

*Estimação de Whittle para réplicas de processos INAR***Isabel Marques da Silva**¹*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: Réplicas independentes de séries temporais são uma classe particular de medidas repetidas, em que a mesma resposta é medida ao longo do tempo em vários sujeitos (unidades experimentais). Neste caso consideram-se réplicas independentes de séries de contagem que são modeladas através de processos AutoRegressivos de valor INteiro de ordem p , INAR(p). Neste trabalho, generalizam-se, para este tipo de dados, alguns dos métodos de estimação referidos na literatura e propõe-se um método de estimação no domínio da frequência que utiliza o critério de Whittle.

Palavras-chave: Critério de Whittle, modelos INAR, réplicas de séries temporais.

Abstract: Replicated time series are a particular type of repeated measures, which consist of time sequence of measurements taken from several subjects (experimental units). In this work, we consider independent replicates of count time series that are modelled by an integer-valued p th-order autoregressive processes, INAR(p). We generalize some of the estimation methods referred in literature and we propose an estimation method in the frequency domain that uses the Whittle criterion.

Keywords: Whittle Criterion, INAR models, replicated time series.

1 Introdução

Nos últimos anos, o estudo de medidas repetidas (Crowder e Hand, 1990) tem sido alvo de um interesse crescente. Esta classe de dados consiste em medições, contagens ou respostas categóricas de vários indivíduos (unidades experimentais) recolhidas ao longo do tempo. Em particular, assume-se que existe correlação entre as observações correspondentes a cada indivíduo mas que existe independência entre os diferentes indivíduos. Por outras palavras, consideram-se réplicas independentes da mesma série temporal.

Neste trabalho, consideram-se réplicas independentes de séries de contagens, que são modeladas através do mesmo processo AutoRegressivo de valor Inteiro de ordem p , INAR(p) (Du e Li, 1991 e Gauthier e Latour, 1994). Propõem-se métodos de estimação no domínio do tempo, nomeadamente, Yule-Walker e Mínimos Quadrados Condicionais, e, ainda, um método de estimação no domínio

¹Isabel Marques da Silva agradece ao PRODEP III pelo apoio financeiro.

da frequência baseado no *critério de Whittle* (Whittle, 1953 e 1954). Os diferentes estimadores são comparados através de um estudo de simulação. Aplicam-se, ainda, estes métodos a um conjunto real de réplicas, correspondentes ao número total de movimentos corporais de recém-nascidos.

2 Réplicas Independentes de Processos INAR

Considerem-se r realizações independentes do mesmo processo estocástico,

$$\{X_{k,t}, k = 1, \dots, r; t = 1, \dots, n\},$$

em que a k -ésima réplica é um processo INAR(p) satisfazendo a seguinte equação às diferenças

$$X_{k,t} = \alpha_1 * X_{k,t-1} + \dots + \alpha_p * X_{k,t-p} + e_{k,t}, \quad (1)$$

onde $\{e_{k,t}\} \subseteq \mathbb{N}_0$ é um conjunto de variáveis aleatórias com $E[e_{k,t}] = \mu_e$ e $\text{Var}[e_{k,t}] = \sigma_e^2$ e $0 < \alpha_i < 1, i = 1, \dots, p$. O símbolo $*$ representa a operação *thinning* generalizada (Steutel e Van Harn, 1979 e Gauthier e Latour, 1994), definida por $\alpha_i * X_{k,t-i} = \sum_{j=1}^{X_{k,t-i}} Y_j^{(k,i)}$, onde $\{Y_j^{(k,i)}\} \subseteq \mathbb{N}_0$, denominado série de contagem, é um conjunto de variáveis aleatórias mutuamente independentes e identicamente distribuídas, tal que $E[Y_j^{(k,i)}] = \alpha_i$ e $\text{Var}[Y_j^{(k,i)}] = \sigma_i^2$. Assume-se, ainda, que $X_{k,t}$ e $e_{k,t}$ são independentes de todas séries de contagem.

Uma vez que a distribuição de Poisson é adequada para contar as ocorrências de um certo fenómeno por unidade de tempo, esta é a distribuição correntemente utilizada para as inovações, e neste caso chamamos ao processo obtido INAR de Poisson. Quanto à distribuição das séries de contagem, a mais usual é a distribuição de Bernoulli com parâmetro α e neste caso, a distribuição condicional de $\alpha * X$ dado X é Binomial com parâmetros (X, α) , razão pela qual é denominada operação *thinning* binomial.

Utilizando as propriedades da operação *thinning* (Silva e Oliveira, 2000a e 2000b), mostra-se que a função de autocovariância de um processo INAR(p) estacionário, $R(j) = E[(X_t - \mu_x)(X_{t+j} - \mu_x)]$, com $\mu_x = E[X_t]$, satisfaz um conjunto de equações do tipo Yule-Walker que podem ser escritas através de

$$\begin{cases} R(0) &= V_p + \sum_{i=1}^p \alpha_i R(i), \\ R(|j|) &= \sum_{i=1}^p \alpha_i R(|j-i|), \quad j \geq 1, \end{cases} \quad (2)$$

onde V_p (variância do erro de previsão linear a 1-passo) é definido por

$$V_p = \sigma_e^2 + \mu_x \sum_{i=1}^p \sigma_i^2. \quad (3)$$

A função de densidade espectral ou espectro, $f(\omega)$, de um processo INAR(p) é uma função contínua e positiva, definida como a transformada de Fourier da função de autocovariância, que pode ser escrita como (Latour, 1998)

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} R(j) e^{-i\omega j} = \frac{1}{2\pi} \frac{V_p}{\left|1 - \sum_{j=1}^p \alpha_j e^{-i\omega j}\right|^2}, \quad -\pi \leq \omega \leq \pi, \quad (4)$$

com V_p dado em (3).

Assume-se que o espectro do processo INAR(p) que modela as réplicas pertence a uma dada família de densidades espectrais, $f(\omega, \boldsymbol{\theta})$, com $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \mu_e, \sigma_e^2) \in \Theta$, onde Θ é um subconjunto aberto no espaço Euclidiano \mathbb{R}^{p+2} . O valor de $\boldsymbol{\theta}$ considera-se desconhecido e será estimado utilizando a matriz de observações dada por

$$\mathbf{X}_{r,n} = \begin{bmatrix} \mathbf{x}_{1,n} \\ \mathbf{x}_{2,n} \\ \vdots \\ \mathbf{x}_{r,n} \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,n} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{r,1} & X_{r,2} & \cdots & X_{r,n} \end{bmatrix}.$$

3 Métodos de Estimação

Para efeitos de estimação dos parâmetros do processo INAR(p) que modela as réplicas independentes, considera-se que a operação *thinning* é binomial e que o processo das inovações segue uma distribuição de Poisson ($\mu_e = \sigma_e^2 = \lambda$). Neste caso, o vector de parâmetros a estimar é $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \lambda)$.

3.1 Estimação de Yule-Walker

Considerem-se r réplicas independentes do mesmo processo estocástico, $\{X_{k,t} : k = 1, \dots, r; t = 1, \dots, n\}$. A função de autocovariância amostral da k -ésima réplica é dada por $\hat{R}_k(j) = \frac{1}{n} \sum_{t=1}^{n-j} (X_{k,t} - \bar{X}_k)(X_{k,t+j} - \bar{X}_k)$, $j \in \mathbb{Z}$, onde $\bar{X}_k = \frac{1}{n} \sum_{t=1}^n X_{k,t}$ é a média amostral. De modo a estimar a função de autocovariância do conjunto de réplicas, sugere-se a função de autocovariância média amostral, definida através de

$$\bar{R}(j) = \frac{1}{r} \sum_{k=1}^r \hat{R}_k(j) = \frac{1}{nr} \sum_{k=1}^r \sum_{t=1}^{n-j} (X_{k,t} - \bar{X}_k)(X_{k,t+j} - \bar{X}_k). \quad (5)$$

Assim sendo, propõem-se a generalização dos estimadores de Yule-Walker para o caso de réplicas independentes do mesmo processo INAR(p) de Poisson, através da substituição da função de autocovariância teórica pela função de autocovariância média amostral, (5), nas equações do tipo Yule-Walker dadas em

(2). Então, os estimadores de Yule-Walker de $\alpha_1, \dots, \alpha_p$ obtêm-se resolvendo, em ordem aos parâmetros, o seguinte sistema de equações lineares

$$\bar{\mathbf{R}}_{p-1} \hat{\boldsymbol{\alpha}} = \bar{\mathbf{r}} \Leftrightarrow \begin{bmatrix} \bar{R}(0) & \bar{R}(1) & \cdots & \bar{R}(p-1) \\ \bar{R}(1) & \bar{R}(0) & \cdots & \bar{R}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \bar{R}(p-1) & \bar{R}(p-2) & \cdots & \bar{R}(0) \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} \bar{R}(1) \\ \bar{R}(2) \\ \vdots \\ \bar{R}(p) \end{bmatrix}.$$

Como foi proposto por Al-Osh e Alzaid (1987), para o caso de um processo INAR(1) de Poisson, considera-se que um estimador razoável de λ é dado por

$$\hat{\lambda} = (1 - \hat{\alpha}_1 - \dots - \hat{\alpha}_p) \bar{\mathbf{X}}_{r,n},$$

onde $\bar{\mathbf{X}}_{r,n} = \frac{1}{nr} \sum_{k=1}^r \sum_{t=1}^n X_{k,t}$ é a média amostral sob todas as observações.

De acordo com Du e Li (1991), nos processos INAR(p) de Poisson, $\hat{R}_k(j)$ é um estimador fortemente consistente de $R_k(j)$. Portanto, $\bar{R}(j)$ é um estimador fortemente consistente de $\hat{R}_k(j)$, porque é uma média dos valores de $\hat{R}_k(j)$. Consequentemente, $\hat{\boldsymbol{\alpha}}$ e $\hat{\lambda}$ são também estimadores fortemente consistentes de $\boldsymbol{\alpha}$ e λ , respectivamente.

3.2 Mínimos Quadrados Condicionais

Este método foi proposto por Klimko e Nelson (1978) e consiste na minimização da soma dos quadrados dos desvios ao valor esperado condicional. Silva *et al.* (2003) propuseram este método para a estimação dos parâmetros de réplicas independentes do mesmo processo INAR(1) de Poisson. Neste trabalho generaliza-se para o caso de réplicas independentes de processos INAR(p) de Poisson.

Seja $\mathbf{x}_{k,n} = [X_{k,1}, \dots, X_{k,n}]$ a k -ésima réplica do processo INAR(p) de Poisson, com vector de parâmetros $\boldsymbol{\theta}$, e seja $F_{k,t} = \mathcal{F}(X_{k,1}, \dots, X_{k,t})$ a σ -álgebra gerada por $\{X_{k,1}, \dots, X_{k,t}\}$, para $t = 1, \dots, n$. Então, utilizando as propriedades da operação *thinning* (Silva e Oliveira, 2000a e 2000b e Gauthier e Latour, 1994), prova-se que o valor esperado condicional de $X_{k,t}$ dado $F_{k,t-1}$, para $k = 1, \dots, r$, $t = p+1, \dots, n$, é dado por

$$g(\boldsymbol{\theta}, F_{k,t-1}) = E[X_{k,t} | F_{k,t-1}] = \sum_{i=1}^p \alpha_i X_{k,t-i} + \lambda. \quad (6)$$

Considerando r réplicas do mesmo processo INAR(p) de Poisson, o estimador dos Mínimos Quadrados Condicionais de $\boldsymbol{\theta}$ é tal que minimiza

$$Q(\boldsymbol{\theta}) = \sum_{k=1}^r \sum_{t=p+1}^n (X_{k,t} - g(\boldsymbol{\theta}, F_{k,t-1}))^2 = \sum_{k=1}^r \sum_{t=p+1}^n \left(X_{k,t} - \sum_{i=1}^p \alpha_i X_{k,t-i} - \lambda \right)^2.$$

É fácil demonstrar que a função g dada em (6) e as derivadas parciais $\partial g/\partial\theta_i$ e $\partial^2 g/\partial\theta_i\partial\theta_j$, para $i, j = 1, \dots, p+1, i \neq j$, satisfazem todas as condições de regularidade propostas no Teorema 3.1 de Klimko e Nelson (1978). Portanto, os estimadores dos Mínimos Quadrados Condicionais para réplicas independentes de processos INAR(p) de Poisson são fortemente consistentes. Mais ainda, uma vez que para a distribuição de Poisson $E[|e_{k,t}|^3] < \infty$, pelo Teorema 3.2 dos mesmos autores, estes estimadores são assintoticamente normais.

3.3 Critério de Whittle

O critério proposto por Whittle, como aproximação da log-verosimilhança Gaussiana de um processo estocástico, foi obtido para várias classes de processos por Dzhaparidze e Yaglom (1983). Em particular, os autores introduziram a classe de processos *mixing não Gaussianos* para tratar o caso de modelos não lineares e não Gaussianos. Um processo X_t pertence a esta classe se satisfaz as condições:

- (i) X_t é estritamente estacionário;
- (ii) todos os momentos absolutos de X_t são finitos, i.e., $E[|X_t|^k] < \infty, t \in \mathbb{Z}, k \in \mathbb{N}$;
- (iii) seja $C_k(s_1, \dots, s_{k-1})$ o cumulante de ordem k do processo X_t , então

$$\sum_{s_1=-\infty}^{\infty} \dots \sum_{s_{k-1}=-\infty}^{\infty} |C_k(s_1, \dots, s_{k-1})| < \infty, \quad k = 2, 3, \dots$$

A última condição é uma condição *mixing* sobre X_t , que garante um rápido decréscimo da dependência estatística entre X_t e X_{t+s} , quando $s \rightarrow \infty$.

No caso de processos INAR(p) estacionários tais que $0 < \alpha_i < 1$ para $i = 1, \dots, p$, e $0 < pr(e_t = 0) < 1$, Franke e Subba Rao (1995) estabeleceram que qualquer solução de (1) é uma cadeia de Markov aperiódica e irredutível sobre \mathbb{N}_0^p . Uma vez que uma cadeia de Markov é aperiódica e irredutível se e só se é fortemente *mixing* (Rosenblatt, 1971, p. 207), então os processos INAR(p) satisfazem a condição (iii). Por outro lado, visto que os cumulantes de um processo podem ser escritos à custa dos seus momentos, a condição (ii) também é satisfeita por modelos INAR(p).

Silva e Silva (2003) obtiveram o critério de Whittle como aproximação da log-verosimilhança Gaussiana para os modelos INAR(p). Neste trabalho propõe-se a generalização do critério para réplicas independentes de modelos INAR.

Considere-se, então, a função de log-verosimilhança Gaussiana da k -ésima réplica $\mathbf{x}_{k,n} = [X_{k,1}, X_{k,2}, \dots, X_{k,n}]$,

$$\mathbf{L}(\mathbf{x}_{k,n}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \mathbf{R}_n) - \frac{1}{2} \mathbf{x}_{k,n} [\mathbf{R}_n]^{-1} \mathbf{x}_{k,n}^T,$$

onde

$$\mathbf{R}_n = [R(h, j)]_{h, j=1}^n = [R(n+1-h, n+1-j)]_{h, j=1}^n = \left[\int_{-\pi}^{\pi} e^{i\omega(h-j)} f(\omega) d\omega \right]_{h, j=1}^n,$$

é a matriz Toeplitz², $n \times n$, de covariâncias do vector $\mathbf{x}_{k,n}$, para $k = 1, \dots, r$. Generalizando para o conjunto de réplicas, tem-se que a função de log-verosimilhança Gaussiana da matriz de observações, $\mathbf{X}_{r,n}$, é

$$\mathbf{L}(\mathbf{X}_{r,n}) = -\frac{nr}{2} \log(2\pi) - \frac{r}{2} \log(\det \mathbf{R}_n) - \frac{1}{2} \sum_{k=1}^r \mathbf{x}_{k,n} [\mathbf{R}_n]^{-1} \mathbf{x}_{k,n}^T. \quad (7)$$

Foi demonstrado por Shaman (1975) e Murthy (1973), para os processos ARMA, e por Silva e Silva (2003), para os processos INAR, que a matriz $[\mathbf{R}_n]^{-1}$ pode ser aproximada por

$$\mathbf{R}_n^* = [R^*(h, j)]_{h, j=1}^n = [R^*(h-j)]_{h, j=1}^n = \left[\frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{e^{i\omega(h-j)}}{f(\omega)} d\omega \right]_{h, j=1}^n.$$

Então, a forma quadrática $\mathbf{x}_{k,n} [\mathbf{R}_n]^{-1} \mathbf{x}_{k,n}^T \approx \mathbf{x}_{k,n} \mathbf{R}_n^* \mathbf{x}_{k,n}^T$ pode ser escrita por $\frac{n}{2\pi} \int_{-\pi}^{\pi} \frac{I_k(\omega)}{f(\omega)} d\omega$, onde $f(\omega)$ é a função de densidade espectral do processo INAR(p), dada em (4), e $I_k(\omega)$ é o periodograma de $\mathbf{x}_{k,n}$, $I_k(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_{k,t} e^{-i\omega t} \right|^2$. Mais ainda, Silva e Silva (2003) demonstraram que, quando $n \rightarrow \infty$, $\det \mathbf{R}_n - V_p^n \rightarrow 0$, onde V_p está definido em (3).

Assim, aplicando estes resultados em (7), propõem-se a seguinte aproximação da função de log-verosimilhança Gaussiana

$$\mathbf{L}(\mathbf{X}_{r,n}) \approx -\frac{nr}{2} \log(2\pi) - \frac{nr}{2} \log(V_p) - \frac{n}{4\pi} \sum_{k=1}^r \int_{-\pi}^{\pi} \frac{I_k(\omega)}{f(\omega)} d\omega.$$

Utilizando a fórmula de Kolmogorov proposta por Silva e Silva (2003) para modelos INAR, que associa a variância do erro de previsão linear a 1-passo com a função de densidade espectral, $V_p = 2\pi \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega\right)$, obtém-se

$$\begin{aligned} \mathbf{L}(\mathbf{X}_{r,n}) &\approx -nr \log(2\pi) - \frac{nr}{4\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega - \frac{n}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\sum_{k=1}^r I_k(\omega)}{f(\omega)} \right) d\omega \\ &= -nr \log(2\pi) - \frac{nr}{4\pi} \int_{-\pi}^{\pi} \left(\log f(\omega) + \frac{\bar{I}(\omega)}{f(\omega)} \right) d\omega, \end{aligned}$$

onde $\bar{I}(\omega)$, periodograma médio amostral (Diggle e al-Wasel, 1993) dado por $\bar{I}(\omega) = \frac{1}{r} \sum_{k=1}^r I_k(\omega)$, é um estimador assintoticamente não enviesado mas não consistente da função de densidade espectral do conjunto de réplicas, $f(\omega)$.

²Uma matriz A , $n \times n$, é uma matriz Toeplitz se os valores ao longo das diagonais com “inclinação negativa” são constantes.

Portanto, propõem-se que o vector de parâmetros θ do processo INAR(p) que modela as réplicas seja estimado através da minimização do Critério de Whittle

$$\tilde{\mathbf{L}}(\mathbf{X}_{r,n}) = \frac{nr}{4\pi} \int_{-\pi}^{\pi} \left(\log f(\omega) + \frac{\bar{I}(\omega)}{f(\omega)} \right) d\omega,$$

sendo $\tilde{\theta} = \min_{\theta} \{\tilde{\mathbf{L}}(\mathbf{X}_{r,n})\}$ o estimador de Whittle de θ .

Na prática, obtém-se o estimador de Whittle de θ pela minimização da seguinte versão discreta de $\tilde{\mathbf{L}}(\mathbf{X}_{r,n})$ (proposta por Sesay e Subba Rao, 1992)

$$\tilde{\ell}(\mathbf{X}_{r,n}) = \frac{1}{n} \sum_{k=1}^r \sum_{j=1}^{[n/2]} \left(\log f(\omega_j) + \frac{I_k(\omega_j)}{f(\omega_j)} \right) = \frac{r}{n} \sum_{j=1}^{[n/2]} \left(\log f(\omega_j) + \frac{\bar{I}(\omega_j)}{f(\omega_j)} \right),$$

onde $f(\omega_j)$ é o valor da função de densidade espectral na frequência de Fourier, $\omega_j = 2\pi j/n$, para $j = 1, \dots, [n/2]^3$, e $\bar{I}(\omega_j)$ é a ordenada do periodograma médio amostral na mesma frequência.

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 conjuntos de 1, 5, 10 e 20 réplicas, com 30, 100, 256 e 512 observações, de processos INAR de Poisson de ordens 1, 2, e 3, com diferentes valores nos parâmetros. Para cada conjunto de réplicas, foram estimados os parâmetros pelos três métodos de estimação anteriormente expostos: Yule-Walker (YW), Mínimos Quadrados Condicionais (CLS) e Critério de Whittle (WHT).

Seja, então, $\hat{\theta}(j) = (\hat{\theta}_1(j), \dots, \hat{\theta}_p(j), \hat{\theta}_{p+1}(j)) = (\hat{\alpha}_1(j), \dots, \hat{\alpha}_p(j), \hat{\lambda}(j))$ o vector de estimativas na j -ésima repetição, $j = 1, \dots, 500$, do vector de parâmetros $\theta = (\theta_1, \dots, \theta_p, \theta_{p+1}) = (\alpha_1, \dots, \alpha_p, \lambda)$ do processo INAR(p) de Poisson que modela as réplicas independentes. Em cada réplica é calculado o enviesamento médio e o erro quadrático médio amostrais das estimativas, dados respectivamente por

$$bias_{\theta_i} = \frac{1}{500} \sum_{j=1}^{500} \hat{\theta}_i(j) - \theta_i, \quad mse_{\theta_i}(j) = \frac{1}{500} \sum_{j=1}^{500} (\hat{\theta}_i(j) - \theta_i)^2.$$

Nas réplicas de modelos INAR(1) de Poisson verifica-se que, fixando o número de réplicas e aumentando o número de observações, o enviesamento médio e o erro quadrático médio amostrais dos estimadores de $\theta = (\alpha, \lambda)$ decrescem, o que fornece uma indicação da consistência dos três métodos utilizados. As

³ $[z]$ é o maior inteiro menor ou igual a z , i.e. é a parte inteira de z , para $z \in \mathbb{R}$.

mesmas conclusões podem ser retiradas quando se fixa o número de observações e se aumenta o número de réplicas. Verifica-se que, em geral, o método de Yule-Walker sub-estima α e sobre-estima λ . Esta situação também é observada nas estimativas CLS e WHT quando o valor verdadeiro de α se aproxima de 1. As estimativas CLS e WHT têm comportamentos semelhantes, embora quando o número de observações por réplica é menor que 100, as estimativas WHT apresentam, em média, uma maior dispersão. Por outro lado, quando α está próximo de 1, a dispersão das estimativas WHT é menor que o das estimativas CLS.

n° de réplicas	n° de obs	$\hat{\alpha}$			$\hat{\lambda}$		
		YW	CLS	WHT	YW	CLS	WHT
1	30	0.5769 (0.1497)	0.6006 (0.1549)	0.6512 (0.1792)	2.7730 (1.0463)	2.6125 (1.0837)	2.1946 (0.6808)
	100	0.6609 (0.0794)	0.6689 (0.0794)	0.6786 (0.0819)	2.2649 (0.5442)	2.2081 (0.5476)	2.0389 (0.3329)
	256	0.6869 (0.0458)	0.6897 (0.0458)	0.6934 (0.0458)	2.0874 (0.3138)	2.0697 (0.3137)	2.0386 (0.2083)
5	30	0.6159 (0.0900)	0.7310 (0.0616)	0.7091 (0.0959)	2.5139 (0.5465)	1.7141 (0.3394)	1.9832 (0.2102)
	100	0.6636 (0.0458)	0.6868 (0.0447)	0.6802 (0.0480)	2.2126 (0.3200)	2.0706 (0.3162)	2.1678 (0.1780)
	256	0.6881 (0.0200)	0.6983 (0.0173)	0.6950 (0.0200)	2.0769 (0.1378)	2.0067 (0.1249)	2.0462 (0.0624)
10	30	0.6043 (0.0480)	0.6874 (0.0436)	0.6714 (0.0539)	2.6362 (0.3309)	2.0817 (0.2995)	2.1568 (0.2126)
	100	0.6741 (0.0224)	0.6973 (0.0224)	0.6919 (0.0245)	2.1723 (0.1606)	2.0170 (0.1578)	2.0471 (0.1025)
	256	0.6897 (0.0141)	0.6986 (0.0100)	0.6964 (0.0141)	2.0702 (0.0938)	2.0108 (0.0927)	2.0226 (0.0663)

Tabela 1: Médias amostrais e desvios padrões amostrais (entre parêntesis) para 500 repetições de réplicas de $X_{k,t} = 0.7 * X_{k,t-1} + e_{k,t}$, $e_{k,t} \sim \mathcal{P}o(2)$.

A título de exemplo, apresentam-se na Tabela 1 os valores das estimativas médias e do desvio padrão amostral (entre parêntesis) obtidos pelos três métodos de estimação para 500 repetições de 1, 5 e 10 réplicas independentes com 30, 100 e 256 observações, respectivamente, do modelo INAR(1) de Poisson, $X_{k,t} = 0.7 * X_{k,t-1} + e_{k,t}$, $e_{k,t} \sim \mathcal{P}o(2)$. Os valores para $k = 20$ réplicas não são mostrados, assim como os valores para $n = 512$ observações, porque são análogos aos apresentados.

Nas simulações de réplicas independentes de processos INAR(2) de Poisson podem ser retiradas as mesmas conclusões referentes ao comportamento assintótico quando se fixa o número de réplicas e se aumenta o número de observações ou *vice-versa*. Observa-se ainda que a dispersão apresentada pelos três métodos de estimação é a mesma para os diferentes parâmetros estimados seja qual for o valor verdadeiro dos mesmos e nota-se, também, uma melhoria nos resultados para réplicas com mais de 100 observações. É também notória a

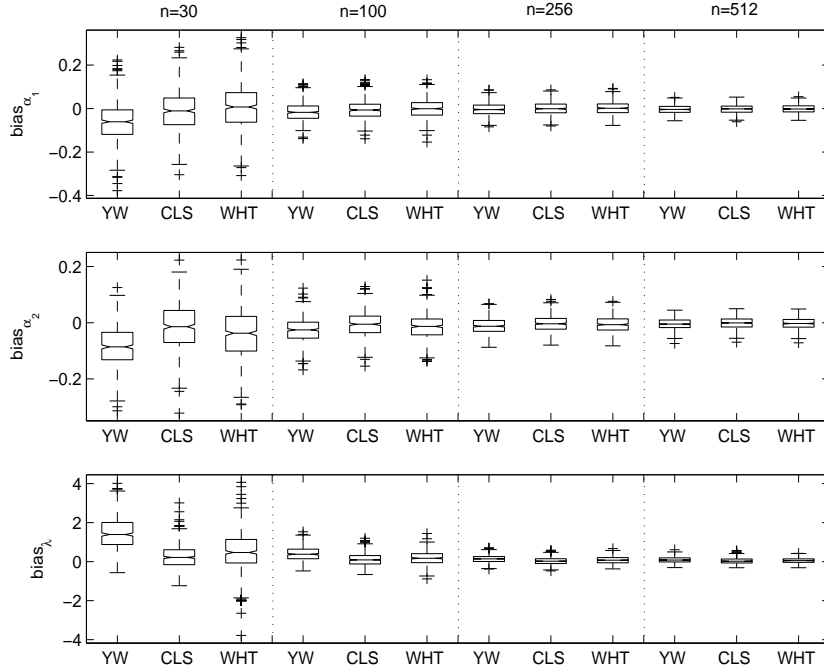


Figura 1: *Boxplots* do enviesamento amostral das estimativas dos parâmetros de um conjunto de 5 réplicas de $X_{k,t} = 0.6 * X_{k,t-1} + 0.2 * X_{k,t-2} + e_{k,t}$, $e_{k,t} \sim \mathcal{Po}(2)$, $k = 1, \dots, 5$ e $t = 1, \dots, n$, para $n = 30, 100, 256, 512$.

sobre-estimação, em média, das estimativas de λ nos três métodos. Para ilustrar estes resultados, apresentam-se na Figura 1, 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 o estimador do parâmetro θ_i na j -ésima repetição, para um conjunto de 5 réplicas de um processo INAR(2) de Poisson com vector de parâmetros $\theta = (0.6, 0.2, 2.0)$, e o número de observações por réplicas aumenta ($n=30, 100, 256, 512$).

Por questões de espaço, e porque nas simulações com réplicas de modelos INAR(3) de Poisson as conclusões são as mesmas que nas ordens inferiores, não são apresentados os resultados referentes às réplicas independentes de processos INAR(3).

5 Aplicação

Neste exemplo de aplicação, são utilizados dados referentes a um estudo epidemiológico que visa estudar os efeitos do consumo de certas substâncias (álcool,

drogas, tabaco, entre outras) durante a gravidez. Uma descrição detalhada do estudo pode ser encontrado em Day *et al.* (1985) e Scher *et al.* (1988).

Foram recolhidos dados do estado do sono neonatal num estudo do efeito do consumo moderado de álcool, durante a gravidez, nos padrões de sono electroencefalográfico (EEG). Como explicado por Stoffer *et al.* (1988) e Stoffer (1990, 1991), o registo do EEG durante o sono de um recém-nascido (não prematuro) é recolhido durante 2 horas, aproximadamente, 24 a 36 horas depois do nascimento. Durante o registo, também são recolhidas observações comportamentais, em intervalos de 1 minuto, por um técnico. Uma dessas observações é o número total de movimentos corporais do recém-nascido (ignorando a sucção).

Neste trabalho consideram-se o número total de movimentos corporais, por minuto, de 8 recém-nascidos. Este conjunto de dados é tratado como sendo 8 réplicas do mesmo modelo INAR(1), com 128 observações por réplica. Os parâmetros do processo INAR(1) de Poisson, (α, λ) , são estimados através dos métodos apresentados anteriormente e apresentam-se na Tabela 2.

Método	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\mu}_x$
YW	0.129	0.260	0.299
CLS	0.144	0.255	0.298
WHT	0.133	0.280	0.323

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

Para um processo INAR(1), X_t , Al-Osh e Alzaid (1987) e Silva e Oliveira (2000a) afirmam que se $X_1 \sim \mathcal{P}o(\lambda/(1-\alpha))$ e $e_t \sim \mathcal{P}o(\lambda)$, então $X_t \sim \mathcal{P}o(\lambda/(1-\alpha))$. Portanto, para verificar o ajuste do modelo INAR(1) às réplicas, aplicou-se um teste χ^2 para comparar as frequências do número de movimentos corporais com a distribuição de Poisson com parâmetro dado por $\hat{\mu}_x = \hat{\lambda}/(1-\hat{\alpha}) \approx 0.3$ (terceira coluna da Tabela 2). Note-se que a média amostral do conjunto das 8 réplicas é 0.299 e a variância amostral é 0.325. Os resultados dos testes χ^2 e a análise dos correlogramas dos resíduos permitem concluir que a classe das réplicas de modelos INAR(1) é razoável na descrição destes dados reais.

6 Comentários Finais

Em muitas situações, existem várias séries temporais correspondentes às medições da mesma resposta, ao longo do tempo, em vários indivíduos. Estas séries podem ser estudadas uma a uma ou, então, podem ser consideradas réplicas de séries temporais. Com esta última aproximação é possível combinar a informação contida em cada uma das réplicas de modo a obter estimadores mais precisos. Neste trabalho são estudadas as réplicas independentes de séries

temporais de contagem modeladas através de processos autoregressivos de valor inteiro, INAR. São apresentados três métodos de estimação para os parâmetros do processo: Yule-Walker, Mínimos Quadrados Condicionais e Critério de Whittle. Os resultados do estudo de simulação indicam que as estimativas apresentam uma melhoria, por exemplo ao nível da diminuição do erro quadrático médio e da variância, para qualquer número de réplicas com, pelo menos, 100 observações. Nota-se, também, que as estimativas de Yule-Walker apresentam, em geral, erros superiores. Porém, estas estimativas servem como valores iniciais de outros métodos de estimação iterativos existentes. Em geral, o comportamento das estimativas obtidas por Mínimos Quadrados Condicionais e pelo Critério de Whittle é semelhante, quer a nível de valores médios quer a nível do erro quadrático médio. Os resultados do exemplo de aplicação sugerem que a análise das séries do número de movimentos corporais como sendo réplicas independentes do mesmo processo INAR(1) é adequada.

Agradecimentos

As autoras estão gratas ao Professor D. S. Stoffer por ter fornecido os dados utilizados na secção § 5.

Referências

- [1] Al-Osh, M.A. e Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, Vol. 8, pp. 261-75.
- [2] Crowder, M.J. e Hand, D.J. (1990). *Analysis of Repeated Measures*. Chapman & Hall.
- [3] Day, N.L., Wagener, D.K. e Taylor, P.M. (1985). *Measurement of substance use during pregnancy: methodologic issues*. Em *Consequences of Maternal Drug Abuse* (Ed. T.M. Pinkert), NIDA Research Monograph 59, pp. 36-47.
- [4] Diggle, P.J. e al-Wasel, I. (1993). *On periodogram-based spectral estimation for replicated time series*. Em *Developments in Time Series Analysis* (Ed. T. Subba Rao), pp. 341-354. Chapman & Hall, London.
- [5] Du, Jin-Guan e Li, Yuan (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, Vol. 12, pp. 129-142.
- [6] Dzhaparidze, K.O. e Yaglom, A.M. (1983). *Spectrum parameter estimation in time series analysis*. Em *Developments in Statistics* (Ed. P.R. Krishnaiah), Vol. 4, Chap. 1, pp. 1-96. Academic Press, New York.
- [7] Franke, J. e Subba Rao, T. (1995). Multivariate first order integer valued autoregressions. *Technical Report*, Math. Dep., UMIST.
- [8] 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, pp. 49-71.
- [9] Klimko, L. e Nelson, P. (1978). On conditional least square estimation for stochastic processes. *Annals of Statistic*, Vol. 6, pp. 629-642.

- [10] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, Vol. 19, pp. 439-455.
- [11] Murthy, D.N.P. (1973). Method of maximum likelihood for stationary time series models. *IEEE Transactions on Automatic Control*, Vol. 18, pp. 397-398.
- [12] Rosenblatt, M. (1971). *Markov process - Structure and asymptotic behavior*. Springer.
- [13] Sesay, S.A.O. e Subba Rao, T. (1992). Frequency-domain estimation of bilinear time series models. *Journal of Time Series Analysis*, Vol. 13, pp. 521-545.
- [14] Shaman, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Annals of Statistics*, Vol. 3, pp. 532-538.
- [15] Scher, M.S., Richardson, G.A., Coble, P.A., Day, N.L. e Stoffer, D.S. (1988). The effects of prenatal alcohol and marijuana exposure: disturbance on neonatal sleep cycling and arousal. *Pediatric Research*, Vol. 24, pp. 101-105.
- [16] Silva, M.E. e Oliveira, V.L. (2000a). Difference equations for the higher order moments and cumulants of the INAR(1) model. *Journal Time of Series Analysis*, aceite para publicação.
- [17] Silva, M.E. e Oliveira, V.L. (2000b). Difference equations for the higher order moments and cumulants of the INAR(p) model. Submetido a *Journal of Time Series Analysis*.
- [18] Silva, N., Pereira, I. e Silva, M.E. (2003). Análise bayesiana de repetições do processo INAR(1). Em *Literacia e Estatística* (Ed. Paula Brito, Adelaide Figueiredo, Fernanda Sousa, Paulo Teles e Fernando Rosado), Actas do X Congresso Anual da Sociedade portuguesa de Estatística, pp.639-650. Edições SPE.
- [19] Silva, I. e Silva, M.E. (2003). Whittle Estimation for INAR processes. *Technical Report DMA/01/03*, Departamento de Matemática Aplicada, Faculdade de Ciências, Universidade do Porto.
- [20] Steutel, F. W. e Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability*, Vol. 7, pp. 893-899.
- [21] Stoffer, D.S., Scher, M.S., Richardson, G.A., Day, N.L. e Coble, P.A. (1988). A Walsh-Fourier Analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *Journal of the American Statistical Association*, Vol. 83, pp. 954-963.
- [22] Stoffer, D.S. (1990). Multivariate Walsh-Fourier Analysis. *Journal of the Time Series Analysis*, Vol. 11, pp. 57-73.
- [23] Stoffer, D.S. (1991). Walsh- Fourier Analysis and its Statistical Applications, *Journal of the American Statistical Association*, Vol. 86, pp. 461-479.
- [24] Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för Matematik*, Vol. 2, pp. 423-434.
- [25] Whittle, P. (1954). *Some recent contributions to the theory of stationary processes*. Em *A Study in the Analysis of Stationary Time Series* (ed. H. Wold), 2nd ed., Appendix 2, pp. 196-228. Almqvist and Wiksell, Stockholm.