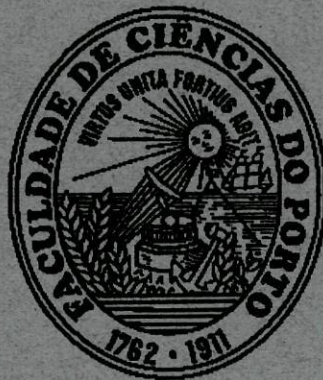


Irene Vitória Ribeiro de Brito

**MÉTODOS DE ESTIMAÇÃO
PARA PROCESSOS RAMIFICADOS COM IMIGRAÇÃO
E PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO**



Departamento de Matemática Aplicada
Faculdade de Ciências da Universidade do Porto
Janeiro de 2003

Irene Vitória Ribeiro de Brito

**MÉTODOS DE ESTIMAÇÃO
PARA PROCESSOS RAMIFICADOS COM IMIGRAÇÃO
E PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO**



*Tese submetida à Faculdade de Ciências da Universidade do Porto
para a obtenção do grau de Mestre em Estatística*

Departamento de Matemática Aplicada
Faculdade de Ciências da Universidade do Porto
Janeiro de 2003

Dissertação orientada por

Maria Eduarda da Rocha Pinto Augusto da Silva

Professora Auxiliar do

**Departamento de Matemática Aplicada da
Faculdade de Ciências da Universidade do Porto**

Aos meus pais

AGRADECIMENTOS

Apresento aqui os meus cordiais agradecimentos à Professora Doutora Maria Eduarda Silva, que me orientou na elaboração deste trabalho, e a Isabel Silva, que me forneceu os programas em Matlab e que também me ajudou no esclarecimento de algumas dúvidas.

RESUMO

Neste trabalho consideram-se duas classes de processos que têm aplicações reais a séries de dados discretos: processos ramificados com imigração - BGWI, e processos autorregressivos de valor inteiro - INAR.

Os processos BGWI são processos estocásticos que têm sido estudados na literatura desde o início do século XX e aplicados em áreas como genética, neurofisiologia, "traffic flow theory" (teoria da circulação do trânsito) e física estatística. Os processos INAR foram propostos recentemente na literatura para modelar séries temporais de contagem. Mostra-se no entanto, que os processos INAR podem ser vistos como processos BGWI.

O objectivo é apresentar e comparar métodos propostos na literatura para resolver o problema da estimação dos parâmetros destes dois tipos de processos. Também se estuda a possibilidade de uma aplicação de métodos, propriedades e resultados do processo ramificado com imigração ao processo autorregressivo de valor inteiro. Pretende-se assim ampliar os conhecimentos sobre a inferência estatística para os processos autorregressivos de valor inteiro.

ABSTRACT

Here we consider two classes of stochastic processes which are used to model series of count data: branching processes with immigration - BGWI, and integer-valued autoregressive processes - INAR.

The BGWI processes have appeared in applications since the beginning of the XX century in genetics, neurophysiology, traffic flow theory and statistical physics. The INAR processes have been proposed recently in the literature to model time series of count data. However, it can be shown that the INAR processes can be viewed as BGWI processes.

The purpose of this study is to apply estimation methods, properties and results established for the BGWI processes to the INAR processes.

ÍNDICE

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| ABREVIATURAS..... | 8 |
| 1. INTRODUÇÃO | 9 |
| 2. PROCESSOS RAMIFICADOS COM IMIGRAÇÃO..... | 12 |
| 2.1. DEFINIÇÕES E PROPRIEDADES | 12 |
| 2.2. ESTIMAÇÃO DOS PARÂMETROS DO PROCESSO RAMIFICADO COM IMIGRAÇÃO | 15 |
| 2.2.1. ESTIMAÇÃO BASEADA NA OBSERVAÇÃO DE UMA ÚNICA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$ | 17 |
| 2.2.1.1. ESTIMADORES BASEADOS NA COMPARAÇÃO DO PROCESSO RAMIFICADO COM IMIGRAÇÃO COM O PROCESSO AUTORREGRESSIVO DE ORDEM 1 | 17 |
| 2.2.1.2. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS | 19 |
| 2.2.1.3. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS PESADOS | 23 |
| 2.2.1.4. COMPARAÇÃO DAS PROPRIEDADES E DOS RESULTADOS OBTIDOS COM OS MÉTODOS DE ESTIMAÇÃO BASEADOS NA OBSERVAÇÃO DE UMA ÚNICA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$ | 27 |
| 2.2.2. ESTIMAÇÃO BASEADA NA OBSERVAÇÃO DE TODAS AS DIMENSÕES DA IMIGRAÇÃO E DA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$ | 31 |
| 2.2.2.1. MÉTODO DE MÁXIMA VEROSIMILHANÇA | 31 |
| 2.2.2.2. ESTIMAÇÃO SEQUENCIAL | 33 |
| 2.2.2.3. ESTIMAÇÃO DE PRECISÃO FIXA | 36 |
| 2.2.2.4. ESTIMAÇÃO SEQUENCIAL DE DUAS ETAPAS | 38 |
| 2.2.2.5. COMPARAÇÃO DAS PROPRIEDADES E DOS RESULTADOS OBTIDOS COM OS MÉTODOS DE ESTIMAÇÃO EM QUE SÃO OBSERVADAS TODAS AS DIMENSÕES DA IMIGRAÇÃO E A REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$ | 40 |
| 2.3. APRECIÇÃO CRÍTICA DOS MÉTODOS..... | 42 |

| | |
|--------------------------------------------------------------------------------------------------------------------------|----|
| 3. PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO | 44 |
| 3.1. DEFINIÇÕES E PROPRIEDADES..... | 44 |
| 3.2. ESTIMAÇÃO DOS PARÂMETROS DOS PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO GENERALIZADOS E NÃO GENERALIZADOS | 52 |
| 3.2.1. ESTIMAÇÃO NO DOMÍNIO DO TEMPO..... | 52 |
| 3.2.1.1. MÉTODO DE YULE-WALKER..... | 52 |
| 3.2.1.2. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS | 55 |
| 3.2.1.3. MÉTODO DE MÁXIMA VEROSIMILHANÇA CONDICIONAL | 58 |
| 3.2.2. ESTIMAÇÃO NO DOMÍNIO DA FREQUÊNCIA - CRITÉRIO DE WHITTLE | 61 |
| 3.3. COMPARAÇÃO DOS MÉTODOS DE ESTIMAÇÃO | 62 |
| 4. RELAÇÃO ENTRE O PROCESSO BGWI E O PROCESSO GINAR..... | 66 |
| 4.1. COMPARAÇÃO E APLICAÇÃO DE MÉTODOS E RESULTADOS..... | 69 |
| 4.2. APLICAÇÃO DO PROCESSO RAMIFICADO COM IMIGRAÇÃO A UMA SÉRIE DE OBSERVAÇÕES DE CONTAGEM..... | 74 |
| 5. CONCLUSÕES | 80 |
| ANEXO | 81 |
| REFERÊNCIAS..... | 87 |

ABREVIATURAS

| | |
|--------|---------------------------------------------------------------------------------------------|
| AICC | “Akaike information corrected criterion” |
| AR | autorregressivo |
| ARMA | autorregressivo e de médias móveis (“autoregressive and moving average”) |
| BGWI | processo ramificado com imigração (“Bienaymé-Galton-Watson process with immigration ”) |
| DARMA | autorregressivo e de médias móveis discreto (“discrete autoregressive and moving average”) |
| et al. | e outros (“et alii”) |
| FAC | função autocorrelação |
| FACP | função autocorrelação parcial |
| GINAR | autorregressivo de valor inteiro generalizado (“generalized integer-valued autoregressive”) |
| i.i.d. | independentes e identicamente distribuídas |
| INAR | autorregressivo de valor inteiro (“integer-valued autoregressive”) |

1. INTRODUÇÃO

Para o estudo de dados de contagem dependentes, que ocorrem com alguma frequência na forma de séries temporais (por exemplo o número de hóspedes num hotel ou o número de acidentes durante um determinado período de tempo), têm sido propostos, na literatura especializada, vários modelos semelhantes aos modelos autorregressivos e de médias móveis, ARMA, Gaussianos para valores contínuos. Alguns desses modelos, entre os quais os modelos autorregressivos e de médias móveis para valores discretos, DARMA, de Jacobs e Lewis (1978a, b, 1983), que se baseiam numa mistura probabilística de variáveis aleatórias, têm uma estrutura complicada e por isso raramente são aplicados na prática. Outros modelos, como os modelos de regressão de Markov, desenvolvidos por Zeger e Qaqish (1988), ou os modelos “parameter-driven” de Cox (1981), podem ser construídos como extensões de modelos lineares generalizados, apresentados por Nelder e Wedderburn (1972). Entre todos esses modelos destacam-se aqui os modelos autorregressivos de valor inteiro, INAR, e os modelos autorregressivos de valor inteiro generalizados, GINAR, que representam uma generalização dos anteriores. Estes modelos têm sido estudados por vários autores como: Al-Osh e Alzaid (1987), Du e Li (1991) e Gauthier e Latour (1992). Os processos INAR com distribuição marginal de Poisson – a distribuição que mais se aplica a dados discretos – parecem ser os modelos que, na prática, têm tido maiores êxitos em aplicações, visto possuírem propriedades semelhantes aos modelos autorregressivos, AR. Para tratar o problema de estimação dos parâmetros dos modelos INAR têm sido apresentados métodos de estimação no domínio do tempo: o método de Yule-Walker, o método de mínimos quadrados condicionais e o método de máxima verosimilhança condicional; e métodos de estimação no domínio da frequência: critério de Whittle. A estimação dos parâmetros nos modelos INAR é dificultada pelo facto de a distribuição marginal do processo, condicionada por observações passadas, ser uma convolução de duas distribuições.

Dion, Gauthier e Latour (1995) provaram que os modelos GINAR(p), e, em particular, os modelos INAR(p), podem ser identificados como processos ramificados multivariados com imigração, também designados por BGWI(p), no caso de p ser igual a 1, como processos ramificados com imigração – BGWI(1). Por essa razão é possível que se encontrem métodos mais eficazes, nos quais se baseia a realização de inferências sobre os processos BGWI, do que os já existentes para os parâmetros dos processos INAR e que possam ser aplicados a estes processos.

Os processos ramificados com imigração aparecem em aplicações no início do século XX. Seneta (1969) fala de aplicações em genética, em neurofisiologia e em “traffic flow theory”

(teoria da circulação do trânsito). Também são conhecidas aplicações em física estatística. O processo ramificado com imigração foi desenvolvido já em 1916 por Smoluchowski. O processo foi usado como modelo para a flutuação do número de partículas num pequeno elemento de volume v , geometricamente bem definido, dentro de um volume muito maior de solução, contendo partículas em movimentação aleatória, estando o sistema em equilíbrio e tendo as observações em v sido feitas, separadamente, em intervalos de tempo constantes. Westgren (1916) aplicou um processo ramificado com imigração às contagens de partículas de ouro num volume fixo de uma solução coloidal feitas em intervalos de tempo de 1.39 segundos. As partículas de ouro na solução são submetidas ao movimento Browniano. A teoria do movimento Browniano foi primeiro apresentada por Einstein (1905) e depois foi desenvolvida por Smoluchowski (1916). A experiência de Westgren confirma a teoria de Einstein-Smoluchowski. Kedem e Chiu (1987) consideraram o problema da modelação da intensidade da chuva usando um processo ramificado com imigração.

Os métodos de estimação conhecidos para o processo ramificado com imigração podem ser divididos em dois casos. No primeiro caso, os métodos baseiam-se na observação de uma única realização do processo. A este caso pertencem o método baseado na comparação do processo ramificado com imigração com o processo autorregressivo de ordem 1, o método de mínimos quadrados condicionais e o método de mínimos quadrados condicionais pesados. O outro caso da estimação baseia-se na observação da realização e da imigração do processo. Os métodos deste caso são: método de máxima verosimilhança, método de estimação sequencial, método de estimação de precisão fixa e método de estimação sequencial de duas etapas.

O objectivo deste trabalho é confrontar os métodos de estimação que têm sido introduzidos na literatura para os processos BGWI e INAR a fim de verificar se alguns métodos ou resultados obtidos para os processos BGWI podem ser aplicados ou adaptados aos processos INAR contribuindo assim para a extensão dos conhecimentos até agora adquiridos para estes processos.

No capítulo 2 apresentam-se os processos ramificados com imigração, BGWI, e os métodos de estimação até agora utilizados para os processos BGWI(1). Por último é feita uma apreciação crítica dos métodos.

No capítulo 3 são definidos os processos autorregressivos de valor inteiro generalizados e não generalizados e expostos os métodos de estimação descobertos para os processos GINAR(p), INAR(p) e INAR(1).

No capítulo 4 expõe-se a relação que existe entre os processos BGWI e os processos GINAR. Os métodos estudados para os processos BGWI(1) e INAR(1) são comparados e faz-se a aplicação dos métodos, dos resultados e das propriedades do processo BGWI(1)

aos processos GINAR(1) e INAR(1). Também se apresenta um exemplo de aplicação do processo ramificado com imigração a uma série de observações de contagem.

Por fim, no último capítulo 5 propõem-se algumas conclusões sobre este trabalho.

2. PROCESSOS RAMIFICADOS COM IMIGRAÇÃO

2.1. DEFINIÇÕES E PROPRIEDADES

Os processos ramificados¹ podem ser usados como modelos para descrever o crescimento e a dinâmica de populações. Estes processos possuem a seguinte estrutura: Seja X_0 a dimensão de uma população inicial e suponha-se que cada elemento dessa população produz k novos elementos com probabilidade

$$p_k, \text{ onde } k = 0, 1, 2, \dots, p_k > 0 \text{ e } \sum_{k=0}^{\infty} p_k = 1 \quad (2.1)$$

A primeira geração de dimensão X_1 é constituída pelos descendentes da população inicial. Os elementos da primeira geração são independentes e têm descendentes de acordo com a lei de probabilidade (2.1). Em geral, a n -ésima geração é constituída pelos descendentes da $(n-1)$ -ésima geração, que são independentes, produzindo cada um deles k novos elementos com probabilidade $p_k, k=0, 1, \dots$. A dimensão da n -ésima geração representa-se por X_n . A sequência de variáveis aleatórias inteiras assim construída, $\{X_n\}_{n=0,1,\dots}$, forma o processo ramificado, que tem a propriedade de ser uma cadeia de Markov. Considerando a distribuição de $X_{n_1}, X_{n_2}, \dots, X_{n_r}, X_n$, onde $n_1 < n_2 < \dots < n_r < n$, a informação importante é o número de elementos da última população, porque o número de descendentes é uma função que só depende da dimensão da população actual.

DEFINIÇÃO DO PROCESSO RAMIFICADO

O processo ramificado $\{X_n\}_{n=0,1,\dots}$ define-se através da expressão

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i}, \quad n \geq 1, \quad (2.2)$$

onde $Y_{n-1,i}$ representa o número de elementos gerado pelo i -ésimo elemento da geração $(n-1)$ e X_0 é uma variável aleatória inteira não negativa. $\{Y_{n-1,i}\}$ são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) de valores inteiros com distribuição

$$P(Y_{n-1,i} = k) = p_k, \quad k = 0, 1, \dots \text{ e } \sum_{k=0}^{\infty} p_k = 1. \quad (2.3)$$

¹ Algumas definições e alguns resultados aqui apresentados baseiam-se no estudo do 8º capítulo do livro de Karlin, S. e Taylor, H. (1975), do 5º capítulo do livro de Guttorp, P. (1991) e do artigo de Dion, J., Gauthier, G. e Latour, A. (1995).

Um dos exemplos de um processo ramificado é a fissão nuclear. Quando um núcleo colide com um neutrão, o núcleo divide-se em várias partículas dando origem a um número aleatório de novos neutrões. Estes por sua vez produzem um número aleatório de neutrões adicionais quando colidem com outros núcleos, etc.. Neste caso, o número inicial de neutrões é $X_0 = 1$. A primeira geração é constituída pelos neutrões que resultaram da colisão do neutrão inicial com um núcleo. A dimensão desta geração designa-se por X_1 . Em geral, a população X_n , ou seja, a n -ésima geração, é constituída pelos neutrões que se formaram a partir das colisões dos X_{n-1} neutrões da $(n-1)$ -ésima geração.

DEFINIÇÃO DO PROCESSO RAMIFICADO COM IMIGRAÇÃO

Adicionando ao processo ramificado (2.2) uma componente de imigração $\{I_n\}_{n=1,2,\dots}$ de variáveis aleatórias inteiras i.i.d. não negativas, independente de $\{X_n\}$, obtém-se o processo ramificado com imigração, ou BGWI(1) - "Bienaymé-Galton-Watson process with Immigration", que se representa por

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i} + I_n, \quad n \geq 1. \quad (2.4)$$

Neste caso cada geração do processo ramificado é ampliada por um número aleatório de componentes de imigração que não se distinguem dos outros elementos da população.

DEFINIÇÃO DO PROCESSO RAMIFICADO MULTIVARIADO

Nos processos ramificados ainda se pode considerar o facto de os elementos da população possuírem características diferentes e por essa razão dividir os elementos em grupos de várias espécies. Seja então $\vec{X}_n = (X_n(1), \dots, X_n(p))^T$ o vector² que contém as contagens de elementos de p espécies diferentes da geração n da população ($X_n(1)$ é o número de elementos da espécie 1, etc.). Designe-se este novo processo "multitype branching process" por processo ramificado multivariado. Uma maneira de o representar é a seguinte:

$$\vec{X}_n = (X_n(1), \dots, X_n(p))^T = \sum_{i=1}^{X_{n-1}(1)} \vec{Y}_{n-1,i}^1 + \dots + \sum_{i=1}^{X_{n-1}(p)} \vec{Y}_{n-1,i}^p = \sum_{k=1}^p \sum_{i=1}^{X_{n-1}(k)} \vec{Y}_{n-1,i}^k, \quad (2.5)$$

onde $\vec{Y}_{n-1,i}^k$ é um vector que contém as contagens dos elementos gerados pelo i -ésimo elemento da espécie k da geração $(n-1)$. Tendo em conta que cada elemento pode produzir elementos de p espécies diferentes, então, representando esse vector por $\vec{Y}_{n-1,i}^k = (Y_{n-1,i}^k(1), \dots, Y_{n-1,i}^k(p))^T$, onde $Y_{n-1,i}^k(1)$ é o número de elementos da espécie 1 produzido

² Por dificuldades técnicas, a notação usada para representar um vector é uma seta e não a letra "bold".

pelo i -ésimo elemento da espécie k da $(n-1)$ -ésima geração, etc., o processo (2.5) escreve-se da forma

$$\vec{X}_n = (X_n(1), \dots, X_n(p))^T = \sum_{i=1}^{X_{n-1}(1)} (Y_{n-1,i}^1(1), \dots, Y_{n-1,i}^1(p))^T + \dots + \sum_{i=1}^{X_{n-1}(p)} (Y_{n-1,i}^p(1), \dots, Y_{n-1,i}^p(p))^T,$$

onde por exemplo o número de elementos da espécie p da geração n é dado por

$$X_n(p) = \sum_{i=1}^{X_{n-1}(1)} Y_{n-1,i}^1(p) + \dots + \sum_{i=1}^{X_{n-1}(p)} Y_{n-1,i}^p(p).$$

DEFINIÇÃO DO PROCESSO RAMIFICADO MULTIVARIADO COM IMIGRAÇÃO

O processo ramificado multivariado com imigração, também designado por BGWI(p), é obtido adicionando ao processo ramificado multivariado $\{\vec{X}_n\} = \{(X_n(1), \dots, X_n(p))^T\}$, que consiste de p tipos de elementos, a imigração $\{\vec{I}_n\}$ independente de $\{\vec{X}_n\}$. Neste modelo todos os elementos são independentes, vivem uma unidade de tempo e produzem novos elementos de p tipos diferentes. Cada elemento da espécie k , $k=1, \dots, p$ dá origem a x_1 elementos da espécie 1, ..., x_p elementos da espécie p , de acordo com a distribuição

$$p^k(\vec{x}) = p^k((x_1, \dots, x_p)^T), \text{ onde } k=1, \dots, p \text{ e } \sum_{\vec{x}} p^k(\vec{x}) = 1. \quad (2.6)$$

Assim, $\{\vec{X}_n\}_{n \geq 1}$ é um processo BGWI(p), se verificar

$$\begin{aligned} \vec{X}_n &= \sum_{k=1}^p \sum_{i=1}^{X_{n-1}(k)} \vec{Y}_{n-1,i}^k + \vec{I}_n, \text{ se } \vec{X}_{n-1} \neq \vec{0} \\ \vec{X}_n &= \vec{I}_n, \text{ se } \vec{X}_{n-1} = \vec{0}, \end{aligned} \quad (2.7)$$

onde:

- $\{\vec{I}_n\}$ é um processo de imigração de vectores aleatórios i.i.d. com valores em \mathbb{N}^p ,
- $\{\vec{Y}_{n-1,i}^k\}$, $k=1, \dots, p$, são sequências independentes de vectores aleatórios i.i.d., independentes de $\{\vec{I}_n\}$, com valores em \mathbb{N}^p e com distribuição dada por (2.6),
- as duas igualdades são válidas em distribuição,
- o processo começa com o primeiro \vec{I}_n não nulo e $\vec{X}_0 = \vec{I}_0$.

Para o processo ramificado com imigração, que é um caso particular do processo BGWI(p) em que só existe uma espécie de elementos ($p=1$), usa-se a notação simplificada, já apresentada em (2.4).

A condição de estacionaridade destes processos vai ser referida no capítulo 4.

2.2. ESTIMAÇÃO DOS PARÂMETROS DO PROCESSO RAMIFICADO COM IMIGRAÇÃO

Considere-se o processo ramificado com imigração

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i} + I_n, \quad n \geq 1, \quad (2.8)$$

onde:

- X_0 é uma variável aleatória inteira não negativa,
- $\{Y_{n-1,i}, i = 1, \dots, n, n = 1, 2, \dots\}$ e $\{I_n, n = 1, 2, \dots\}$ são famílias independentes de variáveis aleatórias i.i.d. não negativas.

Sejam $m = E(Y_{n-1,i})$, $\sigma^2 = V(Y_{n-1,i})$ - média e variância da descendência - e $\lambda = E(I_n)$, $b^2 = V(I_n)$ - média e variância da imigração - os parâmetros que se pretende estimar e suponha-se que estes sejam positivos e finitos.

Alguns métodos propostos na literatura, que foram apresentados por Winnicki (1988) para resolver o problema da estimação, baseiam-se na observação de uma única realização do processo $\{X_0, \dots, X_n\}$ e os resultados referem-se a amostras grandes. Estes métodos, incluindo o método baseado na comparação do processo BGWI(1) com o processo autorregressivo de ordem 1, AR(1), o método de mínimos quadrados condicionais e o método de mínimos quadrados condicionais pesados, são apresentados na secção 2.2.1.. Um outro caso de estimação, apresentado na secção 2.2.2., parte do princípio que, para além da observação da realização do processo, também são observadas todas as dimensões da imigração $\{I_1, \dots, I_n\}$. Os métodos deste caso são: método de máxima verosimilhança, método de estimação sequencial, método de estimação de precisão fixa e método de estimação sequencial de duas etapas.

O problema da inferência estatística é dificultado pelo facto de as propriedades do processo dependerem da escala dos parâmetros desconhecidos, principalmente do parâmetro m . Heathcote (1965, 1966) e, mais tarde, Quine (1970) formularam as seguintes propriedades para o caso subcrítico $m < 1$.

Teorema 2.1: Se $m < 1$, então $\{X_n\}$ é uma cadeia de Markov irreduzível aperiódica, positiva recorrente. Neste caso, $X_n \xrightarrow{d} X$, onde X é uma variável aleatória com suporte $\{k : P(X_1 = k) > 0\}$.

O resultado para o caso “supercrítico”, $m > 1$, que é exposto a seguir, foi obtido por Seneta (1970b).

Teorema 2.2: Se $m > 1$, então $X_n/m^n \xrightarrow{q.c.} V$, onde V é uma variável aleatória com distribuição contínua concentrada em $(0, \infty)$.

As propriedades do processo para o caso crítico, $m=1$, que é o mais complicado, foram estudadas por Pakes (1971,1972) e por Wei e Winnicki (1989) e estão apresentadas no seguinte teorema.

Teorema 2.3: Seja $m=1$ e $\tau = 2\lambda/\sigma^2$, então $\{X_n\}$ é uma cadeia de Markov irreduzível, aperiódica e

- a) se $\tau < 1$, então $\{X_n\}$ é recorrente nulo,
- b) se $\tau > 1$, então $\{X_n\}$ é transiente,
- c) se $\tau = 1$, e $EY_{n-1,i}^2 \log_+ Y_{n-1,i} < \infty$, então $\{X_n\}$ é recorrente nulo.

Assim, quando $m=1$, o processo pode ser transiente ou recorrente nulo e isso depende dos valores de σ^2 e λ . Também se sabe que, se $m=1$, então $X_n/n \xrightarrow{d} Y$, onde Y é uma variável aleatória da distribuição gama (Seneta (1970a), Pakes(1971)).

O estudo dos métodos do primeiro caso de estimação foi feito separadamente, considerando $m < 1$, $m = 1$ e $m > 1$. Os métodos do segundo caso caracterizam-se por terem sido descobertos resultados unificados para $m \leq 1$ e para $m > 0$.

Bercu (2001) aplicou a teoria de controlo adaptativo aos processos ramificados com imigração e conseguiu obter estimadores consistentes com o método de mínimos quadrados pesados para todos os parâmetros m , λ , σ^2 e b^2 do processo BGWI(1) controlado³, sem impor restrições ao domínio de m . A abordagem desenvolvida por este autor pode ter aplicações práticas em física estatística, controlando o número total de partículas contidas num pequeno volume, e em ciências de computadores, com a administração electrónica de documentos no ambiente da “Web”.

³ Ver ANEXO A1, pág. 81.

2.2.1. ESTIMAÇÃO BASEADA NA OBSERVAÇÃO DE UMA ÚNICA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$

2.2.1.1. ESTIMADORES BASEADOS NA COMPARAÇÃO DO PROCESSO RAMIFICADO COM IMIGRAÇÃO COM O PROCESSO AUTORREGRESSIVO DE ORDEM 1

Heyde e Seneta (1972, 1974) foram os primeiros autores que desenvolveram um método de estimação para m e λ supondo que $m < 1$. Os resultados estão expostos nos seguintes teoremas⁴.

Teorema 2.4: Seja $C_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$ e $D_n = \frac{\sum_{i=1}^n \hat{Y}_{i+1} \hat{Y}_i}{\sum_{i=1}^n \hat{Y}_i^2} = \frac{\sum_{i=1}^n (X_{i+1} - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$, onde

$\hat{Y}_i = X_i - C_n$. Então $C_n \xrightarrow{p} \mu \equiv \lambda(1-m)^{-1}$ e $D_n \xrightarrow{p} m$, quando $n \rightarrow \infty$, tal que $C_n(1-D_n)$ e D_n são estimadores consistentes para λ e m , respectivamente. Se a distribuição da descendência, a distribuição da imigração e a distribuição inicial tiverem momentos de terceira ordem finitos, então as convergências em probabilidade podem ser substituídas por convergências quase certas. Isto é, os estimadores respectivos são fortemente consistentes.

Teorema 2.5: Seja $c^2 = b^2 + \sigma^2 \mu$.

$$\lim_{n \rightarrow \infty} P(\alpha^{-1} \sqrt{n}(C_n - \mu) \leq x) = \Phi(x),$$

onde $\alpha^2 = (1-m)^{-2} c^2$, e $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-u^2/2} du$. Se, além disso, todas as distribuições da

descendência, da imigração e a distribuição inicial tiverem momentos de quarta ordem finitos, então

$$\lim_{n \rightarrow \infty} P(\beta^{-1} \sqrt{n}(D_n - m) \leq x) = \Phi(x),$$

onde⁵ $\beta = B(1-m^2)c^{-2}$,

$$B^2 = \gamma\sigma^2 + c^4(1-m^2)^{-1} \text{ e}$$

⁴ Ver Heyde e Seneta (1972: pp. 236-237).

⁵ A expressão para B^2 encontra-se em Heyde e Seneta (1974), que é uma correção da expressão apresentada no teorema que apareceu no artigo de 1972 dos mesmos autores.

$$\gamma = (1 - m^3)^{-1} \left[E(I - \lambda)^3 + \mu E(Y - m)^3 + 3m\sigma^2 c^2 (1 - m^2)^{-1} \right].$$

I e Y são variáveis aleatórias, tendo, respectivamente, as distribuições dos mecanismos da imigração e da descendência.

A natureza dos estimadores e dos resultados pode ser entendida através da comparação com um contexto mais simples de séries temporais.

Note-se que o processo $\{X_n\}$ verifica $E(X_n | X_{n-1}) = mX_{n-1} + \lambda$, $n=1,2,\dots$. Seja agora $Y_n = X_n - \mu$, onde $\mu = \lambda(1-m)^{-1}$ é a média da distribuição estacionária no limite de $\{X_n\}$; então tem-se que $E(Y_n | Y_{n-1}) = mY_{n-1}$ e assim, $Y_n = mY_{n-1} + \delta_{n-1}$, $n=1,2,\dots$, onde $E(\delta_{n-1}) = 0$ e o resíduo δ_{n-1} é não correlacionado com Y_{n-1} . A última relação é semelhante à definição da relação de um processo autorregressivo de ordem 1 (AR(1)) estacionário de segunda ordem:

$$x_n = \rho x_{n-1} + \varepsilon_{n-1}, \quad n = 0, \pm 1, \pm 2, \dots \quad (|\rho| < 1), \quad (2.9)$$

onde $\{\varepsilon_{n-1}\}$ é um ruído branco - uma sequência de variáveis aleatórias i.i.d. normais de média zero. Assim, no processo ramificado com imigração, m corresponde aproximadamente a ρ (ρ é a primeira autocorrelação do processo AR(1)) e os estimadores C_n e D_n têm uma forma análoga aos estimadores de máxima verosimilhança de amostras grandes para as quantidades correspondentes em (2.9), sendo C_n a média amostral e D_n o coeficiente de correlação.

Resumindo, Heyde e Seneta propuseram estimar μ por $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ e m por

$$\hat{m} = \frac{\sum_{k=1}^n (X_{k+1} - \bar{X})(X_k - \bar{X})}{\sum_{k=1}^n (X_k - \bar{X})^2} \quad \text{e mostraram que estes estimadores são consistentes e}$$

assimptoticamente normais. A partir de $\hat{\mu}$ e \hat{m} eles obtiveram um estimador consistente

$$\text{para } \lambda: \hat{\lambda} = \bar{X} \left(1 - \frac{\sum_{k=1}^n (X_{k+1} - \bar{X})(X_k - \bar{X})}{\sum_{k=1}^n (X_k - \bar{X})^2} \right) = \frac{\bar{X} \sum_{k=1}^n (X_k - \bar{X})^2 - \bar{X} \sum_{k=1}^n (X_{k+1} - \bar{X})(X_k - \bar{X})}{\sum_{k=1}^n (X_k - \bar{X})^2}.$$

Atendendo ao caso $m > 1$, Heyde (1970) e Seneta (1970b) afirmaram que m pode ser estimado consistentemente por X_n / X_{n-1} ou por $\sum_{i=1}^n X_i / \sum_{i=1}^n X_{i-1}$. O teorema de Heyde (1975: pp. 49-50) apresenta as distribuições no limite destes estimadores:

Teorema 2.6: Se $m > 1$, então

$$\sqrt{\sum_{k=1}^n X_{k-1}} \left(\sum_{k=1}^n X_k / \sum_{k=1}^n X_{k-1} - m \right) \xrightarrow{d} N(0, \sigma^2)$$

e

$$\sqrt{\sum_{k=1}^n X_{k-1}} (X_n / X_{n-1} - m) \xrightarrow{d} N\left(0, \frac{m}{m-1} \sigma^2\right).$$

2.2.1.2. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS

Klimko e Nelson (1978) desenvolveram um método de mínimos quadrados condicionais para processos estocásticos e aplicaram-no a um processo ramificado com imigração.

Seja $\{Y_n\}$ um processo estocástico, cuja distribuição depende do parâmetro $\alpha = (\alpha_1, \dots, \alpha_p)$, e seja $F_n = \sigma(Y_0, \dots, Y_n)$. Suponha-se que $E(Y_n | F_{n-1}) = Q_n(\alpha)$. O estimador de mínimos quadrados condicionais de α , $\hat{\alpha}$, minimiza $\sum_{k=1}^n (Y_k - Q_k(\alpha))^2 = \sum_{k=1}^n (Y_k - E(Y_k | F_{k-1}))^2$.

Sendo $Y_n = X_n$ o processo ramificado com imigração, então $E(X_k | F_{k-1}) = mX_{k-1} + \lambda$.

Derivando $\sum_{k=1}^n (X_k - mX_{k-1} - \lambda)^2$ em ordem a m e λ e resolvendo depois as equações de

mínimos quadrados, obtêm-se os seguintes estimadores para m e λ :

$$\hat{m}_n = \frac{n \sum_{k=1}^n X_k X_{k-1} - \sum_{k=1}^n X_k \sum_{k=1}^n X_{k-1}}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1} \right)^2}$$

e

$$\hat{\lambda}_n = \frac{\sum_{k=1}^n X_k \sum_{k=1}^n X_{k-1}^2 - \sum_{k=1}^n X_{k-1} \sum_{k=1}^n X_k X_{k-1}}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1} \right)^2}.$$

As propriedades destes estimadores estão resumidas no próximo teorema⁶ - a parte a) foi provada por Klimko e Nelson (1978) e a parte b) por Wei e Winnicki (1987).

Teorema 2.7:

a) Se $m < 1$, então $\hat{m}_n \xrightarrow{q.c.} m$ e $\hat{\lambda}_n \xrightarrow{q.c.} \lambda$. Se, além disso, $EY_{n-1,i}^3 < \infty$ e $EI_n^3 < \infty$,

então $\sqrt{n}(\hat{m}_n - m, \hat{\lambda}_n - \lambda)^T \xrightarrow{d} N(0, \Sigma)$, onde

$$\Sigma_{11} = \sigma^2 (\mu\gamma + d) (1 - m^2)^2 (\mu\sigma^2 + b^2)^{-2} (1 - m^3)^{-1} + 1 - m^2 + 3m\sigma^4 (1 - m^2) (\mu\sigma^2 + b^2)^{-1} (1 - m^3)^{-1},$$

$$\Sigma_{22} = \mu^2 \sigma^2 (\mu\gamma + d) (1 - m^2)^2 (\mu\sigma^2 + b^2)^{-2} (1 - m^3)^{-1} + \mu^2 (1 - m^2) + b^2 - \mu\sigma^2 + 3m\mu^2 \sigma^4 (1 - m^2) (\mu\sigma^2 + b^2)^{-1} (1 - m^3)^{-1},$$

$$\Sigma_{12} = \sigma^2 - \mu(1 - m^2) - \mu\sigma^2 (\mu\gamma + d) (1 - m^2)^2 (\mu\sigma^2 + b^2)^{-2} (1 - m^3)^{-1} - 3m\mu\sigma^4 (1 - m^2) (\mu\sigma^2 + b^2)^{-1} (1 - m^3)^{-1},$$

$$\gamma = E(Y_{n-1,i} - m)^3 \text{ e } d = E(I_n - \lambda)^3.$$

b) Se $m > 1$, então $\hat{m}_n \xrightarrow{q.c.} m$, $|\hat{\lambda}_n| \xrightarrow{p} \infty$ e

$$\sqrt{\sum_{k=1}^n X_{k-1}} (\hat{m}_n - m) \xrightarrow{d} N\left(0, \frac{m^2 + 2m + 1}{m^2 + m + 1} \sigma^2\right).$$

Wei e Winnicki (1987) apresentaram um teorema com as propriedades dos estimadores para o caso crítico:

Teorema 2.8: Se $m=1$, então $\hat{m}_n \xrightarrow{p} m$,

$$\hat{\lambda}_n \xrightarrow{d} \frac{Y(1) \int_0^1 Y^2(t) dt - \frac{1}{2} \left[Y^2(1) \int_0^1 Y(t) dt - \sigma^2 \left(\int_0^1 Y(t) dt \right)^2 \right]}{\int_0^1 Y^2(t) dt - \left(\int_0^1 Y(t) dt \right)^2}$$

e

⁶ Cf. Winnicki (1988: pp. 305-306).

$$\left(\sum_{k=1}^n (X_{k-1} + 1) \right)^{1/2} (\hat{m}_n - m) \xrightarrow{d} \frac{\frac{1}{2} Y^2(1) \left(\int_0^1 Y(t) dt \right)^{1/2} - \left(Y(1) + \frac{1}{2} \sigma^2 \right) \left(\int_0^1 Y(t) dt \right)^{3/2}}{\int_0^1 Y^2(t) dt - \left(\int_0^1 Y(t) dt \right)^2}.$$

Neste teorema, Y representa um processo de difusão⁷.

Yanev e Tchoukova-Dantcheva (1980) estudaram o problema da estimação das variâncias σ^2 e b^2 e propuseram os seguintes estimadores para o caso $m < 1$:

$$\hat{\sigma}_n^2 = \frac{n \sum_{k=1}^n X_k \hat{U}_k^2 - \sum_{k=1}^n X_{k-1} \sum_{k=1}^n \hat{U}_k^2}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1} \right)^2},$$

$$\hat{b}_n^2 = \frac{\sum_{k=1}^n \hat{U}_k^2 \sum_{k=1}^n X_{k-1}^2 - \sum_{k=1}^n X_{k-1} \sum_{k=1}^n X_{k-1} \hat{U}_k^2}{n \sum_{k=1}^n X_{k-1}^2 - \left(\sum_{k=1}^n X_{k-1} \right)^2},$$

onde $\hat{U}_k = X_k - \hat{m}_n X_{k-1} - \hat{\lambda}_n$.

$\hat{\sigma}_n^2$ e \hat{b}_n^2 são estimadores de mínimos quadrados condicionais no sentido de minimizarem

$$\sum_{k=1}^n \left(U_k^2 - E(U_k^2 | F_{k-1}) \right)^2, \text{ onde } U_k = X_k - m X_{k-1} - \lambda.$$

O seguinte teorema, provado por Winnicki (1987) contém as propriedades destes estimadores para os casos não críticos e a distribuição assintótica conjunta de $\hat{\sigma}_n^2$ e \hat{b}_n^2 para $m < 1$.

Teorema 2.9:

a) Se $m < 1$, então $\hat{\sigma}_n^2 \xrightarrow{q.c.} \sigma^2$ e $\hat{b}_n^2 \xrightarrow{q.c.} b^2$. Se além disso $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$, então

$$\sqrt{n} \left(\hat{\sigma}_n^2 - \sigma^2, \hat{b}_n^2 - b^2 \right) \xrightarrow{d} N \left(\vec{0}, V^{-1} \Xi (V^T)^{-1} \right),$$

onde

⁷ Ver ANEXO A2, pág. 81.

$$\Xi = \begin{pmatrix} E(R(X)X^2) & E(R(X)X) \\ E(R(X)X) & E(R(X)) \end{pmatrix}$$

e

$$R(X) = 2\sigma^4 X^2 + (a^4 + 4\sigma^2 b^2 - 3\sigma^4)X + c^4 - b^4,$$

$$a^4 = E(Y_{n-1,i} - m)^4, \quad c^4 = E(I_n - \lambda)^4.$$

b) Se $m > 1$, então

$$\hat{\sigma}_n^2 \xrightarrow{d} \frac{m^2 - 1}{m^2} \sum_{k=0}^{\infty} m^{-2k} \xi_k^2 - 2 \frac{(m^2 - 1)^2}{m^4} \sum_{k=0}^{\infty} m^{-3/2k} \xi_k \sum_{k=0}^{\infty} m^{-5/2k} \xi_k$$

$$+ \frac{(m^2 - 1)^3}{m^3(m^3 - 1)} \left(\sum_{k=0}^{\infty} m^{-3/2k} \xi_k \right)^2,$$

onde $\{\xi_k\}$ são variáveis aleatórias i.i.d. da distribuição $N(0, \sigma^2)$ e

$$|\hat{b}_n^2| \xrightarrow{p} \infty.$$

Em relação ao caso crítico existe o seguinte teorema de Winnicki (1987):

Teorema 2.10: Se $m=1$, $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$, então $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$.

Os últimos teoremas mostram que, com a ajuda do método de mínimos quadrados condicionais, se encontra um estimador consistente para a média da descendência, \hat{m}_n , para todos os casos de m (para os casos não críticos verifica-se a consistência em sentido forte). O estimador para a média da imigração, $\hat{\lambda}_n$, só é um estimador consistente (em sentido forte) para o caso subcrítico; para $m \geq 1$ já não é consistente.

O estimador obtido pelo método dos mínimos quadrados condicionais para a variância da descendência, $\hat{\sigma}_n^2$, é consistente para $m \leq 1$ (para $m < 1$ é consistente em sentido forte).

Considerando o caso "supercrítico", $\hat{\sigma}_n^2$ já não é consistente. O estimador para a variância da imigração, \hat{b}_n^2 , só é consistente (em sentido forte) para o caso subcrítico; para $m \geq 1$ não é consistente.

2.2.1.3. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS PESADOS

Com o objectivo de melhorar a consistência e a variância assintótica de alguns estimadores, Wei e Winnicki (1987) propuseram empregar o método de mínimos quadrados condicionais pesados. Os estimadores de mínimos quadrados condicionais pesados para as médias são obtidos minimizando

$$\sum_{k=1}^n \left(\frac{X_k - E(X_k | F_{k-1})}{\sqrt{\text{Var}(X_k | F_{k-1})}} \right)^2.$$

A seguir explica-se por que motivo foi desenvolvido este método. Reescrevendo o processo ramificado com imigração

$$X_k = mX_{k-1} + \lambda + \varepsilon_k, \quad k=1,2,\dots, \quad (2.10)$$

então o erro $\varepsilon_k = X_k - E(X_k | F_{k-1})$ desta equação, vista como modelo de regressão estocástica, tem variância $E(\varepsilon_k^2 | F_{k-1}) = \sigma^2 X_{k-1} + b^2$. Com a intenção de melhorar os estimadores de mínimos quadrados no sentido de minimizar $E(\varepsilon_k^2 | F_{k-1})$, para os casos em que $X_k \rightarrow \infty$, o termo do erro $\varepsilon_k = X_k - E(X_k | F_{k-1})$ é normalizado. Wei e Winnicki normalizaram ε_k por $\sqrt{X_{k-1} + 1}$ e não por $\sqrt{\text{Var}(X_k | F_{k-1})} = \sqrt{\sigma^2 X_{k-1} + b^2}$ para evitar os parâmetros σ^2 e b^2 , em geral desconhecidos, na equação transformada, que, partindo de (2.10) dividindo por $\sqrt{X_{k-1} + 1}$ e somando e subtraindo $\frac{m}{\sqrt{X_{k-1} + 1}}$, fica com o seguinte

aspecto:

$$\frac{X_k}{\sqrt{X_{k-1} + 1}} = m\sqrt{X_{k-1} + 1} + (\lambda - m) \frac{1}{\sqrt{X_{k-1} + 1}} + \delta_k, \quad (2.11)$$

onde $\delta_k = \frac{\varepsilon_k}{\sqrt{X_{k-1} + 1}}$. A equação (2.11) é uma equação de regressão estocástica em que o

erro já tem variância condicional limitada $\frac{\sigma^2 X_{k-1} + b^2}{X_{k-1} + 1}$.

Heyde e Seneta (1971) afirmaram que para $m \geq 1$, $\delta_n \xrightarrow{d} N(0, \sigma^2)$. Assim, neste caso, os estimadores de mínimos quadrados condicionais pesados seriam, aproximadamente, estimadores de máxima verosimilhança.

Os estimadores de mínimos quadrados condicionais pesados obtidos a partir da equação (2.11) são da forma:

$$\tilde{m}_n = \frac{\sum_{k=1}^n X_k \sum_{k=1}^n \frac{1}{X_{k-1} + 1} - n \sum_{k=1}^n \frac{X_k}{X_{k-1} + 1}}{\sum_{k=1}^n (X_{k-1} + 1) \sum_{k=1}^n \frac{1}{X_{k-1} + 1} - n^2}$$

e

$$\tilde{\lambda}_n = \frac{\sum_{k=1}^n X_{k-1} \sum_{k=1}^n \frac{X_k}{X_{k-1} + 1} - \sum_{k=1}^n X_k \sum_{k=1}^n \frac{X_{k-1}}{X_{k-1} + 1}}{\sum_{k=1}^n (X_{k-1} + 1) \sum_{k=1}^n \frac{1}{X_{k-1} + 1} - n^2}.$$

Para estes estimadores são válidos os seguintes teoremas (Wei e Winnicki (1987)).

Teorema 2.11:

a) Se $m < 1$, então $\tilde{m}_n \xrightarrow{q.c.} m$, $\tilde{\lambda}_n \xrightarrow{q.c.} \lambda$ e $\sqrt{n}(\tilde{m}_n - m, \tilde{\lambda}_n - \lambda) \xrightarrow{d} N(\vec{0}, \Phi)$, onde $\Phi = V^{-1}W(V^T)^{-1}$,

$$V = \begin{pmatrix} E(X) & 1 \\ E\left(\frac{X}{X+1}\right) & E\left(\frac{1}{X+1}\right) \end{pmatrix},$$

$$W = \begin{pmatrix} E(\sigma^2 X + b^2) & E\left(\frac{\sigma^2 X + b^2}{X+1}\right) \\ E\left(\frac{\sigma^2 X + b^2}{X+1}\right) & E\left(\frac{\sigma^2 X + b^2}{(X+1)^2}\right) \end{pmatrix}.$$

a) Se $m > 1$, então

$$\tilde{m}_n \xrightarrow{q.c.} m, \tilde{\lambda}_n \xrightarrow{q.c.} \lambda + \frac{\sum_{i=1}^{\infty} \frac{X_i - mX_{i-1} - \lambda}{X_{i-1} + 1}}{\sum_{i=1}^{\infty} \frac{1}{X_{i-1} + 1}}$$

e

$$\sqrt{\sum_{i=1}^n X_{i-1}} (\tilde{m}_n - m) \xrightarrow{d} N(0, \sigma^2).$$

Teorema 2.12: Se $m=1$, então $\tilde{m}_n \xrightarrow{p} m$, $\tilde{\lambda}_n \xrightarrow{p} \lambda$ e

$$\left(\sum_{k=1}^n (X_{k-1} + 1) \right)^{1/2} (\tilde{m}_n - m) \xrightarrow{d} \frac{Y(1) - \lambda}{\left(\int_0^1 Y(t) dt \right)^{1/2}}.$$

Se além disso $2\lambda > \sigma^2$, $EY_{n-1,i}^{2+\delta} < \infty$ e $EI_n^{2+\delta} < \infty$ para algum $\delta > 0$, então

$$\left(\sum_{k=1}^n \frac{1}{X_{k-1} + 1} \right)^{1/2} (\tilde{\lambda}_n - \lambda) \xrightarrow{d} N(0, \sigma^2).$$

Para melhorar as propriedades dos estimadores $\hat{\sigma}_n^2$ e \hat{b}_n^2 , Winnicki (1987) empregou o método de mínimos quadrados condicionais pesados da maneira que a seguir se apresenta. Uma equação análoga a (2.10), é

$$U_k^2 = \sigma^2 X_{k-1} + b^2 + \eta_k = E(\varepsilon_k^2 | F_{k-1}) + \eta_k$$

com $E(\eta_k | F_{k-1}) = 0$, $E(\eta_k^2 | F_{k-1}) = R(X_{k-1})$ e a equação análoga a (2.11) é

$$\frac{U_k^2}{X_{k-1} + 1} = \sigma^2 + (b^2 - \sigma^2) \frac{1}{X_{k-1} + 1} + \gamma_k, \quad (2.12)$$

onde $E(\gamma_k | F_{k-1}) = 0$, $E(\gamma_k^2 | F_{k-1}) = \frac{R(X_{k-1})}{(X_{k-1} + 1)^2} < R(1) < \infty$.

A partir de (2.12) obtêm-se os seguintes estimadores de mínimos quadrados condicionais pesados:

$$\tilde{\sigma}_n^2 = \frac{\sum_{k=1}^n \frac{1}{(X_{k-1} + 1)^2} \sum_{k=1}^n \frac{\tilde{U}_k^2}{X_{k-1} + 1} - \sum_{k=1}^n \frac{1}{X_{k-1} + 1} \sum_{k=1}^n \frac{\tilde{U}_k}{(X_{k-1} + 1)^2}}{n \sum_{k=1}^n \frac{1}{(X_{k-1} + 1)^2} - \left(\sum_{k=1}^n \frac{1}{X_{k-1} + 1} \right)^2}$$

e

$$\tilde{b}_n^2 = \frac{\sum_{k=1}^n \frac{X_{k-1}}{X_{k-1} + 1} \sum_{k=1}^n \frac{\tilde{U}_k^2}{(X_{k-1} + 1)^2} - \sum_{k=1}^n \frac{X_{k-1}}{(X_{k-1} + 1)^2} \sum_{k=1}^n \frac{\tilde{U}_k^2}{X_{k-1} + 1}}{n \sum_{k=1}^n \frac{1}{(X_{k-1} + 1)^2} - \left(\sum_{k=1}^n \frac{1}{X_{k-1} + 1} \right)^2},$$

onde $\tilde{U}_k = X_k - \tilde{m}_n X_{k-1} - \tilde{\lambda}_n$.

Os próximos dois teoremas, que contêm as propriedades assintóticas destes estimadores, encontram-se em Winnicki (1988).

Teorema 2.13: Suponha que $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$.

a) Se $m < 1$, então $\tilde{\sigma}_n^2 \xrightarrow{q.c.} \sigma^2$, $\tilde{b}_n^2 \xrightarrow{q.c.} b^2$ e

$$\sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2, \tilde{b}_n^2 - b^2) \xrightarrow{d} N(\vec{0}, M^{-1}\Phi(M^T)^{-1}),$$

onde

$$M = \begin{pmatrix} E\left(\frac{X}{X+1}\right) & E\left(\frac{1}{X+1}\right) \\ E\left(\frac{X}{(X+1)^2}\right) & E\left(\frac{1}{(X+1)^2}\right) \end{pmatrix}$$

e

$$\Phi = \begin{pmatrix} E\left(\frac{R(X)}{(X+1)^2}\right) & E\left(\frac{R(X)}{(X+1)^3}\right) \\ E\left(\frac{R(X)}{(X+1)^3}\right) & E\left(\frac{R(X)}{(X+1)^4}\right) \end{pmatrix}.$$

b) Se $m > 1$, então

$$\tilde{\sigma}_n^2 \xrightarrow{q.c.} \sigma^2,$$

$$\tilde{b}_n^2 \xrightarrow{q.c.} b^2 + \frac{\sum_{k=1}^{\infty} \frac{U_k^2 - \sigma^2 X_{k-1} - b^2}{(X_{k-1} + 1)^2}}{\sum_{k=1}^{\infty} \frac{1}{(X_{k-1} + 1)^2}} - 2 \frac{\sum_{k=1}^{\infty} \frac{U_k}{X_{k-1} + 1} \sum_{k=1}^{\infty} \frac{U_k}{(X_{k-1} + 1)^2}}{\sum_{k=1}^{\infty} \frac{1}{X_{k-1} + 1} \sum_{k=1}^{\infty} \frac{1}{(X_{k-1} + 1)^2}} + \left(\frac{\sum_{k=1}^{\infty} \frac{U_k}{X_{k-1} + 1}}{\sum_{k=1}^{\infty} \frac{1}{X_{k-1} + 1}} \right)^2$$

e

$$\sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4).$$

Teorema 2.14: Suponha que $m=1$, $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$, então

a) $\tilde{\sigma}_n^2 \xrightarrow{p} \sigma^2$ e $\sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$.

b) Se, além disso, $2\lambda \leq \sigma^2$, então $\tilde{b}_n^2 \xrightarrow{p} b^2$. Se, por outro lado, $2\lambda > \sigma^2$, então

$$\tilde{b}_n^2 \xrightarrow{d} b^2 + \frac{\sum_{k=1}^{\infty} \frac{U_k^2 - \sigma^2 X_{k-1} - b^2}{(X_{k-1} + 1)^2}}{\sum_{k=1}^{\infty} \frac{1}{(X_{k-1} + 1)^2}}.$$

Assim, os resultados dos últimos teoremas associados aos estimadores obtidos pelo método de mínimos quadrados condicionais pesados indicam que o estimador \tilde{m}_n da média da descendência é consistente para todos os casos de m (para os casos não críticos verifica-se a consistência em sentido forte). O estimador para a média da imigração $\tilde{\lambda}_n$ é consistente para $m \leq 1$ (sendo consistente em sentido forte para $m < 1$). $\tilde{\sigma}_n^2$ verifica a consistência em sentido forte para os casos não críticos, se $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$; para $m=1$ o estimador é consistente. E o estimador \tilde{b}_n^2 é consistente em sentido forte para o caso subcrítico; para o caso “supercrítico” não é consistente e para o caso crítico \tilde{b}_n^2 só é consistente se $2\lambda \leq \sigma^2$.

2.2.1.4. COMPARAÇÃO DAS PROPRIEDADES E DOS RESULTADOS OBTIDOS COM OS MÉTODOS DE ESTIMAÇÃO BASEADOS NA OBSERVAÇÃO DE UMA ÚNICA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$

A teoria de estimação de Heyde e Seneta, baseada na comparação do processo ramificado com imigração com o processo AR(1), só se aplica a processos ramificados com imigração que tenham uma média da descendência com valor inferior a 1. Os estimadores encontrados para a média da descendência e para a média da imigração são consistentes e assintoticamente normais. Para o caso “supercrítico” estes autores propuseram dois estimadores consistentes e assintoticamente normais para a média da descendência, mas diferentes dos estimadores para o caso subcrítico.

A estimação dos parâmetros do processo ramificado com imigração obtida pelo método de mínimos quadrados condicionais e pelo método de mínimos quadrados condicionais pesados é mais completa, porque, em primeiro lugar, são encontrados estimadores, não só para a média da descendência e para a média da imigração, mas também para as variâncias dos dois conjuntos; e, em segundo lugar, o estudo das propriedades de cada estimador é feito para os três casos: $m < 1$, $m = 1$, $m > 1$. Observa-se que muitas vezes as propriedades de um estimador variam muito com o domínio de m .

Note-se que, em todos estes métodos, para se poder aplicar a teoria de estimação, é preciso saber, onde se situa o valor de m . Sabendo que $m < 1$, então os estimadores de mínimos quadrados condicionais e os estimadores de mínimos quadrados condicionais pesados talvez sejam mais vantajosos por serem fortemente consistentes; os estimadores de Heyde e Seneta são consistentes e só são fortemente consistentes, se a distribuição da descendência, a distribuição da imigração e a distribuição inicial tiverem momentos de terceira ordem finitos. As distribuições assintóticas destes estimadores - C_n , \hat{m}_n e \tilde{m}_n - são normais, mas é difícil dizer, qual destes estimadores é assintoticamente mais eficiente, porque as expressões das variâncias assintóticas são muito complicadas para poderem ser comparadas.

Apesar de os estimadores para a média da descendência propostos por Heyde e Seneta para o caso “supercrítico” terem uma expressão simples, serem consistentes e terem uma variância assintótica igual ou até menor do que os estimadores \hat{m}_n e \tilde{m}_n , talvez não seja boa ideia usar esses estimadores para representar a verdadeira média da descendência, porque, da maneira como são construídos, principalmente o estimador X_n/X_{n-1} , eles contêm muito pouca informação sobre a amostra do processo. Mas as propriedades destes estimadores ajudam a concluir que, para o caso “supercrítico”, o estimador \hat{m}_n , por possuir

uma variância assintótica maior do que o estimador $\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_{i-1}}$ de Heyde e Seneta, não é muito eficiente. Por isso deve-se preferir o estimador de mínimos quadrados condicionais pesados \tilde{m}_n , que, efectivamente, tem uma variância assintótica menor do que \hat{m}_n . Assim, uma primeira conclusão que se pode tirar da comparação destes últimos três métodos de estimação, é que os estimadores de mínimos quadrados condicionais e os estimadores de mínimos quadrados condicionais pesados são preferíveis aos estimadores de Heyde e Seneta, e talvez se tenha encontrado uma primeira indicação de que o método de mínimos quadrados condicionais pesados seja mais vantajoso do que o método de mínimos quadrados condicionais; mas, para dizer realmente qual destes dois métodos é melhor, é preciso comparar os resultados obtidos pelos dois métodos para todos os estimadores e para todos os casos de m , o que vai ser feito a seguir.

Caso subcrítico $m < 1$:

Neste caso os estimadores \hat{m}_n , \tilde{m}_n , $\hat{\lambda}_n$ e $\tilde{\lambda}_n$ são todos fortemente consistentes. Já foi visto que é difícil dizer com qual dos dois métodos se obtêm variâncias assintóticas menores para os estimadores, mas uma vantagem do método de mínimos quadrados condicionais pesados é que, para conhecer a forma assintótica dos estimadores, não é preciso conhecer os momentos de terceira ordem da imigração e da descendência; para concluir

que \hat{m}_n e $\hat{\lambda}_n$ têm distribuição assintótica normal, é preciso impor as condições: $EY_{n-1,i}^3 < \infty$ e $EI_n^3 < \infty$.

Para afirmar que os estimadores de mínimos quadrados condicionais pesados das variâncias são fortemente consistentes e que têm distribuição assintótica normal tem de se supor que $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$. Considerando os estimadores das variâncias obtidos pelo método de mínimos quadrados condicionais, que são fortemente consistentes, essas duas condições só são necessárias para concluir que os estimadores têm distribuição assintótica normal. Neste caso, também é difícil ver, quais dos estimadores são assintoticamente mais eficientes, porque a matriz das variâncias-covariâncias está representada na forma de produtos de matrizes que, mesmo calculadas, iriam ter expressões compridas e complicadas.

Caso crítico $m=1$:

Neste caso observa-se que os estimadores obtidos pelo método de mínimos quadrados condicionais pesados têm melhores propriedades assintóticas do que os estimadores obtidos pelo método de mínimos quadrados condicionais, porque os estimadores \tilde{m}_n e $\tilde{\lambda}_n$ são ambos consistentes e, considerando os estimadores \hat{m}_n e $\hat{\lambda}_n$, só \hat{m}_n é que é consistente. Além disso, a distribuição assintótica de \tilde{m}_n é uma função de um processo de difusão com uma expressão mais simples do que a da distribuição assintótica de \hat{m}_n ; e, se $2\lambda > \sigma^2$, $EY_{n-1,i}^{2+\delta} < \infty$ e $EI_n^{2+\delta} < \infty$ para algum $\delta > 0$, então $\tilde{\lambda}_n$ tem distribuição normal no limite.

Quanto aos estimadores das variâncias obtidos pelo método de mínimos quadrados condicionais, só se sabe que se $EY_{n-1,i}^4 < \infty$ e $EI_n^4 < \infty$, então $\hat{\sigma}_n^2$ é consistente; mas, considerando o método de mínimos quadrados condicionais pesados, é sabido que $\tilde{\sigma}_n^2$ é consistente e assintoticamente normal e, se $2\lambda \leq \sigma^2$, então \tilde{b}_n^2 também é consistente.

Caso “supercrítico” $m>1$:

A primeira diferença entre as propriedades dos estimadores obtidos pelos dois métodos é

que $\tilde{\lambda}_n$ converge quase certamente para um valor dado por $\lambda + \frac{\sum_{i=1}^{\infty} \frac{X_i - mX_{i-1} - \lambda}{X_{i-1} + 1}}{\sum_{i=1}^{\infty} \frac{1}{X_{i-1} + 1}}$,

enquanto que $\hat{\lambda}_n$ converge em probabilidade para infinito. Assim, o estimador $\tilde{\lambda}_n$, no limite, vai estar mais próximo do valor verdadeiro do parâmetro do que $\hat{\lambda}_n$.

Analisando os resultados associados aos estimadores das variâncias, pode-se dizer que $\tilde{\sigma}_n^2$ é melhor do que $\hat{\sigma}_n^2$, porque converge fortemente para σ^2 e porque é assintoticamente normal. O resultado obtido para $\hat{\sigma}_n^2$ é difícil de interpretar, mas pode-se dizer que $\hat{\sigma}_n^2$ não é fortemente consistente nem consistente e que $\hat{\sigma}_n^2$ converge em distribuição para uma expressão composta de séries infinitas de variáveis aleatórias normais que são multiplicadas por pesos diferentes, dependentes de m . Considerando os estimadores para a variância da imigração, conclui-se que, apesar de nenhum dos estimadores ser consistente, no limite \tilde{b}_n^2 está mais próximo de b^2 , porque \tilde{b}_n^2 converge quase certamente para um valor, para o qual, teoricamente, ainda é possível calcular a distância entre ele e b^2 , enquanto que \hat{b}_n^2 converge em probabilidade para infinito.

Resumindo e comparando os resultados obtidos pelo método de mínimos quadrados condicionais e pelo método de mínimos quadrados condicionais pesados pode-se concluir que o método de mínimos quadrados condicionais pesados oferece melhores resultados.

Considerando o caso “supercrítico”, enquanto $\hat{\sigma}_n^2$ não verifica a propriedade de consistência, $\tilde{\sigma}_n^2$ já é um estimador fortemente consistente. Uma outra vantagem também é a de o estimador \tilde{m}_n ser mais eficiente do que \hat{m}_n por ter variância assintótica menor, quando $m > 1$. Além disso, para o caso crítico, com o método de mínimos quadrados condicionais pesados obtém-se ainda um estimador consistente para a média da imigração $\tilde{\lambda}_n$ e, se também se verificar $2\lambda \leq \sigma^2$, um estimador consistente para a variância da imigração \tilde{b}_n^2 . Com o método de mínimos quadrados condicionais não é possível obter

estimadores consistentes para estes parâmetros quando $m=1$. Observa-se que o método de mínimos quadrados condicionais pesados também não resolve o problema de estimar os parâmetros λ e b^2 da distribuição da imigração, quando $m > 1$. Wei e Winnicki (1987) provaram que estes dois parâmetros não admitem estimadores consistentes quando $m > 1$.

Apesar do método de mínimos quadrados condicionais pesados ser superior ao método de mínimos quadrados condicionais por se obterem com esse método melhores propriedades, observa-se que a forma dos estimadores de mínimos quadrados condicionais pesados é mais complexa do que a dos estimadores de mínimos quadrados condicionais.

Em geral, pode-se dizer que é muito trabalhoso calcular os estimadores dos dois métodos, mas mais trabalhoso, os do método de mínimos quadrados condicionais pesados, especialmente, os estimadores das variâncias.

2.2.2. ESTIMAÇÃO BASEADA NA OBSERVAÇÃO DE TODAS AS DIMENSÕES DA IMIGRAÇÃO E DA REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$

2.2.2.1. MÉTODO DE MÁXIMA VEROSIMILHANÇA

Suponha-se que $P(X_0 = x_0) = 1$ para algum inteiro positivo específico x_0 e sejam a distribuição da descendência e a distribuição da imigração do processo ramificado com imigração:

$$\begin{aligned} X_1 &= Y_{0,1} + \dots + Y_{0,x_0} + I_1 \\ X_2 &= Y_{1,1} + \dots + Y_{1,X_1} + I_2 \\ &\vdots \\ X_n &= Y_{n-1,1} + \dots + Y_{n-1,X_{n-1}} + I_n \end{aligned}$$

dadas por $p(j, \theta) = P(Y_{n-1,i} = j)$, $j=0,1,\dots$, e $q(i, \varphi) = P(I_n = i)$, $i=0,1,\dots$, respectivamente, onde $\{Y_{n-1,i}, n \geq 1\}$, a sequência de descendentes, e $\{I_n, n \geq 1\}$, a sequência das componentes de imigração, são independentes; os parâmetros θ e φ , dos quais dependem as distribuições, são reais e independentes. O espaço de parâmetros é compacto e contém um rectângulo aberto.

Bhat e Adke (1981) apresentaram um método de máxima verosimilhança para estimar os parâmetros do processo ramificado com imigração subcrítico e "supercrítico", supondo que $Y_{0,1}, \dots, Y_{n-1,X_{n-1}}, I_1, \dots, I_n$ são variáveis aleatórias observáveis. Como as observações X_1, \dots, X_n são obtidas através das dimensões da descendência e das dimensões da imigração e como $X_0 = x_0$ é especificado, então também é "observada" a realização do processo $\{X_0, \dots, X_n\}$.

Uma amostra das primeiras n gerações deste processo é constituída por duas partes:

- i) a amostra aleatória $Y_{0,1}, Y_{0,2}, \dots, Y_{n-1,X_{n-1}}$ da distribuição da descendência $\{p(j, \theta)\}$ de dimensão aleatória $N_{n-1} = x_0 + X_1 + \dots + X_{n-1}$, e
- ii) a amostra aleatória I_1, \dots, I_n da distribuição da imigração $\{q(i, \varphi)\}$ de dimensão fixa n .

Daqui conclui-se que a verosimilhança $L_n(\theta, \varphi)$ de θ e φ , baseada na amostra do processo, é um produto de duas verosimilhanças:

- i) a verosimilhança correspondente à amostra aleatória da distribuição da descendência $\{p(j, \theta)\}$ de dimensão N_{n-1} , e

- ii) a verosimilhança correspondente à amostra aleatória da distribuição da imigração $\{q(i, \varphi)\}$ de dimensão n .

Isto implica que a função de verosimilhança $L_n(\theta, \varphi)$ é da forma:

$$L_n(\theta, \varphi) = \prod_{i=1}^{N_{n-1}} p(Y_{n-1,i}, \theta) \prod_{j=1}^n q(I_j, \varphi).$$

Bhat e Adke (1981) mostraram que, sob determinadas condições⁸, os estimadores de máxima verosimilhança $\hat{\theta}$ e $\hat{\varphi}$ são fortemente consistentes. A distribuição assintótica de $\hat{\theta}$ é caracterizada por $(\hat{\theta} - \theta_0) \sqrt{\{E(N)j(\theta_0)\}} \xrightarrow{L} Z/\sqrt{W}$, onde Z é uma variável normal estandardizada independente da variável aleatória⁹ W e $E(N)$ significa a esperança de N em relação a $\{p(j, \theta_0)\}$ e $\{q(i, \varphi_0)\}$. $(\hat{\varphi} - \varphi_0) \sqrt{\{ni(\varphi_0)\}}$ tem assintoticamente uma distribuição normal estandardizada, onde se admite que $i(\varphi_0)$ - a informação de Fisher em φ contida numa única observação de $\{q(i, \varphi_0)\}$ - seja positiva e finita. Com efeito,

$$\left\{ (\hat{\theta} - \theta_0) \sqrt{\{E(N)j(\theta_0)\}} \right\}, \left\{ (\hat{\varphi} - \varphi_0) \sqrt{\{ni(\varphi_0)\}} \right\} \xrightarrow{L} (Z_1/\sqrt{W}, Z_2),$$

onde Z_1 e Z_2 são variáveis independentes normais estandardizadas e independentemente distribuídas de W .

Estes autores estudaram também a estimação de máxima verosimilhança das probabilidades $p_j = P(Y_{0,1} = j)$, $j=0,1,\dots$, e $q_i = P(I_1 = i)$, $i=0,1,\dots$, e, com a ajuda dos estimadores de máxima verosimilhança dessas probabilidades, obtiveram estimadores para a média da descendência e para a média da imigração.

A função de verosimilhança usada para estimar as probabilidades p_j e q_i , quando $Y_{0,1}, \dots, Y_{n-1, X_{n-1}}, I_1, \dots, I_n$ são as variáveis aleatórias observáveis, é

$$L = \prod_{j=0}^{\infty} p_j^{N_j(n)} \prod_{i=0}^{\infty} q_i^{M_i(n)},$$

onde $N_j(n)$ representa o número de elementos nas primeiras $(n-1)$ gerações que têm j

descendentes de um total de $N_{n-1} = \sum_{j=0}^{n-1} X_j$ elementos nas $(n-1)$ primeiras gerações e

$M_i(n)$ é o número de vezes que i componentes de imigração entram para o processo

⁸ Ver Bhat e Adke (1981: pp. 500-504)

⁹ W é uma variável aleatória positiva tal que quando $n \rightarrow \infty$, $N_n/E(N_n) \xrightarrow{q.c.} W$, onde $W=1$ no caso subcrítico.

durante as n primeiras gerações. $M_i(n)$ e $N_j(n)$ verificam $\sum_{i=0}^{\infty} M_i(n) = n$ e $\sum_{j=0}^{\infty} N_j(n) = N_{n-1}$.

Os estimadores de máxima verosimilhança \hat{p}_j e \hat{q}_i para p_j e q_i são¹⁰:

$$\hat{p}_j = \frac{N_j(n)}{N_{n-1}} \text{ e } \hat{q}_i = \frac{M_i(n)}{n}.$$

Se a média da descendência $m = \sum_{j=0}^{\infty} j p_j$ e a média da imigração $\lambda = \sum_{i=0}^{\infty} i q_i$ são finitas,

então os seus estimadores de máxima verosimilhança \hat{m} e $\hat{\lambda}$ são da forma:

$$\hat{m} = \sum_{j=0}^{\infty} j \hat{p}_j \text{ e } \hat{\lambda} = \sum_{i=0}^{\infty} i \hat{q}_i.$$

Bhat e Adke (1981) mostraram que \hat{p}_j e \hat{m} são fortemente consistentes, que a distribuição assintótica de \hat{p}_j é dada por $\left\{ (\hat{p}_j - p_j) \sqrt{E(N)}, j = 0, 1, \dots \right\} \xrightarrow[n \rightarrow \infty]{d} \left\{ Z_j / \sqrt{W}, j = 0, 1, \dots \right\}$, onde $\{Z_j, j = 0, 1, \dots\}$ é uma sequência Gaussiana com $E(Z_j) = 0$, $Var(Z_j) = p_j(1 - p_j)$ e $Cov(Z_j, Z_i) = -p_j p_i, j, i = 0, 1, \dots$, e que $(\hat{m} - m) \sqrt{E(N)}$ é assintoticamente distribuído como $Z\sigma / \sqrt{W}$, onde $\sigma^2 = \sum_{j=1}^{\infty} j^2 p_j - m^2$ é a variância da distribuição da descendência e supõe-

se que seja positiva e finita. Os mesmos autores provaram também que \hat{q}_i e $\hat{\lambda}$ são não enviesados, fortemente consistentes e assintoticamente normalmente distribuídos.

2.2.2.2. ESTIMAÇÃO SEQUENCIAL

Como se pode verificar, as distribuições no limite e as propriedades dos estimadores vistos até agora variam com o domínio de m : encontram-se resultados diferentes para $m < 1$, $m = 1$ e $m > 1$. Isto significa que, no caso de se querer aplicar esta teoria na prática, é preciso saber, em primeiro lugar, onde se situa o valor da média da descendência. E aqui enfrenta-se um problema: muitas vezes o valor ou o domínio de m pode ser desconhecido. Assim, como tentativa de encontrar uma distribuição no limite e propriedades universais, sem saber localizar o valor de m , alguns autores, motivados pelo trabalho de Lai e Siegmund (1983) para processos autorregressivos de ordem 1 (AR(1)), consideraram o problema da estimação da média da descendência, usando a análise sequencial. A análise sequencial

¹⁰ Ver ANEXO A3, pág. 81 seg..

baseia a inferência sobre m em dimensões de amostra aleatórias em vez de em dimensões de amostra fixas.

Supondo disponível a amostra $\{X_0, (X_j, I_j), j = 1, \dots, n\}$, Sriram, Basawa e Huggins (1991), os primeiros autores que usaram a estimação sequencial para o processo ramificado com imigração unindo o caso subcrítico e o caso crítico, partiram do estimador natural para a média da descendência

$$\hat{m}_n = \left(\sum_{j=1}^n X_{j-1} \right)^{-1} \sum_{j=1}^n (X_j - I_j) \quad (2.13)$$

do processo ramificado com imigração, definido em (2.8), onde X_0 , o estado inicial, tem distribuição arbitrária e é independente de $\{Y_{n-1,i}\}$ e $\{I_n\}$; a distribuição da descendência e a distribuição da imigração não são especificadas.

O estimador \hat{m}_n baseia-se na informação completa sobre as dimensões das gerações $\{X_j\}$ e do processo de imigração $\{I_j\}, j=1, \dots, n$.

As distribuições no limite do estimador \hat{m}_n são muito diferentes para o caso crítico e para o caso subcrítico. Sriram et al. (1991) mostraram que, se $m < 1$, a distribuição é normal:

$\left(\sum_{i=1}^n X_{i-1} \right)^{1/2} (\hat{m}_n - m) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma^2)$, e se $m=1$, a distribuição não é normal:

$\left(\sum_{i=1}^n X_{i-1} \right)^{1/2} (\hat{m}_n - m) \xrightarrow[n \rightarrow \infty]{d} \{Y(1) - \lambda\} / \left\{ \int_0^1 Y(t) dt \right\}^{1/2}$, onde $Y(t)$ é um processo de difusão¹¹.

Para evitar o problema de singularidade no ponto $m=1$, Sriram et al. substituíram o número de observações n no estimador \hat{m}_n por um tempo de paragem N_c , obtido através da lei de paragem

$$N_c = \inf \left\{ n \geq 1 : \sum_{i=1}^n X_{i-1} \geq c\sigma^2 \right\}, \quad (2.14)$$

onde $c > 0$ é escolhido adequadamente. Nesta expressão assume-se que σ^2 seja conhecido.

Se σ^2 é desconhecido, então esse parâmetro pode ser substituído pelo estimador

$$\hat{\sigma}_n^2 = \left\{ \sum_{i=1}^n \frac{X_{i-1}}{(1 + X_{i-1})^2} \right\}^{-1} \sum_{i=1}^n \frac{(X_i - \hat{m}_n X_{i-1} - I_i)^2}{(1 + X_{i-1})^2} \quad (2.15)$$

¹¹ Ver ANEXO A2, pág. 81.

e a lei de paragem associada designa-se por \tilde{N}_c . Por um teorema de Sriram et al., $\hat{\sigma}_n^2$ é fortemente consistente, se $EY_{1,1}^4 < \infty$.

Substituindo agora n por N_c (σ é conhecido), então as distribuições no limite do estimador \hat{m}_{N_c} são normais para $m < 1$ e para $m = 1$, quando $c \rightarrow \infty$. No caso de σ ser desconhecido, o resultado é semelhante. As normalidades assintóticas de \hat{m}_{N_c} e $\hat{m}_{\tilde{N}_c}$ encontram-se nos seguintes teoremas dos mesmos autores.

Teorema 2.15: Para cada $m \leq 1$ fixo e N_c definido em (2.14) tem-se

$$\left(\sum_{i=1}^{N_c} X_{i-1} \right)^{1/2} (\hat{m}_{N_c} - m) \xrightarrow{c \rightarrow \infty} N(0, \sigma^2).$$

Teorema 2.16: Admita que $EY_{1,1}^4 < \infty$ e suponha que o tempo de paragem \tilde{N}_c é definido por $\tilde{N}_c = \inf \left\{ n \geq 2 : \sum_{i=1}^n X_{i-1} \geq c \hat{\sigma}_n^2 \right\}$, onde $\hat{\sigma}_n^2$ é dado por (2.15), então para cada $m \leq 1$ fixo

$$\left(\sum_{i=1}^{\tilde{N}_c} X_{i-1} \right)^{1/2} \frac{(\hat{m}_{\tilde{N}_c} - m)}{\hat{\sigma}_{\tilde{N}_c}} \xrightarrow{c \rightarrow \infty} N(0, 1).$$

Seja \mathcal{F} a classe de funções de distribuição caracterizada por $\mathcal{F} \subset \left\{ F : \int_{-\infty}^{+\infty} (x-m) dF(x) = 0 \text{ e } \int_{-\infty}^{+\infty} (x-m)^2 dF(x) = \sigma_F^2 \in (0, \infty) \text{ para algum } m \in (0, 1] \right\}$ e por verificar

$\sup_F \int_{\{(x-m) > a\}} (x-m)^2 dF(x) = o(1)$ quando $a \rightarrow \infty$ e $\inf_F \sigma_F^2 > 0$, onde $F \in \mathcal{F}$. Sriram et al.

mostraram ainda que o estimador \hat{m}_{N_c} é assintoticamente normalmente distribuído

uniformemente sobre \mathcal{F} : $\limsup_{c \rightarrow \infty} \sup_F \left| P_F \left\{ \left(\sum_{i=1}^{N_c} X_{i-1} \right)^{1/2} \frac{\hat{m}_{N_c} - m}{\sigma} \leq x \right\} - \Phi(x) \right| = 0$, onde Φ

representa a distribuição normal estandardizada e $F \in \mathcal{F}$. A questão, se a uniformidade também é válida para $\hat{m}_{\tilde{N}_c}$, quando σ é desconhecido, fica em aberto. Os autores supõem que a uniformidade continua válida. Sriram (1991) estabeleceu ainda um resultado de consistência uniforme para o estimador \hat{m}_n .

2.2.2.3. ESTIMAÇÃO DE PRECISÃO FIXA

Os estimadores da média da descendência obtidos pelos métodos anteriores, para o caso da imigração ser observável e para o caso de ela não ser observável, são funções não lineares das observações com expressões complicadas. Por isso torna-se difícil calcular os valores esperados e os erros quadráticos médios desses estimadores para um número fixo de observações. Para evitar esta dificuldade, os autores estudaram as propriedades dos estimadores de m , usando a teoria de amostras grandes. Mas a inferência de amostras grandes para m não é muito satisfatória porque, como já foi referido anteriormente, as distribuições assintóticas dos estimadores variam muito, dependendo se $m < 1$, $m = 1$ ou $m > 1$. Apesar de Sriram et al. (1991) terem conseguido obter um estimador com distribuição assintótica normal para o caso subcrítico e para o caso crítico, a natureza do estimador sequencial \hat{m}_{N_c} não facilita o cálculo da esperança e do erro quadrático médio para cada c fixo. Shete e Sriram (1998) afirmaram que é possível mostrar que \hat{m}_{N_c} até é um estimador enviesado de m . Com o fim de melhorar o estimador no sentido de atingir um não enviesamento e de facilitar o cálculo do erro quadrático médio, Shete e Sriram propuseram uma modificação de \hat{m}_{N_c} . O estimador modificado, designado por estimador de precisão fixa, é não enviesado e tem erro quadrático médio limitado. Uma outra vantagem do novo estimador é que para σ^2 desconhecido, $\hat{m}_{\tilde{N}_c}$ também é assintoticamente normal para $m \in (0, 1]$.

Considere-se novamente o processo ramificado com imigração $X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i} + I_n$, $n=1, 2, \dots$, onde $\{Y_{n-1,i}\}$ e $\{I_n\}$ são sequências independentes de variáveis aleatórias inteiras não negativas i.i.d. de médias e variâncias finitas (m, σ^2) e (λ, b^2) . Aqui supõe-se que o estado inicial, além de ser uma variável aleatória inteira independente de $\{Y_{n-1,i}\}$ e $\{I_n\}$, também seja quadraticamente integrável. Shete e Sriram modificaram o estimador sequencial, motivados pela natureza do tempo de paragem

$$N_c = \inf \left\{ n \geq 1 : \sum_{i=1}^n X_{i-1} \geq c\sigma^2 \right\}. \quad (2.16)$$

A partir da relação $\sum_{i=1}^{N_c-1} X_{i-1} < c\sigma^2 \leq \sum_{i=1}^{N_c} X_{i-1}$ é possível definir um multiplicador de correcção

único $\theta_c \in (0,1]$ através da equação $\sum_{i=1}^{N_c-1} X_{i-1} + \theta_c X_{N_c-1} = c\sigma^2$, e, assim, modificar o estimador sequencial \hat{m}_{N_c} , obtendo-se

$$\hat{m}(c) = \frac{\sum_{i=1}^{N_c-1} (X_i - I_i) + \theta_c (X_{N_c} - I_{N_c})}{c\sigma^2}. \quad (2.17)$$

No caso de σ^2 ser desconhecido, define-se primeiro um estimador para σ^2 , diferente do estimador usado por Sriram et al. (1991):

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n [(X_i - \hat{m}_n X_{i-1} - I_i)^2 / X_{i-1}], \quad (2.18)$$

onde \hat{m}_n é dado por (2.13). Seja $\tilde{\sigma}_n^2 = \max\{\hat{\sigma}_n^2, n^{-\alpha}\}$ para algum $\alpha > 0$ e seja n_0 a dimensão da amostra inicial. Definindo agora

$$\tilde{N}_c = \inf \left\{ n \geq n_0 + 1 : \sum_{i=n_0+1}^n X_{i-1} \geq c\tilde{\sigma}_{n_0}^2 \right\} \quad (2.19)$$

e um multiplicador de correcção $\tilde{\theta}_c \in (0,1]$ de maneira única através da equação

$\sum_{i=n_0+1}^{\tilde{N}_c-1} X_{i-1} + \tilde{\theta}_c X_{\tilde{N}_c-1} = c\tilde{\sigma}_{n_0}^2$, então, o estimador sequencial modificado $\hat{m}_{\tilde{N}_c}$, para σ^2

desconhecido, recebe a expressão:

$$\tilde{m}(c) = \frac{\sum_{i=n_0+1}^{\tilde{N}_c-1} (X_i - I_i) + \tilde{\theta}_c (X_{\tilde{N}_c} - I_{\tilde{N}_c})}{c\tilde{\sigma}_{n_0}^2}. \quad (2.20)$$

Shete e Sriram (1998) apresentaram propriedades não assintóticas para $\hat{m}(c)$ e $\tilde{m}(c)$, demonstrando que os estimadores de precisão fixa são não enviesados e têm erro quadrático médio limitado.

As propriedades assintóticas dos estimadores encontram-se nos teoremas expostos a seguir, provados pelos mesmos autores.

Teorema 2.17: Para $\hat{m}(c)$ e N_c , definidos, respectivamente, nas equações (2.17) e (2.16),

e a classe \mathcal{F}^{12} é válido o seguinte:

$$\limsup_{c \rightarrow \infty} \sup_F \left| P_F \left\{ \left(\sum_{i=1}^{N_c} X_{i-1} \right)^{1/2} \frac{\hat{m}(c) - m}{\sigma} \leq x \right\} - \Phi(x) \right| = 0,$$

¹² A definição da classe de funções de distribuição \mathcal{F} encontra-se na página 35.

onde $F \in \mathcal{F}$ e Φ é a função de distribuição normal estandardizada.

Para se obter a distribuição assintótica de $\tilde{m}(c)$ no caso de σ^2 ser desconhecido, suponha-se que n_0 , a dimensão da amostra inicial, dependa de c : $n_0 = n_0(c)$, onde

$$1 < n_0(c) \rightarrow \infty, \text{ mas } n_0(c) = o(c^{1/2}) \text{ quando } c \rightarrow \infty. \quad (2.21)$$

No Teorema 2.18 é usada a consistência forte de $\hat{\sigma}_n^2$. Os autores mostraram com a seguinte proposição que $\hat{\sigma}_n^2$ até é uniformemente fortemente consistente, onde a uniformidade é estabelecida para $m \in (0,1]$.

Proposição 2.1: Para $\hat{\sigma}_n^2$, definido em (2.18), é válido: Se $\sup_F E|Y_{1,1} - m|^4 < \infty$, então para cada $\delta > 0$

$$\limsup_{k \rightarrow \infty} P_F \left\{ \hat{\sigma}_n^2 - \sigma^2 \geq \delta \text{ para algum } n \geq k \right\} = 0,$$

onde $F \in \mathcal{F}$.

Teorema 2.18: Dadas as condições da Proposição 2.1, para $\tilde{m}(c)$ e \tilde{N}_c , definidos, respectivamente, nas equações (2.19) e (2.20), com n_0 definido em (2.21), é válido:

Para cada $0 < m \leq 1$

$$\left(\sum_{i=1}^{\tilde{N}_c} X_{i-1} \right)^{1/2} \frac{(\tilde{m}(c) - m)}{\tilde{\sigma}_{n_0(c)}} \xrightarrow[c \rightarrow \infty]{D} N(0,1),$$

onde $\tilde{\sigma}_n^2 = \max\{\hat{\sigma}_n^2, n^{-\alpha}\}$ e $\hat{\sigma}_n^2$ é dado por (2.18).

Estes últimos resultados assintóticos indicam que para $m \in (0,1]$ a distribuição no limite de $\hat{m}(c)$ é normal uniformemente e a distribuição no limite de $\tilde{m}(c)$ é normal. $\hat{\sigma}_n^2$ é um estimador uniformemente fortemente consistente para σ^2 .

2.2.2.4. ESTIMAÇÃO SEQUENCIAL DE DUAS ETAPAS

Foi visto que as propriedades assintóticas do estimador de precisão fixa continuam desconhecidas para o caso “supercrítico”. Qi e Reeves (2002) basearam-se na lei de

paragem de Sriram et al. (1991) e construíram estimadores sequenciais de duas etapas para a média da descendência, que podem ser vistos como extensões dos estimadores sequenciais de Sriram et al., por serem fortemente consistentes e assintoticamente normais para $m > 0$ e não só para $m \in (0,1]$.

Como nos três métodos de estimação anteriores, suponha-se que $\{X_n\}$ e $\{I_n\}$ do processo ramificado com imigração (2.8) sejam observáveis. Sejam \hat{m}_n o estimador dado por (2.13) e N_c a lei de paragem (2.14) de Sriram et al., onde c é escolhido adequadamente. Nesta definição é assumido que σ^2 seja conhecido. Quando σ^2 é desconhecido, define-se a lei de paragem por

$$N(c) = \inf \left\{ n \geq 1 : \sum_{i=1}^n X_{i-1} \geq c \right\} \quad (2.22)$$

e $\hat{m}_{N(c)}$ é usado como estimador sequencial de m .

Seja agora $G = G(c, N_c) \geq N_c$ uma função de valores inteiros de c e N_c verificando, quando $c \rightarrow \infty$:

$$\frac{G(c, N_c) - N_c}{c^\alpha} \xrightarrow{q.c.} 0 \text{ para algum } \alpha \in \left(0, \frac{1}{2}\right) \text{ se } m \in (0,1] \quad (2.23)$$

ou

$$G(c, N_c) - N_c \xrightarrow{q.c.} \infty, \text{ se } m > 1. \quad (2.24)$$

O estimador sequencial de duas etapas fica definido por \hat{m}_G .

Do teorema seguinte, provado por Qi e Reeves (2002), deduz-se que o estimador sequencial de duas etapas é fortemente consistente e assintoticamente normal.

Teorema 2.19: O estimador \hat{m}_G converge quase certamente para m quando c tende para infinito. Se $G = G(c, N_c)$ verifica (2.23) e (2.24), então

$$\left(\sum_{i=1}^G X_{i-1} \right)^{1/2} \frac{\hat{m}_G - m}{\sigma} \xrightarrow{d} N(0,1).$$

Nota: É fácil de ver que as condições (2.23) e (2.24) são verificadas por uma classe vasta de funções: Seja $g(c)$ uma função de valores inteiros para a qual é válido $\lim_{c \rightarrow \infty} g(c) = \infty$ e

$\lim_{c \rightarrow \infty} \frac{g(c)}{c^\alpha} = 0$ para algum $\alpha \in \left(0, \frac{1}{2}\right)$. Estabelecendo agora a igualdade

$G = G(c, N_c) = g(c) + N_c$, então as condições (2.23) e (2.24) verificam-se automaticamente.

Qi e Reeves mostraram que estas condições também são satisfeitas por $G(c) = G(c, N_c) = \max(g(c), N_c)$, i.e. $G(c) = \inf \left\{ n \geq g(c) : \sum_{i=1}^n X_{i-1} \geq c\sigma^2 \right\}$, onde $g(c)$ verifica

$$\lim_{c \rightarrow \infty} \frac{g(c)}{\log c} = \infty.$$

O resultado do teorema anterior continua verdadeiro, se N_c for substituído por $N(c)$, isto corresponde ao caso da variância ser desconhecida; mas, para se encontrar um intervalo de confiança assintótico para m , é preciso definir um estimador de σ^2 . Shete e Sriram (1998)

provaram que $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \frac{(X_i - \hat{m}_n X_{i-1} - I_i)^2}{X_{i-1}} \xrightarrow[n \rightarrow \infty]{q.c.} \sigma^2$, desde que $EY_{1,1}^4 < \infty$. Daqui Qi e

Reeves chegaram à seguinte consequência:

Corolário 2.1: Suponha que $G = G(c, N(c))$, onde $N(c)$, definido em (2.22), verifica (2.23) e (2.24) quando N_c é substituído por $N(c)$ e que $EY_{1,1}^4 < \infty$, então

$$\left(\sum_{i=1}^G X_{i-1} \right)^{1/2} \frac{\hat{m}_G - m}{\hat{\sigma}_G} \xrightarrow{d} N(0,1).$$

2.2.2.5. COMPARAÇÃO DAS PROPRIEDADES E DOS RESULTADOS OBTIDOS COM OS MÉTODOS DE ESTIMAÇÃO EM QUE SÃO OBSERVADAS TODAS AS DIMENSÕES DA IMIGRAÇÃO E A REALIZAÇÃO DO PROCESSO $\{X_0, \dots, X_n\}$

Destes últimos quatro métodos, que têm em comum ser suposto que todas as dimensões da imigração são observadas, os métodos sequencial, de precisão fixa e sequencial de duas etapas distinguem-se do método de máxima verosimilhança por partirem todos do mesmo estimador para a média da descendência \hat{m}_n (2.13), que é construído à custa das observações $\{X_0, \dots, X_n\}$ e das componentes de imigração $\{I_0, \dots, I_n\}$, e por ser substituída a dimensão n deste estimador por um tempo de paragem que é determinado de uma maneira diferente para os três métodos. Se $m < 1$, então \hat{m}_n é assintoticamente normal e se $m = 1$, então a distribuição no limite depende de um processo de difusão.

Bhat e Adke (1981), que aplicaram o método de máxima verosimilhança a processos ramificados com imigração subcríticos e “super-críticos”, explicam como são determinados

estimadores para os parâmetros θ e φ da distribuição da descendência e da distribuição da imigração, respectivamente. Note-se que, para algumas distribuições particulares (por exemplo Bernoulli, Poisson), estes parâmetros representam as médias dos conjuntos correspondentes (descendência e imigração). Se a distribuição da descendência (ou a distribuição da imigração) for igual à distribuição de Poisson, então $\hat{\theta}$ (ou $\hat{\varphi}$) é um estimador para a média e para a variância da descendência (ou da imigração). Em geral, para outras distribuições, os estimadores das médias e das variâncias da descendência e da imigração vão ser funções dos estimadores $\hat{\theta}$ e $\hat{\varphi}$, dos quais se sabe serem fortemente consistentes e $\hat{\varphi}$ ter assintoticamente uma distribuição normal sob determinadas condições. Assim, os estimadores para θ e φ não são directamente estimadores para as médias ou variâncias; o facto de serem estimadores directos vai depender das distribuições da descendência e da imigração. Bhat e Adke explicam ainda como se obtém estimadores de máxima verosimilhança \hat{p}_j e \hat{q}_i para as probabilidades $p_j = P(Y_{0,1} = j)$, $j=0,1,\dots$, e $q_i = P(I_1 = i)$, $i=0,1,\dots$, e constróem estimadores para a média da descendência e para a média da imigração \hat{m} e $\hat{\lambda}$ que dependem respectivamente de \hat{p}_j e \hat{q}_i através de séries infinitas. Estes autores afirmam que \hat{m} e $\hat{\lambda}$ são estimadores de máxima verosimilhança, se m e λ forem finitos. \hat{m} tem a propriedade de ser fortemente consistente e $\hat{\lambda}$ é não enviesado, fortemente consistente e assintoticamente normal.

Com o método de estimação sequencial, método de estimação de precisão fixa e método de estimação sequencial de duas etapas são obtidos estimadores para a média da descendência para o caso da variância da descendência ser desconhecida e para o caso dela ser conhecida. A vantagem destes estimadores é de, para $m \in (0,1]$ (e para $m > 0$ considerando o método de estimação sequencial de duas etapas) ser possível construir intervalos de confiança assintóticos para m , porque a distribuição no limite desses estimadores é normal. Sriram et al. e Shete e Sriram calcularam também um estimador consistente para a variância que depende das componentes da imigração e das observações do processo. O estimador $\hat{\sigma}_n^2$ de Shete e Sriram tem uma expressão mais simples do que o estimador de Sriram et al. e é uniformemente fortemente consistente (uma propriedade mais forte do que a propriedade do estimador de Shete e Sriram que é só fortemente consistente), se for verificada a condição da Proposição 2.1. Observa-se que os estimadores sequenciais \hat{m}_{N_c} e $\hat{m}_{\bar{N}_c}$ são assintoticamente eficientes como os estimadores de precisão fixa $\hat{m}(c)$ e $\tilde{m}(c)$. Para os estimadores \hat{m}_{N_c} de Sriram et al. e $\hat{m}(c)$ de Shete e

Sriram, que são construídos supondo que σ^2 é conhecida, ainda foi provada a normalidade assintótica uniforme. Esta propriedade é mais forte do que a normalidade assintótica. Shete e Sriram conseguiram mostrar que o estimador de precisão fixa é não enviesado e que tem erro quadrático médio limitado. Os outros autores, excepto Bhat e Adke que afirmaram num teorema que o estimador de máxima verosimilhança $\hat{\lambda}$ é não enviesado, não apresentaram as propriedades não assintóticas dos estimadores.

2.3. APRECIÇÃO CRÍTICA DOS MÉTODOS

Comparando agora os resultados associados aos métodos de estimação, que supõem a possibilidade da observação da realização do processo $\{X_0, \dots, X_n\}$ com os resultados para o caso da estimação em que também são observadas as dimensões da imigração, pode-se dizer que o estudo feito para os primeiros métodos – método de mínimos quadrados condicionais e método de mínimos quadrados condicionais pesados – é mais completo, porque inclui tanto a estimação dos parâmetros da descendência como a estimação dos parâmetros da imigração. Com os métodos do segundo grupo – método sequencial, método de precisão fixa e método sequencial de duas etapas – só são obtidos estimadores para a média e para a variância da descendência; destes métodos só no método de estimação sequencial de duas etapas é que também é considerado o caso “supercrítico”. O método de máxima verosimilhança não é usado para estimar as variâncias σ^2 e b^2 .

Para se poder aplicar os métodos do primeiro grupo é preciso conhecer antes o valor de m ; pelo menos tem de se saber se $m < 1$, $m = 1$ ou $m > 1$. Os métodos do segundo grupo têm a vantagem de os resultados serem mais universais para o domínio de m : os resultados do método de estimação sequencial e do método de estimação de precisão fixa são válidos para processos com $0 < m \leq 1$ e os resultados obtidos pelo método de estimação sequencial de duas etapas, para processos com $m > 0$. Uma outra vantagem destes últimos métodos é também de as distribuições assintóticas não dependerem de ambas as variâncias σ^2 e b^2 . Algumas distribuições no limite, obtidas com os métodos do primeiro grupo, como, por exemplo, a distribuição assintótica dos estimadores de mínimos quadrados condicionais \hat{m} e $\hat{\lambda}$ para o caso subcrítico, dependem de σ^2 e b^2 ; assim, para a aplicação destes resultados, é preciso conhecer ou estimar as variâncias e isso torna o cálculo dos intervalos de confiança assintóticos mais difícil.

Comparando as expressões dos estimadores das médias, observa-se que as do segundo grupo têm uma aparência mais simples; o cálculo dos estimadores de mínimos quadrados

condicionais e dos estimadores de mínimos quadrados condicionais pesados parece ser muito trabalhoso.

O facto de os estimadores do segundo grupo serem construídos também à custa da imigração e não só à custa das observações do processo, pode talvez levar a crer que estes estimadores sejam melhores do que os outros, porque contêm mais informação sobre o processo.

No método de máxima verosimilhança só são apresentados estimadores para as médias, não para as variâncias; a desvantagem é de a distribuição assintótica de \hat{m} não ser normal como nos outros métodos. Segundo Wei e Winnicki (1990: p. 1757), o método de máxima verosimilhança, quando são incluídos na amostra os números de componentes de imigração $\{I_0, \dots, I_n\}$, dá resultados úteis, em comparação com os “estimadores de máxima verosimilhança” baseados na observação do processo $\{X_0, \dots, X_n\}$, obtidos por Heyde e Seneta (1972), que têm expressões complicadas.

De todos estes métodos anteriormente apresentados, só para o método de precisão fixa é que foram estudadas e demonstradas propriedades não assintóticas dos estimadores da média da descendência.

Com base nas comparações feitas anteriormente pode-se concluir que o método de mínimos quadrados condicionais pesados é o método com o qual se obtém melhores resultados, quando só se observa a realização do processo $\{X_0, \dots, X_n\}$. Se os números de componentes de imigração também forem observados, então, para estimar a média da descendência, talvez seja melhor o estimador sequencial de duas etapas.

De certeza que se obteriam melhores métodos e melhores resultados, se se conseguisse adaptar as leis de paragem das estimações sequenciais, de precisão fixa e sequenciais de duas etapas aos estimadores de mínimos quadrados condicionais e aos estimadores de mínimos quadrados condicionais pesados.

Sriram et al. (1991: p. 2233) afirmaram que é possível a extensão da aproximação sequencial para um estimador de m baseado só na observação $\{X_0, \dots, X_n\}$.

3. PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO

3.1. DEFINIÇÕES E PROPRIEDADES

Os processos autorregressivos de valor inteiro, INAR, foram introduzidos na literatura com a finalidade de serem usados para analisar e modelar séries temporais de valores discretos, em particular, para analisar processos de dados de contagem dependentes, como por exemplo, o número de acidentes ou o número de ocorrências de doenças durante um intervalo de tempo. Estes processos baseiam-se na operação “thinning” e são construídos de modo a terem uma dada distribuição marginal discreta e uma estrutura de dependência autorregressiva. Inicialmente, o modelo autorregressivo de valor inteiro de ordem 1, INAR(1), $X_t = m * X_{t-1} + I_t$, foi formulado por McKenzie (1985), que atribuiu a X_{t-1} a distribuição geométrica. Mais tarde, foram usadas outras distribuições discretas para a construção do modelo (por exemplo, Poisson, binomial, binomial negativa) e a dependência de X_t de variáveis aleatórias passadas foi alargada, o que levou à origem do modelo autorregressivo de valor inteiro de ordem p , INAR(p), em que X_t depende de p variáveis aleatórias passadas. Posteriormente, Gauthier e Latour (1992) construíram processos autorregressivos de valor inteiro generalizados, GINAR. Os processos INAR(1) e INAR(p) são casos particulares dos processos GINAR(p). Nesta secção apresentam-se estes processos e as suas propriedades por ordem cronológica do seu aparecimento.

Para definir os processos autorregressivos de valor inteiro é necessário introduzir primeiro a operação “thinning” de Steutel e van Harn (1979).

Definição (Operação “thinning”)

Seja X uma variável inteira não negativa e seja $\{Y_i\}$ uma sequência de variáveis aleatórias de Bernoulli i.i.d. de parâmetro m , independente de X , verificando $P(Y_i = 1) = 1 - P(Y_i = 0) = m$, onde $0 \leq m \leq 1$. A operação “thinning”, representada pelo símbolo ‘*’, é definida por

$$m * X = \sum_{i=1}^X Y_i, \text{ se } X > 0,$$

e

$$m * X = 0, \text{ se } X = 0.$$

(3.1)

A operação ‘ $m * X$ ’ decompõe a variável inteira X numa soma de X componentes binárias Y_i , $i=1, \dots, X$. Assim, $m * X$ transforma X num número inteiro situado no intervalo $[0, X]$. A sequência $\{Y_i\}_{i=1,2,\dots}$ é conhecida por série de contagem de $m * X$. Condicionado por X , $m * X$ segue uma distribuição binomial de parâmetros X e m .

PROCESSO AUTORREGRESSIVO DE VALOR INTEIRO DE ORDEM 1, INAR(1)

O processo INAR(1), processo autorregressivo de valor inteiro de ordem 1, foi proposto por Al-Osh e Alzaid em 1987 para modelar e gerar sequências de processos de contagens dependentes.

Segundo Al-Osh e Alzaid (1987), o processo estocástico em tempo discreto $\{X_t\}_{t=0, \pm 1, \dots}$ representa um INAR(1), se verifica a equação:

$$X_t = m * X_{t-1} + I_t, \quad (3.2)$$

onde:

- $\{X_t\}$ são variáveis aleatórias inteiras não negativas,
- ‘ $*$ ’ simboliza a operação “thinning” definida em (3.1),
- $0 \leq m \leq 1$,
- $\{I_t\}$ é uma sequência de variáveis aleatórias inteiras não negativas e não correlacionadas de média λ e de variância finita b^2 .

Uma possível interpretação do modelo (3.2) é: X_t compõe-se dos sobreviventes dos elementos que estavam presentes no instante de tempo $t-1$, tendo cada elemento a probabilidade de sobrevivência m , mais I_t novos elementos que entram para o processo no intervalo de tempo $(t-1, t]$; I_t também se designa por termo de inovação.

Qualquer processo aleatório que verifique a equação (3.2) é uma cadeia de Markov e $\{X_t\}$ é estacionário de segunda ordem se $0 \leq m < 1$.

O processo INAR(1) tem a propriedade de a distribuição de X_t ser unicamente determinada pela distribuição de I_t . X_t segue uma distribuição de Poisson, se e só se I_t segue uma distribuição de Poisson.

A seguir apresentam-se alguns resultados que foram obtidos por Silva e Oliveira (2000) para o processo estacionário INAR(1) de Poisson.

A média e os momentos de segunda ordem são dados por:

$$E(X_t) = \mu = \frac{\lambda}{1-m}$$

$$\mu(0) = E(X_t^2) = \frac{\lambda(\lambda+1-m)}{(1-m)^2}$$

$$\mu(s) = E(X_t X_{t+s}) = mE(X_t X_{t+s-1}) + \lambda\mu = m\mu(s-1) + \lambda\mu, s > 0.$$

Os cumulantes de primeira e de segunda ordem são da forma:

$$\text{cum}(X_t) = E(X_t) = \mu = \frac{\lambda}{1-m}$$

$$C_2(0) = E(X_t^2) - [E(X_t)]^2 = \text{Var}(X_t) = \frac{\lambda}{1-m}$$

$$C_2(s) = E\{(X_t - E(X_t))(X_{t+s} - E(X_{t+s}))\} = E(X_t X_{t+s}) - \mu^2 = mC_2(s-1) = m^s C_2(0), s > 0.$$

Daqui conclui-se que, para qualquer inteiro k não negativo, a função autocovariância para o "lag" k tem a expressão

$$\gamma(k) = \text{Cov}(X_{t-k}, X_t) = C_2(k) = m^k \gamma(0).$$

A função autocorrelação fica definida por

$$\rho(k) = \frac{C_2(k)}{C_2(0)} = m^k \text{ ou } \rho(k) = \frac{\gamma(k)}{\gamma(0)} = m^k.$$

Esta equação mostra que $\rho(k)$ decresce exponencialmente com o "lag" k e que $\rho(k)$ tem a mesma forma da equação de Yule-Walker do processo AR(1). A diferença é que, no modelo INAR(1), a função autocorrelação $\rho(k)$ é sempre positiva.

A função densidade espectral é dada por:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-i\omega k} = \frac{\lambda(m+1)}{2\pi(1-2m\cos\omega+m^2)}, -\pi \leq \omega \leq \pi.$$

Estes últimos resultados revelam que as propriedades de segunda ordem deste processo são análogas às propriedades de um processo AR(1).

PROCESSO AUTORREGRESSIVO DE VALOR INTEIRO DE ORDEM P, INAR(p)

O processo INAR(p) $\{X_t\}$, definido por Du e Li (1991), é uma extensão do processo INAR(1) e formula-se:

$$X_t = \sum_{i=1}^p m_i * X_{t-i} + I_t = m_1 * X_{t-1} + \dots + m_p * X_{t-p} + I_t, \quad (3.3)$$

onde:

- $\{X_t\}$ são variáveis aleatórias inteiras não negativas,
- '*' simboliza a operação "thinning" definida em (3.1),
- $m_i \in [0,1]$ para $i=1, \dots, p$,
- $\{I_t\}$ é uma sequência de variáveis aleatórias inteiras não negativas i.i.d. de uma distribuição discreta,
- todas as séries de contagem de $m_i * X_{t-i}$, $i=1, \dots, p$, são mutuamente independentes e independentes de $\{I_t\}$.

Du e Li (1991) mostraram com o seguinte teorema, em que condições existe uma solução estacionária de (3.3).

Teorema 3.1:

Seja $\{I_t\}$ uma sequência de variáveis aleatórias inteiras não negativas com $E(I_t) = \lambda$, $Var(I_t) = b^2$ e $m_i \in [0,1]$ ($i=1, \dots, p$). Se as raízes de

$$x^p - m_1 x^{p-1} - \dots - m_{p-1} x - m_p = 0$$

estão dentro do círculo unitário, então existe uma única série aleatória estacionária $\{X_t\}$ de valores inteiros não negativos, que satisfaz

$$X_t = m_1 * X_{t-1} + \dots + m_p * X_{t-p} + I_t$$

$$Cov(X_s, I_t) = 0 \quad (s < t).$$

O modelo (3.3) também pode ser definido na forma de espaço de estados, que facilita depois a representação e os cálculos dos momentos. Para isso utiliza-se a operação "thinning" vectorial. Seja A uma matriz $p \times p$ com entradas a_{ij} , que verificam $0 \leq a_{ij} \leq 1$.

Então, a operação "thinning" vectorial aplicada ao vector aleatório de valores inteiros

$\vec{X} = (X_1, \dots, X_p)^T$, denotada por $A * \vec{X}$, ou escrita de forma extensa:

$$A * \vec{X} = \begin{pmatrix} a_{11} & \dots & \dots & a_{1p} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{p1} & \dots & \dots & a_{pp} \end{pmatrix} * \begin{pmatrix} X_1 \\ \vdots \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} a_{11} * X_1 + \dots + a_{1p} * X_p \\ \vdots \\ \vdots \\ a_{p1} * X_1 + \dots + a_{pp} * X_p \end{pmatrix},$$

resulta num vector de valores inteiros em que a expressão da i -ésima componente é dada por:

$$(A * \vec{X})_i = \left(\sum_{j=1}^p a_{ij} * X_j \right). \quad (3.4)$$

Em (3.4) assume-se que todas as séries de contagens de todas as expressões $a_{ij} * X_j$, $i, j=1, \dots, p$, são independentes.

O modelo (3.3), escrito na forma de espaço de estados, usando a operação “thinning” vectorial, fica então com o seguinte aspecto:

$$\begin{aligned} \vec{X}_t &= A * \vec{X}_{t-1} + \vec{C}I_t \\ X_t &= \vec{H}\vec{X}_t \end{aligned} \quad (3.5)$$

onde $\vec{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})^T$, A é uma matriz de dimensão $p \times p$ da forma

$$A = \begin{pmatrix} m_1 & m_2 & \dots & m_{p-1} & m_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad (3.6)$$

$\vec{C} = (1 \ 0 \ \dots \ 0)^T$ e $\vec{H} = (1 \ 0 \ \dots \ 0)$ são vectores com p elementos.

Escrevendo a primeira equação de (3.5) de forma extensa, obtém-se:

$$\begin{aligned} \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ \vdots \\ X_{t-p+1} \end{pmatrix} &= \begin{pmatrix} m_1 & m_2 & \dots & m_{p-1} & m_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} * \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ \vdots \\ X_{t-p} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} I_t = \\ &= \begin{pmatrix} m_1 * X_{t-1} + \dots + m_p * X_{t-p} + I_t \\ X_{t-1} \\ \vdots \\ \vdots \\ X_{t-p+1} \end{pmatrix}. \end{aligned}$$

A condição¹³ exigida por Du e Li (1991) para que o processo INAR(p), dado por (3.5), seja estacionário, é que o raio espectral da matriz A , designado por $\rho(A)$, seja menor do que 1:

¹³ Ver Teorema 3.1.

$\rho(A) < 1$, o que equivale a escrever: $\max\{|x|: x \text{ é valor próprio da matriz } A\} < 1$. Assim, todas as raízes x do polinómio característico da matriz A : $x^p - m_1 x^{p-1} - \dots - m_{p-1} x - m_p = 0$, devem estar dentro do círculo unitário para que o processo seja estacionário.

Considere-se um processo INAR(p) estacionário $\{X_t\}$ que verifica (3.5), onde $\{I_t\}$ é uma sequência de variáveis aleatórias i.i.d., independente de $\{X_t\}$, com distribuição de Poisson de parâmetro λ . Então o valor esperado do processo é dado por:

$$\bar{\mu} = E(\bar{X}_t) = (I - A)^{-1} \bar{C} \lambda,$$

onde I representa a matriz identidade de dimensão $p \times p$.

A forma escalar da última equação tem a seguinte expressão:

$$\mu = E(X_t) = \frac{\lambda}{1 - \sum_{i=1}^p m_i}.$$

Du e Li (1991) chegaram à conclusão de que as estruturas de correlação do processo INAR(p) são iguais às de um processo AR(p) e, assim, que o modelo INAR é semelhante ao modelo AR não só na forma, mas também no domínio estacionário e na estrutura de correlação. Por essa razão estes autores aplicaram a teoria e métodos de estimação desenvolvidos para o modelo AR ao modelo INAR.

PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO GENERALIZADOS, GINAR

Mais tarde, foram construídos processos autorregressivos de valor inteiro generalizados, modelos GINAR, que representam uma generalização dos modelos INAR. A generalização resulta da substituição das variáveis aleatórias de Bernoulli da série de contagem em (3.1) por outras variáveis aleatórias não negativas de valores inteiros i.i.d., obtendo-se, assim, a operação "thinning" generalizada, que irá ser representada pelo símbolo ' \circ ' e que é definida

através de $m_k \circ X = \sum_{i=1}^X Y_{n-1,i}^k$, onde $m_k = E(Y_{n-1,i}^k)$, $0 \leq m_k < \infty$, $k=1, \dots, p$ e $m_p \neq 0$.

Gauthier e Latour (1992) definiram o processo GINAR(p), processo autorregressivo de valor inteiro generalizado de ordem p , da seguinte forma

$$X_n = \sum_{k=1}^p \sum_{i=1}^{X_{n-k}} Y_{n-1,i}^k + I_n, \quad n \geq p, \quad (3.7)$$

ou, alternativamente, através da expressão autorregressiva

$$X_n = \sum_{k=1}^p m_k \circ X_{n-k} + I_n = m_1 \circ X_{n-1} + \dots + m_p \circ X_{n-p} + I_n, \quad n \geq p, \quad (3.8)$$

onde:

- $\{I_n\}$ são variáveis aleatórias inteiras i.i.d. com valores em \mathbb{N} ,
- $\{Y_{n-1,i}^k\}$ são sequências independentes de variáveis aleatórias i.i.d., de médias m_k e variâncias σ_k^2 , com valores em \mathbb{N} e independentes de $\{I_n\}$,
- a igualdade é válida em distribuição.

Em particular, o processo GINAR(1), escreve-se da forma

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i}^1 + I_n,$$

ou, simplificando a notação,

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n-1,i} + I_n, \quad (3.9)$$

e usando a operação “thinning” generalizada,

$$X_n = m \circ X_{n-1} + I_n.$$

Em 1994, Gauthier e Latour (1994) definiram o processo GINAR(p) de um modo diferente. O

índice n , sujeito à condição $n \geq p$, é substituído por $t \in Z$, isto é, $X_t = \sum_{k=1}^p m_k \circ X_{t-k} + I_t$,

onde $t \in Z$, e as constantes $\{m_k\}_{k \in \{1, \dots, p\}}$ são tais que $\forall k \in \{1, 2, \dots, p-1\}$, $m_k \in [0, 1)$,

$m_p \in (0, 1)$ e $\sum_{k=1}^p m_k < 1$. As séries de contagem $\{Y_t^{(k)}\}_{t \in \mathbb{N}}$ associadas a $m_k \circ X_{t-k}$, para

$k=1, \dots, p$, são independentes entre si e independentes de I_t , têm média finita m_k e variância

finita σ_k^2 . Os mesmos autores mostraram que a média deste processo é dada por

$$E(X_t) = \frac{\lambda}{1 - \sum_{i=1}^p m_i}.$$

A seguir, explica-se o que Latour (1998) entende por existência de um processo GINAR(p).

Considere-se um processo estacionário de valores inteiros $\{X_t\}_{t \in Z}$. O processo $\{X_t\}_{t \in Z}$ é

um processo GINAR(p), se e só se existem uma sequência $\{I_t\}_{t \in Z}$ de variáveis aleatórias

inteiras não negativas i.i.d. de média $\lambda > 0$ e variância $b^2 > 0$ e p operadores de Steutel e

van Harn generalizados $m_j \circ$, $j=1, \dots, p$, que satisfazem

$$X_t = \sum_{j=1}^p m_j \circ X_{t-j} + I_t, \quad \forall t \in Z. \quad (3.10)$$

Latour (1998) formulou um critério que garante a existência de um processo GINAR(p) causal estacionário:

Proposição 3.1:

Dados p operadores de Steutel e van Harn generalizados ' $m_i \circ$ ', $i=1, \dots, p$, tal que $\sum_{i=1}^p m_i < 1$,

e uma sequência $\{I_t\}_{t \in Z}$ de variáveis aleatórias inteiras não negativas i.i.d., então existe um processo estacionário $\{X_t\}_{t \in Z}$ de valores inteiros não negativos:

$$X_t = \sum_{i=1}^p m_i \circ X_{t-i} + I_t, \quad t \in Z$$

e $Cov(X_s, I_t) = E(X_s I_t) - E(X_s)E(I_t) = 0, \quad s < t$.

Latour observou que na literatura até 1998 não se afirma que o processo GINAR (e assim também o processo INAR) é um processo AR; Du e Li (1991), por exemplo, escreveram que o processo INAR(p) é semelhante ao processo AR(p). Latour (1998) demonstrou o seguinte corolário que garante que o processo GINAR(p) tem uma representação de um processo AR(p).

Corolário 3.1:

Seja $\{X_t\}_{t \in Z}$ um processo GINAR(p) que verifica (3.10). Então $\{X_t\}_{t \in Z}$ é um processo AR(p) que pode ser escrito da forma

$$X_t - \mu = \sum_{k=1}^p m_k (X_{t-k} - \mu) + \varepsilon_t,$$

onde $\{\varepsilon_t\}_{t \in Z}$ é um processo de ruído branco de variância $b_\varepsilon^2 = \mu \sum_{i=1}^p \sigma_i^2 + b^2$ e σ_i^2 é a variância da série de contagem envolvida no i -ésimo operador de Steutel e van Harn, $i=1, \dots, p$.

3.2. ESTIMAÇÃO DOS PARÂMETROS DOS PROCESSOS AUTORREGRESSIVOS DE VALOR INTEIRO GENERALIZADOS E NÃO GENERALIZADOS

Nesta secção apresentam-se os métodos de estimação propostos na literatura e as propriedades dos estimadores para os parâmetros dos processos GINAR(p), INAR(p) e INAR(1).

Latour (1998) e Gauthier e Latour (1994) propuseram dois métodos para estimar os parâmetros do processo GINAR(p): o método de estimação de Yule-Walker e o método de mínimos quadrados condicionais, que foi desenvolvido por Klimko e Nelson (1978).

Para a estimação dos parâmetros do processo INAR(p) e, em particular, do processo INAR(1) são conhecidas duas abordagens: a estimação no domínio do tempo e a estimação no domínio da frequência. Al-Osh e Alzaid (1987) apresentaram alguns métodos de estimação no domínio do tempo para o processo INAR(1), supondo que $\{I_t\}$ segue uma distribuição de Poisson de parâmetro λ : método de estimação de Yule-Walker, método de mínimos quadrados condicionais e método de máxima verosimilhança condicional. Estes autores observaram que o problema da estimação para o processo INAR(1) é mais complicado do que o problema da estimação para o processo AR(1) porque a distribuição condicional de X_t , dado X_{t-1} no processo INAR(1) é uma convolução da distribuição de $\{I_t\}$ e da distribuição binomial de parâmetro de escala m e de parâmetro de índice X_{t-1} . Silva e Oliveira (2000) ocuparam-se com o problema da estimação no domínio da frequência, usando o critério de Whittle.

Para os modelos INAR(p), Du e Li (1991) propuseram o método de Yule-Walker e o método de mínimos quadrados condicionais de Klimko e Nelson (1978). Franke e Subba Rao (1995) aplicaram o método de máxima verosimilhança condicional aos processos INAR(p).

3.2.1. ESTIMAÇÃO NO DOMÍNIO DO TEMPO

3.2.1.1. MÉTODO DE YULE-WALKER

Considere-se a amostra $\{X_t, t = 1, \dots, n\}$ de observações do processo GINAR(p) definido em

(3.10). Seja $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ a média amostral deste processo e defina-se por

$\hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})$ a função autocovariância amostral para o “lag” k , e por

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \text{ a função autocorrelação amostral para o “lag” } k.$$

Atendendo aos comentários de Hall e Heyde (1980: p. 194), Latour afirmou que, para K fixo, $\hat{\gamma}_k$ converge quase certamente para $\gamma(k)$, a função autocovariância do processo, $k=1, \dots, K$, e \bar{X} converge quase certamente para $E(X_i) = \mu$, $k=1, \dots, K$. E, assim, também se tem que $\hat{\rho}_k$ converge quase certamente para ρ_k , a função autocorrelação do processo, $k=1, \dots, K$.

Latour (1998) demonstrou que o processo GINAR(p) é um processo AR(p). As equações de Yule-Walker para as autocovariâncias do processo GINAR(p) são idênticas às equações de Yule-Walker para as autocovariâncias do processo AR(p), que podem ser escritas da forma

$$\gamma(k) = \sum_{i=1}^p m_i \gamma(i) + b_\varepsilon^2, \quad k=0$$

e (3.11)

$$\gamma(k) = \sum_{i=1}^p m_i \gamma(k-i), \quad 1 \leq k \leq p,$$

onde b_ε^2 é a variância do processo de ruído branco que aparece na representação do processo GINAR(p) como processo AR(p) (ver Corolário 3.1). Estas equações de Yule-Walker podem ser vistas como ligações entre os parâmetros m_i e as autocovariâncias do modelo. Dadas as autocovariâncias, é possível encontrar estimadores para os parâmetros m_i , $i=1, \dots, p$, e b_ε^2 . Os estimadores de Yule-Walker \hat{m}_i são obtidos a partir das equações (3.11) escritas da forma

$$\sum_{i=0}^p \hat{m}_i \hat{\gamma}_{k-i} = -\delta_{0,k} \hat{b}_\varepsilon^2, \quad (3.12)$$

para $k=0, \dots, p$, onde $\hat{m}_0 = -1$ e $\delta_{0,k}$ é o símbolo delta de Kronecker, que vale 1, se $k=0$ e 0 se $k \neq 0$. \hat{b}_ε^2 é um estimador para a variância que se encontra no Corolário 3.1. Dividindo as autocovariâncias das equações (3.12), para $k=1, \dots, p$, por $\hat{\gamma}_0$, então o novo sistema de equações pode ser representado de uma forma compacta:

$$\hat{\Gamma} \hat{\mathbf{m}} = \hat{\boldsymbol{\rho}}, \quad (3.13)$$

onde $\hat{\Gamma} = [\hat{\rho}_{|i-j|}]_{p \times p}$, $\hat{m} = (\hat{m}_1, \dots, \hat{m}_p)^T$ e $\hat{\rho} = (\hat{\rho}_1, \dots, \hat{\rho}_p)^T$. Para $j=1, \dots, p$, os estimadores \hat{m}_j são funções lineares de $\hat{\rho}_k$, $k=1, \dots, p$, e \hat{b}_ε^2 é uma função linear de $\hat{\gamma}_k$, $k=1, \dots, p$. Daí deduz-se que \hat{m}_j converge quase certamente para m_j , $j=1, \dots, p$, e \hat{b}_ε^2 converge quase certamente para b_ε^2 . Latour (1998) mostrou assim, que os estimadores obtidos pelo método de Yule-Walker são fortemente consistentes.

Du e Li (1991) aplicaram o método de Yule-Walker ao processo INAR(p) definido em (3.3) e, para além dos resultados encontrados para o processo GINAR(p), apresentaram ainda um estimador para $\lambda = E(I_t)$ dado por

$$\hat{\lambda} = (1 - \hat{m}_1 - \dots - \hat{m}_p) \bar{X} \quad (3.14)$$

e um estimador para $b^2 = Var(I_t)$, que pode ser calculado com a ajuda da fórmula

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

onde $\hat{I}_t = X_t - \hat{m}_1 X_{t-1} - \dots - \hat{m}_p X_{t-p}$ e $\bar{I}_n = \frac{1}{n-p} \sum_{t=p+1}^n \hat{I}_t$.

Estes autores indicaram que se pode provar a consistência forte de $\hat{\lambda}$ e \hat{b}^2 , usando a consistência forte de $\hat{\rho}_k$.

O estimador de Yule-Walker para o parâmetro m do processo INAR(1), definido em (3.2), que é obtido substituindo na equação (3.12) k por 1 e p por 1 e resolvendo a nova equação em ordem a \hat{m} , exprime-se por:

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

Foi suposto que $\{I_t\}$ segue uma distribuição de Poisson de parâmetro λ ; então, um estimador para a média λ , pode ser obtido através de

$$\hat{\lambda} = \frac{1}{n} \sum_{t=1}^n \hat{I}_t,$$

onde $\hat{I}_t = X_t - \hat{m}X_{t-1}$ é calculado para $t=1, \dots, n$. Note-se que esta expressão para $\hat{\lambda}$ é igual à expressão (3.14) obtida por Du e Li, em que $p=1$: $\hat{\lambda} = (1 - \hat{m})\bar{X}$.

Park e Oh (1997) demonstraram que os estimadores de Yule-Walker \hat{m} e $\hat{\mu}$ ($\hat{\mu} = \bar{X}$) do processo INAR(1) têm distribuição assintoticamente normal:

$$\sqrt{n} \left(\begin{pmatrix} \hat{\mu} \\ \hat{m} \end{pmatrix} - \begin{pmatrix} \mu \\ m \end{pmatrix} \right) \xrightarrow{d} N(\vec{0}, W),$$

onde W é uma matriz simétrica dada por

$$W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = \begin{pmatrix} \frac{\lambda(1+m)}{(1-m)^2} & 0 \\ 0 & \frac{m(1-m)^2}{\lambda} + (1-m^2) \end{pmatrix}.$$

Observa-se que \hat{m} e $\hat{\mu}$ são assintoticamente independentes.

Já se sabe que os estimadores de Yule-Walker $\hat{m}_1, \dots, \hat{m}_p$, $\hat{\lambda}$ e \hat{b}^2 para os parâmetros do processo INAR(p) são fortemente consistentes. Como o processo INAR(1) é um modelo particular do processo INAR(p) em que $p=1$, então, transferindo essas propriedades para os parâmetros do processo INAR(1), tem-se: \hat{m} , $\hat{\lambda} = (1 - \hat{m})\bar{X}$ e $\hat{b}^2 = \frac{1}{n-1} \sum_{t=2}^n (\hat{I}_t - \bar{I}_n)^2$ onde

$\hat{I}_t = X_t - \hat{m}X_{t-1}$ e $\bar{I}_n = \frac{1}{n-1} \sum_{t=2}^n \hat{I}_t$, são estimadores fortemente consistentes.

3.2.1.2. MÉTODO DE MÍNIMOS QUADRADOS CONDICIONAIS

O método de mínimos quadrados condicionais utilizado por Gauthier e Latour (1994), por Du e Li (1991) e por Al-Osh e Alzaid (1987) foi desenvolvido por Klimko e Nelson (1978) e baseia-se na minimização da soma dos quadrados dos desvios à volta da esperança condicional.

Sejam $F_t = F(X_1, \dots, X_t)$, uma σ -álgebra gerada por X_1, \dots, X_t , $\vec{\beta} = (m_1, \dots, m_p, \lambda)^T$ e $g(\vec{\beta}, F_{t-1}) = E(X_t | F_{t-1}) = \lambda + m_1 X_{t-1} + \dots + m_p X_{t-p}$, onde $t > p$. O estimador de mínimos

quadrados condicionais $\hat{\vec{\beta}}$ de $\vec{\beta}$ obtém-se minimizando

$$Q_n(\vec{\beta}) = \sum_{t=p+1}^n (X_t - g(\vec{\beta}, F_{t-1}))^2 = \sum_{t=p+1}^n \left(X_t - \sum_{i=1}^p m_i X_{t-i} - \lambda \right)^2.$$

Assim, os estimadores de mínimos quadrados condicionais para λ e m_i , $i=1, \dots, p$, encontram-se resolvendo as equações

$$\frac{\partial Q_n(\vec{\beta})}{\partial m_i} = 0, \quad i=1, \dots, p,$$

$$\frac{\partial Q_n(\vec{\beta})}{\partial \lambda} = 0$$

Gauthier e Latour (1994) mostraram que os estimadores de mínimos quadrados condicionais $\hat{m}_1^*, \dots, \hat{m}_p^*$ e $\hat{\lambda}^*$ são estimadores fortemente convergentes para os parâmetros

m_1, \dots, m_p e λ do processo GINAR(p), se $\gamma(0) > 0$ e se $\sum_{h=0}^{\infty} |\varphi_k(h)| < \infty$, onde

$\gamma(0) = \mu \sum_{i=1}^p \sigma_i^2 + \sum_{i=1}^p m_i \gamma(i) + b^2$ é a variância de X_t , $\gamma(k) = \sum_{i=1}^p m_i \gamma(k-i)$, $k > 0$, e

$$\varphi_k(h) = \text{Cov}(X_t X_{t+k}, X_{t+h} X_{t+k+h}).$$

Para os parâmetros do processo INAR(p), Du e Li (1991) provaram que $\hat{m}_1^*, \dots, \hat{m}_p^*$ e $\hat{\lambda}^*$ são fortemente convergentes.

Na seguinte proposição de Latour (1998) é estabelecido em que condições os estimadores de mínimos quadrados condicionais para os parâmetros m_j , $j=1, \dots, p$ e λ são assintoticamente normais.

Proposição 3.2:

Sejam $\{X_t, t=1, \dots, N\}$ observações do processo GINAR(p), tal que $E(I_t^3) < \infty$ e $E\{Y_t^{(i,j)}\}^3 < \infty$, $1 \leq i \leq p$, $1 \leq j$, $t < \infty$. Seja $\vec{\beta} = (m_1, \dots, m_p, \lambda)^T$ e designe-se por $\hat{\vec{\beta}}^*$ o estimador de mínimos quadrados condicionais de $\vec{\beta}$. Então

$$\sqrt{n}(\hat{\vec{\beta}}^* - \vec{\beta}) \xrightarrow{d} N_{p+1}(\vec{0}, V^{-1} W V^{-1}),$$

onde

$$V = \begin{pmatrix} \Gamma_p + E(X_t)^2 \bar{1}_p \bar{1}_p^T & E(X_t) \bar{1}_p \\ E(X_t) \bar{1}_p^T & 1 \end{pmatrix},$$

em que $\Gamma_p = \{\gamma(i-j)\}_{1 \leq i, j \leq p}$ e $\bar{1}_p = \underbrace{(1, \dots, 1)}_{p \text{ vezes}}$, e os elementos da matriz W são

$$W_{ij} = \sum_{k=1}^p \sigma_k^2 E(X_{p-i+1} X_{p-j+1} X_{p-k+1}) + b^2 V_{ij}, \quad 1 \leq i, j \leq p+1.$$

Du e Li (1991) mostraram que os estimadores de mínimos quadrados condicionais $\hat{m}_1^*, \dots, \hat{m}_p^*$ e $\hat{\lambda}^*$ podem ser escritos da forma

$$\hat{\lambda}^* = \bar{X}_n^{(0)} - \sum_{j=1}^p \hat{m}_j^* \bar{X}_n^{(j)} \quad (3.15)$$

$$\hat{\Gamma}^* \hat{\vec{m}}^* = \hat{\vec{\rho}}^*, \quad (3.16)$$

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

$$\hat{\vec{m}}^* = (\hat{m}_1^*, \dots, \hat{m}_p^*)^T, \quad \hat{\Gamma}^* = \left[\hat{\rho}_{|i-j|}^* \right]_{p \times p} \quad \text{e} \quad \hat{\vec{\rho}}^* = (\hat{\rho}_1^*, \dots, \hat{\rho}_p^*)^T.$$

De (3.13), (3.14), (3.15) e (3.16) é fácil de ver que, para n grande, as diferenças $\hat{\Gamma}^* - \hat{\Gamma}$, $\hat{\vec{\rho}}^* - \hat{\vec{\rho}}$ e $\bar{X}_n^{(j)} - \bar{X}$ vão estar próximas de zero, isso implica que $\hat{\vec{m}}^*$ vai estar próximo de $\hat{\vec{m}}$ e $\hat{\lambda}^*$, de $\hat{\lambda}$. Daqui deduz-se que, para amostras grandes, os estimadores de Yule-Walker para $\vec{m} = (m_1, \dots, m_p)^T$ e λ são semelhantes aos estimadores de mínimos quadrados condicionais.

Al-Osh e Alzaid (1987) calcularam os estimadores de mínimos quadrados condicionais para os parâmetros m e λ do processo INAR(1). Derivando $\sum_{t=2}^n (X_t - m_1 X_{t-1} - \lambda)^2$ em ordem a m e λ , igualando as derivadas a zero e resolvendo em ordem a m e λ , obtém-se:

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

e

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

Al-Osh e Alzaid concluíram que estes estimadores são fortemente consistentes. E se $E(I_t^3) < \infty$, esta condição é verificada para a distribuição de Poisson, então $(\hat{m}^*, \hat{\lambda}^*)$ são assintoticamente normalmente distribuídos e a distribuição assintótica pode ser obtida a partir do resultado da Proposição 3.2.

Park e Oh (1997) encontraram a distribuição assintótica conjunta dos estimadores de mínimos quadrados condicionais \hat{m}^* e $\hat{\lambda}^*$ do processo INAR(1), sendo

$$\hat{\mu}^* = \frac{1}{n(1-m)} \left(\sum_{t=1}^n X_t - \hat{m}^* \sum_{t=1}^n X_{t-1} \right)$$

o estimador de mínimos quadrados condicionais para $\mu = E(X_t) = \frac{\lambda}{1-m}$:

$$\sqrt{n} \left(\begin{pmatrix} \hat{m}^* \\ \hat{\mu}^* \end{pmatrix} - \begin{pmatrix} m \\ \mu \end{pmatrix} \right) \xrightarrow{d} N(\vec{0}, V^{-1} W V^{-1}),$$

onde

$$V^{-1} W V^{-1} = \begin{pmatrix} \frac{m(1-m)^2}{\lambda} + (1-m^2) & -m - \frac{2\lambda(1+m)}{(1-m)} \\ -m - \frac{2\lambda(1+m)}{(1-m)} & \frac{\lambda}{(1-m)^2} \left(\frac{4\lambda}{(1-m)} + 1 \right) (m+1) \end{pmatrix}.$$

3.2.1.3. MÉTODO DE MÁXIMA VEROSIMILHANÇA CONDICIONAL

Franke e Subba Rao (1995) escreveram os processos INAR(p), estudados por Du e Li (1991), como processos INAR(1) p -variados e aplicaram a teoria de Billingsley (1961) aos estimadores de máxima verosimilhança condicional dos parâmetros destes processos, que verificam a propriedade de Markov.

Franke e Subba Rao definiram o processo INAR(1) p -variado, uma série temporal $\{\vec{X}_t, -\infty < t < \infty\}$ com valores em \mathbb{N}^p_0 da forma

$$\vec{X}_t = A * \vec{X}_{t-1} + \vec{I}_t, \quad (3.17)$$

onde A é uma matriz de dimensão $p \times p$ com entradas $0 \leq a_{ij} \leq 1$ ($i, j=1, \dots, p$), \vec{I}_t são variáveis aleatórias i.i.d. com valores em \mathbb{N}^p_0 , $-\infty < t < \infty$, e ‘*’ é a operação “thinning” vectorial.

Os mesmos autores demonstraram o seguinte teorema que pode ser usado para verificar se um processo INAR(1) multivariado é estacionário.

Teorema 3.2:

Seja $\{\vec{X}_t\}$, verificando (3.17), uma cadeia de Markov irreduzível, aperiódica em \mathbb{N}^p_0 . Se

$E\|\vec{I}_t\| < \infty$ e se o maior valor próprio da matriz A for menor do que 1, então existe um processo INAR(1) p -variado estritamente estacionário que verifica a equação (3.17).

Considere-se o processo INAR(p), definido em (3.3), mas com a diferença de $m_i \in (0,1)$ em vez de $m_i \in [0,1]$, para $i=1, \dots, p$, e represente-se por θ o parâmetro do qual depende a distribuição da inovação $\{I_t\}$. A partir de uma amostra finita $\vec{X} = (X_0, \dots, X_n)$ do processo encontra-se o estimador de máxima verosimilhança condicional (X_0 é dado) $\hat{\beta} = (\hat{m}_1, \dots, \hat{m}_p, \hat{\theta})$ para $\beta = (m_1, \dots, m_p, \theta)$, maximizando o logaritmo da função de verosimilhança condicional

$$l_n(\vec{X}, \beta | X_0) = \sum_{t=1}^n \log P_\beta(\vec{X}_t | \vec{X}_{t-1}).$$

Para que os resultados de Billingsley (1961) para estimadores dos parâmetros de processos de Markov também sejam válidos para os estimadores dos parâmetros do processo INAR(p), é necessário que o processo verifique a condição de estacionaridade do Teorema 3.2 e outras seis condições de regularidade¹⁴. Franke e Subba Rao provaram que, sob essas condições, o estimador de máxima verosimilhança condicional $\hat{\beta} = (\hat{m}_1, \dots, \hat{m}_p, \hat{\theta})$ é assintoticamente normal:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{d} N(\vec{0}, \Sigma^{-1}(\beta)), \quad (3.18)$$

onde $\Sigma^{-1}(\beta)$ é a inversa da matriz de informação de Fisher $\Sigma(\beta) = (\sigma_{uv}(\beta))_{u,v=1, \dots, p+1}$ com entradas dadas por $\sigma_{uv}(\beta) = E\left(\frac{\partial}{\partial \beta_u} \log P_\beta(\vec{X}_2 | \vec{X}_1) \frac{\partial}{\partial \beta_v} \log P_\beta(\vec{X}_2 | \vec{X}_1)\right)$, $u, v=1, \dots, p+1$.

Al-Osh e Alzaid (1987) aplicaram o método de máxima verosimilhança condicional e não condicional ao processo INAR(1).

A função de verosimilhança associada a uma amostra de $(n+1)$ observações do processo INAR(1) escreve-se da forma

$$L(\vec{X}, m, \lambda) = \left(\prod_{t=1}^n P_t(X_t) \right) \frac{[\lambda/(1-m)]^{X_0}}{X_0!} \exp[-\lambda/(1-m)], \quad (3.19)$$

onde $P_t(X_t) = \exp(-\lambda) \sum_{i=0}^{\min(X_{t-1}, X_t)} \frac{\lambda^{X_t-i}}{(X_t-i)!} \binom{X_{t-1}}{i} m^i (1-m)^{X_{t-1}-i}$, para $t=1, \dots, n$ e

$\vec{X} = (X_0, \dots, X_n)$. Considerando os estimadores de máxima verosimilhança condicional de m e λ quando X_0 é dado, a função de verosimilhança condicional calcula-se assim:

¹⁴ Ver Franke e Subba Rao (1995: p. 8).

$$L(\vec{X}, m, \lambda | X_0) = \prod_{t=1}^n P_t(X_t).$$

Usando o método que Sprott (1983) descreveu para maximizar $L(\vec{X}, m, \lambda | X_0)$ em ordem a m e λ , quando $P_t(X)$ é independente de t (parâmetro do índice das observações), então obtêm-se as seguintes expressões para as derivadas do logaritmo da função de verosimilhança condicional em ordem a λ e m :

$$S_\lambda = \sum_{t=1}^n \frac{\partial \log P_t(X_t)}{\partial \lambda} = \sum_{t=1}^n P_t(X_t - 1)/P_t(X_t) - n = \sum_{t=1}^n H(t) - n$$

e

$$S_m = \sum_{t=1}^n \frac{\partial \log P_t(X_t)}{\partial m} = \sum_{t=1}^n [(X_t - mX_{t-1}) - \lambda H(t)]/m(1-m).$$

Igualando as duas expressões a zero tem-se: $\sum_{t=1}^n X_t = m \sum_{t=1}^n X_{t-1} + n\lambda$. Resolvendo esta equação em ordem a m (λ) e substituindo em S_λ (S_m), encontra-se um estimador de máxima verosimilhança condicional para λ (m).

Os estimadores de máxima verosimilhança não condicional obtêm-se a partir da função de verosimilhança (3.19). Derivando o logaritmo dessa função em ordem a λ e m , encontra-se:

$$S_\lambda^0 = \frac{\partial \log L}{\partial \lambda} = S_\lambda + \frac{X_0}{\lambda} - \frac{1}{1-m} \quad (3.20)$$

e

$$S_m^0 = \frac{\partial \log L}{\partial m} = S_m + \frac{1}{1-m} \left(X_0 - \frac{\lambda}{1-m} \right). \quad (3.21)$$

Os estimadores de máxima verosimilhança obtêm-se igualando $S_\lambda^0 = 0$ e $S_m^0 = 0$; daqui resulta que

$$\lambda = \left(\sum_{t=0}^n X_t - m \sum_{t=1}^n X_t \right) / \left(n + \frac{1+m}{1-m} \right),$$

e usando esta equação para eliminar λ em (3.21) ou, resolvendo a equação em ordem a m , para eliminar m em (3.20). Analisando estas três últimas equações, observa-se que os estimadores de máxima verosimilhança não condicional não são tão simples de calcular como os estimadores de máxima verosimilhança condicional.

Já foi visto que os estimadores de máxima verosimilhança condicional $\hat{m}_1, \dots, \hat{m}_p$ e $\hat{\theta}$ (θ é um parâmetro do qual depende a distribuição da inovação) para os parâmetros do processo

INAR(p) são assintoticamente normais, se forem verificadas sete condições¹⁵. Assim conclui-se que, se o processo INAR(1) verificar essas sete condições, então os estimadores \hat{m} e $\hat{\theta}$, $\hat{\beta} = (\hat{m}, \hat{\theta})$, são assintoticamente normais com distribuição assintótica dada por (3.18), em que $p=1$.

3.2.2. ESTIMAÇÃO NO DOMÍNIO DA FREQUÊNCIA-CRITÉRIO DE WHITTLE

Whittle (1953) propôs um método de estimação para processos Gaussianos em que a verosimilhança do processo é representada através das suas propriedades espectrais. Para processos Gaussianos, maximizar o logaritmo da função de verosimilhança é assintoticamente equivalente a minimizar o critério de Whittle que é dado por

$$\hat{L}_n(\vec{\theta}) = \frac{n}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f(\omega, \vec{\theta}) + \frac{I_n(\omega)}{f(\omega, \vec{\theta})} \right\} d\omega, \quad (3.22)$$

onde $f(\omega, \vec{\theta})$ representa a função de densidade espectral não normalizada do processo,

$I_n(\omega)$ é o periodograma dado por $I_n(\omega) = (2\pi n)^{-1} \left| \sum_{t=1}^n X_t e^{i\omega t} \right|^2$ e $\vec{\theta}$ contém os parâmetros

que se pretendem estimar; no caso dos processos INAR(p) tem-se $\vec{\theta} = (m_1, \dots, m_p, \lambda)$.

Whittle (1954) e Walker (1964) mostraram que o estimador de Whittle $\hat{\theta}$, que é obtido minimizando (3.22), é fracamente consistente e tem distribuição assintoticamente normal. Chandler (1995) mostrou que o critério de Whittle também pode ser usado para estimar os parâmetros de uma série não Gaussiana. Rice (1979) provou que, nesse caso, os estimadores também são consistentes e assintoticamente normais, mas é difícil obter resultados assintóticos, porque a variância assintótica de $(\hat{\theta} - \vec{\theta}^0)$, $\vec{\theta}^0$ contém os verdadeiros valores dos parâmetros, depende do triespectro, o espectro de quarta ordem.

A minimização do critério de Whittle é conseguida substituindo $\hat{L}_n(\vec{\theta})$ em (3.22) por

$$\hat{l}_n(\vec{\theta}) = \frac{1}{n} \sum_{j=1}^{\lfloor n/2 \rfloor} \left\{ \log f(\omega_j, \vec{\theta}) + \frac{I_n(\omega_j)}{f(\omega_j, \vec{\theta})} \right\},$$

¹⁵ A condição do Teorema 3.2 e seis condições de regularidade (v. n. 14).

onde $f(\omega_j, \vec{\theta})$ é a função de densidade espectral no ponto de frequência $\omega_j = 2\pi j/n$, $I_n(\omega_j)$ é a ordenada do periodograma no mesmo ponto de frequência e $[n/2] = n/2$, se n é par, e $[n/2] = (n-1)/2$, se n é ímpar.

3.3. COMPARAÇÃO DOS MÉTODOS DE ESTIMAÇÃO

Gauthier e Latour (1994) e Latour (1998) aplicaram o método de Yule-Walker e o método de mínimos quadrados condicionais ao processo GINAR(p) para encontrar estimadores para os parâmetros m_1, \dots, m_p e λ . Os estimadores de Yule-Walker $\hat{m}_1, \dots, \hat{m}_p$, \bar{X} , $\hat{\rho}_k$, $\hat{\gamma}_k$ e \hat{b}_ε^2 são fortemente consistentes. Para que os estimadores de mínimos quadrados condicionais sejam fortemente consistentes são necessárias duas condições: 1) $\gamma(0) > 0$ e 2)

$\sum_{h=0}^{\infty} |\varphi_k(h)| < \infty$. Sabendo que $E(I_t^3) < \infty$ e $E\{Y_t^{(i,j)}\}^3 < \infty$, então os estimadores de mínimos

quadrados condicionais são assintoticamente normais. A distribuição assintótica dos estimadores de Yule-Walker não foi apresentada pelos autores Latour e Gauthier. Destes dois métodos é preferível o método de mínimos quadrados condicionais, apesar de serem necessárias algumas condições para que se verifiquem as propriedades assintóticas, porque permite a construção de intervalos de confiança assintóticos para os parâmetros m_1, \dots, m_p e λ .

Dos três métodos estudados para o processo INAR(p), o método de mínimos quadrados condicionais oferece melhores resultados assintóticos do que o método de Yule-Walker e do que o método de máxima verosimilhança condicional. Os estimadores de mínimos quadrados condicionais $\hat{\lambda}^*$ e $\hat{m}_1^*, \dots, \hat{m}_p^*$ são fortemente consistentes e, se $E(I_t^3) < \infty$, estes estimadores são assintoticamente normais. Para que os estimadores de máxima verosimilhança $\hat{m}_1, \dots, \hat{m}_p$ e $\hat{\theta}$ sejam assintoticamente normais é preciso que se verifiquem sete condições já referidas. Franke e Subba Rao (1995) não escreveram se estes estimadores são consistentes. Dos estimadores $\hat{m}_1, \dots, \hat{m}_p$, $\hat{\lambda}$ e \hat{b}^2 , obtidos pelo método de Yule-Walker, só se sabe que eles são fortemente consistentes. Assim como Gauthier e Latour (1994) e Latour (1998) não apresentaram a distribuição no limite dos estimadores de Yule-Walker, Du e Li (1991) também não apresentaram. Viu-se que para amostras grandes, os estimadores de Yule-Walker $\hat{m}_1, \dots, \hat{m}_p$ e $\hat{\lambda}$ são muito semelhantes aos estimadores de mínimos quadrados condicionais. Franke e Subba Rao não apresentaram a

forma explícita dos estimadores de máxima verosimilhança condicional, mas, analisando o método como são encontrados estes estimadores, pode-se concluir que é muito mais complicado e trabalhoso calcular estes estimadores do que calcular os estimadores de Yule-Walker e do que calcular os estimadores de mínimos quadrados condicionais.

O modelo INAR(1) é o mais simples, que contém menos parâmetros, entre os modelos do grupo dos processos autorregressivos de valor inteiro. Por isso apareceram na literatura mais estudos sobre a estimação dos parâmetros deste modelo do que sobre a estimação dos parâmetros dos modelos INAR(p) e GINAR(p). Os estimadores obtidos pelo método de Yule-Walker \hat{m} , $\hat{\lambda}$ e \hat{b}^2 são fortemente consistentes e os estimadores \hat{m} e $\hat{\mu}$ têm distribuição assintótica normal e são assintoticamente independentes. Comparando estes dois estimadores com os estimadores \hat{m}^* e $\hat{\mu}^*$, obtidos pelo método de mínimos quadrados condicionais, que também são assintoticamente normais, pode observar-se que a distribuição, no limite, dos estimadores para m , obtidos pelos dois métodos, é igual; isto confirma o resultado de Du e Li de os estimadores de Yule-Walker e os estimadores de mínimos quadrados condicionais para m e λ serem semelhantes para amostras grandes; mas da comparação das variâncias assintóticas de $(\hat{\mu} - \mu)$ e $(\hat{\mu}^* - \mu)$, conclui-se que o estimador de Yule-Walker $\hat{\mu} = \bar{X}$, a média amostral, é assintoticamente mais eficiente do que o estimador de mínimos quadrados condicionais. Também se sabe que os estimadores de mínimos quadrados condicionais \hat{m}^* e $\hat{\lambda}^*$ são fortemente consistentes e, se $E(|I_t|^3) < \infty$, então $(\hat{m}^*, \hat{\lambda}^*)$ têm distribuição assintótica normal.

Supondo que $\{I_t\}$ segue uma distribuição de Poisson de parâmetro λ , então os estimadores de máxima verosimilhança condicional \hat{m} e $\hat{\lambda}$ são assintoticamente normais se forem verificadas as já referidas sete condições. A propriedade de consistência não foi estudada para estes estimadores. Os estimadores de Whittle \hat{m} e $\hat{\lambda}$ só são fracamente consistentes. A distribuição assintótica destes estimadores é normal, mas é difícil calcular a variância assintótica, porque ela depende do espectro de quarta ordem. Isso torna difícil a comparação da eficiência destes estimadores com a eficiência dos estimadores obtidos pelos outros métodos. O mesmo acontece com os estimadores de máxima verosimilhança condicional: a variância assintótica desses estimadores é calculada com a ajuda da matriz de informação de Fisher, que é difícil de obter. As expressões dos estimadores de Yule-Walker e dos estimadores de mínimos quadrados condicionais são mais simples do que as expressões dos estimadores dos outros métodos.

O estudo de Silva e Oliveira (2000), que aplicaram o critério de Whittle para estimar os parâmetros m e λ ($\vec{\theta} = (m, \lambda)$) do processo INAR(1) de Poisson e compararam os resultados e as propriedades não assintóticas dos estimadores de Whittle com os resultados e as propriedades dos estimadores de máxima verosimilhança condicional, indica que as propriedades não assintóticas (viés, erro quadrático médio, curtose, assimetria) dos estimadores de máxima verosimilhança condicional são muito parecidas com as propriedades dos estimadores de Whittle; isto acontece porque o método de Whittle também parte da função de verosimilhança do processo, mas com a diferença de ser representada através das propriedades espectrais. O método de estimação de Whittle é mais vantajoso do que o método de máxima verosimilhança condicional, porque é mais simples e mais rápido ao efectuar os cálculos.

O estudo feito por Alzaid e Al-Osh (1987), em que são comparadas as propriedades não assintóticas dos estimadores de Yule-Walker, dos estimadores de mínimos quadrados condicionais e dos estimadores de máxima verosimilhança condicional, revela que o método de máxima verosimilhança condicional, apesar de ser o método mais trabalhoso de todos, oferece melhores resultados não assintóticos. O pior destes três métodos, no sentido de se atingirem viés e erros quadrático médios mais altos, é o método de Yule-Walker.

Destas últimas considerações conclui-se que para estimar os parâmetros m e λ do processo INAR(1), o método de máxima verosimilhança condicional e o método de estimação de Whittle, sendo mais simples e mais rápido efectuar os cálculos deste último método, levam a melhores propriedades não assintóticas do que o método de Yule-Walker e do que o método de mínimos quadrados condicionais. E o método que apresenta os piores resultados não assintóticos é o método de Yule-Walker.

Analisando as propriedades assintóticas, parece que as conclusões estão em contradição com as conclusões tiradas dos resultados não assintóticos, porque os estimadores de Whittle só são fracamente consistentes e não se sabe se os estimadores de máxima verosimilhança condicional são consistentes, enquanto que os estimadores obtidos pelos outros métodos são todos fortemente consistentes. A normalidade assintótica dos estimadores de máxima verosimilhança condicional não se verifica tão facilmente como a normalidade assintótica dos estimadores de mínimos quadrados condicionais: no método de máxima verosimilhança condicional é preciso que se verifiquem sete condições, no método de mínimos quadrados condicionais só é necessária uma condição. Quanto à eficiência assintótica dos estimadores só se pode dizer que o estimador de Yule-Walker \hat{m} e o estimador de mínimos quadrados condicionais \hat{m}^* são igualmente eficientes e, considerando os estimadores para μ , viu-se que \bar{X} é mais eficiente do que $\hat{\mu}^*$. Assim, pode-se concluir que para estimar os parâmetros m e μ , o método de Yule-Walker tem

propriedades assintóticas mais desejáveis do que o método de mínimos quadrados condicionais. E o método de mínimos quadrados condicionais tem propriedades assintóticas mais acessíveis do que o método de máxima verosimilhança condicional.

4. RELAÇÃO ENTRE O PROCESSO BGWI E O PROCESSO GINAR

Dion, Gauthier e Latour (1995) mostraram que os processos GINAR(p), $p \geq 1$, podem ser identificados como processos ramificados multivariados com imigração, BGWI(p).

Se $p=1$, é fácil de ver que o modelo GINAR(1) é igual ao modelo BGWI(1) e, como o modelo INAR(1) para $t \geq 1$ é um caso particular do modelo GINAR(1), conclui-se que o processo INAR(1) também é um processo BGWI(1). A seguir, mostra-se que o processo GINAR(p), pode ser visto como um processo BGWI(p). Escrevendo a expressão (3.8) do modelo GINAR(p) definido em (3.7):

$$X_n = \sum_{k=1}^p m_k \circ X_{n-k} + I_n = m_1 \circ X_{n-1} + \dots + m_p \circ X_{n-p} + I_n,$$

na forma de espaço de estados, obtém-se:

$$\vec{X}_n = A \circ \vec{X}_{n-1} + \vec{C}I_n, \quad (4.1)$$

onde $\vec{X}_n = (X_n, X_{n-1}, \dots, X_{n-p+1})^T$, $\vec{C} = (1, 0, \dots, 0)^T$ e A representa a matriz (3.6):

$$A = \begin{pmatrix} m_1 & m_2 & \dots & m_{p-1} & m_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Efectuando cálculos, a igualdade (4.1) também pode ser escrita de uma forma equivalente:

$$\begin{aligned} (X_n, X_{n-1}, \dots, X_{n-p+1})^T &= (m_1 \circ X_{n-1} + \dots + m_p \circ X_{n-p} + I_n, X_{n-1}, \dots, X_{n-p+1})^T = \\ &= \left(\sum_{i=1}^{X_{n-1}} Y_{n-1,i}^1 + \dots + \sum_{k=1}^{X_{n-p}} Y_{n-1,i}^p + I_n, X_{n-1}, \dots, X_{n-p+1} \right)^T = \sum_{i=1}^{X_{n-1}} (Y_{n-1,i}^1, 1, 0, \dots, 0)^T + \\ &\sum_{i=1}^{X_{n-2}} (Y_{n-1,i}^2, 0, 1, 0, \dots, 0)^T + \dots + \sum_{i=1}^{X_{n-p+1}} (Y_{n-1,i}^{p-1}, 0, \dots, 0, 1)^T + \sum_{i=1}^{X_{n-p}} (Y_{n-1,i}^p, 0, \dots, 0)^T + (I_n, 0, \dots, 0)^T \end{aligned} \quad (4.2)$$

Estabelecendo a forma dos vectores \vec{I}_n e $\vec{Y}_{n-1,i}^k$ da seguinte maneira:

$$\vec{I}_n = (I_n, 0, \dots, 0)^T$$

$$\vec{Y}_{n-1,i}^k = (Y_{n-1,i}^k, 0, \dots, 0)^T + e_{k+1} \text{ para } k=1, \dots, p, \text{ onde } e_j = (\delta_{j1}, \dots, \delta_{jp})^T \text{ e } \delta_{ji} \text{ representa o símbolo}$$

delta de Kronecker, a igualdade (4.2) transforma-se na expressão:

$$\vec{X}_n = (X_n, X_{n-1}, \dots, X_{n-p+1})^T = \sum_{i=1}^{X_{n-1}} \vec{Y}_{n-1,i}^1 + \sum_{i=1}^{X_{n-2}} \vec{Y}_{n-1,i}^2 + \dots + \sum_{i=1}^{X_{n-p}} \vec{Y}_{n-1,i}^p + \vec{I}_n. \quad (4.3)$$

Definindo agora o vector \vec{Z}_n por $\vec{Z}_n = (Z_n(1), \dots, Z_n(p))^T$, onde as componentes são determinadas através da igualdade $Z_n(j) = X_{n-j+1}$ com $1 \leq j \leq p$, obtém-se a partir da equação (4.3):

$$\begin{aligned} \vec{Z}_n &= (Z_n(1), \dots, Z_n(p))^T = (X_n, X_{n-1}, \dots, X_{n-p+1})^T = \\ &= \sum_{i=1}^{Z_{n-1}(1)} \vec{Y}_{n-1,i}^1 + \sum_{i=1}^{Z_{n-1}(2)} \vec{Y}_{n-1,i}^2 + \dots + \sum_{i=1}^{Z_{n-1}(p)} \vec{Y}_{n-1,i}^p + \vec{I}_n = \sum_{k=1}^p \sum_{i=1}^{Z_{n-1}(k)} \vec{Y}_{n-1,i}^k + \vec{I}_n, \end{aligned}$$

que verifica a igualdade (2.7) da definição do processo BGWI(p) para $n \geq p-1$. E, assim, chega-se à conclusão, que é possível identificar o processo $\{\vec{Z}_n\}_{n \geq p-1}$, obtido a partir do processo GINAR(p), como um processo BGWI(p) e $X_n = Z_n(1)$. E sendo o processo INAR(p) um caso particular do GINAR(p), então esse processo também pode ser identificado como um processo BGWI(p).

Estes últimos resultados estão resumidos na seguinte proposição de Dion, Gauthier e Latour (1995).

Proposição 4.1: Dado um processo GINAR(p) $\{X_n\}$, definido em (3.7):

$X_n = \sum_{k=1}^p \sum_{i=1}^{X_{n-k}} Y_{n-1,i}^k + I_n$, o processo $\{\vec{Z}_n\}_{n \geq p-1}$, definido por $Z_n(j) = X_{n+1-j}$, $1 \leq j \leq p$, é um

processo BGWI(p) com vectores aleatórios de imigração $\{\vec{I}_n = (I_n, 0, \dots, 0)^T\}$ e com matriz M - "offspring mean matrix" - dada por

$$M = \begin{pmatrix} m_1 & 1 & & 0 \\ m_2 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ m_p & \vdots & & 0 \end{pmatrix},$$

com entradas $m_{kj} = E(Y_{n-1,i}^k(j))$, $k, j=1, \dots, p$, que podem ser encontradas com a ajuda da fórmula $E(\vec{Y}_{n-1,i}^k) = (m_k, 0, \dots, 0)^T + e_{k+1} = (m_{k1}, \dots, m_{kd})^T$.

Lewis (1942) e Leslie (1945, 1948) obtiveram os dois resultados seguintes, que relacionam os valores próprios da matriz M - as raízes do polinómio característico associado a M - com os valores esperados das variáveis $Y_{n-1,i}^k$, $k=1, \dots, p$.

Proposição 4.2:

1. O polinómio característico da matriz M é $x^p - \sum_{k=1}^p m_k x^{p-k} = 0$.
2. Se ρ é o valor próprio máximo de M , então $\sum_{k=1}^p m_k \leq 1 \Leftrightarrow \rho \leq 1$ e $\sum_{k=1}^p m_k \geq 1 \Leftrightarrow \rho \geq 1$.

A condição $\sum_{k=1}^p m_k < 1$ tem como consequência que todas as raízes do polinómio característico da matriz M estejam dentro do círculo unitário.

Dion, Gauthier e Latour (1995) provaram a seguinte proposição que indica em que condições o processo BGWI(p) admite uma distribuição estacionária no limite.

Proposição 4.3:

Seja $\{\vec{Z}_n^{(i)}\}$ um processo BGWI(p) definido em (2.7), onde i significa que $\vec{Z}_0 = i$. Suponha-se que esse processo é não singular e que $E(\vec{I}_n) = \vec{\lambda} = (\lambda, 0, \dots, 0)^T \neq \vec{0}$ é finito. Se $\rho < 1$, então $\vec{Z}_n^{(i)} \xrightarrow{d} \vec{Z} \quad \forall i \in \mathbb{N}^p$ (\vec{Z} é uma variável aleatória cuja distribuição não depende de i) e $E(\vec{Z}_n^{(i)}) \rightarrow E(\vec{Z}) = [(I - M)^{-1}]^T \vec{\lambda}$, que é finito.

Dion, Gauthier e Latour usaram as últimas proposições para determinar um critério para a estacionaridade dos processos GINAR que é apresentado no seguinte teorema.

Teorema 4.1:

Seja $\{X_n\}_{n \geq 0}$ um processo GINAR(p) definido em (3.7) e seja $\vec{Z}_n = (X_n, X_{n-1}, \dots, X_{n-p+1})^T$ para $n \geq p-1$. Se $\sum_{k=1}^p m_k < 1$ e $0 < E(I_n) = \lambda < \infty$, então $\vec{Z}_n \xrightarrow{d} \vec{Z}$ para qualquer distribuição inicial de $\vec{Z}_{p-1} = (X_{p-1}, \dots, X_0)^T$ e $E(\vec{Z}_n) \rightarrow E(\vec{Z}) = (\mu_1, \dots, \mu_p)^T$, onde $\mu_i = \lambda \left(1 - \sum_{k=1}^p m_k\right)^{-1}$, $i=1, \dots, p$.

Assim, o processo GINAR(p) admite uma distribuição estacionária única no limite, se $\sum_{k=1}^p m_k < 1$. Pela Proposição 4.2, $\sum_{k=1}^p m_k < 1$ implica que as raízes x do polinómio

característico ou valores próprios da matriz M estejam dentro do círculo unitário e essas são iguais às raízes do polinómio característico da matriz A (3.6), porque, como uma matriz é a transposta da outra, $\det(xI - M) = \det(xI - A)$. Assim, esta condição de estacionaridade,

$\sum_{k=1}^p m_k < 1$, é igual à condição que Du e Li (1991) obtiveram para garantir a estacionaridade

dos processos INAR(p)¹⁶ e se $p=1$, igual à condição¹⁷ obtida por Al-Osh e Alzaid (1987).

Observa-se que o valor esperado $E(\vec{Z})$, para o qual converge o valor esperado do processo BGWI(p)¹⁸ e do processo GINAR(p)¹⁹, coincide com o valor esperado do processo

estacionário INAR(p)²⁰, porque $[(I - M)^{-1}]^T \vec{\lambda} = (I - A)^{-1} \vec{C}\lambda$.

4.1. COMPARAÇÃO E APLICAÇÃO DE MÉTODOS E RESULTADOS

Na secção anterior viu-se que o processo GINAR(1) é equivalente a um processo BGWI(1). À descendência e à imigração do processo BGWI(1) correspondem, respectivamente, a série de contagem e a inovação do processo GINAR(1). O modelo INAR(1), anteriormente apresentado, é um modelo particular do processo BGWI(1), em que $E(I_n) = \lambda$, $Var(I_n) = b^2 = \lambda$, $E(Y_{n-1,i}) = m$ e $Var(Y_{n-1,i}) = \sigma^2 = m(1 - m)$. Para os processos ramificados com imigração foram propostos métodos de estimação que se apoiam na amostra da realização do processo $\{X_0, \dots, X_n\}$ e métodos que partem do princípio de que, além das observações da realização do processo, também deve ser observável o processo de imigração. Os métodos encontrados para os modelos GINAR(p), INAR(p) e INAR(1) baseiam-se todos somente na observação de uma realização do processo. Métodos, em que os estimadores são calculados utilizando também as inovações, não foram construídos. Da análise dos artigos sobre os processos INAR(1), INAR(p) e GINAR(p), parece concluir-se que, nestes processos, a inovação não é observável. Nenhum autor referiu a observação ou a possibilidade da observação da inovação. O facto de as componentes da inovação serem

¹⁶ Cf. pág. 48 seg..

¹⁷ Cf. pág. 45.

¹⁸ Ver Proposição 4.3.

¹⁹ Ver Teorema 4.1.

²⁰ Cf. pág.49.

observáveis ou não, depende do tipo de observações. Por exemplo, quando as observações representam os números de hóspedes num hotel em determinados instantes de tempo, então, neste caso, é possível observar as entradas de novos hóspedes. Franke e Seligmann (1993) aplicaram um modelo INAR(1) às contagens diárias de ataques epiléticos de um paciente. Neste exemplo, a observação das inovações deve ser quase impossível. Assim, nos casos de as inovações não serem observáveis ou de não se conseguir observar as inovações, só os métodos, os resultados e as propriedades do capítulo 2.2.1. - estimadores baseados na comparação do processo ramificado com imigração com o processo AR(1), método de mínimos quadrados condicionais e método de mínimos quadrados condicionais pesados - é que podem ser transferidos para os processos GINAR(1) e INAR(1). Quando também se consegue observar as inovações dos processos GINAR(1) e INAR(1), além dos métodos do capítulo 2.2.1., também podem ser aplicados os métodos do capítulo 2.2.2. - método de máxima verosimilhança, estimação sequencial, estimação de precisão fixa e estimação sequencial de duas etapas.

A seguir comparam-se os métodos, os resultados e as propriedades desses métodos (que se apoiam só na observação de uma realização do processo) encontrados para os processos BGWI(1) e INAR(1).

O estudo dos métodos de estimação para os parâmetros do processo INAR(1) foi feito, supondo que a série de contagem segue uma distribuição de Bernoulli de parâmetro m , $E(Y_i) = m$ e $Var(Y_i) = m(1 - m)$, que a inovação segue uma distribuição de Poisson de parâmetro λ , $E(I_i) = \lambda = Var(I_i)$ e que o processo é estacionário. Neste caso, o parâmetro m do processo, a média da série de contagem, só admite valores no intervalo $[0,1)$.

Os estimadores apresentados por Heyde e Seneta (1972) para os parâmetros do processo BGWI(1) correspondem aos estimadores de Yule-Walker dos parâmetros do processo INAR(1). O estimador da média da descendência²¹, D_n , só difere do estimador para o parâmetro do processo INAR(1), \hat{m} (média da série de contagem), no denominador: a variância amostral de D_n é calculada com n observações e a variância amostral de \hat{m} , com $n+1$ observações. Sabendo que $C_n = \bar{X} \xrightarrow{p} \mu = \lambda(1 - m)^{-1}$ e $D_n \xrightarrow{p} m$, Heyde e

Seneta encontraram um estimador consistente para λ : $C_n(1 - D_n) \xrightarrow{p} \frac{\lambda}{(1 - m)}(1 - m) = \lambda$.

Al-Osh e Alzaid (1987) seguiram um raciocínio diferente para encontrar um estimador para λ : $\hat{\lambda}$ foi calculado como sendo a média amostral de \hat{I}_t , em que $\hat{I}_t = X_t - \hat{m}X_{t-1}$.

²¹ Ver pág. 17.

Comparando os dois estimadores de λ , observa-se que são quase iguais; a diferença está no número de observações com que são calculados os estimadores: no estimador $\hat{\lambda}$ de Al-Osh e Alzaid (1987), dois somatórios são calculados com $n+1$ observações, que é uma consequência de a variância amostral já ter sido calculada com esse número de observações, e não com n observações, como no estimador $C_n(1-D_n)$. Para amostras muito grandes esta diferença pode ser ignorada e, assim, pode afirmar-se que os estimadores de Yule-Walker do processo INAR(1) são iguais aos estimadores propostos por Heyde e Seneta para o processo BGWI(1).

Heyde e Seneta mostraram que os estimadores $C_n = \bar{X}$, D_n e $C_n(1-D_n)$ para μ , m e λ são consistentes e, se a distribuição da descendência, a distribuição da imigração e a distribuição inicial tiverem momentos de terceira ordem finitos, então esses estimadores são fortemente consistentes; para os mesmos parâmetros dos processos GINAR(1) e INAR(1) a consistência forte dos estimadores verifica-se automaticamente.

O caminho seguido por Heyde e Seneta (1972) e por Latour (1998) – que estudou os estimadores de Yule-Walker para o processo GINAR(p) – para obter os estimadores foi o mesmo. O processo foi identificado como um processo AR e algumas fórmulas conhecidas para o processo AR(p), por exemplo, a fórmula da função autocorrelação, foram aproveitadas para construir os estimadores.

Heyde e Seneta (1972) apresentaram as distribuições assintóticas de $C_n = \bar{X}$ e de D_n .

As variâncias assintóticas de $\sqrt{n}(C_n - \mu)$ e $\sqrt{n}(D_n - m)$ dependem dos parâmetros m e σ^2 da distribuição da descendência e dos parâmetros λ e b^2 da distribuição da imigração. A distribuição da descendência e a distribuição da imigração do processo BGWI(1) não foram especificadas.

Park e Oh (1997) obtiveram a distribuição assintótica conjunta de $\sqrt{n}(\hat{\mu} - \mu, \hat{m} - m)$, supondo que as variáveis da série de contagem seguem uma distribuição de Bernoulli de parâmetro m , $E(Y_i) = m$ e $Var(Y_i) = m(1-m)$, e que a inovação segue uma distribuição de Poisson de parâmetro λ , $E(I_n) = \lambda = Var(I_n)$

Considerando que a imigração do processo BGWI(1) segue também uma distribuição de Poisson de parâmetro λ e que a descendência segue uma distribuição de Bernoulli de parâmetro m , então as variâncias assintóticas dos estimadores D_n e C_n devem ser iguais às variâncias assintóticas dos estimadores de Yule-Walker \hat{m} e $\hat{\mu}$; isto significa que a

variância assintótica²² de $\sqrt{n}(D_n - m) = \beta^2$ deve ser igual a²³ $w_{22} = \frac{m(1-m)^2}{\lambda} + (1-m^2)$ e que a variância assintótica²⁴ de $\sqrt{n}(C_n - \mu) = \alpha^2$ deve ser igual a²⁵ $w_{11} = \frac{\lambda(1+m)}{(1-m)^2}$.

Fazendo as substituições $\sigma^2 = m(1-m)$ e $b^2 = \lambda$ nas fórmulas para β^2 e para α^2 , verifica-se, depois de alguns cálculos²⁶, que β^2 e α^2 são iguais, respectivamente, às variâncias assintóticas w_{22} e w_{11} .

As distribuições assintóticas de C_n e D_n foram apresentadas por Heyde e Seneta de uma forma geral, não especificando as distribuições da descendência e da imigração. As expressões α^2 e β^2 podem ser aproveitadas para calcular as variâncias assintóticas dos estimadores $\hat{\mu}$ e \hat{m} para os parâmetros do processo GINAR(1), em que a série de contagem segue uma distribuição de média $E(Y_i) = m$ e de variância $Var(Y_i) = \sigma^2$; a distribuição da inovação é caracterizada por $E(I_i) = \lambda$ e $Var(I_i) = b^2$.

O método de mínimos quadrados condicionais desenvolvido por Klimko e Nelson (1978) foi aplicado por Al-Osh e Alzaid (1987) ao processo INAR(1) e, por Wei e Winnicki (1987) ao processo BGWI(1). As expressões dos estimadores \hat{m}^* e $\hat{\lambda}^*$ para os parâmetros m e λ do processo INAR(1) são iguais às expressões dos estimadores dos parâmetros correspondentes do processo BGWI(1). Para $m < 1$ todos os estimadores são fortemente consistentes.

A variância assintótica²⁷ de $\sqrt{n}(\hat{m}^* - m)$ calculada por Park e Oh (1997) - $\frac{m(1-m)^2}{\lambda} + (1-m^2)$ - deverá ser igual à variância assintótica²⁸ Σ_{11} do Teorema 2.7 obtida por Wei e Winnicki (1987), se se considerar, como no método de Yule-Walker, que $\sigma^2 = Var(Y_{n-1,i}) = m(1-m)$ e $b^2 = Var(I_n) = \lambda$. Substituindo estas últimas expressões em Σ_{11} e efectuando cálculos, confirma-se o que já se esperava²⁹: $\Sigma_{11} = \frac{m(1-m)^2}{\lambda} + (1-m^2)$.

As propriedades associadas ao método de mínimos quadrados condicionais para os

²² Cf. Teorema 2.5 (pág. 17 seg.).

²³ Cf. pág. 55.

²⁴ Cf. Teorema 2.5.

²⁵ Cf. pág. 55.

²⁶ Ver Anexo A4, pág. 82 sgg..

²⁷ Cf. pág. 58.

²⁸ Ver pág. 20.

²⁹ Ver Anexo A5, pág. 84 seg..

parâmetros do processo INAR(1) podem ainda ser completadas por alguns resultados encontrados para o processo BGWI(1):

A distribuição assintótica conjunta de $\sqrt{n}(\hat{m}_n - m, \hat{\lambda}_n - \lambda)^T$ pode ser transferida para o processo INAR(1); basta substituir as quantidades b^2 e σ^2 da matriz de variâncias-covariâncias, respectivamente, por λ e por $m(1-m)$. Os resultados encontrados para o processo BGWI(1) em que $m < 1$ e $m = 1$ podem ser todos transferidos para o processo INAR(1), onde os resultados para o caso crítico, $m = 1$, se adaptam a um processo INAR(1) não estacionário. Assim obtêm-se também estimadores para as variâncias $Var(Y_i)$ e $Var(I_i)$ do processo INAR(1) e alguns resultados assintóticos desses estimadores.

Como os resultados do método de mínimos quadrados condicionais foram estudados para um processo BGWI(1), em que a distribuição da imigração e a distribuição da descendência são gerais para o caso subcrítico, para o caso crítico e para o caso “supercrítico”, $m < 1$, $m = 1$ e $m > 1$, então todos os resultados apresentados no capítulo 2.2.1.2. são válidos para o processo GINAR(1) estacionário, isto corresponde ao caso $m < 1$, e para o processo GINAR(1) não estacionário, $m \geq 1$.

No capítulo 2 chegou-se à conclusão, que, dos três métodos apresentados para estimar os parâmetros do processo BGWI(1) a partir de uma única realização do processo $\{X_0, \dots, X_n\}$, o método de mínimos quadrados condicionais pesados é o método que leva a melhores resultados, porque se encontram mais estimadores consistentes e estimadores mais eficientes do que com o método baseado na comparação do processo BGWI(1) com o processo AR(1) e o método de mínimos quadrados condicionais. Assim, aplicando este método ao processo GINAR(1), e em particular ao processo INAR(1), consegue-se um melhoramento das propriedades assintóticas dos estimadores até agora encontrados para os parâmetros destes processos. Todos os resultados apresentados em 2.2.1.3. são válidos para o modelo GINAR(1). Os resultados obtidos para o caso subcrítico podem ser aproveitados para o processo INAR(1) estacionário e os resultados obtidos para o caso crítico são válidos para o processo INAR(1) não estacionário.

Afinal não foram encontrados muitos mais métodos que possam ser aplicados aos processos GINAR(1) e INAR(1) do que os já existentes para estes processos, considerando o caso de só ser observada a realização do processo. Só a aplicação do método de mínimos quadrados condicionais pesados é que parece ainda não ter sido estudada para os processos GINAR(1) e INAR(1). Mesmo assim, só com os métodos encontrados para os processos BGWI(1), é possível ampliar as propriedades e os resultados existentes para a estimação dos parâmetros dos modelos INAR(1) e GINAR(1).

Também é possível aplicar métodos e resultados dos processos INAR(1) aos processos BGWI(1), como, por exemplo, o método de máxima verosimilhança condicional e o método de Whittle. Menciona-se aqui o facto que afinal o método de máxima verosimilhança condicional até já foi estudado para o processo BGWI(1) por Al-Osh e Alzaid (1988).

É muito raro encontrar na literatura abordagens sobre o problema de estimação dos parâmetros do processo BGWI de ordem superior a 1; só Dion, Gauthier e Latour (1995) é que apresentaram estimadores fortemente convergentes para a média, a matriz M , e a matriz de covariâncias do processo BGWI(p). Talvez seja possível aplicar os métodos propostos para os processos GINAR(p) e INAR(p) a processos BGWI(p).

4.2. APLICAÇÃO DO PROCESSO RAMIFICADO COM IMIGRAÇÃO A UMA SÉRIE DE OBSERVAÇÕES DE CONTAGEM

Guttorp (1991) aplicou um processo BGWI(1) ao movimento Browniano.

Suponha que observa ao microscópio um pequeno volume bem definido de uma solução coloidal e que conta o número de partículas nesse volume em intervalos de tempo discretos $\Delta, 2\Delta, \dots$. Admitindo que os movimentos das partículas são independentes e que o processo está em estado estacionário, então a série das observações pode ser modelada por um processo ramificado com imigração, BGWI(1), em que a descendência segue uma distribuição de Bernoulli e a imigração, uma distribuição de Poisson.

Guttorp (1991) aplicou aos dados da experiência de Westgren (1916) - 1598 observações que representam as contagens de partículas de ouro num volume pequeno bem definido de água, feitas em intervalos de $\Delta = 1.39$ segundos - o método de mínimos quadrados condicionais pesados e obteve as seguintes estimativas para os parâmetros m e λ : $\hat{m} = 0.613$ e $\hat{\lambda} = 0.555$. As partículas de ouro na solução obedecem à lei do movimento Browniano. Para relacionar a lei do movimento Browniano com o processo ramificado com imigração, Smoluchowski (1916) calculou a probabilidade de uma partícula, presente no volume observado num dado instante, desaparecer depois de t unidades de tempo. Esta probabilidade, que se exprime $P=1-m$, foi designada "probability after-effect". Guttorp calculou a probabilidade teórica através da fórmula

$$P = 2(1 - \Phi(\beta)) + \frac{2}{\beta(2\pi)^{1/2}} \left(1 - \exp\left(-\frac{\beta^2}{2}\right) \right),$$

onde Φ é a função distribuição normal e $\beta = h(2Dt)^{-1/2}$, onde D é o coeficiente de difusão das partículas e h é a densidade do volume, e obteve: $1-m=0.394$; usando o estimador de

mínimos quadrados condicionais pesados, encontrou: $1-\hat{m}=0.387$. Como se pode verificar, a estimativa condiz bem com a teoria.

Quando Westgren fez a experiência, o valor da constante de Avogadro N ainda não estava bem definido. Assim, o objectivo de Westgren foi a determinação dessa constante. Como a constante de Avogadro depende da probabilidade P , Guttorp calculou intervalos de confiança para P e N ; e a partir daí, fez corresponder à estimativa $1-\hat{m}$ uma estimativa para N : $N = 6.29 \times 10^{23} \text{ mol}^{-1}$. Nessa altura o valor da constante de Avogadro N situava-se em $6.02252 \times 10^{23} \text{ mol}^{-1}$, o que indica que a aproximação obtida com o método de mínimos quadrados condicionais pesados é satisfatória.

A seguir é feita uma breve análise da série de Westgren³⁰ através da interpretação do cronograma e dos gráficos da função autocorrelação (FAC) e da função autocorrelação parcial (FACP) e depois são estimados os parâmetros do modelo BGWI(1), usando os métodos anteriormente estudados.

Cronograma

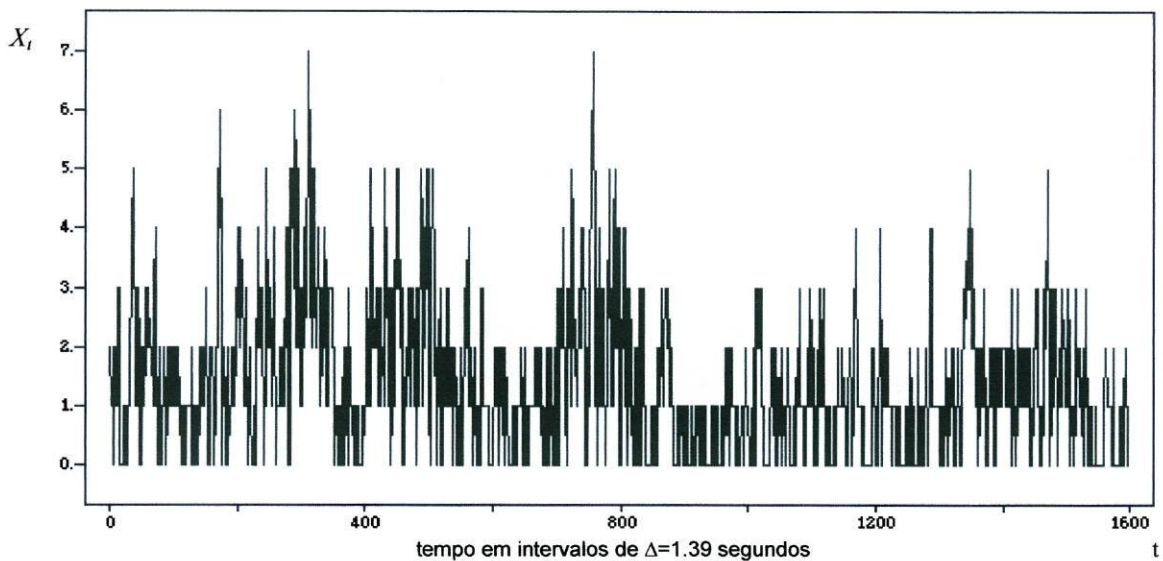


Figura 4.1: Contagens de partículas de ouro numa solução coloidal

Da análise do cronograma da série de 1598 dados observa-se que os valores (contagens) 0,1 e 2 ocorrem mais frequentemente do que os restantes valores mais elevados (3,4,5,6 e 7). Estes valores: 0,1 e 2, estão próximos da média amostral $\bar{X} = 1.4249$. Por um lado, isto indica que a série parece ser estacionária e daí conclui-se que o parâmetro m do modelo

³⁰ Os dados desta série encontram-se no livro de Guttorp (1991).

BGWI(1) aplicado deverá ser menor do que 1. Por outro lado, como a série apresenta movimentos oscilatórios, principalmente durante os primeiros 900 intervalos de tempo, ela parece ser não estacionária. A restante parte da série apresenta menos oscilações e as amplitudes destas oscilações são menores do que as oscilações da primeira metade da série.

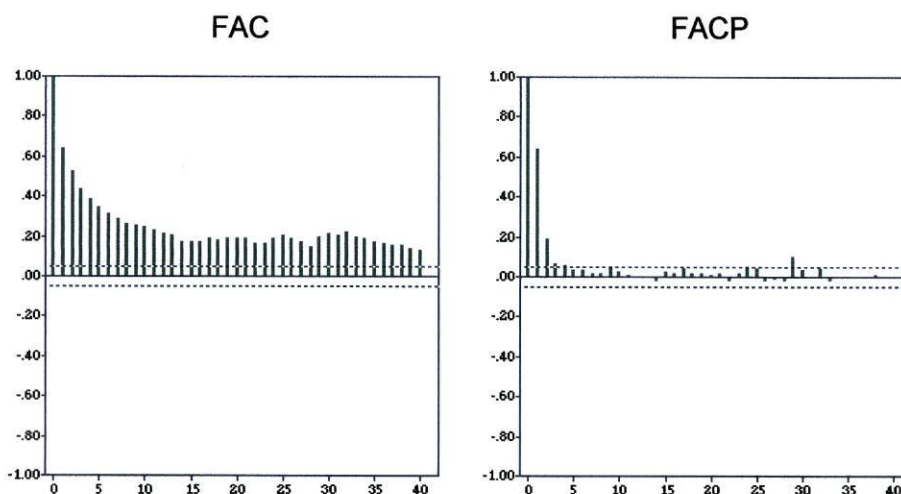


Figura 4.2: Função autocorrelação amostral e função autocorrelação parcial amostral de X_t

Observando o gráfico da FAC do processo, verifica-se que a função autocorrelação decresce muito lentamente para zero. Este comportamento é típico para processos AR não estacionários; isso indica que esta série talvez não seja estacionária.

Como se pode ver do gráfico da FACP, a função autocorrelação parcial apresenta valores significativos nos “lags” 1, 2, 3, 4, 9, 29 e, aumentando o número de “lags”, em “lags” mais elevados. Guttorp utilizou um processo BGWI(1) para descrever a série, mas o gráfico da FACP de um processo BGWI(1), que é idêntico a um processo INAR(1), só devia ter um valor significativo no “lag” 1. No gráfico da FACP da Figura 4.2 ainda existem muitos “lags” superiores a 1 com função autocorrelação parcial significativa. Neste caso, um modelo de ordem superior a $p=1$ talvez se adaptasse melhor a esta série.

Aplicando às observações um critério de selecção de ordem para modelos INAR³¹ (à ordem mais adequada corresponde o valor AICC mínimo), então a ordem que minimiza o critério, considerando como ordem máxima o valor 20, é $p=14$. Segundo este critério, um modelo INAR(14)/BGWI(14) seria adequado para descrever a série, o que não está de acordo com

³¹ Ver Silva e Silva (2002).

a conclusão tirada a partir da análise do gráfico da FACP. Assim, como é difícil determinar a ordem óptima do modelo, escolhe-se como ordem $p=1$ e continua-se a estudar a série calculando as estimativas dos parâmetros do modelo BGWI(1)/INAR(1) com os diferentes métodos³², que estão apresentadas na seguinte tabela, e, em anexo³³, encontram-se as estimativas para os parâmetros dos modelos de ordens 2, 4, 9 e 14.

No método de Whittle foram consideradas também restrições sobre os parâmetros para garantir a estacionaridade do processo.

| $p=1$ | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{m} | 0.6447 | 0.6447 | 0.6118 | 0.6447 | 0.6458 | 0.6458 |
| $\hat{\lambda}$ | 0.5054 | 0.5057 | 0.5526 | 0.5054 | 0.5356 | 0.5354 |

Tabela 4.1: Estimativas dos parâmetros do modelo BGWI(1)/INAR(1)

Como se pode verificar, as estimativas obtidas com os diferentes métodos são muito semelhantes, sendo idênticas até à primeira casa decimal e, na maior parte dos casos, até à segunda casa decimal, excepto as estimativas de mínimos quadrados condicionais pesados e, considerando $\hat{\lambda}$, as estimativas obtidas pelo método de Whittle. Observa-se que as estimativas obtidas com o método de Yule-Walker e com o método de máxima verosimilhança condicional são iguais. A estimativa de mínimos quadrados condicionais para m é igual às estimativas obtidas com estes dois últimos métodos e a estimativa para λ só difere das estimativas de Yule-Walker e máxima verosimilhança condicional na quarta casa decimal.

Uma maneira de avaliar a qualidade dos modelos é fazer uma análise aos resíduos. Os resíduos $\{\varepsilon_t\}$, que se calculam da forma: $\varepsilon_t = X_t - \hat{m}X_{t-1} - \hat{\lambda}$, devem ser variáveis aleatórias não correlacionadas de média 0.

Aplicando o critério AICC de selecção de ordem às variáveis $\{\varepsilon_t\}$, então a ordem deve ser igual a 0, se os resíduos são não correlacionados.

³² Os métodos aqui aplicados foram programados em MATLAB por Isabel Silva

³³ Ver ANEXO A6, pág. 85 seg..

| | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|--------------------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| $\bar{\varepsilon}_t$ | 0.00034 | 0.00037 | -0.0047 | 0.00033 | -0.0314 | -0.0312 |
| Ordem de ε_t | 0 | 0 | 0 | 0 | 0 | 0 |

Tabela 4.2: Média e ordem dos resíduos $\{\varepsilon_t\}$, onde $\varepsilon_t = X_t - \hat{m}X_{t-1} - \hat{\lambda}$

Nesta tabela observa-se que as médias dos resíduos estão próximas de zero. As médias dos resíduos obtidos com os métodos de Whittle estão mais afastadas de zero do que as médias associadas aos outros métodos e a média dos resíduos obtidos com o método de máxima verosimilhança condicional é a que está mais próxima de zero.

As estimativas \hat{m} podem ser avaliadas através da comparação das estimativas para a probabilidade P , $\hat{P} = 1 - \hat{m}$, com o valor teórico dessa probabilidade calculado por Gutterp, $P=1-m=0.394$. A Tabela 4.3 contém os valores de \hat{P} e os erros absolutos $|\hat{P} - P|$ calculados pelos diferentes métodos.

| | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{P} | 0.3553 | 0.3553 | 0.3882 | 0.3553 | 0.3542 | 0.3542 |
| $ \hat{P} - P $ | 0.0387 | 0.0387 | 0.0058 | 0.0387 | 0.0398 | 0.0398 |

Tabela 4.3: Estimativas para a continuação do efeito da probabilidade, $\hat{P} = 1 - \hat{m}$, e erros absolutos entre o valor teórico da continuação do efeito da probabilidade, $P=1-m=0.394$, e \hat{P}

Interpretando esta tabela, conclui-se que o método, com o qual se encontra uma estimativa para \hat{P} mais próxima do valor teórico P , é o método de mínimos quadrados condicionais pesados. Os erros absolutos mais altos obtêm-se com os métodos de Whittle.

Analisando os resultados obtidos para esta série de dados, chega-se à conclusão que é difícil dizer quais estimativas são melhores para representar os parâmetros do modelo, supondo que um processo BGWI/INAR de ordem 1 é o mais apropriado para descrever a série.

Por um lado, a estimativa \hat{m} obtida com o método de mínimos quadrados pesados parece ser a mais adequada, mas, por outro lado, a média dos resíduos, calculada com as

estimativas desse método, não está tão próxima de zero como as médias obtidas com o método de Yule-Walker, o método de mínimos quadrados e o método de máxima verossimilhança condicional. Considerando só a análise dos resíduos, o método de máxima verossimilhança condicional parece ser o mais adequado para estimar os parâmetros do modelo.

Os resultados que se encontram nas tabelas em anexo, que contêm as estimativas para os parâmetros dos modelos BGWI/INAR de ordens 2, 4, 9 e 14, revelam que, em geral, as estimativas obtidas com os métodos estudados são muito semelhantes; só as estimativas obtidas com o método de mínimos quadrados condicionais pesados é que se distinguem das estimativas dos outros métodos, às vezes, logo a partir da primeira casa decimal. Observa-se que as estimativas do método de Yule-Walker e do método de máxima verossimilhança condicional são idênticas. Já foi visto que o método de Yule-Walker e o método de mínimos quadrados condicionais são assintoticamente equivalentes e as estimativas destes métodos estão realmente muito próximas. Aumentando a ordem do modelo, verifica-se que das estimativas \hat{m}_i , $1 \leq i \leq p$, só \hat{m}_1 e \hat{m}_2 é que têm valores significativos; as restantes estimativas aproximam-se de zero. Isso significa que nos modelos de ordem superior a dois, duas observações passadas, X_{t-1} e X_{t-2} , é que têm a maior influência na observação X_t .

5. CONCLUSÕES

Ao terminar este trabalho chega-se à conclusão que aos resultados (métodos, propriedades, etc.) sobre a inferência estatística conhecidos para os processos GINAR(1) e INAR(1) pode-se juntar o seguinte, encontrado para os processos BGWI(1):

- Método de mínimos quadrados condicionais pesados
- Novos estimadores, não só para a média da série de contagem e para a média da inovação, mas também para as variâncias respectivas
- Propriedades assintóticas para processos GINAR(1) e INAR(1) estacionários e não estacionários
- Intervalos de confiança assintóticos para a média da série de contagem e para a média da inovação, porque as distribuições assintóticas destes últimos parâmetros, o estimador para a variância da série de contagem e o estimador para a variância da inovação são conhecidos.

Através da comparação dos métodos conclui-se ainda que entre o método de Yule-Walker, o de mínimos quadrados condicionais e o de mínimos quadrados condicionais pesados, é com este último que se atingem propriedades assintóticas mais desejáveis. Comparando o método de Yule-Walker, o de mínimos quadrados condicionais, o de máxima verosimilhança condicional e o de Whittle, obtêm-se propriedades não assintóticas mais aceitáveis com os métodos de Whittle e de máxima verosimilhança condicional.

Encontraram-se também novos métodos de estimação que podem ser aplicados a processos INAR(1) e GINAR(1) quando se conhecem as observações e as inovações.

ANEXO

A1) O processo BGWI(1) controlado é da forma

$$X_n = \sum_{i=1}^{X_{n-1}+U_{n-1}} Y_{n-1,i} + I_n, \quad n \geq 1,$$

onde:

- $\{Y_{n-1,i}\}$ e $\{I_n\}$ são sequências independentes de variáveis aleatórias inteiras i.i.d.,
- as variáveis iniciais X_0 e U_0 são variáveis aleatórias inteiras, quadraticamente integráveis e independentes de $\{Y_{n-1,i}\}$ e $\{I_n\}$,
- $\{U_{n-1}\}$ é uma sequência de variáveis aleatórias inteiras adaptadas a \bar{F} , tal que para qualquer $n \geq 1$, $X_{n-1} + U_{n-1} \geq 1$, onde $\bar{F} = (\bar{F}_n)$ com $F_n = \sigma\{X_0, U_0, Y_{k,i}, I_k; 1 \leq k \leq n, i \geq 1\}$.

O controlo adaptativo $\{U_{n-1}\}_{n \geq 1}$ regula a dinâmica do processo $\{X_{n-1}\}_{n \geq 1}$. Se o processo tem pouca “energia”, então $\{U_{n-1}\}$ gera descendentes, se por outro lado existe “energia” a mais, então $\{U_{n-1}\}$ elimina descendentes. O controlo $\{U_{n-1}\}$ obriga o processo $\{X_{n-1}\}$ a seguir uma trajectória de referência dada.

A2) Um processo de difusão é um processo de Markov de variáveis contínuas em tempo contínuo.

Para enunciar o Teorema 2.8, os autores partiram do processo de difusão definido na proposição seguinte (Wei e Winnicki (1989)).

Proposição: Seja $Y_n(t) = \frac{1}{n} X_{[nt]}$, $t \geq 0$, então a distribuição do processo Y_n , $n=1,2,\dots$

converge para a distribuição de um processo de difusão não negativo Y com $Y(0)=0$ e gerador A dado por

$$Af(x) = \frac{1}{2} \sigma^2 x f''(x) + \lambda f'(x), \quad f \in C_c^\infty[0, \infty),$$

onde $C_c^\infty[0, \infty)$ é o espaço das funções infinitamente diferenciáveis em $[0, \infty)$ com suporte compacto.

A3) Os estimadores \hat{p}_j e \hat{q}_i encontrados por Bhat e Adke podem ser obtidos a partir da função de verosimilhança da seguinte maneira:

A expressão para o logaritmo da função de verosimilhança é da forma

$$l = \log L = \sum_{j=0}^{\infty} N_j(n) \log(p_j) + \sum_{i=0}^{\infty} M_i(n) \log(q_i).$$

Do sistema de derivadas desta função em ordem a p_j e q_i

$$\begin{cases} \frac{\partial l}{\partial p_j} = \frac{N_j(n)}{p_j} = 0 \\ \frac{\partial l}{\partial q_i} = \frac{M_i(n)}{q_i} = 0 \end{cases}$$

obtém-se a igualdade

$$N_j(n)q_i = M_i(n)p_j. \quad (1a)$$

Aplicando agora somatórios, onde j varia de 0 até ∞ , às expressões dos dois lados

desta última equação, então, como $\sum_{j=0}^{\infty} p_j = 1$ e $\sum_{j=0}^{\infty} N_j(n) = N_{n-1}$, chega-se à igualdade

$q_i = \frac{M_i(n)}{N_{n-1}}$. E substituindo em (1a) q_i por esta última expressão, obtém-se o estimador

de máxima verosimilhança para p_j : $\hat{p}_j = \frac{N_j(n)}{N_{n-1}}$.

O estimador \hat{q}_i calcula-se de uma forma análoga usando somatórios, onde i varia de 0

até ∞ , e as equivalências $\sum_{i=0}^{\infty} p_i = 1$ e $\sum_{i=0}^{\infty} M_i(n) = n$.

A4) A variância assintótica de $\sqrt{n}(C_n - \mu)$ é dada por $\alpha^2 = (1-m)^{-2} c^2$, onde

$c^2 = b^2 + \sigma^2 \mu$. Tendo em conta que $\mu = \frac{\lambda}{1-m}$, $b^2 = \text{Var}(I) = \lambda$ e

$\sigma^2 = \text{Var}(Y) = m(1-m)$, obtém-se:

$$\alpha^2 = (1-m)^{-2} c^2 = (1-m)^{-2} \left(\lambda + \frac{m(1-m)\lambda}{1-m} \right) = \frac{1+m}{(1-m)^2} \lambda,$$

que é igual a w_{11} , a variância assintótica de $\sqrt{n}(\hat{\mu} - \mu)$, onde $\hat{\mu}$ representa o estimador de Yule-Walker.

A seguir, mostra-se que a variância assintótica de $\sqrt{n}(D_n - m)$, β^2 , é igual à

variância assintótica de $\sqrt{n}(\hat{m} - m)$, w_{22} , obtida com o estimador de Yule-Walker \hat{m} .

Suponha-se que I segue uma distribuição de Poisson de parâmetro λ e que Y segue uma distribuição de Bernoulli de parâmetro m , então tem-se as seguintes equivalências:

$$\mu = \frac{\lambda}{1-m}$$

$$c^2 = \lambda(1+m), \quad c^4 = \lambda^2(1+m)^2$$

$$E(I) = \lambda, \quad Var(I) = \lambda$$

$$E(I^2) = Var(I) + [E(I)]^2 = \lambda + \lambda^2$$

$E(I^3)$ calcula-se com a ajuda da função de geradora de momentos:

A função geradora de momentos de uma variável aleatória X com distribuição de Poisson de parâmetro λ é:

$$\Psi_X(t) = E(e^{tX}) = \sum_{X=0}^{\infty} e^{tX} e^{-\lambda} \frac{\lambda^X}{X!} = e^{-\lambda} \sum_{X=0}^{\infty} \frac{(\lambda e^t)^X}{X!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}.$$

$$\begin{aligned} E(I^3) &= \left. \frac{d^3}{dt^3} \Psi_X(t) \right|_{t=0} = \left. \frac{d^2}{dt^2} (\lambda e^t e^{\lambda(e^t-1)}) \right|_{t=0} = \left. \frac{d}{dt} (\lambda e^t e^{\lambda(e^t-1)} (1 + \lambda e^t)) \right|_{t=0} = \\ &= \left. (\lambda e^t e^{\lambda(e^t-1)} (1 + \lambda e^t) + \lambda^2 e^{2t} e^{\lambda(e^t-1)} (1 + \lambda e^t) + \lambda e^t e^{\lambda(e^t-1)} \lambda e^t) \right|_{t=0} = \\ &= \lambda + 3\lambda^2 + \lambda^3 \end{aligned}$$

$$\begin{aligned} E(I - \lambda)^3 &= E[(I^2 - 2\lambda I + \lambda^2)(I - \lambda)] = E(I^3) - 3\lambda E(I^2) + 3\lambda^2 E(I) - \lambda^3 = \\ &= \lambda + 3\lambda^2 + \lambda^3 - 3\lambda^2 - 3\lambda^3 + 3\lambda^3 - \lambda^3 = \lambda \end{aligned}$$

$$m = E(Y), \quad \sigma^2 = Var(Y) = m(1-m)$$

$$E(Y^2) = Var(Y) + [E(Y)]^2 = m(1-m) + m^2 = m$$

A função geradora de momentos de uma variável aleatória X com distribuição de Bernoulli de parâmetro m é:

$$\Psi_X(t) = E(e^{tX}) = \sum_{X=0}^{\infty} e^{tX} m^X (1-m)^{1-X} = 1 - m + me^t.$$

$$E(Y^3) = \left. \frac{d^3}{dt^3} \Psi_X(t) \right|_{t=0} = \left. \frac{d^2}{dt^2} (me^t) \right|_{t=0} = \left. \frac{d}{dt} (me^t) \right|_{t=0} = m$$

$$\begin{aligned} E(Y - m)^3 &= E[(Y^2 - 2mY + m^2)(Y - m)] = E(Y^3) - 3mE(Y^2) + 3m^2E(Y) - m^3 = \\ &= m - 3m^2 + 2m^3 \end{aligned}$$

Substituindo agora todas as expressões anteriores em $\beta^2 = B^2(1-m^2)^2 c^{-4}$, onde

$$B^2 = \gamma \sigma^2 + c^4(1-m^2)^{-1} \quad \text{e} \quad \gamma = (1-m^3)^{-1} [E(I-\lambda)^3 + \mu E(Y-m)^3 + 3m\sigma^2 c^2(1-m^2)^{-1}],$$

obtém-se:

$$\begin{aligned}
\beta^2 &= \left\{ \frac{m(1-m)}{(1-m^3)} \left[\lambda + \frac{\lambda}{1-m} (m-3m^2+2m^3) + \frac{3m^2(1-m)\lambda(1+m)}{(1-m^2)} \right] + \frac{\lambda^2(1+m)^2}{(1-m^2)} \right\} \frac{(1-m^2)^2}{\lambda^2(1+m)^2} \\
&= \left\{ \frac{m(1-m)}{(1-m^3)} \left[\lambda + \lambda(-2m^2+m) + 3m^2\lambda \right] + \frac{\lambda^2(1+m)^2}{(1-m^2)} \right\} \frac{(1-m^2)^2}{\lambda^2(1+m)^2} \\
&= \left\{ \frac{m(1-m)[1-2m^2+m+3m^2]}{(1-m^3)\lambda} + \frac{(1+m)^2}{(1-m^2)} \right\} \frac{(1-m^2)^2}{(1+m)^2} \\
&= \left[\frac{m-m^4}{(1-m^3)\lambda} + \frac{(1+m)^2}{(1-m^2)} \right] \frac{(1-m^2)^2}{(1+m)^2} = \left[\frac{m}{\lambda} + \frac{(1+m)}{(1-m)} \right] (1-m)^2 \\
&= \frac{m(1-m)^2}{\lambda} + (1-m^2) = w_{22}
\end{aligned}$$

A5) Substituindo em Σ_{11} as seguintes expressões que já foram calculadas em A4):

$$\mu = \frac{\lambda}{1-m},$$

$$\sigma^2 = \text{Var}(Y) = m(1-m),$$

$$b^2 = \text{Var}(I) = \lambda,$$

$$\begin{aligned}
\gamma &= E(Y-m)^3 = E[(Y^2 - 2mY + m^2)(Y-m)] = E(Y^3) - 3mE(Y^2) + 3m^2E(Y) - m^3 = \\
&= m - 3m^2 + 2m^3
\end{aligned}$$

e

$$\begin{aligned}
d &= E(I-\lambda)^3 = E[(I^2 - 2\lambda I + \lambda^2)(I-\lambda)] = E(I^3) - 3\lambda E(I^2) + 3\lambda^2 E(I) - \lambda^3 = \\
&= \lambda + 3\lambda^2 + \lambda^3 - 3\lambda^2 - 3\lambda^3 + 3\lambda^3 - \lambda^3 = \lambda,
\end{aligned}$$

chega-se à conclusão, depois de alguns cálculos, que Σ_{11} é igual à variância assintótica de $\sqrt{n}(\hat{m}^* - m)$ obtida por Park e Oh (1997):

$$\begin{aligned}
\Sigma_{11} &= \sigma^2(\mu\gamma + d)(1-m^2)^2(\mu\sigma^2 + b^2)^{-2}(1-m^3)^{-1} + 1-m^2 \\
&\quad + 3m\sigma^4(1-m^2)(\mu\sigma^2 + b^2)^{-1}(1-m^3)^{-1} \\
&= m(1-m) \left[\frac{\lambda}{1-m} (m-3m^2+2m^3) + \lambda \right] (1-m^2)^2 \left[\frac{\lambda}{1-m} m(1-m) + \lambda \right]^{-2} (1-m^3)^{-1} \\
&\quad + 1-m^2 + 3m^3(1-m)^2(1-m^2) \left[\frac{\lambda}{1-m} m(1-m) + \lambda \right]^{-1} (1-m^3)^{-1}
\end{aligned}$$

$$\begin{aligned}
&= m(1-m) \left[\frac{\lambda m(1-3m+2m^2) + \lambda(1-m)}{1-m} \right] (1-m)^2 (1+m)^2 \frac{1}{\lambda^2 (m+1)^2 (1-m^3)} \\
&+ 1-m^2 + \frac{3m^3(1-m)^2(1-m)(1+m)}{\lambda(1+m)(1-m^3)} \\
&= \frac{m[m(1-3m+2m^2) + 1-m](1-m)^2}{\lambda(1-m^3)} + 1-m^2 + \frac{3m^3(1-m)^2(1-m)}{\lambda(1-m^3)} \\
&= \frac{(1-m)^2 [m(m-3m^2+2m^3+1-m) + 3m^3(1-m)]}{\lambda(1-m^3)} + 1-m^2 \\
&= \frac{(1-m)^2 [-3m^3+2m^4+m+3m^3-3m^4]}{\lambda(1-m^3)} + 1-m^2 \\
&= \frac{(1-m)^2 (m-m^4)}{\lambda(1-m^3)} + 1-m^2 \\
&= \frac{(1-m)^2 m}{\lambda} + (1-m^2)
\end{aligned}$$

A6)

| $p=2$ | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{m}_1 | 0.5162 | 0.5165 | 0.4768 | 0.5162 | 0.5179 | 0.5174 |
| \hat{m}_2 | 0.1993 | 0.1992 | 0.2196 | 0.1993 | 0.1989 | 0.1996 |
| $\hat{\lambda}$ | 0.4047 | 0.4050 | 0.4324 | 0.4047 | 0.3461 | 0.3458 |

Tabela 1A : Estimativas dos parâmetros do modelo BGWI(2)/INAR(2)

| $p=4$ | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{m}_1 | 0.4969 | 0.4969 | 0.4565 | 0.4969 | 0.4980 | 0.4978 |
| \hat{m}_2 | 0.1511 | 0.1510 | 0.1468 | 0.1511 | 0.1508 | 0.1508 |
| \hat{m}_3 | 0.0410 | 0.0415 | 0.0486 | 0.0410 | 0.0413 | 0.0421 |
| \hat{m}_4 | 0.0643 | 0.0643 | 0.0996 | 0.0643 | 0.0654 | 0.0648 |
| $\hat{\lambda}$ | 0.3507 | 0.3511 | 0.3541 | 0.3507 | 0.2833 | 0.2832 |

Tabela 2A: Estimativas dos parâmetros do modelo BGWI(4)/INAR(4)

| $p=9$ | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{m}_1 | 0.4898 | 0.4907 | 0.4384 | 0.4898 | 0.4917 | 0.4903 |
| \hat{m}_2 | 0.1456 | 0.1460 | 0.1392 | 0.1456 | 0.1463 | 0.1439 |
| \hat{m}_3 | 0.0307 | 0.0309 | 0.0266 | 0.0307 | 0.0296 | 0.0313 |
| \hat{m}_4 | 0.0356 | 0.0337 | 0.0579 | 0.0356 | 0.0342 | 0.0358 |
| \hat{m}_5 | 0.0125 | 0.0119 | 0.0584 | 0.0125 | 0.0123 | 0.0113 |
| \hat{m}_6 | 0.0237 | 0.0224 | 0.0207 | 0.0237 | 0.0227 | 0.0243 |
| \hat{m}_7 | 0.0103 | 0.0097 | 0.0094 | 0.0103 | 0.0121 | 0.0124 |
| \hat{m}_8 | -0.0050 | 0.0001 | 0.0009 | -0.0050 | 0 | 0.0055 |
| \hat{m}_9 | 0.0520 | 0.0492 | 0.0443 | 0.0520 | 0.0509 | 0.0469 |
| $\hat{\lambda}$ | 0.2897 | 0.2921 | 0.2934 | 0.2897 | 0.2271 | 0.2249 |

Tabela 3A: Estimativas dos parâmetros do modelo BGWI(9)/INAR(9)

| $p=14$ | Método de Yule-Walker | Método de mínimos quadrados condicionais | Método de mínimos quadrados condicionais pesados | Método de máxima verosimilhança condicional | Método de Whittle (com restrições) | Método de Whittle (sem restrições) |
|-----------------|-----------------------|------------------------------------------|--------------------------------------------------|---------------------------------------------|------------------------------------|------------------------------------|
| \hat{m}_1 | 0,4878 | 0,491 | 0.4387 | 0.4878 | 0,4929 | 0,4848 |
| \hat{m}_2 | 0,1451 | 0,1495 | 0.1469 | 0.1451 | 0,1404 | 0,14 |
| \hat{m}_3 | 0,0302 | 0,0271 | 0.0249 | 0.0302 | 0,0253 | 0,0259 |
| \hat{m}_4 | 0,0351 | 0,029 | 0.0550 | 0.0351 | 0,0298 | 0,04 |
| \hat{m}_5 | 0,0126 | 0,0104 | 0.0657 | 0.0126 | 0,0133 | 0,0154 |
| \hat{m}_6 | 0,0219 | 0,0181 | 0.0181 | 0.0219 | 0,0229 | 0,0226 |
| \hat{m}_7 | 0,0088 | 0,0072 | 0.0072 | 0.0088 | 0,0124 | 0,0138 |
| \hat{m}_8 | -0,0098 | 0,0001 | 0.0001 | -0.0098 | 0 | 0,0119 |
| \hat{m}_9 | 0,0345 | 0,0285 | 0.0285 | 0.0345 | 0,0313 | 0,0029 |
| \hat{m}_{10} | 0,0238 | 0,0197 | 0.0197 | 0.0238 | 0,0177 | 0,0152 |
| \hat{m}_{11} | 0,0084 | 0,0069 | 0.0069 | 0.0084 | 0,0032 | 0,0123 |
| \hat{m}_{12} | 0,0088 | 0,0073 | 0.0073 | 0.0088 | 0,0027 | 0,0127 |
| \hat{m}_{13} | 0,0194 | 0,016 | 0.0160 | 0.0194 | 0,0043 | 0,0165 |
| \hat{m}_{14} | -0,0237 | 0,0001 | 0.0001 | -0.0237 | 0 | 0,0027 |
| $\hat{\lambda}$ | 0,277 | 0,2695 | 0.2287 | 0.2770 | 0,224 | 0,2059 |

Tabela 4A: Estimativas dos parâmetros do modelo BGWI(14)/INAR(14)

REFERÊNCIAS

- Al-Osh, M.A. and Alzaid, A.A. (1987). First order integer-valued autoregressive (INAR(1)) processes. *J. Time Ser. Anal.*, **8**, 261-275.
- Al-Osh, M. and Alzaid, A. (1988). On the maximum likelihood estimation for a subcritical branching process with immigration. *Pak. J. Statist.*, **4A**, 147-156.
- Bercu, B. (2001). Weighted estimation and tracking for branching processes with immigration. *IEEE Trans. Automat. Control*, **46**, No.1, 43-50.
- Bhat, B.R. and Adke, S.R. (1981). Maximum likelihood estimation for branching processes with immigration. *Adv. Appl. Prob.*, **13**, 498-509.
- Billingsley, P. (1961). *Statistical Inference for Markov processes*. Chicago, University of Chicago Press.
- Chandler, K. (1995). A spectral method for estimating parameters in rainfall models. Research Report 142, Department of Statistical Science, UCL.
- Cox, D. (1981). Statistical analysis of time series: some recent developments. *Scand. J. Statist.*, **8**, 93-115.
- Dion, J., Gauthier, G. and Latour, A. (1995). Branching processes with immigration and integer-valued time series. *Serdica Mathematical Journal*, **21**, 123-136.
- Du, J.-G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal.*, **12**, 129-142.
- Einstein, A. (1905). Über die von der molekular-kinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.*, **17**, 549-560.
- Franke, J. and Seligmann, T. (1993). Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In *Developments in Time Series Analysis* (e.d. T. S. Rao). New York, Chapman and Hall.

- Franke, J. and Subba Rao, T. (1995). Multivariate first order integer valued autoregressions. Technical Report, Maths Dep. UMIST.
- Gauthier, G. and Latour, A. (1992). GENAR: Processus autoregressif à valeurs entières généralisé. Rapport de recherche no. 187, Département de mathématiques et informatique, UQAM, Montréal, 31 p.
- Gauthier, G. and Latour, A. (1994). Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Ann. Sci. Math. Québec*, **18**, 49-71.
- Guttorp, P. (1991). *Statistical Inference for Branching Processes*. New York, Wiley.
- Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. New York, Academic Press.
- Heathcote, C.R. (1965). A branching process allowing immigration. *J. R. Statist. Soc.*, **B27**, 138-143.
- Heathcote, C.R. (1966). Correction and comments on the paper "A branching process allowing immigration". *J. R. Statist. Soc.*, **B28**, 213-217.
- Heyde, C.C. (1970). Extension of a result of Seneta for the supercritical Galton-Watson process. *Ann. Math. Stat.*, **41**, 739-742.
- Heyde, C.C. (1975). Remarks on efficiency in estimation for branching procoesses. *Biometrika*, **62**, 49-55.
- Heyde, C.C. and Seneta, E. (1971). Analogues of classical limit theorems for the supercritical Galton-Watson process with immigration. *Math. Biosc.*, **11**, 249-259.
- Heyde, C.C. and Seneta, E. (1972). Estimation theory for growth and immigration rates in a multiplicative process. *J. Appl. Prob.*, **9**, 235-250.
- Heyde, C.C. and Seneta, E. (1974). Notes on "Estimation theory for growth and immigration rates in a multiplicative process". *J. Appl. Prob.*, **11**, 572-577.

- Jacobs, P. and Lewis, P. (1978a). Discrete time series generated by mixtures I: correlation and runs properties. *J. Roy. Statist. Soc.*, **B 40**, 94-105.
- Jacobs, P. and Lewis, P. (1978b). Discrete time series generated by mixtures II: asymptotic properties. *J. Roy. Statist. Soc.*, **B 40**, 222-228.
- Jacobs, P. and Lewis, P. (1983). Stationary discrete autoregressive moving average time series generated by mixtures. *J. Time Ser. Anal.*, **4**, 19-36.
- Karlin, S. and Taylor, H. (1975). *A first course in Stochastic Processes*. Second Edition, London, Academic Press.
- Kedem, B. and Chiu, L.S. (1987). On the lognormality of rain date. *Proc. Natl. Acad. Sci. USA*, **84**, 901-905.
- Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Stat.*, **6**, 629-642.
- Lai, T.L. and Siegmund, D. (1983). Fixed accuracy estimation of an autoregressive parameter. *Ann. Statist.*, **11**, 478-485.
- Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *J. Time Ser. Anal.*, **19**, No. 4, 439-455.
- Leslie, P.H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183-212.
- Leslie, P.H. (1948). Some further notes on the use of matrices in population mathematics. *Biometrika*, **35**, 213-245.
- Lewis, E.G. (1942). On the generation and growth of a population. *Sankhya*, **6**, 93-96.
- McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resour. Bull.*, **21**, 645-650.
- Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *J. R. Statist. Soc.*, **A135**, 370-384.

- Pakes, A.G. (1971). On the critical Galton-Watson process with immigration. *J. Austral. Math. Soc.*, **12**, 476-482.
- Pakes, A.G. (1972). Further results on the critical Galton-Watson process with immigration. *J. Austral. Math. Soc.*, **13**, 277-290.
- Park, Y. and Oh, C.W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers*, **38**, 287-302.
- Qui, Y. and Reeves, J. (2002). On sequential estimation for branching processes with immigration. *Stochastic Process. Appl.*, **100**, 41-51.
- Quine, M.P. (1970). The multitype Galton-Watson process with immigration. *J. Appl. Prob.*, **7**, 411-422.
- Rice, J. (1979). On the estimation of the parameters of a power spectrum. *J. Mult. Analysis*, **9**, 378-392.
- Seneta, E. (1969). Functional equations and the Galton-Watson process. *Adv. Appl. Prob.*, **1**, 1-42.
- Seneta, E. (1970a). An explicit limit theorem for the critical Galton-Watson process with immigration. *J. R. Statist. Soc.*, **B32**, 149-152.
- Seneta, E. (1970b). A note on the supercritical Galton-Watson process with immigration. *Math. Biosc.*, **6**, 305-311.
- Shete, S. and Sriram, T.N. (1998). Fixed precision estimator of the offspring mean in branching processes. *Stochastic Process. Appl.*, **77**, 17-33.
- Silva, I. e Silva, M.E. (2002). Selecco de ordem em modelos INAR. Submetido para publicao nas *Actas do X Congresso Anual da SPE*.
- Silva, M.E. and Oliveira, V.L. (2000). Difference equations for the higher order moments and cumulants of the INAR(1) model. Preprint CMAUP 9/00.

- Smoluchowski, M.v. (1916). Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Phys. Z.*, **17**, 557,585.
- Sprott, D.A. (1983). Estimating the parameters of a convolution by maximum likelihood. *J. Amer. Statist. Ass.*, **78**, 457-460.
- Sriram, T.N., Basawa, I.V. and Huggins, R.M. (1991). Sequential estimation for branching processes with immigration. *Ann. Statist.*, **19**, 2232-2243.
- Steutel, F.W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Prob.*, **7**, 893-899.
- Walker, A.M. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary non-deterministic series. *J. Aust. Math. Soc.*, **4**, 107-128.
- Wei, C.Z. and Winnicki, J. (1987). A unified estimation theory for the branching process with immigration. Technical Report, Univ. Maryland.
- Wei, C.Z. and Winnicki, J. (1989). Some asymptotic results for the branching process with immigration. *Stochastic Process. Appl.*, **31**, 261-281.
- Wei, C.Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *Ann. Statist.*, **18**, No. 4, 1757-1773.
- Westgren, A. (1916). Die Veränderungsgeschwindigkeit der lokalen Teilchenkonzentration in kolloiden Systemen (Erste Mitteilung). *Ark. Mat. Astron. Fys.*, **11**, No. 14, 1-25.
- Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.*, **2**, 423-434.
- Whittle, P. (1954). Some recent contributions to the theory of stationary processes. In *A study in the Analysis of Stationary Time Series, 2nd Edn* (ed. H. Wold). Stockholm: Almqvist and Wiksell. Appendix 2, 196-228.
- Winnicki, J. (1987). Estimation theory for the offspring and immigration distributions variances in a branching process with immigration. Technical Report, Columbia Univ..

Winnicki, J. (1988). Estimation theory for the branching process with immigration. *Contemp. Math.*, **80**, 301-322.

Yanev, N.M. and Tchoukova-Dantcheva, S. (1980). On the statistics of branching processes with immigration. *C.R. Acad. Bulgare Sci.*, **33**, 463-471.

Zeger, S. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics*, **44**, 1019-1031.

“Das schönste Glück des denkenden Menschen ist, das Erforschliche erforscht zu haben und das Unerforschliche ruhig zu verehren. ”

Johann Wolfgang von Goethe (1749-1832)

(A maior felicidade da pessoa que pensa é ter investigado o investigável e respeitar calmamente aquilo que não se pode investigar.)