é uma matriz identidade  $n \times n$ ,  $0_{n \times m}$  uma matriz de zeros e  $\otimes$  o produto de Kronecker definido como segue. Sejam  $A$  e  $B$  matrizes  $p \times q$  e  $m \times n$  respectivamente. Então  $A \otimes B$  é uma matriz  $pm \times qn$  dada por

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1q}B \\ a_{21}B & a_{22}B & \dots & a_{2q}B \\ \vdots & \vdots & \ddots & \dots \\ a_{p1}B & a_{p2}B & \dots & a_{pq}B \end{bmatrix}$$

A distribuição assintótica do estimador de Yule-Walker é um processo INAR(p) dado no teorema seguinte:

**Teorema:** Seja  $\{X_t\}$  um processo INAR(p), satisfazendo as suas condições exigidas, e  $\hat{\alpha}$  o estimador Yule-Walker de  $\alpha$ . Então

$$N^{1/2}(\hat{\alpha} - \alpha) \text{ é um } (0_p, D^T \sum_{R_r} D)$$

onde  $0_n$  é um vetor de  $n$  zeros,  $\sum_{R_r}$  é dado em 3.3 e  $D^T$  é definido por 3.4.

#### 3.4.2.4 Mínimos Quadrados Condicionais (MQC)

O estimador de mínimos quadrados condicionais busca estimadores de  $\alpha_j$  através da minimização da soma do quadrado dos erros. Seja o modelo 3.1 e  $F_t$  representando

toda a informação até ao momento  $t$ , temos:

$$\begin{aligned} F_t &= F(X_1, X_2, \dots, X_t) \\ \theta &= (\lambda, \alpha_1, \dots, \alpha_p) \\ g(\theta, F_{t-1}) &= E(X_t | F_{t-1}) \\ &= \lambda + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p}. \end{aligned}$$

A soma do quadrado dos erros, denotada por  $Q_n(\theta)$ , pode se escrever como:

$$Q_n(\theta) = \sum_{t=p+1}^n \{X_t - g(\theta, F_t)\}^2.$$

O estimador de mínimos quadrados condicionais,  $\hat{\theta}$ , obtido minimizando a quantidade dada (acima), é,

$$Q_n(\hat{\theta}) = \min_{\theta} \{Q_n(\theta)\}.$$

Através das derivadas parciais em relação a cada componente de  $\theta$  igualada a zero é obtido o estimador. A sua forma explícita é dada por:

$$\hat{\lambda} = \bar{X}_n^0 - \sum_{j=1}^p \hat{\alpha}_j \bar{X}_n^{(j)}$$

e

$$\hat{\Gamma} \hat{\alpha} = \hat{\rho}$$

em que

$$\begin{aligned} \bar{X}_n^{(j)} &= \frac{1}{n-p} \sum_{t=p+1}^n X_{t-j} \\ \hat{\gamma}_{k-j} &= \frac{1}{n-p} \sum_{t=p+1}^n (X_{t-j} - \bar{X}_n^{(j)}) (X_{t-1} - \bar{X}_n^{(k)}) \\ \hat{\rho}_{k-j} &= \frac{\hat{\gamma}_{k-j}}{\hat{\gamma}_0}, \end{aligned}$$

e

$$\hat{\Gamma} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \hat{\rho}_1 & 1 & \ddots & \ddots & \vdots \\ \hat{\rho}_2 & \hat{\rho}_1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \cdots & 1 \end{bmatrix}.$$

Pode-se notar que a medida que o tamanho da série  $n$  aumenta, os estimadores de mínimos quadrados condicionais e de Yule-Walker tornam-se equivalentes.

### 3.4.2.5 Estimadores de Máxima verosimilhança Condicional

Nesta secção se apresentam os resultados de Al-Osh e Alzaid (1987) para o processo INAR(1) com distribuição marginal de Poisson e logo se obtém os estimadores dos parâmetros para o processo INAR(2) definido por Du e Li (1991).

Supõe-se que  $X_0$  é observada, neste caso a função de verosimilhança condicionada a que  $X_0$  é da forma:

$$L(X, \alpha, \lambda | X_0) = \prod_{t=1}^n P_{X_t/X_{t-1}}(X_t) \quad (3.5)$$

onde  $X = (X_0, X_1, \dots, X_n)$  e

$$P_{X_t/X_{t-1}}(X_t) = \sum_{i=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i} \alpha^i (1-\alpha)^{X_{t-1}-i} \frac{e^{-\lambda} \lambda^{X_t-i}}{(X_t-i)!}.$$

Tomando os logaritmos de ambos os lados da equação 3.5 e derivando em relação a  $\alpha$  e  $\lambda$ , tem-se:

$$S_\alpha = \frac{1}{\alpha(1-\alpha)} \sum_{t=1}^n \left( X_t - \alpha X_{t-1} - \lambda \frac{P_{X_t/X_{t-1}}(X_t - 1)}{P_{X_t/X_{t-1}}(X_t)} \right)$$

e

$$S_\lambda = \sum_{t=1}^n \left( \frac{P_{X_t/X_{t-1}}(X_t - 1)}{P_{X_t/X_{t-1}}(X_t)} \right) - n.$$

Realizando simples operações algébricas no sistema de equações de verosimilhança condicional  $S_\alpha = 0$  e  $S_\lambda = 0$ , temos que:

$$\sum_{t=1}^n X_t = \alpha \sum_{t=1}^n X_{t-1} + n\lambda. \quad (3.6)$$

A equação 3.6 pode ser utilizada para eliminar um dos parâmetros, digamos  $\alpha$ , então  $S_\alpha$  pode se escrever como uma função unicamente de  $\lambda$  e o estimador de máxima verosimilhança de  $\lambda$  se obtém utilizando os métodos numéricos para resolver  $S_\lambda$ .

A seguir se aplica o método de máxima verosimilhança condicional ao processo INAR(2) definido por Du e Li (1991). A probabilidade condicional de  $X_t$  dados  $X_{t-1}, X_{t-2}$  apresenta a seguinte estrutura:

$$P_{X_t/X_{t-2}, X_{t-1}}(X_t) = \sum_{i=0}^{\min(X_{t-1}, X_t)} \sum_{j=0}^{\min(X_{t-2}, X_{t-1})} P(Z_1 = i)P(Z_2 = j)P(\varepsilon_t = X_t - i - j)$$

onde as variáveis  $Z_1$  e  $Z_2$  tem distribuição binomial com parâmetros  $(X_{t-1}, \alpha_1)$  e  $(X_{t-2}, \alpha_2)$  respectivamente e  $\varepsilon_t$  tem distribuição de Poisson com parâmetro  $\lambda$ .

A função de verosimilhança condicional de  $(\alpha_1, \alpha_2, \lambda)$  dado  $X_0$  e  $X_1$ , pode escrever-se como:

$$L(\alpha_1, \alpha_2, \lambda; X/X_0, X_1) = \prod_{t=2}^n P_{X_t/X_{t-2}, X_{t-1}}(X_t) \quad (3.7)$$

Tomando os logaritmos em ambos os lados da equação 3.7 e diferenciando em relação a  $\alpha_1, \alpha_2$  e  $\lambda$ , obtém-se:

$$S_{\alpha_1} = \frac{1}{\alpha_1(1 - \alpha_1)} \sum_{t=2}^n \left( \frac{P_{X_t/X_{t-2}, X_{t-1}}(X_t)i}{P_{X_t/X_{t-2}, X_{t-1}}(X_t)} - \alpha_1 X_{t-1} \right)$$

$$S_{\alpha_2} = \frac{1}{\alpha_2(1 - \alpha_2)} \sum_{t=2}^n \left( \frac{P_{X_t/X_{t-2}, X_{t-1}}(X_t)j}{P_{X_t/X_{t-2}, X_{t-1}}(X_t)} - \alpha_2 X_{t-2} \right)$$

$$S_\lambda = \frac{1}{\lambda} \sum_{t=2}^n \left( \frac{P_{X_t/X_{t-2}, X_{t-1}}(X_t)(X_t - i - j)}{P_{X_t/X_{t-2}, X_{t-1}}(X_t)} - \lambda \right)$$

Realizando simples operações algébricas em  $S_{\alpha_1}$  e  $S_{\alpha_2}$ , obtém-se:

$$\sum_{t=2}^n \left\{ \frac{P_{X_t/X_{t-2}, X_{t-1}}(X_t)(i + j)}{P_{X_t/X_{t-2}, X_{t-1}}(X_t)} \right\} = \alpha_1 \sum_{t=2}^n X_{t-1} + \alpha_2 \sum_{t=2}^n X_{t-2}$$

Agora, utilizando a última equação e  $S_\lambda = 0$ , se obtém  $\lambda$  em termos de  $\alpha_1$  e  $\alpha_2$ , isto é:

$$\lambda = \frac{\sum_{t=2}^n (X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2})}{n - 1}.$$

Portanto, as equações  $S_{\alpha_1}$  e  $S_{\alpha_2}$  podem escrever-se como funções unicamente de  $\alpha_1$  e  $\alpha_2$ . Os estimadores de máxima verosimilhança condicional se obtém utilizando métodos numéricos para resolver  $S_{\alpha_1}$  e  $S_{\alpha_2}$ .

A estimação da matriz de covariância assintótica para os estimadores dos parâmetros pode ser obtida calculando a segunda derivada da função de verosimilhança. Os estimadores de máxima verosimilhança condicional do processo INAR(2) tem distribuição assintótica normal. Por outro lado, o método de máxima verosimilhança condicional se pode generalizar para obter estimadores de parâmetros do processo INAR(p) definido por Du e Li (1991). Para este processo, tem que ter em conta que a probabilidade condicional de  $X_t$  dado  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ , no processo INAR(p) é a convolução da distribuição de  $\varepsilon_t$  e as binomiais com parâmetros  $(X_{t-1}, \alpha_1), (X_{t-2}, \alpha_2), \dots, (X_{t-p}, \alpha_p)$ , respetivamente.

Considera-se a probabilidade conjunta, condicionada aos  $p$  primeiros valores como:

$$L(\alpha_1, \alpha_2, \dots, \alpha_p, \lambda) = \prod_{t=p+1}^N P(X_t | X_{t-1}, \dots, X_{t-p})(X_{t-1}, \dots, X_{t-p})$$

em que

$$\begin{aligned} P(X_t | X_{t-1}, \dots, X_{t-p}) &= \sum_{i_1=0}^{\min(X_{t-1}, X_t)} \binom{X_{t-1}}{i_1} \alpha_1^{i_1} (1 - \alpha_1)^{(X_{t-1}-i_1)} \times \\ &\quad \sum_{i_2=0}^{\min(X_{t-2}, X_t-i_1)} \binom{X_{t-2}}{i_2} \alpha_2^{i_2} (1 - \alpha_2)^{(X_{t-2}-i_2)} \times \\ &\quad \vdots \times \\ &\quad \sum_{i_p=0}^{\min(X_{t-p}, X_t-(i_1+\dots+i_{p-1}))} \binom{X_{t-p}}{i_p} \alpha_p^{i_p} (1 - \alpha_p)^{(X_{t-p}-i_p)} \times \\ &\quad \frac{e^{-\lambda} \lambda^{X_t-(i_1+\dots+i_p)}}{[X_t-(i_1+\dots+i_p)]!} \end{aligned}$$

A estimativa de Máxima Verosimilhança Condicional para o vetor dos parâmetros  $(\alpha_1, \alpha_2, \dots, \alpha_p, \lambda)'$  é obtida através da maximização de  $L$  com a probabilidade apresentada acima.

### 3.4.3 Método de quase-verosimilhança utilizando a abordagem de modelos lineares generalizados

Seja  $\{Y_t : t \in \mathbb{N}\}$  uma série de contagem e seja  $\{X_t : t \in \mathbb{N}\}$  um vetor de covariável de dimensão  $r$ , variável no tempo, i.e,  $X_t = (X_{t,1}, \dots, X_{t,r})^T$ . Utiliza-se uma média latente do processo,  $\{\lambda_t : t \in \mathbb{N}\}$  para modelar a média condicional da série de contagem, tal que  $E(Y_t | F_{t-1}) = \lambda_t$ . Sendo  $F_t$  o histórico do processo conjunto  $\{Y_t, \lambda_t, X_{t+1}, t \in \mathbb{N}\}$  no período  $t$ , interessa-se em modelos apresentando a seguinte forma geral

$$g(\lambda_t) = \beta_0 + \sum_{k=1}^p \beta_k \tilde{g}(Y_{t-k}) + \sum_{l=1}^q \alpha_l g(\lambda_{t-l}) + \eta^T X_t, \quad (3.8)$$

onde  $g, \tilde{g} : \mathbb{R}^+ \rightarrow \mathbb{R}$  são funções de ligação e transformação respetivamente e  $\eta = (\eta_1, \dots, \eta_r)^T$  corresponde o efeito das covariáveis. Se considerarmos a situação em que  $g(x) = \tilde{g}(x) = x$  e  $\eta = 0$ , obtemos do modelo precedente

$$\lambda_t = \beta_0 + \sum_{k=1}^p \beta_k Y_{t-k} + \sum_{l=1}^q \alpha_l \lambda_{t-l}.$$

Assumindo que  $Y_t$  dado o passado, tem distribuição de Poisson, obtém-se um modelo INteger Ordinary Generalized AutoRegressive Conditional Heteroscedasticity de ordem  $p$  e  $q$  (*INGARCH*( $p, q$ )). O modelo 3.8, sob a hipótese  $E(Y_t | F_{t-1}) \sim \text{Poisson}(\lambda_t)$ , implica que

$$P(Y_t = y | F_{t-1}) = \frac{\lambda_t^y \exp(-\lambda_t)}{y!}, \quad y = 0, 1, \dots,$$

Assim,  $\text{VAR}(Y_t | F_{t-1}) = E(Y_t | F_{t-1}) = \lambda_t$  e ajusta-se o modelo da forma 3.8 pela quase Máxima Verosimilhança Condicional, tendo como vetor dos parâmetros da regressão  $\theta = (\beta_0, \beta_1, \dots, \beta_p, \alpha_1, \dots, \alpha_q, \eta_1, \dots, \eta_r)^T$ . O espaço dos parâmetros para o modelo *INGARCH* é dado por

$$\Theta = \left\{ \theta \in \mathbb{R}^{p+q+r+1} : \beta_0 > 0, \beta_1, \dots, \beta_p, \alpha_1, \dots, \alpha_q, \eta_1, \dots, \eta_r \geq 0, \sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j < 1 \right\}$$

Neste caso,  $\beta_0$  deve ser positivo e todos os outros parâmetros não negativos de modo que a média latente do processo seja positivo. As outras soluções garantem que o modelo ajustado tenha solução estacionária. Para um conjunto de observações  $y =$

$(y_1, \dots, y_n)^T$ , função de quase verosimilhança condicional é dada por

$$l(\theta) = \sum_{t=1}^n \log p_t(y_t; \theta) \propto \sum_{t=1}^n (y_t \ln(\lambda_t(\theta)) - \lambda_t(\theta)),$$

onde,  $p_t(y; \theta) = P(Y_t = y | F_{t-1})$  é a função de densidade de probabilidade de uma distribuição de Poisson. O estimador de quase máxima verosimilhança  $\hat{\theta}_n$  de  $\theta$  é

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} l(\theta).$$

### 3.4.4 Propriedades de correlação

Para o estudo das propriedades de correlação dos processos INAR(p) faz-se o uso do pressuposto que dado  $X_t = x_t$ , a variável aleatória  $\alpha_i \circ x_t$  é independente do passado histórico do processo  $\{X_{t-k}\}$  e de  $\alpha_j \circ X_{t-k}$  para  $i, j = 1, 2, \dots, p$  e  $k \geq 1$ . Agora, pode-se ver de 3.1 que

$$E(X_n) = \sum_{i=1}^p \alpha_i E(X_{n-i}) + \mu_\varepsilon \quad (3.9)$$

onde  $\mu_\varepsilon = E(\varepsilon_n)$ . Sob o pressuposto de estacionaridade fraca temos

$$\mu_X \equiv E(X_n) = \frac{\mu_\varepsilon}{(1 - \sum_{i=1}^p \alpha_i)}.$$

Por derivação da autocovariância na defasagem  $k$ ,  $\gamma(k)$ , temos

$$\begin{aligned} \gamma(k) &= \text{Cov}(X_{n-k}, X_n) \\ &= \sum_{i=1}^p \text{Cov}(X_{n-k}, \alpha_i \circ X_{n-i}) + \delta_k(0) \sigma_\varepsilon^2 \\ &= \sum_{i=1}^p \gamma(k-i, \alpha_i) + \delta_k(0) \sigma_\varepsilon^2 \end{aligned} \quad (3.10)$$

onde  $\gamma(l, \alpha_i) \equiv \text{Cov}(X_{n-l}, \alpha_i \circ X_n)$  e  $\delta_k(0) = 1$  se  $k = 0$  e zero, caso contrário.

É óbvio, da definição de  $\gamma(l, \alpha_i)$  que

$$\gamma(l, \alpha) = \alpha \gamma(l) \quad \text{para } l \geq 0. \quad (3.11)$$

Agora, usando um argumento condicional e o pressuposto de que  $X = x$ , o vetor  $(\alpha_1 \circ x, \alpha_2 \circ x, \dots, \alpha_p \circ x)$  tem uma distribuição multinomial com parâmetros  $(\alpha_1, \alpha_2, \dots, \alpha_p, x)$  temos

$$\text{Cov}(\alpha_j \circ X_{n-k}, \alpha_i \circ X_{n-l}) = \begin{cases} \alpha_j \gamma(k-l, \alpha_i) & \text{se } k < l \\ \alpha_j \alpha_i \delta(0) + \alpha_j (\delta(i) - \alpha_i) \mu_x & \text{se } k = l \\ \alpha_i \gamma(l-k, \alpha_j) & \text{se } k > 0 \end{cases} \quad (3.12)$$

Agora  $\gamma(-l, \alpha)$  pode ser determinado recursivamente usando (3.4). Começando com  $l = 1$ , temos

$$\begin{aligned} \gamma(-1, \alpha_i) &= \alpha_1 \alpha_i \gamma(0) + \alpha_1 (\delta_i(1) - \alpha_i) \mu_x + \alpha_i \sum_{j=2}^p \gamma(\alpha_j, j-1) \\ &= \alpha_i \gamma(-1) + \alpha_1 (\delta_i(1) - \alpha_i) \mu_x \end{aligned} \quad (3.13)$$

Agora, seja  $\mu(l, \alpha_i) \equiv \gamma(l, \alpha_i) - \alpha_i \gamma(l)$ ; segundo a equação (3.11),  $\mu(l, \alpha_i) = 0$  para  $l \geq 0$ , e pela equação (3.13) temos  $\mu(-1, \alpha_i) = \alpha_1 (\delta_i(1) - \alpha_i) \mu_x$ . Do mesmo modo,

$$\begin{aligned} \mu(-2, \alpha_i) &= \sum_{j=1}^p \text{Cov}(\alpha_j \circ X_{n-j}, \alpha_i \circ X_{n-2}) - \alpha_i \gamma(-2) \\ &= \alpha_1 \gamma(-1, \alpha_i) + \alpha_2 \alpha_i \gamma(0) + \alpha_2 (\delta_i(2) - \alpha_i) \mu_x + \alpha_i \sum_{j=3}^p \text{Cov}(\alpha_j \circ X_{n-j}, X_{n-2}) - \alpha_i \gamma(-2) \\ &= \alpha_1 \mu(-1, \alpha_i) + \alpha_2 (\gamma_i(2) - \alpha_i) \mu_x \end{aligned}$$

Usando este procedimento de iteração, obtemos

$$\mu(-l, \alpha_i) = \sum_{j=1}^{l-1} \alpha_j \mu(j-l, \alpha_i) + \mu_{i,l} \quad (3.14)$$

onde  $\mu_{i,l} = \alpha_i (\delta_i(l) - \alpha_i) \mu_x$ . Isto significa que  $\mu(-l, \alpha_i)$  é uma função linear da média do processo  $\mu_x$ . Utilizando a definição de  $\mu(l, \alpha_i)$ , pode-se ver que  $\gamma(k-i, \alpha_i)$  em (3.10) satisfaz

$$\gamma(k-i, \alpha_i) = \alpha_i \gamma(k-i) + \mu(k-i, \alpha_i) \quad (3.15)$$

onde  $\mu(k-i, \alpha_i)$  é determinado por 3.14 para  $k < i$  e  $\mu(k-i, \alpha_i) = 0$  para  $k \geq i$  como

demonstrado na discussão a seguir à equação 3.13. Agora, Usando (3.14), temos

$$\gamma(k) = \sum_{i=1}^p \alpha_i \gamma(k-i) + \sum_{i=k+1}^p \mu(k-i, \alpha_i) + \delta_k(0) \sigma_\varepsilon^2 \quad (3.16)$$

Pode-se ver de 3.16 que a autocovariância do processo INAR(p) é da mesma forma que o processo Gaussiano  $ARMA(p, p-1)$ . Este comportamento de  $\gamma(k)$  deve-se à estrutura de dependência mútua entre os componentes de  $X_n$ , i.e.  $\alpha_i \circ X_{n-i}$  para  $i = 1, 2, \dots, p$ , aparecendo em diferentes períodos como discutido na secção precedente.



# Capítulo 4

## Seleção de ordem nos modelos INAR

### 4.1 Introdução

Apesar da maioria dos métodos de estimação descritos na literatura suporem o conhecimento prévio da ordem do modelo, na prática, a escolha de uma ordem do modelo é uma etapa importante na modelação estatística, uma vez que desta depende uma descrição apropriada dos dados. Assim, este é um problema tratado por diversos autores e várias técnicas para o efeito foram desenvolvidas e propostas, dependendo do tipo e tamanho da série em estudo, nomeadamente, o método gráfico, procedimento de testes de hipótese e critérios automáticos.

Verifica-se que, se a ordem escolhida for inferior da verdadeira, a estimação dos parâmetros é inconsistente e por outro lado, se esta for superior, regista-se o aumento da variância dos estimadores. Então, o objetivo do estudo de seleção de ordem consiste na análise do risco que se assume devido à inconsistência dos estimadores e ao aumento das variâncias destes estimadores. A formulação destes critérios se baseia na atribuição de um custo ou penalização através da introdução de variáveis adicionais, de modo que o equilíbrio possa ser garantido.

## 4.2 Seleção de ordem autorregressiva

Seja o processo discreto estacionário AR(p),

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} \quad (4.1)$$

onde  $\{\varepsilon_t\}$  é uma sequência de variáveis aleatórias não correlacionadas, e todas as raízes do polinômio autorregressivo estão fora do círculo unitário, isto é,  $A(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$  para todo  $z \in \mathbb{C}$  tal que  $|z| \leq 1$ .

O critério automático de seleção de ordem Erro de Predição Final, FPE, proposto por Akaike (1969, 1970), se define como um estimador da predição a 1-passo à frente do erro médio quadrático na realização de um processo  $Y_t$ , independente daquele observado e utilizado na estimação dos parâmetros.

Seja  $\hat{\alpha} = [\hat{\alpha}_1, \cdots, \hat{\alpha}_p]^T$  e  $\hat{\sigma}_\varepsilon^2$  os estimadores de máxima verosimilhança dos parâmetros do processo  $Y_t$ . O preditor linear 1-passo à frente de  $X_{N+1}$  é dado por

$$\hat{X}_{N+1} = \hat{\alpha}_1 X_N + \cdots + \hat{\alpha}_p X_{N+1-p}.$$

Então o erro médio quadrático da predição é

$$E[(X_{N+1} - \hat{X}_{N+1})^2] = \sigma_\varepsilon^2 + E[(\hat{\alpha} - \alpha)^T R_p (\hat{\alpha} - \alpha)],$$

onde

$$[R_p]_{i,j} = E[X_i X_j], \quad i, j = 1, \cdots, p,$$

é a matriz de covariância do processo AR(p) e  $\alpha = (\alpha_1 \cdots \alpha_p)^T$  é o vetor dos coeficientes. Recorrendo à distribuição normal assintótica dos estimadores de máxima verosimilhança,  $\sqrt{n}(\hat{\alpha} - \alpha) \sim \mathcal{N}(0, \sigma_\varepsilon^2 R_p^{-1})$ , segue-se que  $(N/\sigma_\varepsilon^2)(\hat{\alpha} - \alpha)^T R_p (\hat{\alpha} - \alpha) \sim \chi_p^2$ . Ainda  $N\hat{\sigma}_\varepsilon^2/\sigma_\varepsilon^2 \sim \chi_{N-p}^2$ .

Então a predição do erro médio quadrático pode ser estimado por

$$FPE(p) = \hat{\sigma}_\varepsilon^2 \left( \frac{N+p}{N-p} \right),$$

que define o critério proposto por Akaike. Assim, para implementar este método, calcula-se o FPE para cada modelo candidato e se seleciona o modelo cujo este valor é mínimo.

Outro critério de seleção de ordem é o Critério de Informação de Akaike, AIC, também proposto por Akaike (1973, 1974)[11], que se vê como uma aproximação não enviesada da informação esperada de Kullback-Leibler (Kullback & Leibler, 1951) de um modelo ajustado relativa ao verdadeiro modelo.

Nota-se que se  $X$  é um conjunto de  $n$  observações, gerado por um determinado modelo, com função de distribuição  $h$  e cujo vetor dos parâmetros é  $\theta_0$ , e se  $g$  é a distribuição do modelo candidato a representar este conjunto de dados, com vetor dos parâmetros  $\theta$ , a *informação de Kullback-Leibler* entre estes dois modelos se define por

$$I(g, h) = \int_{-\infty}^{+\infty} h(\theta_0 | X) \log \left( \frac{h(\theta_0 | X)}{g(\theta | X)} \right) dX \quad (4.2)$$

$$= \int_{-\infty}^{+\infty} h(\theta_0 | X) \log(h(\theta_0 | X)) dX - \int_{-\infty}^{+\infty} h(\theta_0 | X) \log(g(\theta | X)) dX \quad (4.3)$$

$$= E_0 [\log(h(\theta_0 | X))] - E_0 [\log(g(\theta | X))] \quad (4.4)$$

onde  $E_0$  representa a esperança sob o verdadeiro modelo.

A informação de Kullback-Leibler pode ser aproximada pela *discrepância de Kullback-Leibler*, isto é,

$$d(h, g) = E_0 [-2 \log(g(\theta | X))] \quad (4.5)$$

pois a primeira esperança em 4.4 tem um valor constante independentemente da forma da distribuição candidata  $g(\theta | X)$ .

Se  $Y$  é uma realização independente do verdadeiro processo e  $X$ , um conjunto de  $N$  observações de um processo Gaussiano AR de média nula, com o vetor dos parâmetros  $\theta_0 = (\alpha_1, \dots, \alpha_p, \sigma_\varepsilon^2)$  e estimador de máxima verosimilhança  $\hat{\theta}_X$ , a função log-verosimilhança associada a  $X$  é

$$\log(L(\theta; X)) = l(\theta; X) = -\frac{2}{N} \log(2\pi\sigma_\varepsilon^2) - \frac{1}{2} \log(\det R_p) - \frac{1}{2\sigma_\varepsilon^2} S_X(\theta),$$

onde  $R_p$  é uma matriz de covariância do processo AR(p) e

$$S_X(\theta) = \sum_{j=1}^N \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

é a soma ponderada do quadrado dos erros de predição a 1-passo à frente.

Segundo Akaike (1973, 1974),  $-2 \log(L(\theta; X))$  pode ser considerado como um estimador enviesado da discrepância de Kullback-Leibler, 4.5, e pode escrever-se que

$$-2 \log(L(\theta; Y)) = -2 \log(L(\theta; X)) + \frac{1}{\sigma_\varepsilon^2} (S_Y(\theta) - S_X(\theta))$$

Como  $E[S_X(\hat{\theta}_X)] = E[S_Y(\hat{\theta}_Y)]$ , então

$$E[S_Y(\hat{\theta}_X)] - E[S_X(\hat{\theta}_X)] = E[S_Y(\hat{\theta}_X)] - E[S_Y(\hat{\theta}_Y)].$$

Considerando a expansão em série de Taylor de  $S_Y(\hat{\theta}_X)$  e tendo em conta que  $X$  e  $Y$  são mutuamente independentes, vê-se que

$$\begin{aligned} E[S_Y(\hat{\theta}_X)] - E[S_X(\hat{\theta}_X)] &\approx \sigma_\varepsilon^2 E[\hat{\theta}_X - \theta_0]^T J(\theta_0) E[\hat{\theta}_X - \theta_0] \\ &+ \sigma_\varepsilon^2 E[\hat{\theta}_Y - \theta_0]^T J(\theta_0) E[\hat{\theta}_Y - \theta_0] \\ &= 2\sigma_\varepsilon^2 k, \end{aligned}$$

onde  $J$  é uma matriz de informação de Fisher e  $k$  é o número de parâmetros do vetor dos estimadores, ou seja,  $k = p + 1$  no caso autorregressivo com variância de ruído desconhecido.

Assim, seleciona-se a ordem  $p$  que minimiza o seguinte critério

$$AIC(p) = N \log(\hat{\sigma}_\varepsilon^2) + 2(p + 1)$$

que é uma estimativa não enviesada da discrepância esperada de Kullback-Leibler, quando o tamanho da amostra é relativamente grande e a ordem  $p$  é comparativamente pequena. Define-se ainda uma versão corrigida deste critério, denotado AICC, que pode ser escrito como a soma do AIC e um termo não estocástico e toma em

consideração o tamanho da amostra, essencialmente, através do aumento da penalidade relativa para modelos de amostras pequenas. Utilizou-se o critério de Whittle [22] para aproximar a verosimilhança através da função de densidade espectral do modelo e viu-se que

$$E[-2\log(L(\theta))] \approx d(f, g) = N \log(2\pi) + \frac{N}{2\pi} \int_{-\pi}^{+\pi} \log(g_{\theta}(\omega)) + \frac{f_{\theta_0}(\omega)}{g_{\theta}(\omega)} d\omega$$

onde  $d(.,.)$  é discrepância de Kullback-Leibler utilizado como aproximação da informação de Kullback-Leibler,  $f_{\theta_0}(\omega)$  e  $g_{\theta}(\omega)$  são densidades dos modelos verdadeiro e candidato, com parâmetros  $\theta_0$  e  $\theta$ , respectivamente.

Se se assume que as aproximações incluem o verdadeiro modelo e, como os estimadores de Yule-Walker são normais, Hurvich e Tsai (1989) [9] mostraram que a esperança da discrepância de Kullback-Leibler entre a densidade espectral,  $f$ , e a estimativa de um candidato  $\hat{f}$  é dada por

$$E[d(f, \hat{f})/N] = E[\log(\hat{\sigma}_{\varepsilon}^2)] + E\left[\frac{\sigma_{\varepsilon}^2 + (\hat{\alpha} - \alpha)^T R_p (\hat{\alpha} - \alpha)}{\hat{\sigma}_{\varepsilon}^2}\right]$$

onde  $\hat{\sigma}_{\varepsilon}^2$  é um estimador de  $\sigma_{\varepsilon}^2$ ,  $\hat{\alpha}$  é o estimador dos parâmetros de Yule-Walker e  $R_p$  é a matriz de covariância do modelo candidato. Utilizando a distribuição assintótica normal dos estimadores de Yule-Walker,  $\sqrt{N}(\hat{\alpha} - \alpha) \sim \mathcal{N}(0, \sigma_{\varepsilon}^2 R_p^{-1})$  e  $N\hat{\sigma}_{\varepsilon}^2/\sigma_{\varepsilon}^2 \sim \chi_{N-p}^2$ , para  $N$  grande, a esperança da discrepância de Kullback-Leibler pode ser escrito por

$$E[d(f, \hat{f})/N] = E[\log(\hat{\sigma}_{\varepsilon}^2)] + \frac{1 + p/N}{1 - (p+2)/N}$$

Assim o critério AICC é dado por

$$AICC(p) = N \log(\hat{\sigma}_{\varepsilon}^2) + N \frac{1 + p/N}{1 - (p+2)/N}$$

e o valor de  $p$  que minimiza o critério é escolhido.

### 4.3 Seleção de ordem nos modelos INAR

Suponhamos que se pretende ajustar uma série temporal de valores inteiros não negativos, através de um modelo INAR e que a melhor ordem  $p$  deve ser selecionada. A interpretação de  $\sigma^2$  como variância do erro de predição a 1-passo ou, ainda, como variância dos resíduos para os modelos AR não tem analogia com o caso de modelos INAR, pelo que não se pode selecionar a ordem verdadeira de um modelo INAR recorrendo ao uso dos critérios expostos anteriormente.

Assim, em caso de utilização do modelo INAR, a solução [22] consiste em deduzir um critério do tipo AICC em que, através da função de densidade espectral, considera-se uma aproximação da função de verosimilhança proposta por Whittle (1953) e usualmente designada por *critério de Whittle*.

Se  $X_1, \dots, X_n$  são observações de um processo estacionário, cujas funções de autocovariância e de densidade espectral são  $R(k)$  e  $f(\omega)$ , respetivamente, e seja  $g(\omega)$  uma função par, não negativa e integrável em  $[-\pi, \pi]$ , uma aproximação para a log-verosimilhança,  $l(g)$ , é tal que (Hurvich & Tsai (1989)), conforme consultado em Silva e Silva [21],

$$-2l(g) \simeq n \log(2\pi) + \frac{n}{2\pi} \int_{-\pi}^{\pi} \log(g(\omega)) + \frac{I_n(\omega)}{g(\omega)} d\omega$$

onde  $I_n(\omega)$  é o periodograma definido por  $I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{i\omega t} \right|^2$ . Como o periodograma é um estimador assintoticamente cêntrico da função de densidade espectral,  $f(\omega)$ , tem-se que

$$E[-2l(g)] \simeq d(f, g) = n \log(2\pi) + \frac{n}{2\pi} \int_{-\pi}^{\pi} \left( \log(g(\omega)) + \frac{f(\omega)}{g(\omega)} \right) d\omega,$$

é uma aproximação do índice de Kullback-Leibler.

Suponha-se que o modelo aproximante é da forma INAR(p), cujo vetor de parâmetros  $\hat{\alpha} = [-1, \hat{\alpha}_1, \dots, \hat{\alpha}_p]^T$ ,  $\hat{\mu}_\varepsilon$  e  $\hat{\sigma}_\varepsilon^2$  é estimado, por exemplo, através do método de mínimos

quadrados condicionais. O espectro estimado por

$$\hat{f}(\omega) = \frac{1}{2\pi} \left| \frac{\hat{V}_p}{1 - \sum_{k=1}^p \hat{\alpha}_k e^{-i\omega k}} \right|^2, \text{ com } \hat{V}_p = \hat{\sigma}_\varepsilon^2 + \bar{X} \left( \sum_{k=1}^p \hat{\beta}_k \right),$$

onde  $\hat{\beta}(k)$  é o estimador da variância de série de contagem envolvida no  $k$ -ésimo operador *thinning*,  $\alpha(k) \circ X_{t-k}$ ,  $k = 1, \dots, p$  e  $\bar{X}$  é a média amostral.

Admitindo que a família de modelos aproximante inclui o modelo verdadeiro, que é do tipo INAR(p) com coeficientes  $\alpha = [-1, \alpha_1, \dots, \alpha_p]^T$ , então as equações do tipo de Yule-Walker são satisfeitas,  $R_p \alpha = [-V_p \ 0 \ \dots \ 0]^T$ , escrevendo  $R_p$  na sua forma escalar como

$$\begin{aligned} R(0) &= V_p + \sum_{i=1}^p \alpha_i R(i) \\ R(k) &= \sum_{i=1}^p \alpha_i R(k-1), k \geq 1 \end{aligned}$$

e na vetorial por

$$R_p \alpha = \begin{pmatrix} R(0) & R(1) & \dots & R(p) \\ R(1) & R(0) & \dots & R(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ R(p) & R(p-1) & \dots & R(0) \end{pmatrix} \begin{pmatrix} -1 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} -V_p \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\text{e } V_p = \sigma_\varepsilon^2 + \mu_x \sum_{k=1}^p \beta_k.$$

Então, utilizando a fórmula de Kolmogorov adaptada aos processos INAR(p),  $V_p = 2\pi \exp \left[ \int_{-\pi}^{\pi} \log(f(\omega)) d\omega \right]$  [22] para os modelos AR, e as propriedades das equações de

Yule-Walker, tem-se que

$$\begin{aligned}
 E \left[ \frac{d(\hat{f}, f)}{n} \right] &= E[\log(2\pi)] + \frac{1}{2\pi} E \left[ \int_{-\pi}^{\pi} \left( \log \hat{f}(\omega) + \frac{f(\omega)}{\hat{f}(\omega)} \right) d\omega \right] \\
 &= E \left[ \log(\hat{V}_p) \right] + E \left[ \int_{-\pi}^{\pi} \frac{f(\omega)}{\hat{V}_p} \left| 1 - \sum_{k=1}^p \hat{\alpha}_k e^{i\omega k} \right|^2 d\omega \right] \\
 &= E \left[ \log(\hat{V}_p) \right] + E \left[ \frac{1}{\hat{V}_p} \hat{\alpha}^T R_p \hat{\alpha} \right] \\
 &= E \left[ \log(\hat{V}_p) \right] + E \left[ \frac{V_p + (\hat{\alpha} - \alpha)^T R_p (\hat{\alpha} - \alpha)}{\hat{V}_p} \right]
 \end{aligned}$$

Tentou-se determinar, ou pelo menos aproximar, o valor esperado de

$$\left[ \frac{V_p + (\hat{\alpha} - \alpha)^T R_p (\hat{\alpha} - \alpha)}{\hat{V}_p} \right]$$

através do método Delta ou do desenvolvimento em série de Taylor dessa expressão, utilizando a distribuição normal assintótica dos estimadores dos mínimos quadrados condicionais dos parâmetros de um modelo INAR(p) [22]. No entanto, como a matriz de covariância é bastante complexa, não foi possível obter um termo de penalização satisfatório no sentido de que as aproximações obtidas dependem dos parâmetros  $(\alpha_i, \mu_\epsilon, \sigma_\epsilon^2)$ , da ordem do modelo a ajustar e do número de observações disponíveis.

Assim, considera-se o seguinte critério

$$AICC_{inar}(p) = n \log(\hat{V}_p) + n \frac{1 + p/n}{1 - (p + 2)/n} \quad (4.6)$$

cujo termo de penalização seja correspondente ao critério existente para os modelos AR.

# Capítulo 5

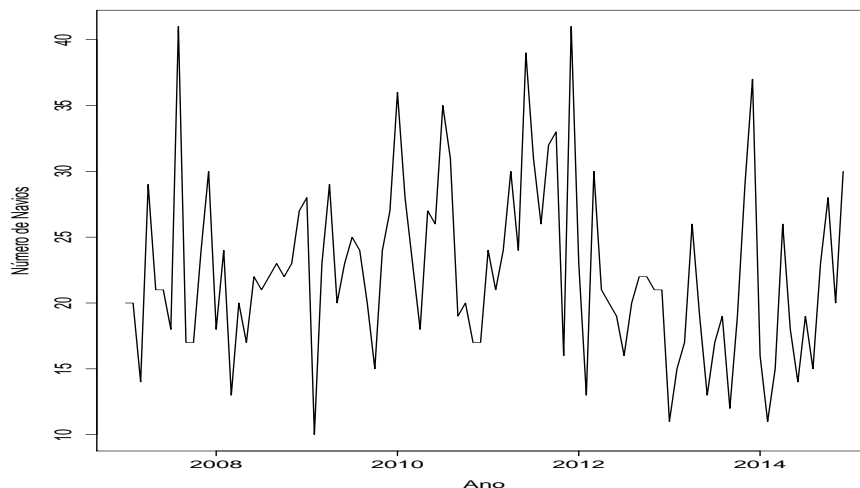
## Aplicação a dados reais

Neste capítulo aplicamos os métodos e técnicas dos modelos INAR, apresentados no capítulo precedente, a dois conjuntos de dados reais sobre os setores dos transportes marítimos e a atividade dos seguros de Cabo Verde, utilizando o software R.

### 5.1 Transporte marítimo em Cabo Verde

Sendo a República de Cabo Verde um país insular, o transporte marítimo sempre teve um papel importante no seu desenvolvimento económico e social, contribuindo para a minimização da descontinuidade territorial e facilitando assim, a movimentação de pessoas e produtos a um menor custo. No que diz respeito ao comércio exterior, o transporte marítimo tem sido o principal meio de aprovisionamento do mercado nacional, de produtos cuja produção nacional é insuficiente, bem como na exportação da produção local.

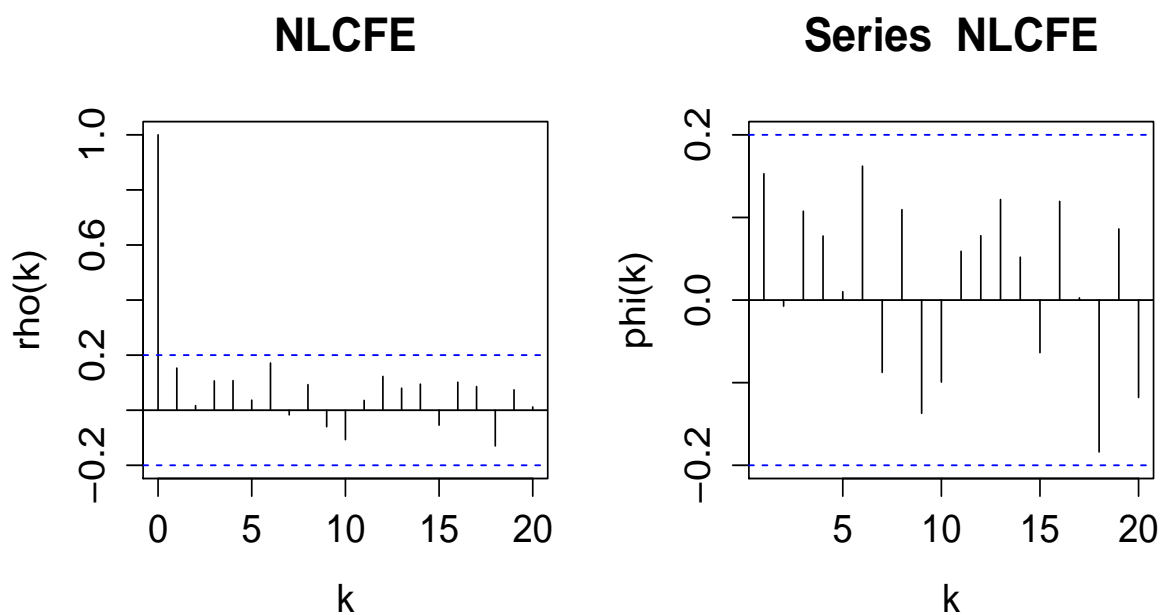
Nesta secção consideramos o número mensal de navios estrangeiros de longo curso e o número mensal de navios nacionais de cabotagem fundeados nos portos de Cabo Verde cujos dados são recolhidos e compilados pela Empresa Nacional de Administração dos Portos, e disponíveis a partir de Janeiro de 2007 até Dezembro de 2014. A figura 5.1 apresenta o número de navios de longo curso estrangeiros fundeados nos portos.



**Figura 5.1:** *Número de navios estrangeiros de longo curso fundeados*

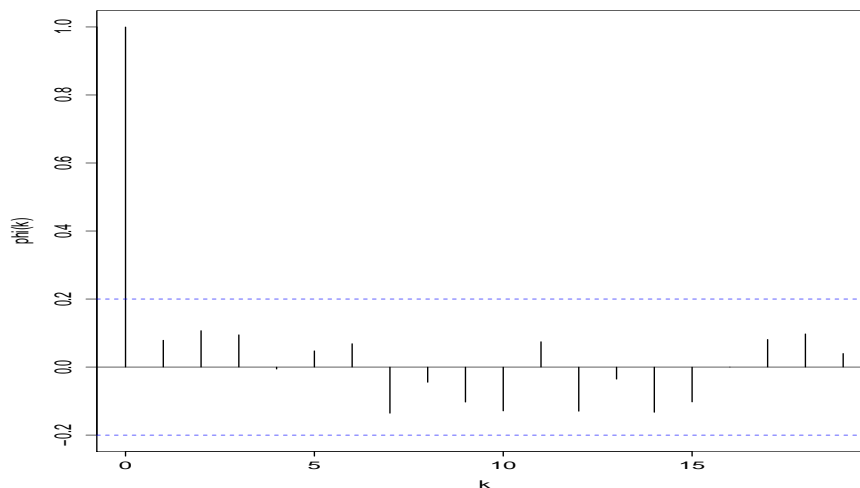
As funções de autocorrelação amostral e autocorrelação amostral são apresentados na figura 5.2 e a sua análise sugere que o modelo possível para esta série pode se definir como um ruído aleatório, apresentado da seguinte forma:

$$X_t = \mu + \varepsilon_t \quad \text{com } \mu = 22 \quad (5.1)$$



**Figura 5.2:** *Funções de autocorrelação e autocorrelação parcial do número de navios estrangeiros de longo curso fundeados nos portos de Cabo Verde*

A função de autocorrelação dos erros para o modelo ajustado do número de navios de longo curso fundeados nos portos de Cabo Verde mostra o comportamento típico de variáveis aleatórias independentes e identicamente distribuídas.



**Figura 5.3:** Função de autocorrelação dos erros do modelo 5.1

Uma vez que estes erro tem distribuição de Poisson, o modelo ajustado é

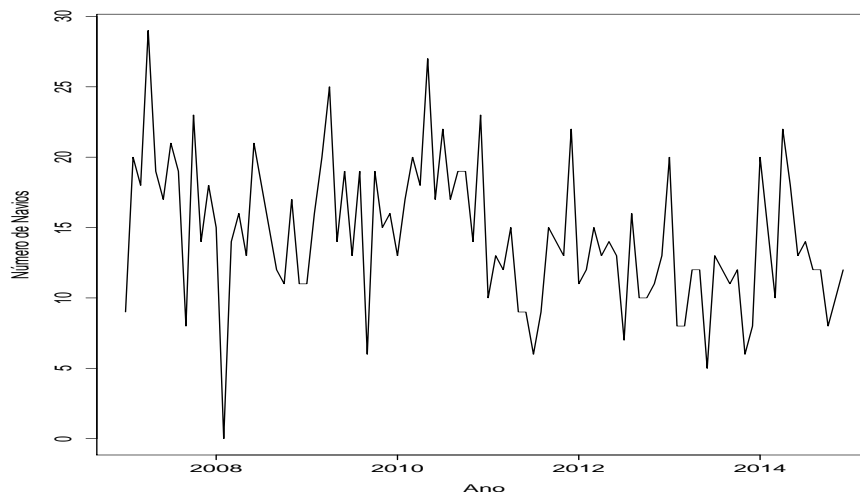
$$X_t = 22 + \varepsilon_t$$

e os erros  $\varepsilon_t$  apresentam distribuição de Poisson com média  $\lambda = 5.094$ .

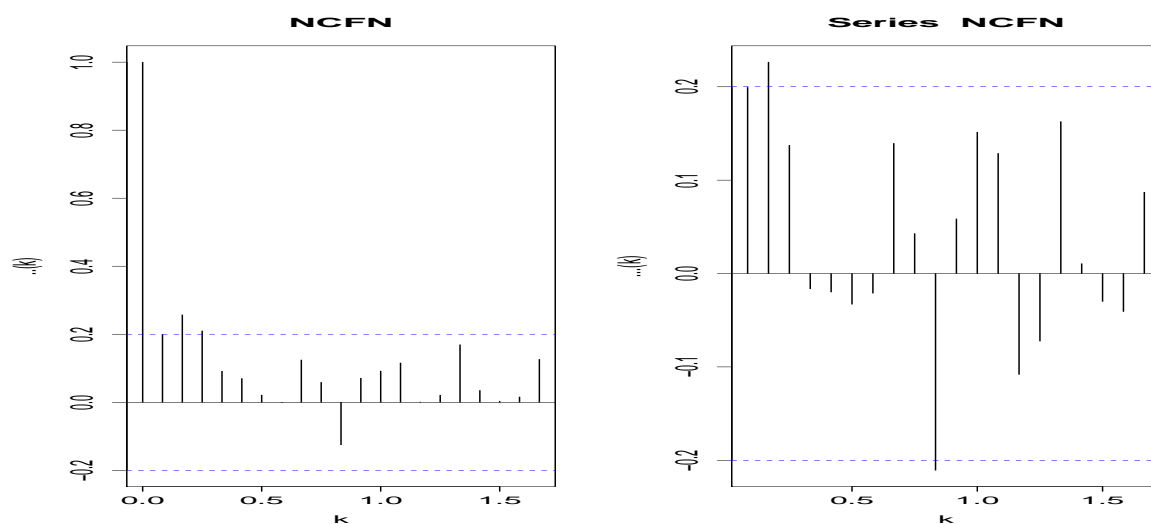
Analisamos também a série do número mensal de navios nacionais de cabotagem fundeados nos portos de Cabo Verde, durante este período, apresentada no gráfico seguinte.

As funções de autocorrelação e autocorrelação parcial dos dados sobre o número de navios nacionais de cabotagem fundeados nos portos nacionais, apresentados na figura 5.5, e o critério  $AICC_{inar}$  apontam para uma relação de ordem 2 para o ajustamento do modelo INAR.

As estimativas dos parâmetros do modelo INAR(2), relativas ao número de navios de cabotagem fundeados, são apresentados na tabela a seguir. Nota-se que os métodos de Yule-Walker e de Mínimos Quadrados Condicionais fornecem resultados similares, apresentando principalmente estimativas de  $\alpha$  com valores superiores aos correspondentes obtidos através dos outros métodos. Inversamente, a estimativa de Máxima



**Figura 5.4:** Número de navios nacionais de cabotagem fundeados nos portos de Cabo Verde



**Figura 5.5:** Funções de autocorrelação e autocorrelação parcial do número de navios nacionais de cabotagem fundeados nos portos de Cabo Verde

Método	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\lambda}$	$\hat{\mu}_\varepsilon$	$\hat{\sigma}_\varepsilon^2$
YW	0.19944	0.21737	8.39542	3.58490	7.66890
MQC	0.21725	0.21651	8.03722	3.57656	7.76762
MVC	0.13319	0.15789	10.18069	3.62130	7.72587
QMV	0.15500	0.21300	9.09100	3.59213	7,64992

**Tabela 5.1:** Estimativa dos coeficientes do modelo INAR(2) ajustado sobre o número mensal de navios de cabotagem nacionais fundeados

Verosimilhança Condicional de  $\hat{\mu}_\varepsilon$  apresenta valor superior aos outros enquanto a diferença entre  $\hat{\alpha}_1$  e  $\hat{\alpha}_2$  obtidos pelo método de Quase Máxima Verosimilhança é maior.

A bondade de ajuste dos modelos é avaliada pelos resíduos, através do critério defi-

nido por

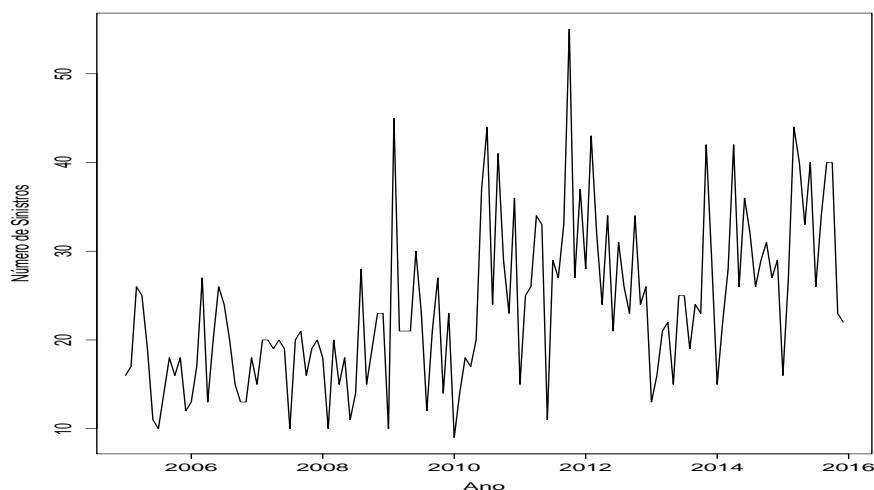
$$r_t^{(M)} = X_t - \hat{\alpha}_1^{(M)} X_{t-1} - \hat{\alpha}_2^{(M)} X_{t-2} - \hat{\mu}_\varepsilon^{(M)}$$

onde  $t = 3, \dots, 96$ , e  $M$  representa o método utilizado. Os resultados do modelo INAR(2) permitem concluir que, neste caso particularmente, a estimação pelo método de Mínimos Quadrados Condicionais é o mais apropriado para o ajuste do número mensal de navios de cabotagem fundeados nos portos de Cabo Verde.

## 5.2 Atividades de seguros

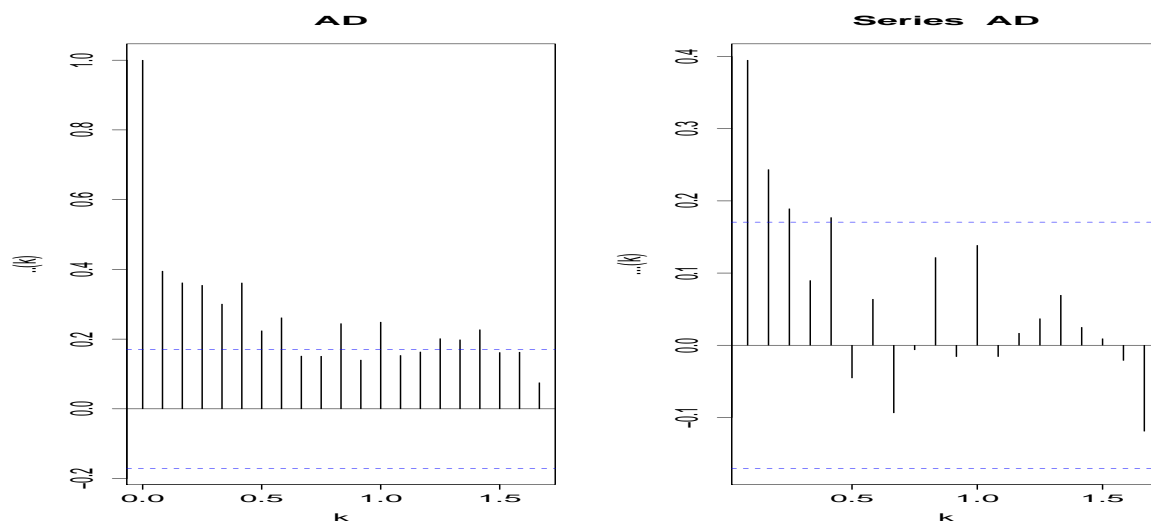
Propomos também aplicar a metodologia dos modelos INAR aos dados do setor de seguros nomeadamente, o número de sinistros comunicados no mês, nos ramos de *acidentes e doença e automóveis*, disponíveis no Banco de Cabo Verde (BCV), de Janeiro de 2005 a Dezembro de 2015.

A análise da autocorrelação e autocorrelação parcial do número de sinistros comunicados mensalmente no ramo de acidentes e doenças sugerem o ajustamento da série por um modelo INAR(5).



**Figura 5.6:** Número de sinistros comunicados no mês no ramo de acidentes e doença

Nota-se, da tabela 5.2 a similaridade dos estimadores dos métodos de Yule-Walker e de Mínimos Quadrados Ordinários enquanto o método de Máxima verosimilhança Condicional apresenta estimativas maiores para  $\hat{\mu}_\varepsilon$  e  $\hat{\sigma}_\varepsilon^2$ . Analisando o critério de verificação, concluímos que o método de Quase Máxima verosimilhança é mais adequado



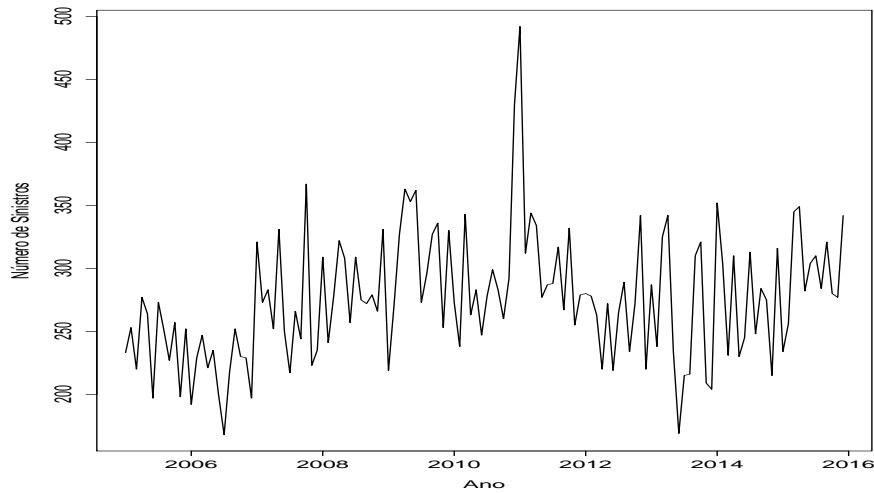
**Figura 5.7:** Funções de autocorrelação e autocorrelação parcial do número de sinistros comunicados no mês - ramo de acidentes e doença

para o ajustamento do modelo INAR(5).

Parâmetros	YW	MQC	QMV
$\hat{\alpha}_1$	0.39504	0.39678	0.22020
$\hat{\alpha}_2$	0.20548	0.21203	0.12980
$\hat{\alpha}_3$	0.13040	0.12612	0.14800
$\hat{\alpha}_4$	0.03476	0.03125	0.06950
$\hat{\alpha}_5$	0.10878	0.11214	0.16370
$\hat{\lambda}$	2.98070	2.58912	6.45640
$\hat{\mu}_\varepsilon$	6.03253	6.05785	5.76581
$\hat{\sigma}_\varepsilon^2$	27.70951	27.65054	27.75941

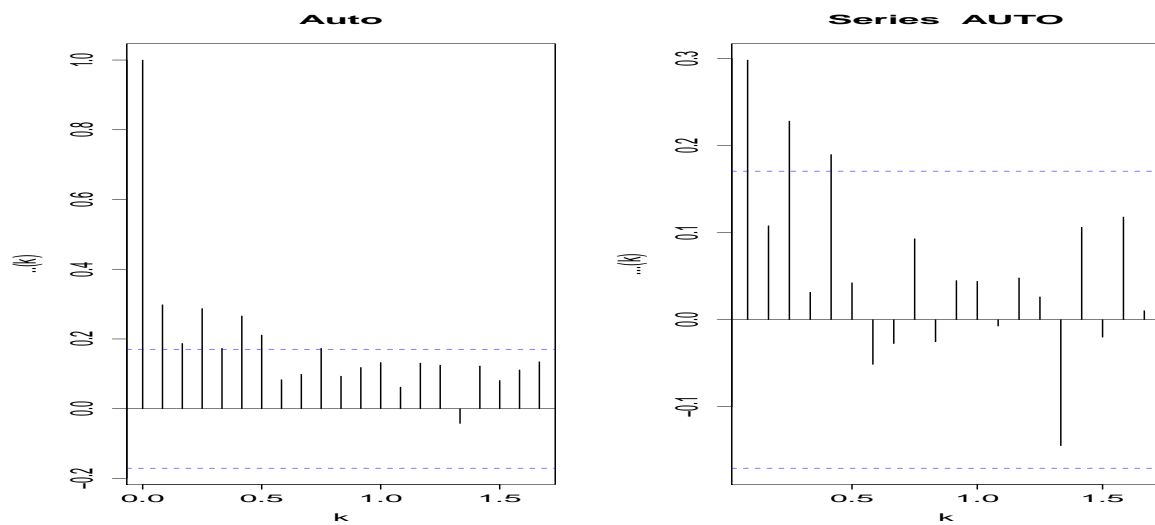
**Tabela 5.2:** Estimativa dos coeficientes do modelo INAR(5) ajustado sobre o número mensal de sinistros comunicados no ramo acidentes e doenças

Apresentamos, na figura seguinte, o número mensal de sinistros comunicados referente ao ramo *transportes automóveis*.



**Figura 5.8:** Número de sinistros comunicados no mês - ramo transportes automóveis

A análise do critério de seleção de ordem  $AICC_{inar}$  (4.6) e das funções de autocorrelação e autocorrelação parcial permite ajustar os dados por um modelo INAR(5). Na tabela 5.3 apresentamos as estimativas dos parâmetros.



**Figura 5.9:** Funções de autocorrelação e autocorrelação parcial do número de sinistros comunicados no mês no ramo de transportes automóveis

Parâmetros	YW	MQC	QMV
$\hat{\alpha}_1$	0.29822	0.29686	0.00301
$\hat{\alpha}_2$	0.09837	0.09321	0.00000
$\hat{\alpha}_3$	0.20226	0.21034	0.00313
$\hat{\alpha}_4$	0.00875	0.00897	0.00002
$\hat{\alpha}_5$	0.14581	0.15124	0.00276
$\hat{\lambda}$	67.90337	62.31874	275.0000
$\hat{\mu}_\varepsilon$	36.18693	36.10169	39.21758
$\hat{\sigma}_\varepsilon^2$	842.8730	867.1602	1026.208

**Tabela 5.3:** Estimativa dos coeficientes do modelo INAR(5) ajustado sobre o número mensal de sinistros comunicados no ramo transportes automóveis

As estimativas de  $\alpha$  pelo método de Quase Máxima Verosimilhança são muito inferiores às resultantes dos outros métodos e e inversamente, para os parâmetros  $\hat{\mu}_\varepsilon$  e  $\hat{\sigma}_\varepsilon^2$ , apresenta estimativas superiores. Pela análise dos resíduos conclui-se de que o método de Quase Máxima Verosimilhança é mais robusto para a realização deste ajustamento.

# Conclusões

Tendo esta dissertação como objetivo a análise de séries temporais discretas, estudamos um tipo particular de processo para séries de contagem nomeadamente, os modelos autorregressivos com valores inteiros não negativos (INAR). Apresentamos, no capítulo introdutório, e de forma descritiva, o âmbito geral de modelação das séries temporais. Definimos o conceito de estacionaridade e apresentamos alguns métodos de modelação e estimação dos componentes "não estacionários" de um modelo clássico de decomposição de uma série. Vimos igualmente alguns métodos usuais de análise e modelação de séries temporais com variáveis contínuas como os processos autorregressivos ou Média Móvel (Moving Average) e de previsão.

Após estabelecermos a base de análise de séries temporais, apresentamos os modelos autorregressivos com valores inteiros não negativos (INAR) com distribuição marginal de Poisson, os métodos de estimação dos parâmetros, as suas propriedades assintóticas e a sua aplicação a dados reais. A função de autocorrelação de um processo INAR é idêntica à do modelo AR apesar de terem uma média e variância com distribuição discreta. Os modelos INAR devem também satisfazer as condições de estacionaridade e invertibilidade.

Da aplicação da metodologia dos modelos INAR em séries económicas de Cabo Verde, verificamos que, a estimação de parâmetros de um processo INAR, usando os métodos de Yule-Walker e Mínimos Quadrados Condicionais, apresenta resultados similares. Dentre os métodos de estimação estudados, o de Quase Máxima Verosimilhança demonstra ser geralmente o mais eficiente pois, os estimadores de Yule-Walker e Mínimos Quadrados Condicionais parecem enviesados, enquanto as estimativas dadas pela Quase Máxima Verosimilhança, mais próximos do verdadeiro valor do parâmetro.

A análise do caso prático pode ser mais desenvolvida, estudando outras componentes das séries, como por exemplo, a sazonalidade. Seria interessante comparar estes resultados com os modelos clássicos ARMA, para o caso contínuo e estudar metodologias de modelação de dados de contagem do tipo INARMA, isto é, que permitem estimar, além dos coeficientes autorregressivos, os coeficientes de média móvel.

# Referências Bibliográficas

- [1] Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21, 243-247.
- [2] Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, 22, 203-217.
- [3] Al-Osh, M. A. & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* 8.
- [4] Alzaid, A. A., & Al-Osh, M. (1990). An Integer-Valued pth-Order Autoregressive Structure (INAR(p)) Process. *Journal of Applied Probability*, 27(2), 314-324.
- [5] Boudreault, M. & Charpentier, A. (2011). Multivariate integer-valued autoregressive models applied to earthquake counts. hal-00646848.
- [6] Brockwell, P. J., & Davis, R. A. (1996). Introduction to Time Series and Forecasting. *Springer*.
- [7] Brockwell, P. J. & Davis, R. A. (1991). Time Series: Theory and Methods. Second Edition. *Springer*.
- [8] Bu, R., McCabe, B. & Hadri, K. (2008). Maximum likelihood estimation of higher order integer-valued autoregressive processes. *Journal of Time Series Analysis*, 29, 973-994.
- [9] Hurvich, C. and C.L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika*, 76, 297-307.
- [10] Jin-Guan, D. & Yuan L., (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, 12, 129-142.

- [11] Klimko, L. and P. Nelson (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistic*.
- [12] Kullback, S. and R.A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics*.
- [13] Kurt, B. (1994). Estimation and Testing in Integer-Valued AR(1) Models. *Umea Economic Studies* 335.
- [14] Latour, A. (1998). Existence and Stochastic Structure of a Non-negative Integer-valued Autoregressive Process. *Journal of Time Series Analysis* 19, 439-455.
- [15] Liboschik, T., Fokianos, K. & Fried, R. (2015). tscount: An R package for analysis of count time series following generalized linear models. *SFB 823 Discussion Paper*.
- [16] Lima, T. A. C., (2013). Modelos INAR e RCINAR: estimação e aplicação. *Instituto de Matemática e Estatística da Universidade de São Paulo*.
- [17] M. Bourguignon & K. L. P. Vasconcellos. (2015). Improved estimation for Poisson INAR(1) models. *Journal of Statistical Computation and Simulation*, 85:12, 2425-2441.
- [18] McKenzie, E. (1988). Some ARMA Models for Dependent Sequences of Poisson Counts. *Advances in Applied Probability*, 20(4), 822-835.
- [19] Park, Y. & Oh, C. W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers* 38, 287-302.
- [20] Quddus, M. A. (2007). Time series count data models: an empirical application to traffic accidents. *Journal of Statistical Planning and Inference*.
- [21] Silva, I. & Silva, M. E. (2005). Selecção de ordem em modelos INAR. *Actas do X Congresso Anual da SPE*.
- [22] Silva, I. M. M. (2005). Contributions to the analysis of discrete-valued time series. *Departamento de Matemática Aplicada. Faculdade de Ciências da Universidade do Porto*.

- [23] Silva, I. M. M. & Silva, M. E. (2006). Asymptotic distribution of the Yule-Walker estimator for INAR(p) processes. *Journal of Statistical Planning and Inference*.
- [24] Silva, I. M. M. & Silva, M. E. (2009). Parameter estimation for INAR processes Based on high-order statistics. *REVSTAT - Statistical Journal*, 7, 105-117.
- [25] Silva, M. E., & Oliveira, V. L., (2000). Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of time series analysis*, 25, No. 3.
- [26] Silva, M. E. (2015). Modelling time series of counts: an INAR approach. *Textos de Matemática, DMUC*, 47, 107-122.
- [27] Silva N., Pereira, I., & Silva, M. E. (2009). Forecasting in INAR(1) model. *REVSTAT-Statistical Journal*, 7, Number 1, 119-134.
- [28] Steutel, F.W. & Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, 7, 893 - 899.
- [29] Thyregod, P., Carstensen, J., Madsen, H., & Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements. *Econometrics*, 10, 395-411.
- [30] Torres, J. N. F. (2001). Modelo "INARMA" para series temporales de valores enteros: análisis, propiedades asintomáticas y estimación. *Universidad Complutense de Madrid, Facultad de Ciencias Matemáticas*.
- [31] Zhang, H., Wang, D., Zhu, F. (2009). Inference for INAR(p) processes with signed generalized power series thinning operator. *Journal of Statistical Planning and Inference*.