



Universidade Federal de Pernambuco

Departamento de Estatística

Bacharelado em Estatística

**ABORDAGEM BAYESIANA DOS PROCESSOS
ZERO-INFLADOS INTEIROS AUTORREGRESSIVOS
DE ORDEM p (ZINAR(p))**

Discente: Tharso Augustus Rossiter Araújo Monteiro

Orientador: Prof. Aldo William Medina Garay

Coorientadora : Prof^a Francielle de Lima Medina

Recife, Maio / 2022

Tharso Augustus Rossiter Araújo Monteiro

**ABORDAGEM BAYESIANA DOS PROCESSOS ZERO-INFLADOS
INTEIROS AUTORREGRESSIVOS DE ORDEM p (ZINAR(p))**

Trabalho de conclusão de curso apresentado
ao Departamento de Estatística, Centro
de Ciências Exatas e da Natureza, como
requisito parcial para obtenção do título de
Bacharel em Estatística pela Universidade
Federal de Pernambuco.

Orientador: Prof. Aldo William Medina Garay
Coorientadora : Prof^a Francielle de Lima Medina

Recife

2022

Agradecimentos

Agradeço especialmente:

A Deus, pois sem Ele nada posso fazer;

*Aos meus pais, Ângelo e Raquel Monteiro, por todo
o apoio;*

*Aos professores Aldo Garay e Franciyelle Medina,
assim como todo o corpo docente do DE-UFPE, pelo
excelente curso que formou meu pensamento
científico.*

Tecnologia é ideologia.

Neil Postman

RESUMO

O principal objetivo deste projeto é implementar um pacote computacional no software R para análise exploratória, estimação dos parâmetros e previsão para o modelo ZINAR(p) sob uma perspectiva bayesiana. Este processo se caracteriza por assumir valores inteiros não-negativos, correlacionados e com proporção significativa de zeros.

Ao longo do trabalho, apresentamos o processo ZINAR(p), suas propriedades matemáticas e um método para inferir seus parâmetros por meio de um algoritmo de Monte Carlo via Cadeia de Markov (MCMC). Finalmente, apresentamos o pacote computacional e demonstramos sua aplicação tanto a dados simulados quanto reais.

Palavras-chave: Séries Temporais, INAR, ZINAR, MCMC, Bayesiano.

ABSTRACT

The main objective of this project is to implement a computational package for the R software for exploratory analysis, parameter estimation and prediction for ZINAR(p) models under a Bayesian perspective. This process is characterized by assuming integer, non-negative and correlated values with a significant proportion of zeros.

Throughout the project, we will present the ZINAR(p) process, its mathematical properties and a method for inferring its parameters through a Markov Chain Monte Carlo (MCMC) algorithm. Finally, we present the computational package and demonstrate its application to both simulated and real data.

Keywords: Time Series, INAR, ZINAR, MCMC, Bayesian.

LISTA DE FIGURAS

Figura 1	<i>Output</i> da função explore_zinarp da amostra ZINAR(2) gerada .	22
Figura 2	<i>Output</i> da função explore_zinarp da amostra INAR(1) gerada ...	23
Figura 3	<i>Output</i> da função explore_zinarp para dados da contagem semanal de furtos na UFPE	31
Figura 4	Distribuição dos valores preditos.....	34

LISTA DE TABELAS

Tabela 1	Resultados obtidos da função estimate_zinarp	26
Tabela 2	Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.2$	28
Tabela 3	Coberturas dos intervalos de credibilidade de 95% das amostras da distribuição <i>a posteriori</i> para cada conjunto de parâmetros fixando-se $\rho = 0.2$	29
Tabela 4	Conjunto de dados de furto. Medidas descritivas das cadeias MCMC para os modelos propostos	32
Tabela 5	Aproximação das funções de verossimilhança sob os modelos considerados	33
Tabela 6	Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.4$	36
Tabela 7	Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.6$	37

LISTA DE SIGLAS

UFPE	Universidade Federal de Pernambuco
i.i.d.	Independente e identicamente distribuído
ZIP	<i>Zero-inflated Poisson</i>
INAR(p)	<i>p-Order Integer-valued Autoregressive</i>
ZINAR(p)	<i>p-Order Zero-inflated Integer-valued Autoregressive</i>
MCMC	<i>Markov Chain Monte Carlo</i>
fdp	Função de Densidade de Probabilidade
ACF	<i>Autocorrelation Function</i>
PACF	<i>Partial Autocorrelation Function</i>
CRAN	<i>Comprehensive Archive R Network</i>
SSI	Superintendência de Segurança Institucional

SUMÁRIO

1	INTRODUÇÃO	10
2	ASPECTOS TEÓRICOS	12
2.1	O operador <i>thinning binomial</i>	12
2.2	O processo INAR(p)	12
2.3	A distribuição Poisson Zero-inflada	13
2.4	O processo ZINAR(p)	14
2.5	Inferência bayesiana para o processo ZINAR(p)	16
2.5.1	Inferência bayesiana	16
2.5.2	Distribuições <i>a priori</i>	17
2.5.3	O algoritmo MCMC	17
2.5.4	Previsão	19
3	O PACOTE ZINARP	20
3.0.1	Geração de amostras	20
3.0.2	Análise exploratória de dados	21
3.0.3	Estimação dos parâmetros	24
4	ESTUDO DE SIMULAÇÃO	27
5	APLICAÇÃO	30
6	CONCLUSÃO E TRABALHOS FUTUROS	35
7	APÊNDICE	36

1 INTRODUÇÃO

As séries temporais de valores inteiros não-negativos surgem em diversas situações, como na contagem de pacientes em um hospital, de clientes em uma loja, de mensagens enviadas em um determinado intervalo de tempo, entre outras.

Buscando um modelo apropriado para analisar este tipo de série, Al-Osh e Alzaid (1987) [1] e McKenzie (1988) [2] apresentaram o processo INAR, em que se supõe que os elementos observados em um determinado tempo podem ser tanto os sobreviventes do processo em um tempo anterior quanto o resultado de um processo de inovação, que inicialmente se supunha seguir distribuição Poisson, no entanto, esta suposição nem sempre é razoável, pois a característica de igualdade a entre a média e a variância, inerente a esta distribuição, nem sempre é encontrada nos dados reais.

Modificações aos pressupostos iniciais do processo INAR surgiram em diversos artigos : Al-Osh e Alzaid (1988) [3] e Du e Li (1991) [4] propuseram modelos com distribuição marginal geométrica. Processos com distribuição marginal Binomial Negativa foram apresentados por Freeland e McCabe (2005) [5] e Ristić et al. (2009) [6].

Outros modelos foram sugeridos para os cenários em que os dados apresentam uma alta proporção de zeros. Barreto (2015) [7] expandiu o modelo apresentado por Ristić et al. (2009) [6], propondo o processo ZMGINAR, que possui distribuição marginal geométrica zero-modificada, enquanto Jazi et al. (2012) [8] apresentaram o modelo ZINAR(1), em que se supõe que os elementos do processo possuem dependência temporal de ordem 1 e que suas inovações seguem distribuição Poisson Zero-inflada (ZIP). Garay et al. (2020) [9] estenderam este modelo para incluir dependências temporais de ordem p . Ademais, propuseram uma técnica para estimação, sob uma perspectiva bayesiana, dos parâmetros deste

modelo por meio de um algoritmo MCMC (Gamerman et al.(2006) [10]), de maneira similar àquele desenvolvido por Neal e Subba Rao (2007) [11], aplicando a inferência bayesiana de forma inédita neste contexto.

O objetivo principal desse trabalho é a elaboração e documentação, na linguagem R, de um pacote chamado **ZINARp**, que será disponibilizado de maneira gratuita para toda a comunidade. Seu propósito é a realização de análise exploratória e simulação dos processos $\text{ZINAR}(p)$, assim como a estimação, sob perspectiva bayesiana, dos parâmetros deste modelo. Conforme nosso conhecimento, é a primeira vez que um pacote computacional é elaborado para estes fins.

2 ASPECTOS TEÓRICOS

A seguir, apresentamos algumas definições e conceitos que serão utilizados no desenvolvimento do trabalho.

2.1 O operador *thinning binomial*

Seja X uma variável aleatória de valor inteiro não-negativo e $\alpha \in [0, 1]$. Como apresentado por Steutel e Harn (1979) [12], o operador “ \circ ”, denominado *thinning binomial*, é definido como a soma de variáveis aleatórias:

$$\alpha \circ Y = \sum_{i=1}^Y V_i, \quad (2.1)$$

em que $\{V_i\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), independentes de Y , que seguem uma distribuição Bernoulli com parâmetro α .

Assim, se $y > 0$:

$$\alpha \circ Y | Y = y \sim \text{Bin}(y; \alpha),$$

em que $\text{Bin}(n, p)$ denota a distribuição binomial com parâmetros n e p . Se $y = 0$, $\alpha \circ Y | Y = y$ segue distribuição degenerada em zero.

2.2 O processo INAR(p)

O processo INAR(p) é definido por:

$$Y_t = \sum_{i=1}^p \alpha_i \circ Y_{t-i} + Z_t, \quad t \in \mathbb{Z},$$

em que $\alpha_i \in [0, 1]$, $i = 1, \dots, p$ e $\{Z_t\}$ é uma sequência i.i.d de variáveis de valores inteiros não-negativos, definida como inovações. Inicialmente, supõe-se que Z_t segue distribuição Poisson.

Como sugerido por Jazi et al. (2012) [8] e Garay et al. (2020) [9], assume-se que :

1. Para todo t , Z_t é independente das variáveis aleatórias V_i na equação (2.1).
2. Z_t tem média μ e variância σ^2 finita, para todo t .
3. V_i e V_j são mutuamente independentes para todo $i \neq j$.

2.3 A distribuição Poisson Zero-inflada

As distribuições zero-infladas são geradas por meio de um processo em duas etapas : na primeira etapa, define-se se a variável aleatória assumirá, com probabilidade ρ , o valor zero, chamado "zero estrutural". Caso não assuma, uma segunda aleatorização é feita na segunda etapa, na qual a variável pode assumir valor zero, chamado "zero de excesso" ou um valor inteiro positivo.

Como apresentado em Cohen (1963) [13], Z_t é uma variável aleatória discreta que segue distribuição Poisson zero-inflada, denotada $Z_t \sim ZIP(\rho, \lambda)$, com distribuição de probabilidade dada por:

$$P(Z_t = k) = \rho \mathbb{I}_{\{0\}}(k) + (1 - \rho) \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots,$$

em que $\lambda > 0$, $0 \leq \rho \leq 1$ e $\mathbb{I}_B(\cdot)$ representa a função indicadora no conjunto B .

O valor esperado e a variância de uma variável Z_t com distribuição ZIP são dados, respectivamente, por:

$$E[Z_t] = (1 - \rho)\lambda \text{ e } Var[Z_t] = \lambda(1 - \rho)(1 + \rho\lambda).$$

Com o intuito de propor um modelo adequado para o ajuste em casos nos quais os dados de contagem apresentam excesso de zeros, Garay et al. (2020) [9] propuseram uma extensão do modelo apresentado por Jazi et al. (2012) [8], considerando um processo inteiro autorregressivo de ordem p , com inovações que seguem distribuição ZIP, denominado processo ZINAR(p).

2.4 O processo ZINAR(p)

Garay et al. (2020) [9] descrevem o processo ZINAR(p) como uma série temporal $\{Y_t\}$ de valores inteiros não-negativos que satisfaz:

$$Y_t = \sum_{i=1}^p \alpha_i \circ Y_{t-i} + Z_t$$

e

$$\text{Cov}(Y_s, Z_t) = 0, \text{ para } s < t ,$$

em que “ \circ ” é o operador *thinning binomial*, definido na Eq. (2.1) e $Z_t \sim \text{ZIP}(\rho, \lambda)$ para todo $t \in \mathbb{Z}$.

Derkhan et al. (2012) [14] verificaram que $\sum_{i=1}^p \alpha_i < 1$ é condição suficiente para estacionariedade e ergodicidade do processo INAR(p). Conforme Garay et al. (2020) [9], como os processos INAR(p) e ZINAR(p) possuem a mesma estrutura estocástica, esta condição também é suficiente para garantir a estacionariedade do processo ZINAR(p).

A média μ e a função de covariância $\gamma(\cdot)$ de um processo ZINAR(p) são dadas, respectivamente, por :

$$\mu = \frac{\lambda(1-\rho)}{1 - \sum_{i=1}^p \alpha_i} \text{ e } \gamma(h) = \sum_{i=1}^p \alpha_i \gamma(h-i) + \mathbb{I}_{\{0\}}(h) \lambda(1-\rho)(1+\rho\lambda) .$$

Sejam Y_{p+1}, \dots, Y_{p+n} , $n > p$, realizações de um processo ZINAR(p), $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \rho, \lambda)^T$, em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T$. A função de densidade de probabilidade conjunta de (Y_1, \dots, Y_{p+n}) dado $\boldsymbol{\theta}$ é definida por:

$$\pi(\mathbf{y} \mid \boldsymbol{\theta}) = \pi(y_1, \dots, y_p \mid \boldsymbol{\theta}) \prod_{t=p+1}^{p+n} \pi(y_t \mid Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p}, \boldsymbol{\theta}), \quad (2.2)$$

em que

$$\begin{aligned} & \pi(y_t \mid Y_{t-1} = y_{t-1}, \dots, Y_{t-p} = y_{t-p}, \boldsymbol{\theta}) \\ &= \sum_{k_1=0}^{\min\{y_{t-1}, y_t\}} \binom{y_{t-1}}{k_1} \alpha_1^{k_1} (1 - \alpha_1)^{y_{t-1}-k_1} \\ &\times \sum_{k_2=0}^{\min\{y_{t-2}, y_t-k_1\}} \binom{y_{t-2}}{k_2} \alpha_2^{k_2} (1 - \alpha_2)^{y_{t-2}-k_2} \\ &\times \vdots \\ &\times \sum_{k_p=0}^{\min\{y_{t-p}, y_t-k_1 \dots -k_{p-1}\}} \binom{y_{t-p}}{k_p} \alpha_p^{k_p} (1 - \alpha_p)^{y_{t-p}-k_p} \\ &\times \left[\rho \mathbb{I}_{\{0\}}(y_t - k_1 \dots - k_p) + (1 - \rho) \frac{e^{-\lambda} \lambda^{y_t-k_1 \dots -k_p}}{[y_t - k_1 \dots - k_p]!} \right]. \end{aligned}$$

Não é possível estimar os parâmetros do modelo diretamente por meio da maximização da função de verossimilhança dada pela Eq. (2.2), pois para os p primeiros valores, $\pi(y_1, \dots, y_p \mid \boldsymbol{\theta})$ é desconhecido. Assim, é comum utilizar a chamada análise condicional, em que $\pi(y_1, \dots, y_p \mid \boldsymbol{\theta})$ é ignorado. Para mais detalhes, veja Hamilton (2020) [15], Bu et al. (2008) [16], Prado e West (2010) [17] e Jazi et al. (2012) [8].

2.5 Inferência bayesiana para o processo ZINAR(p)

2.5.1 Inferência bayesiana

Na inferência bayesiana, os parâmetros de um modelo são considerados variáveis aleatórias, não valores fixos e desconhecidos, como no ponto de vista frequentista (Mayer (1969) [18]). Este conceito permite a associação de uma distribuição de probabilidade a $\boldsymbol{\theta}$, o conjunto de parâmetros do modelo.

Para calcular a chamada distribuição *a posteriori* $\pi(\boldsymbol{\theta}|\mathbf{Y})$, utiliza-se a função de verossimilhança $\pi(\mathbf{Y}|\boldsymbol{\theta})$ em conjunto com uma distribuição a-priori $\pi(\boldsymbol{\theta})$, que representa uma informação prévia ou um grau de crença em relação à distribuição de probabilidade de $\boldsymbol{\theta}$. Utilizando o teorema de Bayes, tem-se que:

$$\pi(\boldsymbol{\theta} | \mathbf{Y}) = \frac{\pi(\mathbf{Y} | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{h(\mathbf{Y})}, \boldsymbol{\theta} \in \Theta,$$

em que Θ é o espaço paramétrico do vetor de parâmetros $\boldsymbol{\theta}$ e $h(\mathbf{Y})$ é a distribuição marginal de \mathbf{Y} , constante em relação a $\boldsymbol{\theta}$.

O cálculo de $h(\mathbf{Y})$ pode ser bastante complexo. Para evitar esta operação, uma alternativa é utilizar a distribuição *a posteriori* não-normalizada, representada por:

$$\pi(\boldsymbol{\theta} | \mathbf{Y}) \propto \pi(\mathbf{Y} | \boldsymbol{\theta})\pi(\boldsymbol{\theta}) .$$

O desenvolvimento de algoritmos do tipo Monte Carlo via Cadeia de Markov (MCMC) tem sido um grande avanço para a inferência bayesiana. Esses métodos nos permitem obter amostras da distribuição *a posteriori*. As estatísticas desta amostra podem ser utilizadas para estimar $\boldsymbol{\theta}$ com precisão.

2.5.2 Distribuições *a priori*

Considerando a mesma estratégia definida em Garay et al. (2020), [9] supomos independência entre os parâmetros do modelo e consideramos as seguintes distribuições *a priori*:

$$\alpha_i \sim U(0, 1); \text{ para } i = 1, \dots, p, \text{ satisfazendo } \sum_{i=1}^p \alpha_i < 1;$$

$$\rho \sim U(0, 1);$$

$$\lambda \sim \text{Gama}(A_\lambda, B_\lambda),$$

em que $U(a, b)$ denota a distribuição uniforme com parâmetros $-\infty < a < b < \infty$ e $\text{Gama}(c, d)$ denota a distribuição Gama com parâmetros $c > 0$, $d > 0$ e valor esperado c/d .

Deste modo, garante-se, primeiramente, que a distribuição *a priori* seja fracamente informativa, refletindo a ausência de informação advinda de experimentos prévios, e também própria, evitando a uma distribuição *a posteriori* imprópria.

2.5.3 O algoritmo MCMC

Conforme estratégia proposta por Neal e Subba Rao [11], Garay et al. [9] desenvolveram um algoritmo MCMC [10] que considera simultaneamente os dados observados e variáveis latentes. Seu objetivo é a estimação dos parâmetros do modelo por meio de estatísticas de amostras da distribuição *a posteriori*, como a média, desvio padrão, moda, etc.

Assim, seja $\mathbf{y} = \{y_1, y_2, \dots, y_{n+p}\}$ uma realização de um processo ZINAR(p), com $n > p$. Assumindo p conhecido, definem-se as seguintes variáveis latentes:

- $S_{t,i} = \alpha_i \circ Y_{t-i}$,

em que $S_{t,i}|Y_{t-i} = y_{t-i}$ segue distribuição $Bin(y_{t-i}, \alpha_i)$ se $y_{t-1} > 0$ e degenerada em zero se $y_{t-1} = 0$. Além disso, as variáveis aleatórias associadas à distribuição de $S_{t,i}|Y_{t-i} = y_{t-i}$, α_i , $i = 1, \dots, p$, são independentes.

Conforme Ando [19] e Garay et al. [20], também consideramos a variável aleatória latente binária denominada W_t , tal que :

- $W_t|\rho \sim Ber(\rho)$,

em que $Ber(\rho)$ denota a distribuição Bernoulli com parâmetro ρ .

Assim, temos que $Z_t = Y_t - \sum_{i=1}^p S_{t,i}$ segue distribuição ZIP, isto é, $Z_t|(W_t = 0, \lambda) \sim Poisson(\lambda)$ e $Z_t|(W_t = 1, \lambda)$ segue distribuição degenerada em zero.

Pela definição do processo, $Z_t = Y_t - \sum_{i=1}^p S_{t,i}$. Sejam $\mathbf{S}_t = (S_{t,1}, S_{t,2}, \dots, S_{t,p})$ e $\mathbf{y}_{(t-p)} = \{y_{t-1}, y_{t-2}, \dots, y_{t-p}\}$, para $t = p+1, \dots, n$. Considerando as distribuições *a priori* para $\boldsymbol{\theta}$ descritas na seção 2.5.2, define-se $\mathbf{S} = (\mathbf{S}_{p+1}, \dots, \mathbf{S}_{p+n})$, $\mathbf{W} = (\mathbf{W}_{p+1}, \mathbf{W}_{p+2}, \dots, \mathbf{W}_{p+n})$ e $\mathbf{Z} = (Z_{p+1}, Z_{p+2}, \dots, Z_{p+n})$. A distribuição conjunta de \mathbf{S} , \mathbf{Z} , \mathbf{W} e $\boldsymbol{\theta}$ é dada por:

$$\begin{aligned} \pi(\mathbf{z}, \mathbf{s}, \mathbf{w}, \boldsymbol{\theta} \mid \mathbf{y}) &\propto \prod_{t=p+1}^{p+n} \pi(z_t, \mathbf{s}_t, w_t \mid \mathbf{y}_{(t-p)}, y_t, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\ &\propto \prod_{t=p+1}^{p+n} \left\{ \rho^{w_t} \left[\frac{e^{-\lambda} \lambda^{z_t}}{z_t!} (1 - \rho) \right]^{1-w_t} \right. \\ &\quad \left. \times \prod_{i=1}^p \binom{y_{t-i}}{s_{t,i}} \alpha_i^{s_{t,i}} (1 - \alpha_i)^{y_{t-i} - s_{t,i}} \right\} \lambda^{A_\lambda - 1} e^{-B_\lambda \lambda} \end{aligned} \quad (2.3)$$

As distribuições condicionais completas para os parâmetros dos processos podem ser obtidas por meio da Eq. (2.3) :

- $\alpha_i | \mathbf{z}, \mathbf{s}, \mathbf{w}, \mathbf{y}, \boldsymbol{\alpha}_{[-i]}, \rho, \lambda \sim \text{Beta} \left(\sum_{t=p+1}^{p+n} s_{t,i} + 1; \sum_{t=p+1}^{p+n} (y_{t-i} - s_{t,i}) + 1 \right)$,

para $i = 1, \dots, p$, em que $\boldsymbol{\alpha}_{[-i]} = (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_p)^T$ e $\text{Beta}(\cdot; \cdot)$ denota a distribuição Beta.

- $\lambda | \mathbf{z}, \mathbf{s}, \mathbf{w}, \mathbf{y}, \boldsymbol{\alpha}, \rho \sim \text{Gama} \left(\sum_{t=p+1}^{p+n} z_t(1 - w_t) + A_\lambda; \sum_{t=p+1}^{p+n} (1 - w_t) + B_\lambda \right)$

e

- $\rho | \mathbf{z}, \mathbf{s}, \mathbf{w}, \mathbf{y}, \boldsymbol{\alpha}, \lambda \sim \text{Beta} \left(\sum_{t=p+1}^{p+n} w_t + 1; \sum_{t=p+1}^{p+n} (1 - w_t) + 1 \right)$.

Enquanto as distribuições condicionais completas dos parâmetros do modelo possuem formas conhecidas, o mesmo não ocorre para as variáveis latentes Z_t , $S_{t,i}$ e W_t , dificultando sua amostragem. Para contornar este problema, utiliza-se um algoritmo Metropolis-Hastings (Metropolis et al. [21]), que envolve tanto a amostragem dos parâmetros a partir de suas distribuições condicionais completas, como das distribuições de $S_{t,i} | y_{t-i}$ e $W_t | z_t$, levando em consideração a restrição $Z_t = Y_t - \sum_{i=1}^p S_{t,i}$. O algoritmo completo está descrito em Garay et al. [9].

2.5.4 Previsão

Uma estratégia para a previsão de valores futuros de um processo ZINAR(p) é descrita por Garay et al. [9]. Seja $\mathbf{Y}_m^{pred} = (Y_{T+1}, \dots, Y_{T+m})$, em que $T = n + p$. O método consiste na expansão do algoritmo MCMC para geração de amostras da função de verossimilhança marginal $\pi(\mathbf{y}_m^{pred} | \mathbf{y})$.

Para cada iteração do algoritmo MCMC, atualiza-se o valor de \mathbf{Y}_m^{pred} , de y_{T+1} a y_{T+m} , a partir da geração de propostas para as variáveis latentes w_{T+k} , z_{T+k} e $s_{T+k,i}$. Em seguida, define-se $y_{T+k} = \sum_{i=1}^p s_{T+k,i} + z_{T+k}$. O algoritmo completo está descrito em Garay et al. [9].

A seguir, apresentamos o pacote computacional **ZINARp**, detalhamos e ilustramos o uso de seus comandos.

3 O PACOTE ZINARp

O pacote **ZINARp** está disponível por meio do Comprehensive R Archive Network (CRAN), acessível pelo link <https://cran.r-project.org/>. Sua instalação e uso podem ser feitos no software R [22] por meio dos comandos descritos no Listing 3.1.

O pacote **ZINARp** depende dos seguintes pacotes : **progress**, **stats**, **utils**, **graphics**.

Listing 3.1: Instalação do pacote ZINARp

```
install.packages( 'ZINARp' )
library( ZINARp )
```

O pacote possui três funções principais : a geração de amostras, a análise exploratória e a estimação de parâmetros de processos INAR(p)/ZINAR(p).

A seguir, são descritas as funções apresentadas no pacote e seus respectivos argumentos:

3.0.1 Geração de amostras

- Descrição

O comando **simul.zinarp** gera uma realização de tamanho n do processo INAR(p) ou ZINAR(p). Seus argumentos são:

- **n** : O tamanho da cadeia simulada. Deve ser um valor inteiro positivo.
- **alpha** : O vetor $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)^T$, em que $\alpha_i \in [0, 1]$, $\forall i = 1, \dots, p$, com $\sum_{i=1}^p \alpha < 1$.
- **lambda** : Taxa λ , que representa o parâmetro da distribuição Poisson. $\lambda > 0$.

- **pii** : A probabilidade ρ de ocorrência do zero estrutural (estado perfeito) na distribuição Poisson zero-inflada. Caso não especificada, assume valor 0, seguindo uma distribuição Poisson, isto é, a cadeia seguirá uma distribuição $\text{INAR}(p)$.
- Exemplo

Um uso da função **simul_zinarp** está descrito no Listing 3.2. Utilizamos a função **set.seed** para garantir a reprodutibilidade do exemplo.

Foram simuladas duas realizações de tamanho 100 : (a) uma de um processo $\text{ZINAR}(2)$ com $\alpha = (0.25, 0.15)^T$, $\lambda = 2$ e $\rho = 0.4$ e outra (b) de um processo $\text{INAR}(1)$ com $\alpha = 0.1$ e $\lambda = 2$.

Listing 3.2: (a) Simulação de um processo $\text{ZINAR}(2)$ (b) Simulação de um processo $\text{INAR}(1)$

```

set.seed(1234)
# (a)
sim_zinar2 <- simul_zinarp(n = 100, alpha = c(0.25,0.15),
                           lambda = 2, pii = 0.4)

#(b)
sim_inar1 <- simul_zinarp(n = 100, alpha = 0.1,
                           lambda = 2, pii = 0)

```

3.0.2 Análise exploratória de dados

- Descrição

O comando **explore_zinarp** é utilizado para a análise exploratória dos processos $\text{INAR}(p)$ e $\text{ZINAR}(p)$. Apenas um argumento é requerido:

- **x** : um vetor contendo uma série temporal de valores inteiros não-negativos.

- Exemplos

O *output* da função é um painel com quatro gráficos : o gráfico de série temporal, o gráfico de barras da frequência relativa dos valores da série, além dos gráficos das funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

No Listing 3.3, está descrito o uso da função para análise exploratória das cadeias geradas na seção 3.0.1 . Seus *outputs* estão disponíveis na Figura 1 e na Figura 2.

Listing 3.3: Uso da função `explore_zinarp`

```
explore_zinarp(sim_zinar2)
explore_zinarp(sim_inar1)
```

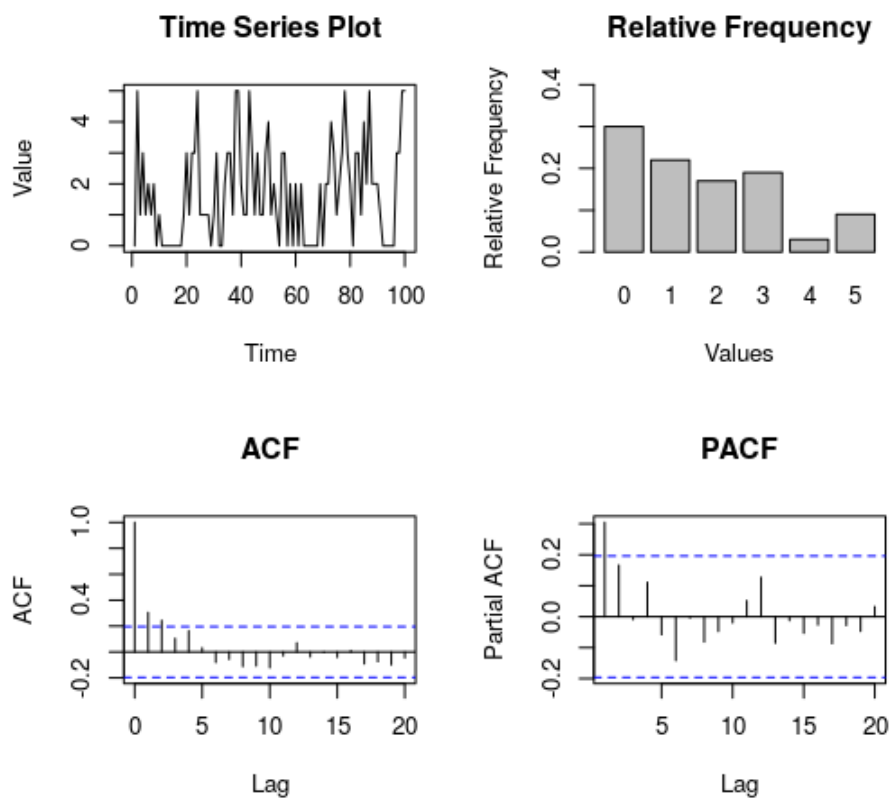


Figura 1: *Output* da função `explore_zinarp` da amostra ZINAR(2) gerada

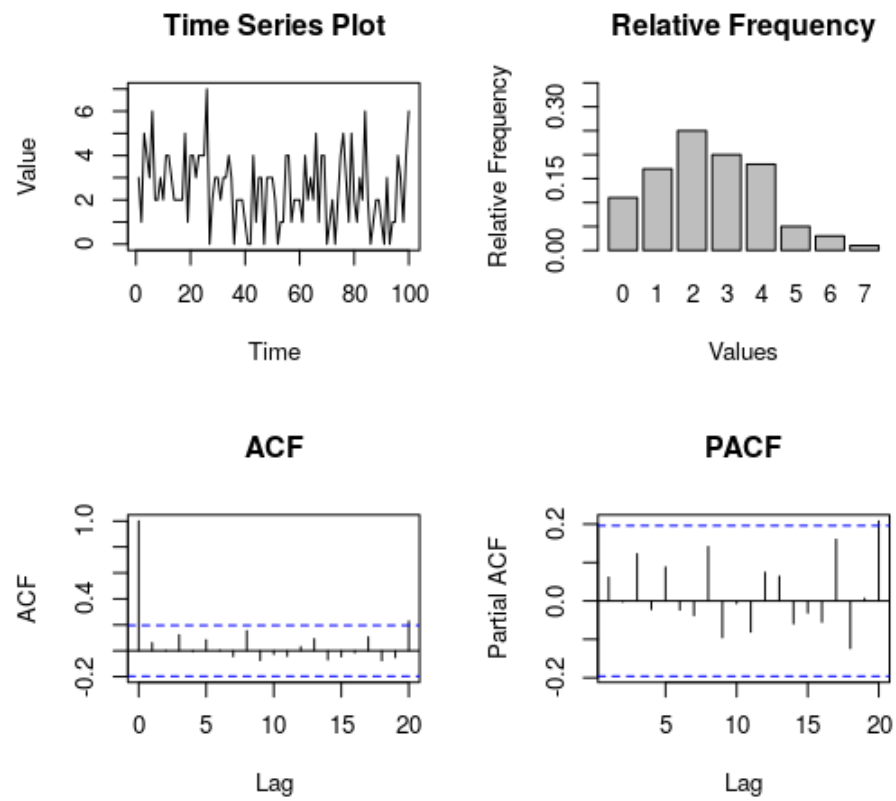


Figura 2: *Output* da função **explore_zinarp** da amostra INAR(1) gerada

3.0.3 Estimação dos parâmetros

O comando **estimate_zinarp** utiliza o algoritmo descrito por Garay et al. [9] para gerar amostras de uma distribuição *a posteriori* para os parâmetros do modelo. Utilizam-se os hiperparâmetros $A_\lambda = B_\lambda = 0.001$.

Os parâmetros da função são:

- **x** : Um vetor contendo uma série temporal de valores discretos não-negativos.
- **p** : A ordem do processo INAR/ZINAR.
- **iter** : Número de iterações do algoritmo MCMC. Caso não especificado, assume valor 5000.
- **thin** : Fator *thinning* para amostras da distribuição *a posteriori*. Caso não especificado, assume valor 2.
- **burn** : Fração de *burn-in* para as amostras geradas. Deve pertencer ao intervalo (0,1). Assume valor 0.1 por padrão.
- **innovation** : A distribuição a ser adotada para a inovação, "Poisson", para processos INAR ou "ZIP", para processos ZINAR. Caso não especificada, assume valor "Poisson".

A função **estimate_zinarp** retorna as cadeias com as amostras das distribuições condicionais completas para os parâmetros $\theta = (\alpha, \lambda, \rho)$ do modelo INAR(p)/ZINAR(p).

No Listing 3.4, apresenta-se um exemplo de estimação dos parâmetros das cadeias geradas pela função **estimate_zinarp**. Foram calculados a média, desvio padrão (SD) e intervalo de credibilidade de 95% (IC). Utilizaram-se 100000 iterações, *burn-in* de 0.1 e fator *thinning* de 2. Os resultados obtidos estão representados na Tabela 1.

Listing 3.4: Uso da função `estimate_zinarp`

```

ezinar2 <- estimate_zinarp(sim_zinar2,
                           p = 2, iter = 100000,
                           thin = 2,
                           burn = 0.1,
                           innovation = 'ZIP')

mean(ezinar2$alpha[,1])
mean(ezinar2$alpha[,2])
mean(ezinar2$lambda)
mean(ezinar2$rho)

sd(ezinar2$alpha[,1])
sd(ezinar2$alpha[,2])
sd(ezinar2$lambda)
sd(ezinar2$rho)

quantile(ezinar2$alpha[,1], 0.025)
quantile(ezinar2$alpha[,1], 0.975)

quantile(ezinar2$alpha[,2], 0.025)
quantile(ezinar2$alpha[,2], 0.975)

quantile(ezinar2$lambda, 0.025)
quantile(ezinar2$lambda, 0.975)

quantile(ezinar2$rho, 0.025)
quantile(ezinar2$rho, 0.975)

einar1 <- estimate_zinarp(sim_inar1, p = 1,
                          iter = 100000, thin = 2,
                          burn = 0.1,

```

```
innovation = 'Poisson')
```

```
mean(einar1$alpha)
```

```
mean(einar1$lambda)
```

```
sd(einar1$alpha)
```

```
sd(einar1$lambda)
```

```
quantile(einar1$alpha, 0.025)
```

```
quantile(einar1$alpha, 0.975)
```

```
quantile(einar1$lambda, 0.025)
```

```
quantile(einar1$lambda, 0.975)
```

Modelo	Parâmetro	Valor Real	Média MCMC	SD MCMC	IC
ZINAR(2)	α_1	0.25	0.2180	0.0914	(0.0443, 0.3951)
	α_2	0.15	0.2466	0.0948	(0.0530, 0.4237)
	λ	2.00	1.7596	0.3317	(1.1467, 2.4381)
	ρ	0.40	0.4663	0.1028	(0.2522, 0.6529)
<hr/>					
INAR(1)	α_1	0.10	0.1137	0.0720	(0.0065, 0.2712)
	λ	2.00	2.2137	0.2277	(1.7562, 2.644)

Tabela 1: Resultados obtidos da função **estimate_zinarp**

4 ESTUDO DE SIMULAÇÃO

Realizamos um estudo de simulação com o intuito de verificar a eficácia do algoritmo proposto para estimação bayesiana dos parâmetros dos processos ZINAR(p). Todos os procedimentos foram realizados utilizando o software R [22].

Consideramos o modelo ZINAR(2) com parâmetros $\alpha_1 \in \{0.1, 0.4, 0.6\}$, $\alpha_2 \in \{0.05, 0.1, 0.2\}$, $\lambda \in \{1, 2, 3\}$ e $\rho \in \{0.2, 0.4, 0.6\}$. Para cada combinação de parâmetros, foram consideradas 100 réplicas de tamanho $n = 80$.

Em cada cenário, a estimação dos parâmetros foi feita por meio de 5000 iterações do algoritmo MCMC proposto. Foi utilizado *burn-in* de 0.1 e *thinning* de ordem 2.

Fixando-se $\rho = 0.2$, a Tabela 2 exibe a média das médias (M) e do desvio padrão (SD) das amostras da distribuição *a posteriori*. A tabela Tabela 3 exibe a cobertura (C) das amostras da distribuição condicional completa. A cobertura se refere à porcentagem de intervalos de credibilidade de 95% que contêm o valor real do parâmetro.

Observa-se que o algoritmo MCMC proposto produz estimativas precisas em quase todos os casos. Há interação entre os parâmetros, especialmente para valores maiores de α_1 . Resultados para $\rho = 0.4$ e $\rho = 0.6$ estão disponíveis no apêndice e apresentam resultados semelhantes.

α_1	α_2	λ	ρ	$M(\alpha_1)$	$M(\alpha_2)$	$M(\lambda)$	$M(\rho)$	$SD(\alpha_1)$	$SD(\alpha_2)$	$SD(\lambda)$	$SD(\rho)$
0.1	0.05	1		0.1587	0.1297	0.8804	0.2345	0.0785	0.0711	0.2021	0.1254
		2		0.1541	0.1249	1.7651	0.2333	0.0750	0.0683	0.2678	0.1039
		3		0.1490	0.1233	2.6286	0.2327	0.0673	0.0603	0.2948	0.0804
	0.1	1		0.1618	0.1701	0.9301	0.2477	0.0801	0.0809	0.217	0.1309
		2		0.1556	0.1616	1.7947	0.2467	0.0735	0.0777	0.2942	0.105
		3		0.1446	0.1453	2.8267	0.2335	0.0729	0.0732	0.3354	0.0899
	0.2	1		0.1474	0.2144	0.9777	0.2468	0.0807	0.0914	0.2332	0.1362
		2		0.1479	0.2106	1.8939	0.2296	0.0754	0.0867	0.3037	0.1111
		3		0.1559	0.2031	2.8550	0.2291	0.0753	0.0845	0.3766	0.0972
	0.05	1		0.3660	0.1258	1.0067	0.2493	0.0952	0.0761	0.2411	0.1389
		2		0.3874	0.1423	1.7800	0.2385	0.0873	0.0706	0.3146	0.1184
		3		0.3679	0.1112	2.9227	0.2270	0.0887	0.0711	0.4320	0.1060
0.4	0.1	1		0.3791	0.1600	1.0342	0.2629	0.0963	0.0839	0.2609	0.1449
		2	0.2	0.3732	0.1686	1.9191	0.2353	0.0912	0.0823	0.3691	0.124
		3		0.3690	0.1562	2.9333	0.2353	0.0922	0.0820	0.4527	0.1159
	0.2	1		0.3936	0.2258	1.0319	0.2425	0.0954	0.0936	0.2718	0.1495
		2		0.3860	0.2179	1.9862	0.2316	0.0944	0.0904	0.3931	0.1381
		3		0.3627	0.2255	2.9891	0.2330	0.1024	0.0944	0.5710	0.1305
	0.05	1		0.5563	0.1397	0.9494	0.2566	0.0823	0.0795	0.2629	0.1565
		2		0.5602	0.1361	1.9033	0.2522	0.0825	0.0779	0.4034	0.1449
		3		0.4897	0.1172	2.7176	0.2322	0.1092	0.0760	0.6291	0.1287
	0.1	1		0.5614	0.1619	1.0714	0.2724	0.0816	0.0862	0.3098	0.1626
		2		0.5415	0.1691	1.9675	0.2525	0.0928	0.0888	0.4568	0.1562
		3		0.3839	0.1102	2.1487	0.1759	0.1123	0.0690	0.6632	0.1163
0.6	0.2	1		0.5632	0.2379	1.0847	0.2811	0.0851	0.0950	0.3643	0.1811
		2		0.3951	0.1724	1.5994	0.2006	0.1051	0.0832	0.5722	0.1403
		3		0.0848	0.0428	0.8073	0.0493	0.0356	0.0246	0.2805	0.0418

Tabela 2: Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.2$

α_1	α_2	λ	ρ	$C(\alpha_1)$	$C(\alpha_2)$	$C(\lambda)$	$C(\rho)$
0.1	0.05	1	0.2	88%	88%	90%	94%
		2	0.2	92%	91%	90%	98%
		3	0.2	89%	89%	89%	90%
	0.1	1	0.2	89%	87%	91%	97%
		2	0.2	90%	91%	92%	95%
		3	0.2	93%	94%	89%	100%
	0.2	1	0.2	98%	93%	97%	99%
		2	0.2	95%	95%	95%	97%
		3	0.2	93%	97%	93%	99%
	0.05	1	0.2	96%	96%	95%	99%
		2	0.2	96%	86%	89%	98%
		3	0.2	96%	98%	95%	96%
	0.1	1	0.2	98%	95%	94%	99%
		2	0.2	97%	92%	94%	98%
		3	0.2	93%	96%	94%	96%
	0.2	1	0.2	97%	93%	91%	100%
		2	0.2	98%	96%	93%	100%
		3	0.2	93%	97%	94%	98%
0.4	0.05	1	0.2	92%	92%	93%	100%
		2	0.2	96%	90%	94%	96%
		3	0.2	91%	93%	96%	95%
	0.1	1	0.2	97%	94%	97%	99%
		2	0.2	92%	95%	95%	100%
		3	0.2	74%	80%	78%	79%
	0.2	1	0.2	96%	93%	90%	98%
		2	0.2	73%	77%	79%	80%
		3	0.2	18%	18%	20%	19%

Tabela 3: Coberturas dos intervalos de credibilidade de 95% das amostras da distribuição *a posteriori* para cada conjunto de parâmetros fixando-se $\rho = 0.2$

5 APLICAÇÃO

Para evidenciar casos para os quais o ajuste do modelo proposto é adequado, assim como a aplicação do pacote **ZINARp**, consideramos um banco de dados original de furtos nas dependências da Universidade Federal de Pernambuco (UFPE). Todas as ocorrências são descritas, registradas e organizadas pela Superintendência de Segurança Institucional, SSI/UFPE, um escritório encarregado do planejamento, execução e avaliação dos projetos e atividades relacionadas à segurança institucional da universidade.

Os dados consistem na contagem semanal de furtos nas dependências da UFPE, durante um período de 424 semanas, de janeiro de 2008 a dezembro de 2015. As duas últimas observações foram removidas para propósito de comparação da previsão dos valores. Estes dados estão sendo utilizados pela primeira vez na literatura científica.

Na Figura 3, está ilustrado o *output* da função **explore_zinarp**, do pacote **ZINARp**. Observa-se no primeiro painel o gráfico de série temporal dos dados, seguido do gráfico de barras das frequências relativas e dos gráficos das funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

O primeiro e o segundo gráfico indicam uma presença significativa de zeros, que representam 67.25% da série. Além disso, o gráfico da PACF sugere uma dependência temporal de ordem $p \geq 1$. Consideramos os modelos ZINAR(1), ZINAR(2), INAR(1) e INAR(2).

Utilizando os métodos descritos na seção 3.0.3, estimamos os parâmetros para os modelos propostos. Foram utilizadas 100000 iterações, com fator *thinning* de 2 e 10% de *burn-in*. Os resultados estão descritos na Tabela 4.

Sejam M_1, M_2, M_3, M_4 os modelos ZINAR(1), ZINAR(2), INAR(1) e INAR(2), respectivamente. Para fins de comparação, calculamos o fator de Bayes, dado pela

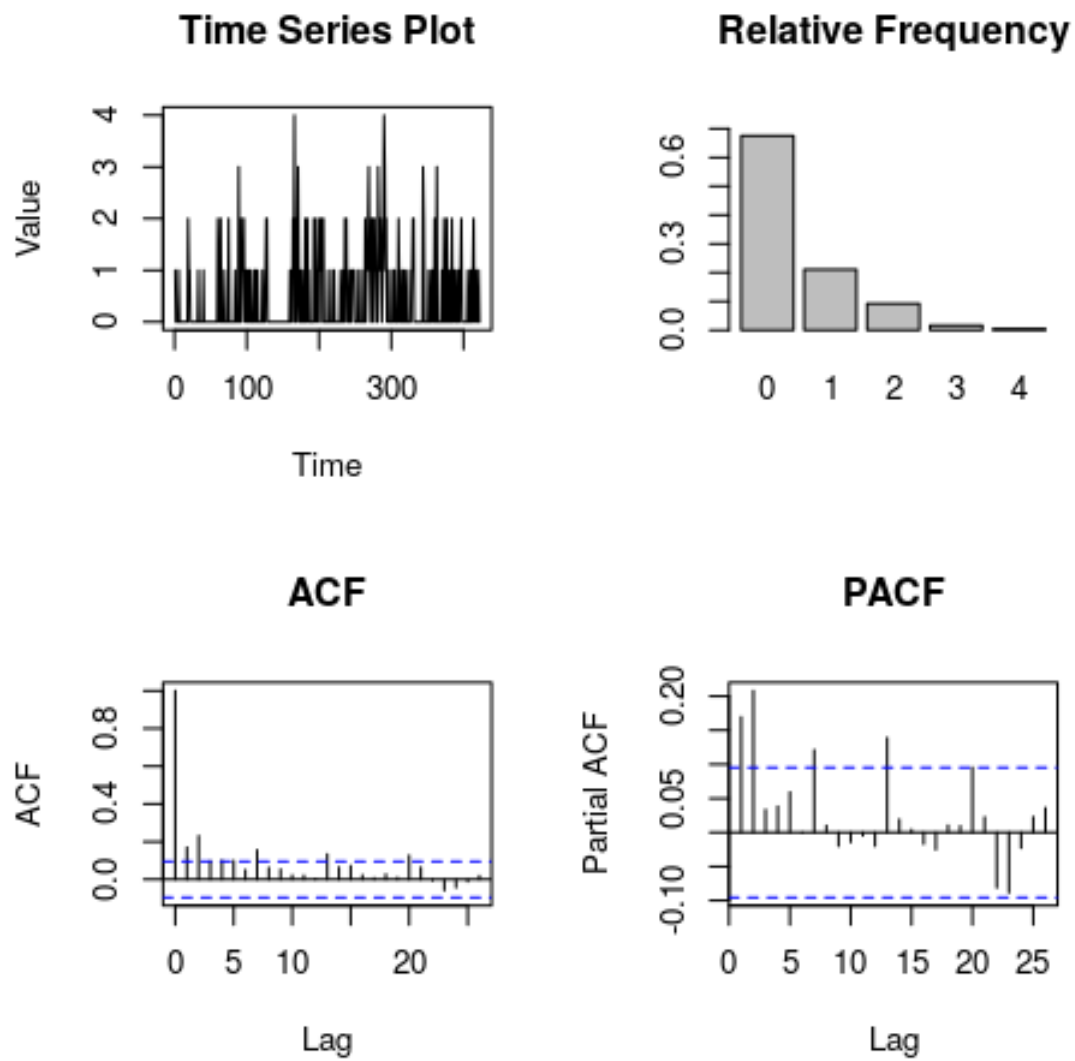


Figura 3: *Output* da função **explore_zinarp** para dados da contagem semanal de furtos na UFPE

Modelo	Parâmetro	Média da cadeia MCMC	Desvio padrão da cadeia MCMC
ZINAR(1)	α_1	0.132	0.045
	λ	0.695	0.105
	ρ	0.407	0.083
ZINAR(2)	α_1	0.131	0.044
	α_2	0.190	0.046
	λ	0.623	0.119
	ρ	0.471	0.097
INAR(1)	α_1	0.137	0.043
	λ	0.401	0.035
INAR(2)	α_1	0.123	0.042
	α_2	0.189	0.045
	λ	0.322	0.035

Tabela 4: Conjunto de dados de furto. Medidas descritivas das cadeias MCMC para os modelos propostos

razão entre as densidades marginais $\pi(\mathbf{y}|M_k)$ sob cada modelo. Utiliza-se a aproximação proposta por Kaas e Raftery [23]:

$$\pi(\mathbf{y}|M_k) \approx \left\{ \frac{1}{L} \sum_{l=1}^L \frac{1}{\pi(\mathbf{y}|\boldsymbol{\theta}_k^{(l)}, M_k)} \right\}^{-1}, \quad (5.1)$$

em que $\pi(\mathbf{y}|\boldsymbol{\theta}_k^{(l)}, M_k)$ representa a função de verossimilhança do vetor \mathbf{y} condicionado a $\boldsymbol{\theta}_k^{(l)}$, o l -ésimo valor amostrado da distribuição *a posteriori* sob o k -ésimo modelo.

Como discutido na seção 2.4, não é possível lidar diretamente com a função de verossimilhança de um processo ZINAR(p). Para resolver este problema, considera-se que os p primeiros valores do processo são independentes e seguem distribuição ZIP.

O valor calculado da aproximação das densidades marginais para os modelos considerados estão descritos na Tabela 5.

Modelo	$\pi(\mathbf{y} M_k)$
ZINAR(1)	3.119×10^{-165}
ZINAR(2)	5.986×10^{-293}
INAR(1)	4.107×10^{-171}
INAR(2)	7.767×10^{-304}

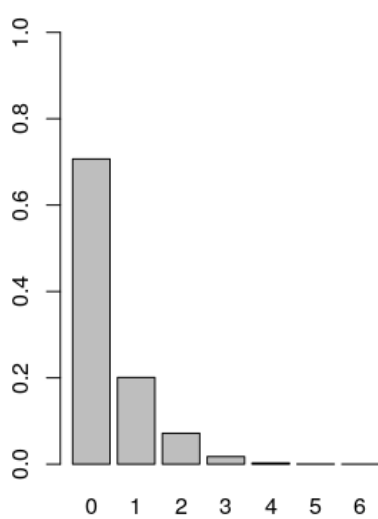
Tabela 5: Aproximação das funções de verossimilhança sob os modelos considerados

O fator de Bayes do modelo ZINAR(1), em comparação aos modelos ZINAR(2), INAR(1) e INAR(2) é, respectivamente : 5.21×10^{127} , 7.59×10^5 , 4.01×10^{138} . Logo, de acordo com este critério, o modelo mais apropriado para os dados considerados é o ZINAR(1).

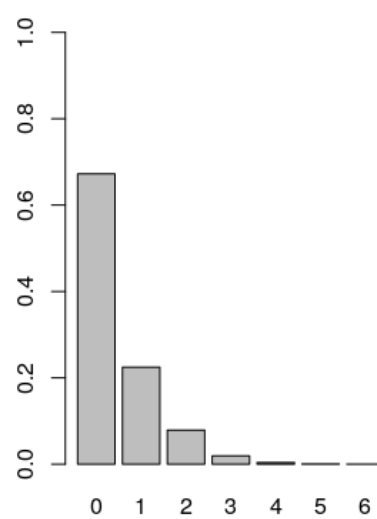
Sob modelo ZINAR(1), a estimativa da proporção de inovações que assumem valor zero ($P(Z_t = 0)$) é dada por:

$$\frac{1}{L} \sum_{l=1}^L \{\hat{\rho}^{(l)} + (1 - \hat{\rho}^{(l)})e^{-\hat{\lambda}^{(l)}}\} = 0.7103$$

Utilizando este modelo, foi realizada uma análise preditiva para os últimos dois valores da série de temporal utilizando o método sugerido por Garay et al. [9]. Foram geradas 100000 previsões para cada semana. As distribuições preditivas dos valores estão representadas, respectivamente, na Figura 4a e Figura 4b. Ambas distribuições preditivas atribuem maior probabilidade ao valor zero, que é valor real da série em abas as semanas.



(a) Distribuição dos valores preditos na semana #423



(b) Distribuição dos valores preditos na semana #424

Figura 4: Distribuição dos valores preditos

6 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho de conclusão de curso, foi apresentado o modelo $\text{ZINAR}(p)$ como uma generalização do modelo $\text{ZINAR}(1)$ para dependências temporais de ordem maior ou igual a 1. As características do modelo, assim como algumas de suas propriedades matemáticas foram expostas. Em seguida, foi exposto um algoritmo MCMC para estimação dos parâmetros do processo sob uma perspectiva bayesiana.

Após a fundamentação teórica, demonstrou-se o uso e aplicação do pacote **ZINARp**, disponível através do software R [22], para análise exploratória, simulação e estimação de processos $\text{ZINAR}(p)$. Sua eficiência foi verificada através de sua aplicação tanto a dados simulados como reais.

Conforme Garay et al. [9], o modelo proposto pode ser expandido por meio da adição de médias móveis, entretanto esta proposta foge ao escopo deste trabalho.

7 APÊNDICE

α_1	α_2	λ	ρ	$M(\alpha_1)$	$M(\alpha_2)$	$M(\lambda)$	$M(\rho)$	$SD(\alpha_1)$	$SD(\alpha_2)$	$SD(\lambda)$	$SD(\rho)$	$C(\alpha_1)$	$C(\alpha_2)$	$C(\lambda)$	$C(\rho)$	
0.1	0.05	1	0.4	0.1463	0.1268	0.8952	0.3603	0.076	0.0705	0.2469	0.1388	95%	88%	89%	92%	
		2	0.4	0.1274	0.1066	1.8097	0.3834	0.0637	0.0583	0.288	0.0952	94%	92%	86%	93%	
		3	0.4	0.1344	0.0923	2.7419	0.4043	0.0584	0.0508	0.3389	0.0796	91%	93%	87%	89%	
	0.1	1	0.4	0.1458	0.1421	0.851	0.3383	0.0742	0.0744	0.2295	0.1359	90%	92%	88%	88%	
		2	0.4	0.1331	0.1305	1.8394	0.3775	0.0649	0.0661	0.2972	0.0988	93%	93%	89%	93%	
		3	0.4	0.1196	0.1206	2.7977	0.3941	0.0576	0.058	0.3375	0.0814	95%	95%	86%	95%	
	0.2	1	0.4	0.1513	0.214	0.8579	0.3335	0.0759	0.0863	0.2446	0.1395	91%	95%	88%	91%	
		2	0.4	0.1489	0.1938	1.8323	0.3567	0.0679	0.0763	0.3076	0.1066	89%	90%	87%	92%	
		3	0.4	0.1375	0.2008	2.7325	0.3663	0.062	0.0716	0.3358	0.0878	91%	96%	84%	92%	
	0.4	0.05	1	0.4	0.358	0.1199	0.9321	0.3482	0.0949	0.0744	0.2752	0.1547	95%	95%	97%	98%
			2	0.4	0.3531	0.1173	1.8584	0.37	0.0865	0.0671	0.3301	0.119	95%	93%	90%	92%
			3	0.4	0.3436	0.1174	2.839	0.3886	0.0776	0.063	0.3873	0.0989	86%	93%	93%	94%
0.1		1	0.4	0.3562	0.1671	0.8701	0.3214	0.094	0.0814	0.2551	0.1506	97%	92%	91%	92%	
		2	0.4	0.354	0.1477	1.8861	0.3491	0.0855	0.0739	0.3447	0.121	93%	93%	92%	94%	
		3	0.4	0.3513	0.1503	2.9015	0.377	0.0836	0.0735	0.4051	0.1051	94%	92%	94%	92%	
0.2		1	0.4	0.3786	0.2028	0.9602	0.3427	0.0947	0.0879	0.2759	0.16	96%	95%	92%	95%	
		2	0.4	0.3674	0.214	1.9634	0.3511	0.0903	0.0869	0.3961	0.1397	96%	97%	95%	97%	
		3	0.4	0.3697	0.2048	2.9069	0.339	0.0867	0.0824	0.4496	0.1205	89%	93%	94%	91%	
0.6		0.05	1	0.4	0.5511	0.1113	0.9518	0.365	0.0825	0.0723	0.2868	0.1674	90%	96%	96%	98%
			2	0.4	0.5485	0.1271	1.791	0.36	0.0772	0.0719	0.3751	0.1442	93%	90%	89%	96%
			3	0.4	0.53	0.1232	2.8293	0.3685	0.0827	0.073	0.4675	0.1273	91%	95%	96%	99%
	0.1	1	0.4	0.5491	0.1514	0.9941	0.3666	0.0824	0.0806	0.3082	0.1651	95%	94%	98%	98%	
		2	0.4	0.5444	0.1515	1.8723	0.3616	0.0808	0.0798	0.4167	0.1494	92%	94%	93%	98%	
		3	0.4	0.5033	0.1588	2.7556	0.3301	0.089	0.0796	0.5096	0.1332	87%	91%	93%	91%	
	0.2	1	0.4	0.5563	0.2213	0.9992	0.3258	0.0841	0.0891	0.3319	0.1804	91%	92%	94%	98%	
		2	0.4	0.5143	0.2142	1.9335	0.3083	0.1015	0.0955	0.5534	0.1721	90%	94%	98%	97%	
		3	0.4	0.333	0.1403	2.0985	0.2205	0.1198	0.0798	0.7357	0.1319	65%	71%	75%	69%	

Tabela 6: Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.4$

α_1	α_2	λ	ρ	$M(\alpha_1)$	$M(\alpha_2)$	$M(\lambda)$	$M(\rho)$	$SD(\alpha_1)$	$SD(\alpha_2)$	$SD(\lambda)$	$SD(\rho)$	$C(\alpha_1)$	$C(\alpha_2)$	$C(\lambda)$	$C(\rho)$
0.1	0.05	1	0.6	0.1344	0.1114	0.9106	0.5111	0.0739	0.0665	0.3034	0.1395	95%	91%	89%	92%
		2	0.6	0.1403	0.1115	1.6154	0.5023	0.0638	0.0562	0.3494	0.0949	90%	85%	82%	82%
		3	0.6	0.1308	0.077	2.5687	0.5387	0.0547	0.0428	0.4069	0.0776	94%	89%	88%	85%
	0.1	1	0.6	0.1361	0.1437	0.8454	0.4752	0.0749	0.0763	0.2771	0.1413	93%	95%	84%	88%
		2	0.6	0.1378	0.1307	1.6244	0.5143	0.0643	0.0622	0.34	0.094	90%	91%	83%	87%
		3	0.6	0.1279	0.1175	2.4772	0.4949	0.0514	0.0515	0.3533	0.0704	88%	93%	80%	82%
	0.2	1	0.6	0.1564	0.1974	0.7624	0.4347	0.0785	0.0856	0.269	0.1537	93%	96%	77%	86%
		2	0.6	0.1409	0.2134	1.6193	0.5015	0.063	0.0718	0.3591	0.104	86%	90%	84%	85%
		3	0.6	0.1331	0.2032	2.5653	0.5015	0.0544	0.0645	0.4035	0.0838	90%	93%	84%	86%
	0.05	1	0.6	0.3537	0.128	0.7758	0.4405	0.0942	0.0708	0.2786	0.157	89%	84%	83%	87%
		2	0.6	0.3522	0.1022	1.7906	0.5154	0.0808	0.0598	0.3657	0.1025	95%	95%	90%	88%
		3	0.6	0.373	0.1016	2.772	0.5376	0.071	0.0559	0.4451	0.0882	91%	92%	87%	90%
0.4	0.1	1	0.6	0.3486	0.1575	0.6706	0.3943	0.0957	0.08	0.2445	0.1589	94%	89%	77%	80%
		2	0.6	0.3694	0.1392	1.7065	0.4983	0.0821	0.0705	0.3839	0.1156	94%	94%	85%	88%
		3	0.6	0.3684	0.136	2.5731	0.4988	0.0707	0.0609	0.4181	0.0889	90%	88%	82%	81%
	0.2	1	0.6	0.3789	0.2139	0.8674	0.4306	0.094	0.0896	0.311	0.165	95%	96%	84%	85%
		2	0.6	0.3656	0.2191	1.7188	0.4741	0.0834	0.0803	0.4119	0.1314	97%	91%	90%	89%
		3	0.6	0.381	0.2158	2.5858	0.5133	0.0774	0.0754	0.4799	0.1072	92%	91%	87%	86%
	0.05	1	0.6	0.5348	0.139	0.8072	0.4621	0.0865	0.0764	0.3034	0.1651	90%	86%	84%	89%
		2	0.6	0.5544	0.1094	1.8238	0.5343	0.0732	0.0647	0.4335	0.1275	94%	92%	92%	95%
		3	0.6	0.5476	0.1073	2.7906	0.5555	0.0676	0.0609	0.517	0.1047	90%	91%	91%	94%
	0.1	1	0.6	0.5404	0.1638	0.8144	0.459	0.087	0.0823	0.3098	0.1733	93%	94%	85%	92%
		2	0.6	0.5563	0.1396	1.7166	0.4843	0.0758	0.0717	0.4015	0.1345	92%	92%	86%	91%
		3	0.6	0.5424	0.1333	2.7599	0.5227	0.0741	0.0701	0.4914	0.1136	91%	97%	88%	92%
0.6	0.2	1	0.6	0.5604	0.221	0.8575	0.4033	0.0839	0.0885	0.3425	0.1894	96%	94%	87%	90%
		2	0.6	0.5461	0.2134	1.8276	0.4401	0.0856	0.09	0.5153	0.1681	95%	97%	93%	90%
		3	0.6	0.5348	0.1892	2.7593	0.4645	0.1057	0.0875	0.6548	0.1529	89%	96%	91%	95%

Tabela 7: Resumo dos resultados do algoritmo MCMC baseado em 100 amostras simuladas para cada conjunto de parâmetros de um modelo ZINAR(2), fixando-se $\rho = 0.6$

REFERÊNCIAS

- [1] AL-OSH, M. A.; ALZAID, A. A. First-order integer-valued autoregressive (INAR (1)) process. *Journal of Time Series Analysis*, Wiley Online Library, v. 8, n. 3, p. 261–275, 1987.
- [2] MCKENZIE, E. Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, Cambridge University Press, v. 20, n. 4, p. 822–835, 1988.
- [3] ALZAID, A.; AL-OSH, M. First-order integer-valued autoregressive (INAR (1)) process: distributional and regression properties. *Statistica Neerlandica*, Wiley Online Library, v. 42, n. 1, p. 53–61, 1988.
- [4] JIN-GUAN, L. Y. D. The integer-valued autoregressive (INAR (p)) model. *J. Time Ser. Anal*, v. 12, p. 129–142, 1991.
- [5] FREELAND, R. K.; MCCABE, B. Asymptotic properties of CLS estimators in the Poisson AR (1) model. *Statistics & probability letters*, Elsevier, v. 73, n. 2, p. 147–153, 2005.
- [6] RISTIĆ, M. M.; BAKOUCH, H. S.; NASTIĆ, A. S. A new geometric first-order integer-valued autoregressive (NGINAR (1)) process. *Journal of Statistical Planning and Inference*, Elsevier, v. 139, n. 7, p. 2218–2226, 2009.
- [7] BARRETO-SOUZA, W. Zero-modified geometric INAR (1) process for modelling count time series with deflation or inflation of zeros. *Journal of Time Series Analysis*, Wiley Online Library, v. 36, n. 6, p. 839–852, 2015.

- [8] JAZI, M. A.; JONES, G.; LAI, C.-D. First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, Wiley Online Library, v. 33, n. 6, p. 954–963, 2012.
- [9] GARAY, A. M. et al. Bayesian analysis of the p-order integer-valued AR process with zero-inflated Poisson innovations. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 90, n. 11, p. 1943–1964, 2020.
- [10] GAMERMAN, D.; LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. [S.l.]: CRC Press, 2006.
- [11] NEAL, P.; RAO, T. S. MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis*, Wiley Online Library, v. 28, n. 1, p. 92–110, 2007.
- [12] STEUTEL, F. W.; HARN, K. van. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, JSTOR, p. 893–899, 1979.
- [13] COHEN, A. C. *Estimation in mixtures of discrete distributions*. [S.l.]: Statistical Pub. Society, 1963.
- [14] DOUKHAN, P.; FOKIANOS, K.; LI, X. On weak dependence conditions: The case of discrete valued processes. *Statistics & Probability Letters*, Elsevier, v. 82, n. 11, p. 1941–1948, 2012.
- [15] HAMILTON, J. D. *Time series analysis*. [S.l.]: Princeton university press, 2020.
- [16] BU, R.; MCCABE, B.; HADRI, K. Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of time series analysis*, Wiley Online Library, v. 29, n. 6, p. 973–994, 2008.
- [17] PRADO, R.; WEST, M. *Time series: modeling, computation, and inference*. [S.l.]: Chapman and Hall/CRC, 2010.

- [18] MAYER, W. V. *Introduction to Probability and Statistics*. [S.l.]: JSTOR, 1969.
- [19] PETTIT, L. *Bayesian Model Selection and Statistical Modeling by ANDO, T.* [S.l.]: Wiley Online Library, 2012.
- [20] GARAY, A. M. et al. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, Elsevier, v. 55, n. 3, p. 1304–1318, 2011.
- [21] METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The journal of chemical physics*, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953.
- [22] R Core Team. R: A language and environment for statistical computing. Vienna, Austria, 2022.
- [23] KASS, R. E.; RAFTERY, A. E. Bayes factors. *Journal of the american statistical association*, Taylor & Francis, v. 90, n. 430, p. 773–795, 1995.