

*Estatísticas de ordem superior dos modelos INAR***Isabel Silva***Fac. de Engenharia da Univ. do Porto, Dep. de Engenharia Civil e Fac. de Ciências da Univ. do Porto, Dep. de Matemática Aplicada***Maria Eduarda Silva***Fac. de Ciências da Universidade do Porto, Departamento de Matemática Aplicada*

Resumo: Nas últimas décadas, as estatísticas de ordem superior de um processo estocástico têm sido aplicadas em várias áreas, especialmente onde se suspeita de desvios de gaussianidade e linearidade. Neste trabalho são considerados processos Auto-Regressivos de valor INteiro, INAR, que são não lineares. São propostos dois métodos para a estimação dos parâmetros destes modelos, que utilizam estatísticas de ordem superior. São exibidos os resultados de um estudo de Monte Carlo que investiga o desempenho dos estimadores propostos. Por último, os métodos são aplicados a um conjunto de réplicas correspondentes ao número de movimentos corporais de recém-nascidos.

Palavras-chave: modelos INAR, momentos e cumulantes de ordem superior, equações recursivas de terceira ordem, método dos mínimos quadrados.

Abstract: During the last decades, the high-order statistics of a stochastic process have been applied in several fields, specially in problems where is conjectured a lack of Gaussianity and/or non-linearity. In this work, the INteger-valued AutoRegressive, INAR, process, which is a non-linear process, is considered. Two estimation methods for the parameters of an INAR model that use high-order statistics are proposed. A Monte Carlo study, to investigate the performance of the estimators, is carried out. Finally, these methods are applied to a set of observations consisting in the number of body movements of newborn babies.

Keywords: INAR models, high-order moments and cumulants, third-order recursive equations, least square method.

1 Introdução

Nas últimas décadas, as estatísticas de ordem superior (HOS) têm sido alvo de interesse crescente em diversos campos de aplicação. Por estatísticas de ordem superior entendem-se os momentos e cumulantes de ordem superior a dois, no domínio do tempo, e as correspondentes transformadas de Fourier multidimensionais (denominadas poli-espectro), no domínio da frequência. Neste trabalho é considerado, unicamente, o domínio do tempo.

Seja $\{Y_t, t \in \mathbb{Z}\}$ um processo estocástico estacionário até a ordem k , e seja

$M(\theta_1, \dots, \theta_k)$ a função geradora de momentos conjunta de $Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}$. Então o k -ésimo momento conjunto de $Y_t, Y_{t+s_1}, \dots, Y_{t+s_{k-1}}$, é uma função de $k-1$ variáveis definida por

$$\mu_Y(s_1, \dots, s_{k-1}) = E[Y_t Y_{t+s_1} \dots Y_{t+s_{k-1}}],$$

com $\mu_Y = E[Y_t]$. Seja, agora, $K(\theta_1, \dots, \theta_k) = \log(M(\theta_1, \dots, \theta_k))$ a função geradora de cumulantes. O k -ésimo cumulante conjunto de $Y_t, Y_{t+s_1}, \dots, Y_{t+s_{k-1}}$, é o coeficiente de $\theta_1 \dots \theta_k$ na série de Taylor de $K(\theta_1, \dots, \theta_k)$, é também uma função de $k-1$ variáveis e denota-se por $C_Y(s_1, \dots, s_{k-1})$. Note-se que, se um processo estocástico é estacionário, então são satisfeitas as seguintes relações de simetria (Nikias e Petropulu, 1993)

$$\mu_Y(m, k) = \mu_Y(k, m) = \mu_Y(-k, m-k) = \mu_Y(k-m, -m), \quad m, k > 0.$$

Os cumulantes satisfazem relações análogas. Assim, dados os valores dos momentos (cumulantes) de terceira ordem no “triângulo” infinito limitado pelas rectas $m=0$ e $m=k$ para $m, k > 0$, estes ficam conhecidos em todo o espaço (figura 1).

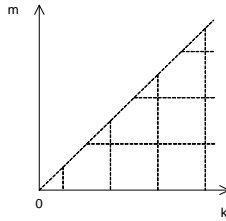


Figura 1: Parte da região ilimitada que determina completamente os momentos (cumulantes) de 3ª ordem.

Uma vez que os modelos INAR são não lineares, a informação contida nos momentos e cumulantes de segunda ordem não é suficiente para descrever completamente estes processos. Assim, neste trabalho é apresentada a caracterização de terceira ordem dos processos INAR. Adicionalmente, são propostos dois métodos de estimação que utilizam esta caracterização. O primeiro, é uma generalização da estimação de Yule-Walker, usando cumulantes de terceira ordem. O segundo utiliza o método dos mínimos quadrados sobre os momentos de terceira ordem. O desempenho das estimativas obtidas através dos dois métodos propostos são comparados por um estudo de simulação. Finalmente, são apresentados os resultados da aplicação destes métodos a um conjunto real de observações, correspondentes ao número total de movimentos corporais de recém-nascidos.

2 Caracterização de terceira ordem dos modelos INAR

Seja $\{X_t, t \in \mathbb{Z}\}$ um processo INAR(p) que satisfaz a seguinte equação às diferenças (Du e Li, 1991 e Latour, 1998)

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

onde $\alpha_i \geq 0$ para $i = 1, \dots, p-1$, e $\alpha_p > 0$ e o processo das inovações, $\{e_t\}$, é um conjunto de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) de valor inteiro não negativo, seguindo uma distribuição discreta com $E[e_t] = \mu_e$, $\text{Var}[e_t] = \sigma_e^2$ e $E[e_t^3] = \gamma_e$. O símbolo $*$ representa a operação *thinning* generalizada (Steutel e Van Harn, 1979 e Gauthier e Latour, 1994), definida por $\alpha_i * X_{t-i} = \sum_{j=1}^{X_{t-i}} Y_j^{(i)}$, onde $\{Y_j^{(i)}\}$, denominado série de contagem, é um conjunto de variáveis aleatórias i.i.d. de valor inteiro não negativo, seguindo uma distribuição discreta tal que $E[Y_j^{(i)}] = \alpha_i$, $\text{Var}[Y_j^{(i)}] = \sigma_i^2$ e $E[(Y_j^{(i)})^3] = \gamma_i$. Assume-se, ainda, que X_t e e_t são independentes de todas séries de contagem.

Usualmente, $\{X_t\}$ diz-se um processo INAR de Poisson se as séries de contagem são distribuídas segundo a distribuição de Bernoulli, com parâmetro α_i , e o processo das inovações segue a distribuição de Poisson.

A caracterização de ordem superior dos processos INAR, em termos de momentos, cumulantes e espectros, foi obtida por Silva e Oliveira (2004, 2005). Para completar esta caracterização, a forma escalar dos momentos e cumulantes de terceira ordem em $(0, 0)$ são aqui apresentados. Assim, utilizando a definição e as propriedades da operação *thinning* generalizada (Silva e Silva, 2003), mostra-se que os momentos de terceira ordem de um processo INAR(p) estacionário satisfazem um conjunto de equações do tipo Yule-Walker que podem ser escritas através de

$$\begin{aligned} \mu_X(0, 0) &= \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \alpha_i \alpha_j \alpha_k \mu_X(i-j, i-k) + 3 \sum_{i=1}^p \sum_{j=1}^p \alpha_j \sigma_i^2 \mu_X(i-j) \\ &\quad + 3\mu_e \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j \mu_X(i-j) + 3\mu_X(\sigma_e^2 + \mu_e^2) \sum_{i=1}^p \alpha_i + \gamma_e \\ &\quad + 3\mu_X \mu_e \sum_{i=1}^p \sigma_i^2 + \mu_X \sum_{i=1}^p (\gamma_i - 3\alpha_i \sigma_i^2 - \alpha_i^3), \end{aligned} \quad (2a)$$

$$\mu_X(0, k) = \sum_{i=1}^p \alpha_i \mu_X(0, k-i) + \mu_e \mu_X(0), \quad k > 0, \quad (2b)$$

$$\begin{aligned} \mu_X(k, k) &= \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j \mu_X(k-i, k-j) + \sum_{i=1}^p \sigma_i^2 \mu_X(k-i) \\ &\quad + 2\mu_e \mu_X(k) - \mu_X(\mu_e^2 - \sigma_e^2), \quad k > 0, \end{aligned} \quad (2c)$$

$$\mu_X(k, m) = \sum_{i=1}^p \alpha_i \mu_X(k, m-i) + \mu_e \mu_X(k), \quad m > k > 0. \quad (2d)$$

Para os cumulantes, tem-se que

$$\begin{aligned} C_X(0, 0) &= \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \alpha_i \alpha_j \alpha_k \mu_X(i-j, i-k) + 3 \sum_{i=1}^p \sum_{j=1}^p \alpha_j \sigma_i^2 \mu_X(i-j) + \gamma_e \\ &\quad + 3(\mu_e - \mu_X) \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j \mu_X(i-j) + 3\mu_X(\mu_e - \mu_X) \sum_{i=1}^p \sigma_i^2 + 2\mu_X^3 \\ &\quad - 6\mu_e \mu_X^2 \sum_{i=1}^p \alpha_i - 3\mu_e(\mu_e^2 + \sigma_e^2) + \mu_X \sum_{i=1}^p (\gamma_i - 3\alpha_i \sigma_i^2 - \alpha_i^3), \quad (3a) \end{aligned}$$

$$C_X(0, k) = \sum_{i=1}^p \alpha_i C_X(0, k-i), \quad k > 0, \quad (3b)$$

$$C_X(k, k) = \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j C_X(k-i, k-j) + \sum_{i=1}^p \sigma_i^2 C_X(k-i), \quad k > 0, \quad (3c)$$

$$C_X(k, m) = \sum_{i=1}^p \alpha_i C_X(k, m-i), \quad m > k > 0. \quad (3d)$$

Como notado por Silva e Oliveira (2005), os momentos e cumulantes de terceira ordem de um processo INAR satisfazem equações do tipo Yule-Walker análogas às satisfeitas pelos processos bilineares, confirmando a estrutura não linear dos processos INAR.

3 Estimação dos parâmetros de processos INAR através de estatísticas de ordem superior

3.1 Equações recursivas de terceira ordem

As equações Recursivas de Terceira Ordem (TOR) descrevem os momentos e/ou cumulantes de terceira ordem dos processos estocásticos. Este termo foi utilizado inicialmente por Raghuvver e Nikias (1985, 1986), que estabeleceram as equações TOR para modelos AR com processo ruído branco não gaussiano e, posteriormente, as utilizaram para estimar os parâmetros do processo.

Silva e Silva (2003) obtiveram as equações recursivas de terceira ordem para os modelos INAR(p), definidos em (1), em termos de momentos e cumulantes. Em particular, tem-se que a equação recursiva de terceira ordem satisfeita por

(3b), (3c) e (3d), é dada por

$$C_X(-k, -m) - \sum_{i=1}^p \alpha_i C_X(i - k, i - m) = \delta(k) \sum_{i=1}^p \sigma_i^2 C_X(i - m), \quad (4)$$

onde $0 \leq k \leq m, m \neq 0$, e $\delta(\cdot)$ é a função delta de Kronecker, definida por $\delta(0) = 1, \delta(x) = 0$ se $x \neq 0$.

Assim, fazendo $k = m > 0$, a equação (4) satisfaz a seguinte representação matricial

$$\mathbf{C}_{3,x} \boldsymbol{\alpha} = \mathbf{c}_{3,x}, \quad (5)$$

ou equivalentemente

$$\begin{bmatrix} C_X(0,0) & C_X(1,1) & \cdots & C_X(p-1,p-1) \\ C_X(0,1) & C_X(0,0) & \cdots & C_X(p-2,p-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(0,p-1) & C_X(0,p-2) & \cdots & C_X(0,0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} C_X(0,1) \\ C_X(0,2) \\ \vdots \\ C_X(0,p) \end{bmatrix},$$

onde $\mathbf{C}_{3,x}$ é a matriz Toeplitz não simétrica dos cumulantes de terceira ordem, de dimensão $p \times p$.

Então, dada uma realização, $\{X_1, \dots, X_n\}$, de um processo INAR(p), a equação TOR dada em (5), pode ser utilizada para obter estimativas dos parâmetros do processo através do seguinte procedimento:

- dividir os dados em B blocos com M observações por bloco, de modo que $n = BM$,
- em cada bloco, obter as estimativas (enviesadas) dos cumulantes de terceira ordem

$$\hat{C}_X^{(i)}(k, k) = \frac{1}{M} \sum_{j=1}^{M-k} (X_j^{(i)} - \bar{X}^{(i)})(X_{j+k}^{(i)} - \bar{X}^{(i)})^2, \quad k = 0, \dots, p-1,$$

$$\hat{C}_X^{(i)}(0, k) = \frac{1}{M} \sum_{j=1}^{M-k} (X_j^{(i)} - \bar{X}^{(i)})^2 (X_{j+k}^{(i)} - \bar{X}^{(i)}), \quad k = 1, \dots, p,$$

onde $X_j^{(i)}$ é a j -ésima observação do i -ésimo bloco e $\bar{X}^{(i)}$ é a média amostral do i -ésimo bloco, $i = 1, \dots, B$, dada por $\bar{X}^{(i)} = \frac{1}{M} \sum_{j=1}^M X_j^{(i)}$,

- obter as estimativas “totais” dos cumulantes de terceira ordem, calculando a média sobre todos os blocos,

$$\hat{C}_X(k, k) = \frac{1}{B} \sum_{i=1}^B \hat{C}_X^{(i)}(k, k), \quad \hat{C}_X(0, k) = \frac{1}{B} \sum_{i=1}^B \hat{C}_X^{(i)}(0, k),$$

para $k = 0, \dots, p-1$, e $k = 1, \dots, p$, respectivamente.

- substituir os cumulantes de terceira ordem teóricos pelos estimados, na equação (5), de modo a obter

$$\hat{\mathbf{C}}_{3,x}\hat{\boldsymbol{\alpha}} = \hat{\mathbf{c}}_{3,x}, \quad (6)$$

- resolver (6) em ordem a $\hat{\boldsymbol{\alpha}}$.

Note-se que o número de blocos pode ser igual a 1. Por outro lado, em situações em que estão disponíveis réplicas independentes da mesma série temporal de valores inteiros não negativos, este método pode ser aplicado directamente considerando que cada réplica corresponde a um bloco.

As estimativas para μ_e e σ_e^2 podem ser obtidas, respectivamente, através de $\hat{\mu}_e = \bar{X}(1 - \sum_{i=1}^p \hat{\alpha}_i)$, $\hat{\sigma}_e^2 = \hat{V}_p - \bar{X} \sum_{i=1}^p \hat{\alpha}_i(1 - \hat{\alpha}_i)$, onde $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$, $\hat{V}_p = \hat{R}(0) - \sum_{i=1}^p \hat{\alpha}_i \hat{R}(i)$, e $\hat{R}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})$.

Note-se que a variância assintótica de $(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ depende do cumulante de sexta ordem do processo, que é difícil de obter. Assim, o desempenho dos estimadores TOR é investigado através de um estudo de simulação, cujos resultados são apresentados na secção 4.

3.2 Mínimos quadrados utilizando estatísticas de ordem superior

Nesta secção propõem-se um método de estimação de parâmetros de processos INAR baseado em momentos de terceira ordem e que utiliza o método dos mínimos quadrados para minimizar os erros entre as observações e o modelo ajustado aos dados. Foi utilizado inicialmente por Al-Smadi e Alshamali (2002) na estimação dos parâmetros de processos ARMA não gaussianos com ruído branco aditivo gaussiano.

Seja $\{X_t\}$ um processo INAR(p) cujos momentos de terceira ordem estão definidos em (2). Note-se que $\mu_X(0, k)$, pode ser escrito na seguinte forma matricial,

$$\boldsymbol{\mu}_{3,x} = \mathbf{M}_{3,x}\boldsymbol{\alpha} + \mu_e \mu_X(0)\mathbf{1}_p, \quad (7)$$

onde $\boldsymbol{\mu}_{3,x} = [\mu_X(0, 1) \ \cdots \ \mu_X(0, p)]^T$ é o vector dos momentos de terceira ordem do processo, $\boldsymbol{\alpha} = [\alpha_1 \ \cdots \ \alpha_p]^T$ é o vector de coeficientes, $\mu_X(0) = E[X_t^2]$ é o momento de segunda ordem, $\mathbf{1}_p = [1 \ \cdots \ 1]^T$ é um vector $p \times 1$, e

$$\mathbf{M}_{3,x} = \begin{bmatrix} \mu_X(0, 0) & \mu_X(1, 1) & \cdots & \mu_X(p-1, p-1) \\ \mu_X(0, 1) & \mu_X(0, 0) & \cdots & \mu_X(p-2, p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_X(0, p-1) & \mu_X(0, p-2) & \cdots & \mu_X(0, 0) \end{bmatrix} \quad (8)$$

é a matriz Toeplitz não simétrica dos momentos de terceira ordem.

Suponha-se que é ajustado aos dados um modelo INAR(p), cujos momentos de terceira ordem são $\mathbf{H}\boldsymbol{\theta}$, onde

$$\mathbf{H} = [\mathbf{M}_{3,x} \quad \mu_X(0)\mathbf{1}_p], \quad \boldsymbol{\theta} = [\alpha_1 \ \cdots \ \alpha_p \ \mu_e]^T.$$

O erro quadrático entre os momentos de terceira ordem do modelo ajustado e os do modelo dado pelas observações, $\boldsymbol{\mu}_{3,x}$, é

$$L(\boldsymbol{\theta}) = (\boldsymbol{\mu}_{3,x} - \mathbf{H}\boldsymbol{\theta})^T (\boldsymbol{\mu}_{3,x} - \mathbf{H}\boldsymbol{\theta}). \quad (9)$$

Então, é proposto estimar os parâmetros do processo através da minimização de $L(\boldsymbol{\theta})$. Na prática, os estimadores são calculados pela minimização de uma versão amostral do erro,

$$\boldsymbol{\theta}^* = \min_{\boldsymbol{\theta}} \{\hat{L}(\boldsymbol{\theta})\} = \min_{\boldsymbol{\theta}} \{(\hat{\boldsymbol{\mu}}_{3,x} - \hat{\mathbf{H}}\boldsymbol{\theta})^T (\hat{\boldsymbol{\mu}}_{3,x} - \hat{\mathbf{H}}\boldsymbol{\theta})\}, \quad (10)$$

obtida pela substituição, em $\boldsymbol{\mu}_{3,x}$ e \mathbf{H} , dos momentos teóricos pelos amostrais,

$$\hat{\mu}_X(0) = \frac{1}{n} \sum_{t=1}^n X_t^2, \quad \hat{\mu}_X(0, k) = \frac{1}{n} \sum_{t=1}^{n-k} X_t^2 X_{t+k}, \quad \hat{\mu}_X(k, k) = \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}^2.$$

Como anteriormente, o desempenho dos estimadores obtidos por (10) é investigado através de um estudo de simulação, cujos resultados são apresentados na secção seguinte.

4 Estudo de simulação

De modo a verificar e comparar o desempenho dos métodos de estimação propostos na secção anterior, fez-se um estudo empírico das características assintóticas utilizando métodos de Monte-Carlo. Foram simuladas 500 repetições de processos INAR de Poisson de ordens $p = 1, 2, 3$, com diferentes valores nos parâmetros e amostras de dimensão $n = 60, 300, 600, 1200$. Em cada caso, foram estimados os parâmetros pelos dois métodos de estimação anteriormente apresentados: equações recursivas de terceira ordem com número de blocos igual a $B = 1, 2, 4, 6$, denotados por TOR_1b, TOR_2b, TOR_4b e TOR_6b, respectivamente, e mínimos quadrados utilizando estatísticas de ordem superior, LS_HOS.

Para as estimativas obtidas através das equações TOR, de uma forma geral verifica-se que fixando o número de blocos e aumentando a dimensão de cada bloco, o enviesamento médio e o desvio padrão amostrais das estimativas decrescem, o que indica que os estimadores são não enviesados e consistentes. As mesmas conclusões são retiradas quando se fixa a dimensão da amostra e se aumenta o número de blocos a considerar. Da mesma maneira, para as estimativas obtidas através do método LS_HOS, incrementando o número de observações por amostra, observa-se um decréscimo do enviesamento e desvio padrão das estimativas, indicando a consistência e o não enviesamento destes estimadores.

Na Tabela 1, apresentam-se os valores das estimativas médias e do desvio padrão amostral (entre parêntesis) obtidos pelos dois métodos de estimação para 500 repetições de modelos INAR(1) de Poisson, $X_t = \alpha * X_{t-1} + e_t$, $e_t \sim \mathcal{Po}(1)$, com $\alpha = 0.1, 0.5, 0.9$, e dimensões de amostra $n = 60, 300$ e 600 . Os valores

para $n = 1200$ observações, $\lambda = 3$ e número de blocos $B = 2, 6$, não são mostrados, porque são análogos aos apresentados. Analisando a tabela observa-se que as estimativas obtidas por LS_HOS são, claramente, melhores (em termos de menor enviesamento e desvio padrão) que as obtidas através de TOR, embora para amostras de dimensão grande, por exemplo $n = 1200$, a diferença seja menor. Quando a dimensão da amostra é pequena ($n < 100$), e sobretudo nas estimativas TOR, verifica-se que o α é (em média) sub-estimado, enquanto que o λ é sobre-estimado. Quando o valor do parâmetro α se aproxima da região de não estacionariedade, as estimativas obtidas pelas equações TOR apresentam um aumento acentuado do desvio padrão e do enviesamento médio. Contudo, estes valores decrescem quando o número de observações aumenta. É de notar que as estimativas LS_HOS não apresentam estas características quando os valores dos parâmetros são próximos da região não estacionariedade.

Tabela 1: Médias amostrais e desvios padrões amostrais (entre parêntesis) das estimativas para 500 repetições de modelos INAR(1) de Poisson.

n.º de obs.	(α, λ)	$\hat{\alpha}$			$\hat{\lambda}$		
		TOR 1b	TOR 4b	LS_HOS	TOR 1b	TOR 4b	LS_HOS
60	(0.1, 1)	0.0699	0.0054	0.0720	1.0345	1.1082	1.0116
		(0.3072)	(0.3257)	(0.1279)	(0.3821)	(0.4074)	(0.1772)
	(0.5, 1)	0.2548	0.2053	0.4307	1.5151	1.5735	1.1068
300	(0.1, 1)	(1.2765)	(1.8781)	(0.1260)	(3.1344)	(3.9096)	(0.2670)
		0.4611	0.6284	0.8069	4.9469	4.6427	1.7322
	(0.9, 1)	(6.6959)	(6.8138)	(0.0793)	(62.9861)	(71.5684)	(0.7302)
600	(0.1, 1)	0.0938	0.0800	0.0917	1.0034	1.0187	1.0036
		(0.1170)	(0.1171)	(0.0612)	(0.1426)	(0.1432)	(0.0802)
	(0.5, 1)	0.4699	0.4368	0.4878	1.0636	1.1297	1.0210
1200	(0.1, 1)	(0.1474)	(0.1543)	(0.0579)	(0.3050)	(0.3174)	(0.1180)
		0.7915	0.9174	0.8822	1.9776	0.8050	1.1424
	(0.9, 1)	(1.7573)	(2.5485)	(0.0288)	(16.8157)	(23.5895)	(0.2912)
3000	(0.1, 1)	0.0976	0.0904	0.0993	1.0018	1.0097	0.9983
		(0.0784)	(0.0787)	(0.0440)	(0.0960)	(0.0963)	(0.0607)
	(0.5, 1)	0.4791	0.4643	0.4915	1.0414	1.0711	1.0128
6000	(0.1, 1)	(0.1028)	(0.1075)	(0.0406)	(0.2133)	(0.2237)	(0.0829)
		0.7827	0.7773	0.8924	2.1458	2.2400	1.0604
	(0.9, 1)	(1.0487)	(2.5231)	(0.0180)	(10.2181)	(25.1472)	(0.1816)

Quando a ordem do processo aumenta, observa-se o mesmo tipo de comportamento: as estimativas TOR apresentam uma maior dispersão de valores que as estimativas LS_HOS, dispersão que diminui quando o número de observações a considerar aumenta. Para ilustrar estes resultados, apresentam-se na Figura 2, os *boxplots* do enviesamento amostral, $bias_{\theta_i}(j) = \hat{\theta}_i(j) - \theta_i$, $i = 1, 2, 3$; $j = 1, \dots, 500$, onde θ_i é o valor verdadeiro do parâmetro e $\hat{\theta}_i(j)$ é a estimativa do parâmetro θ_i na j -ésima repetição, para um processo INAR(2) de Poisson com vector de parâmetros $\theta = (0.4, 0.3, 1.0)$, e considerando dimensões de amostras $n = 300, 1200$.

Por questões de espaço, e porque nas simulações com modelos INAR(3) de Poisson as conclusões são as mesmas que nas ordens inferiores, não são apresen-

tados os resultados referentes a esta ordem.

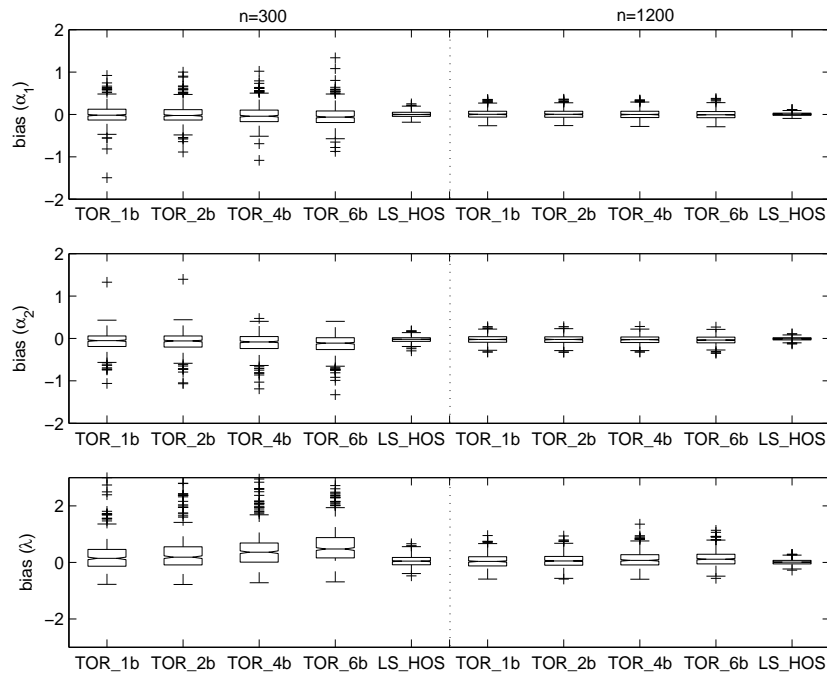


Figura 2: *Boxplots* do enviesamento amostral das estimativas dos parâmetros do modelo INAR(2), $X_t = 0.4 * X_{t-1} + 0.3 * X_{t-2} + e_t$, $e_t \sim \mathcal{P}o(1)$.

5 Aplicação

Nesta secção consideram-se 8 séries referentes ao número total de movimentos corporais, por minuto, de 8 recém-nascidos (Stoffer, 1991). Silva e Silva (2004) trataram este conjunto de dados como 8 réplicas do mesmo modelo INAR(1) de Poisson, $X_{k,t} = \alpha * X_{k,t-1} + e_{k,t}$, $e_{k,t} \sim \mathcal{P}o(\lambda)$, $k = 1, \dots, 8$, $t = 1, \dots, 128$, e os parâmetros foram estimados utilizando estimação de Yule-Walker (YW), o método dos Mínimos Quadrados Condicionais (CLS) e estimação de Whittle (WHT).

Neste caso, os métodos baseados em HOS são utilizados para obter estimativas dos parâmetros de modo a comparar com os outros métodos de estimação. Como foi referido na secção 3.1, a extensão do método TOR à situação das réplicas é directa: cada réplica é considerada como sendo um bloco. Por outro lado, a adaptação do método LS_HOS é feita considerando as ideias do método TOR: estimam-se os momentos de terceira ordem em cada réplica, obtêm-se os

estimadores “totais” (fazendo a média sobre as 8 réplicas) e constrói-se a matriz $\hat{\mathbf{H}}$ e o vector $\hat{\boldsymbol{\mu}}_{3,x}$, de modo a minimizar a versão amostral de (9). As estimativas HOS apresentam-se na Tabela 2, juntamente com as estimativas obtidas por Silva e Silva (2004).

Tabela 2: Estimativas dos parâmetros de 8 réplicas do mesmo processo INAR(1), correspondentes ao número de movimentos corporais, por minuto, de recém-nascidos.

Método	$\hat{\alpha}$	$\hat{\lambda}$
TOR	0.106	0.267
LS_HOS	0.132	0.263
YW	0.129	0.260
CLS	0.144	0.255
WHT	0.133	0.280

Verifica-se que as estimativas HOS (parte superior da tabela) estão de acordo com estimativas obtidas por outros métodos de estimação (parte inferior da tabela). Mais uma vez, a análise dos correlogramas dos resíduos obtidos a partir das estimativas HOS permitem concluir que a classe das réplicas de modelos INAR(1) é razoável na descrição destes dados reais.

6 Comentários finais

As estatísticas de ordem superior, nomeadamente momentos e cumulantes de terceira ordem, fornecem informação adicional na análise de processos não gaussianos e não lineares, como é o caso dos modelos INAR, permitindo uma melhor caracterização dos processos.

Neste caso, a informação adicional é utilizada na estimação dos parâmetros dos modelos INAR. Assim, são propostos dois métodos de estimação baseados em HOS. Verifica-se que o método dos mínimos quadrados usando estatísticas de ordem superior apresenta um ótimo desempenho, mesmo em situações onde a dimensão da amostra é pequena e os coeficientes se encontram próximos da região de não estacionaridade. O método das equações TOR apresenta a desvantagem de ser necessário a escolha do número de blocos. Contudo, uma vantagem adicional dos dois métodos propostos é a fácil adaptação dos mesmos a situações de medidas repetidas, especialmente em réplicas independentes do mesmo processo INAR.

Agradecimentos

As autoras estão gratas ao Professor D. S. Stoffer por ter fornecido os dados utilizados na secção § 5. A primeira autora agradece ao PRODEP III pelo apoio financeiro.

Referências

- [1] Al-Smadi, A. e Alshamali, A. (2002). Fitting ARMA models to linear non-Gaussian processes using higher order statistics. *Signal Processing*, Vol. 82, p. 1789-1793.
- [2] Du, J.-G. e Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, Vol. 12, p. 129-142.
- [3] Gauthier, G. e Latour, A. (1994). Convergence forte des estimateurs des paramètres d'un processus GENAR(p). *Annales des Sciences Mathématiques du Québec*, Vol. 18, p. 49-71.
- [4] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, Vol. 19, p. 439-455.
- [5] Nikias, C. L. e Petropulu, A. P. (1993). *Higher-order Spectral Analysis, a Non-linear Signal Processing Framework*. PTR Prentice Hall.
- [6] Raghuvver, M. R. e Nikias, C. L. (1985). Bispectrum estimation: a parametric approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, p. 1213-1230.
- [7] Raghuvver, M. R. e Nikias, C. L. (1986). Bispectrum estimation via AR modeling. *Signal Processing*, Vol. 10, p. 35-48.
- [8] Silva, M. E. e Oliveira, V. L. (2004). Difference equations for the higher order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis*, Vol. 25, p. 317-333.
- [9] Silva, M. E. e Oliveira, V. L. (2005). Difference equations for the higher order moments and cumulants of the INAR(p) model. *Journal of Time Series Analysis*, Vol. 26, p. 17-36.
- [10] Silva, I. e Silva, M. E. (2003). High-order statistics for INAR models. *Technical Report DMA/01/03*, Departamento de Matemática Aplicada, Faculdade de Ciências, Universidade do Porto.
- [11] Silva, I. e Silva, M. E. (2004). Estimação de Whittle para réplicas de processos INAR. Em *Estatística com Acaso e Necessidade* (Rodrigues, P. M. M., Rebelo, E. L. e Rosado, F., eds.), p. 767-778. Edições SPE.
- [12] Steutel, F. W. e Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability*, Vol. 7, p. 893-899.
- [13] Stoffer, D. S. (1991). Walsh-Fourier analysis and its statistical applications. *Journal of the American Statistical Association*, Vol. 86, p. 461-479.