



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

Maria Júlia Neves Gregório

**ABORDAGEM BAYESIANA PARA O PROCESSO RCINAR(1): ESTIMAÇÃO,
INFERÊNCIA E PREVISÃO**

Recife

2025

Maria Júlia Neves Gregório

**ABORDAGEM BAYESIANA PARA O PROCESSO RCINAR(1): ESTIMAÇÃO,
INFERÊNCIA E PREVISÃO**

Trabalho apresentado ao Programa de Pós-graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Estatística.

Área de Concentração: Estatística Matemática

Orientador : Prof. Dr. Aldo William Medina Garay

Coorientadora: Prof^a. Dr^a. Francielle de Lima Medina

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Gregório, Maria Júlia Neves.

Abordagem bayesiana para o processo RCINAR(1): estimação, inferência e previsão / Maria Júlia Neves Gregório. - Recife, 2024.

46 f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Ciências Exatas e da Natureza, Programa de Pós-Graduação em Estatística, 2024.

Orientação: Aldo William Medina Garay.

Coorientação: Francielle de Lima Medina.

Inclui referências e apêndice.

1. RCINAR(1); 2. Distribuição Poisson inflacionada de zeros;
3. Abordagem bayesiana. I. Garay, Aldo William Medina. II. Medina, Francielle de Lima. III. Título.

UFPE-Biblioteca Central

MARIA JÚLIA NEVES GREGÓRIO

**"ABORDAGEM BAYESIANA PARA O PROCESSO RCINAR(1):
ESTIMAÇÃO, INFERÊNCIA E PREVISÃO"**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 30 de outubro de 2024.

BANCA EXAMINADORA

Profa. Dra. Francielle de Lima Medina
Presidente/Coorientadora, UFPE

Prof. Dr. Rodrigo Lambert
Examinador Externo à Instituição, UFMT

Prof. Dr. Vinicius Quintas Souto Maior
Examinador Interno à Instituição, UFPE

À minha madrinha Lourdinha, mesmo que você não esteja mais conosco, sua memória sempre permanecerá.

AGRADECIMENTOS

O desenvolvimento desta dissertação contou com a ajuda de diversas pessoas, dentre as quais agradeço:

Primeiramente à Deus, que me deu oportunidades e força de vontade para superar todos os desafios.

Aos meus pais Jorge e Carmem Lúcia, e minha irmã Laísa, meus tios, primos e afilhados, por todo apoio e incentivo durante todo o período do Mestrado.

Ao meu orientador Aldo e minha coorientadora Francielle, por toda dedicação e paciência e que através dos seus ensinamentos permitiram que eu concluísse esse trabalho.

Aos professores que contribuíram para minha formação.

Aos amigos que fiz durante essa jornada, em especial Clarice, Débora, Fran, Luciano e Pedro, que sempre me incentivaram a persistir e me acolheram de braços abertos.

Aos meus amigos de longa data, que mesmo de longe, sempre se fizeram presentes Carol, Jordana e Isabela.

E a todas pessoas que passaram pela minha vida nesse período e contruíram direta ou indiretamente com esse conquista.

A FACEPE, pelo apoio financeiro.

RESUMO

Neste trabalho, estudamos o modelo autorregressivo de valores inteiros não negativos com coeficiente aleatório e inovação seguindo distribuição de Poisson, conhecido como RCINAR(1). Propomos também uma extensão desse modelo, denominada RCZINAR(1), cuja inovação segue uma distribuição Poisson inflacionada de zeros, sendo especialmente útil para modelar dados de contagem que apresentam um número maior de zeros que o esperado. Desenvolvemos um algoritmo via abordagem Bayesiana para estimação dos parâmetros dos modelos RCINAR(1) e RCZINAR(1). Implementamos estudos de simulação com o intuito de avaliar as metodologias propostas. Além disso, ajustamos o modelo para um conjunto de dados reais, comparamos seus resultados e utilizamos o método Bootstrap em blocos para estimar a distribuição preditiva.

Palavras-chave: RCINAR(1), RCZINAR(1), Distribuição Poisson Inflacionada de zeros, Abordagem Bayesiana

ABSTRACT

In this work, we study the random coefficient integer-valued autoregressive model with innovations following Poisson distribution, known as RCINAR(1). We also propose an extension of this model, called RCZINAR(1), where the innovations follow a zero-inflated Poisson distribution. This extension is particularly useful for modeling count data that exhibit a higher number of zeros than expected. We developed an algorithm through a Bayesian approach to estimate the parameters for the RCINAR(1) and RCZINAR(1) models. We implemented simulation studies to evaluate the proposed methodologies. Furthermore, we fit the model to a set of real data, compared their results, and we use block Bootstrap method to estimate the predictive distribution.

Keywords: RCINAR(1), RCZINAR(1), Zero-inflated Poisson distribution, Bayesian approach.

LISTA DE FIGURAS

Figura 1 – Viés dos estimadores dos parâmetros do modelo RCINAR, para cada parâmetro estimado individualmente, para $\lambda = 2$, $\alpha = -1.5$ e $\sigma^2 = 5$	31
Figura 2 – Viés dos estimadores dos parâmetros do modelo RCINAR, estimados simultaneamente, para $\lambda = 2$, $\alpha = -1.5$ e $\sigma^2 = 5$	32
Figura 3 – Número mensal de casos de Poliomielite nos EUA, registrados entre 1970 e 1983.	36
Figura 4 – ACF e PACF do conjunto de dados Poliomielite.	37
Figura 5 – Histórico das cadeias e densidade posterior marginal aproximada de λ , considerando o modelo RCINAR(1)	39
Figura 6 – Histórico das cadeias e densidade posterior marginal aproximada de α , considerando o modelo RCINAR(1)	39
Figura 7 – Histórico das cadeias e densidade posterior marginal aproximada de σ^2 , considerando o modelo RCINAR(1)	40
Figura 8 – Gráfico de barras da distribuição de previsão das observações 167 e 168, do modelo RCINAR(1)	40
Figura 9 – Gráfico de barras da distribuição de previsão das observações 167 e 168, do modelo RCZINAR(1).	46

LISTA DE TABELAS

Tabela 1 – Viés e EQM das estimativas dos parâmetros do processo RCINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 5$	33
Tabela 2 – Viés e EQM das estimativas dos parâmetros do processo RCINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 2.5$	33
Tabela 3 – Viés e EQM das estimativas dos parâmetros do processo RCZINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 2.5$	34
Tabela 4 – Viés e EQM das estimativas dos parâmetros do processo RCZINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 5$	35
Tabela 5 – Estatísticas Descritivas	36
Tabela 6 – Tabela de resumo MCMC para os modelos INAR(1), ZINAR(1), RCINAR(1), RCZINAR(1) e RCINAR _T (1).	38
Tabela 7 – Densidade marginal estimada para os modelos estudados.	38
Tabela 8 – Estatísticas da distribuição preditiva das observações #167 e #168, no modelo RCINAR. As linhas a partir de $j=9$ possuem todos os valores iguais a zero.	41
Tabela 9 – Estatísticas da distribuição preditiva das observações #167 e #168, no modelo RCZINAR. As linhas a partir de $j=9$ possuem todos os valores iguais a zero.	46

SUMÁRIO

1	INTRODUÇÃO	11
2	CONCEITOS FUNDAMENTAIS	13
2.1	OPERADOR <i>THINNING</i> BINOMIAL	13
2.2	MODELO INAR(1)	14
2.3	DISTRIBUIÇÃO POISSON INFLACIONADA DE ZEROS	15
2.4	MODELO ZINAR(1)	16
2.5	MODELO RCINAR(1)	17
2.5.1	Definição e Propriedades	17
3	PROCESSO RCZINAR(1)	20
3.1	DEFINIÇÃO E PROPRIEDADES	20
3.2	FUNÇÃO DE VEROSSIMILHANÇA	21
3.2.1	Modelo RCINAR	21
3.2.2	Algoritmo MCMC	23
3.2.3	Modelo RCZINAR	27
3.2.4	Algoritmo MCMC	27
3.3	PREVISÃO	29
3.3.1	Abordagem Bootstrap em Blocos Móveis	29
3.4	FATOR DE BAYES	30
4	ESTUDO DE SIMULAÇÃO	31
4.1	CASO RCINAR(1)	31
4.2	CASO RCZINAR(1)	34
5	APLICAÇÃO	36
6	CONSIDERAÇÕES FINAIS	42
	REFERÊNCIAS	43
	Apêndice A - Previsão RCZINAR	46

1 INTRODUÇÃO

A análise de séries temporais é amplamente utilizada em diversas áreas do conhecimento. Por exemplo, Schneckenreither et al. (2023) apresentam um estudo de séries temporais no âmbito de doenças infecciosas, no qual são apresentados alguns métodos que visam fornecer informações sobre os surtos epidêmicos. Outro exemplo é o interesse na taxa de conversão de um anúncio publicitário, como visto em Gong et al. (2015). Além disso, Costa, Ferreira e Cordeiro (2015) discutem a aplicação de modelos de séries temporais em dados climáticos, destacando sua importância na análise e previsão de variáveis meteorológicas, entre outros casos. Em comum, essas séries temporais são utilizadas para analisar e prever o comportamento da variável resposta ao longo do tempo.

As séries temporais podem assumir valores contínuos ou discretos. Séries temporais discretas, podem ser classificadas como séries de contagem, séries binárias, entre outras. Em relação às séries de contagem, um dos modelos mais conhecidos para a modelagem desse tipo de dados é o processo autorregressivo de valores inteiros de primeira ordem, INAR(1), introduzido por McKenzie (1985) e Al-Osh e Alzaid (1987). Os autores consideraram que as inovações seguem uma distribuição Poisson. Posteriormente, Jazi, Jones e Lai (2012) propuseram o processo ZINAR(1), que supõe inovações seguindo uma distribuição Poisson Inflacionada de Zeros. Garay et al. (2020) estenderam o modelo para ZINAR(p) e desenvolveram um algoritmo MCMC (Cadeia de Markov Monte Carlo) para estimar os parâmetros, sob abordagem Bayesiana. Garay et al. (2022), também apresentaram o estudo do modelo INAR(1) com uma classe de distribuições inflacionadas em zeros para a distribuição de inovação, além de discutir características da classe e, estimação dos parâmetros, ainda apresentam métodos Bootstrap para obter intervalos de confiança.

Por outro lado, Zheng, Basawa e Datta (2007) propuseram um processo autorregressivo de valores inteiros não negativos de primeira ordem, com coeficiente aleatório, baseado na transformação *logito*. A literatura possui outros estudos sobre o modelo RCINAR em diferentes condições. Por exemplo, Zhang, Wang e Yang (2020) exploram o processo RCINAR(1) com marginais binomiais negativas generalizadas, originalmente proposto por Weiß (2008). Neste estudo, eles estimam os parâmetros desse processo usando métodos clássicos e Bayesianos. Já Awale e Kashikar (2023) propõem um modelo autorregressivo inteiro de coeficiente aleatório de ordem 1, com estrutura sazonal.

Neste trabalho estudaremos o modelo RCINAR, de ordem 1, por ser um modelo mais flexível, pois permite variações no coeficiente do processo quando supomos diferentes distribuições. Ainda em relação ao coeficiente do processo, o abordaremos de duas formas distintas. Zheng, Basawa e Datta (2007), sugerem o uso da transformação logito do coeficiente, supondo distribuição Normal. Neste trabalho, supomos que a transformação do coeficiente segue distribuição t de Student, pois visa ser uma alternativa mais robusta para estimação do parâmetro de escala. Além disso, também estudaremos o modelo quando a inovação segue uma distribuição Poisson Inflacionada em Zeros, que é útil quando o conjunto de dados apresenta um número maior de zeros do que o esperado. Utilizamos algoritmos Bayesianos para a estimação dos parâmetros do modelo. Além disso, utilizamos o método Bootstrap para o cálculo da distribuição preditiva.

Essa dissertação está dividida da seguinte forma: no Capítulo 2, apresentamos as principais definições necessárias para o desenvolvimento desta dissertação, incluindo o operador thinning binomial, a distribuição Poisson inflacionada em zeros e os modelos INAR, ZINAR e RCINAR. No Capítulo 3, definimos o modelo RCZINAR, além de apresentarmos o algoritmo Gibbs para a estimação dos parâmetros dos processos RCINAR e RCZINAR, além de uma abordagem de predição utilizando o método de bootstrap. No Capítulo 4, realizamos um estudo de simulação referente ao algoritmo proposto no capítulo anterior. No Capítulo 5, é feita uma aplicação à dados reais. Por fim, o Capítulo 6 discute as conclusões obtidas e as possibilidades de trabalhos futuros.

2 CONCEITOS FUNDAMENTAIS

Neste capítulo serão apresentados alguns conceitos, definições e propriedades a serem utilizados ao longo do trabalho.

2.1 OPERADOR *THINNING* BINOMIAL

O operador *thinning* binomial, proposto por Steutel e Van Harn (1979) é definido como:

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

em que Y é uma variável aleatória inteira não negativa, $\phi \in [0, 1]$, $\{X_i\}_{i \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid), que seguem distribuição Bernoulli, com parâmetro ϕ , independente de Y . Note que, para $y > 0$, $\phi \circ Y | Y = y \sim \text{Bin}(y, \phi)$ e para $y = 0$, $\phi \circ Y | Y = y$ é uma variável degenerada em 0. Podemos interpretar o operador *thinning* como a parcela de observações passadas que permaneceram no tempo presente, de acordo com uma certa probabilidade.

Silva e Oliveira (2004), Weiß (2008) entre outros autores discutiram e apresentaram propriedades do operador *thinning*. A seguir são apresentadas algumas dessas propriedades.

Assim, seja Y uma v.a. não-negativa de valores inteiros e α e β valores reais pertencentes ao intervalo $[0,1]$, então:

- (i) $0 \circ Y = 0$,
- (ii) $1 \circ Y = Y$,
- (iii) $(\alpha\beta) \circ Y \stackrel{d}{=} \alpha \circ (\beta \circ Y)$,
- (iv) $E(\alpha \circ Y) = \alpha E(Y)$,
- (v) $E((\alpha \circ Y)^2) = \alpha^2 E(Y^2) + \alpha(1 - \alpha)E(Y)$,
- (vi) $\text{Var}(\alpha \circ Y) = \alpha^2 \text{Var}(Y) + \alpha(1 - \alpha)E(Y)$.

Além do operador *thinning* binomial, outros operadores *thinning* foram propostos na literatura, como por exemplo o operador *thinning* binomial negativo, definido por Ristic, Bakouch e Nastic (2009), que utiliza a distribuição geométrica como sua sequência de variáveis aleatórias iid (Y_i) e o operador *thinning* hipergeométrico definido por Al-Osh e Alzaid (1991).

2.2 MODELO INAR(1)

O processo autoregressivo de valores inteiros não negativos de primeira ordem (INAR(1)), introduzido por McKenzie (1985), é definido por:

$$Y_t = \phi \circ Y_{t-1} + Z_t, t \in \mathbb{Z}, \quad (2.2)$$

em que $\{Z_t\}_{t \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias não negativas assumindo valores inteiros, iid com média μ e variância σ^2 finitas, com Z_t independente de Y_{t-1} , $\forall t \in \mathbb{Z}$.

De acordo com Al-Osh e Alzaid (1987), Z_t são definidas como inovações e são interpretadas como o número de elementos adicionados ao processo, no intervalo de tempo $(t-1, t]$. A função " $\phi \circ Y_{t-1}$ " indica a quantidade total de elementos presentes no instante $t-1$, que permaneceram no processo no instantante t . Podemos ilustrar o modelo com uma série de registros semanais de chegada em uma unidade de terapia intensiva (UTI). Conforme a abordagem proposta, o número total de pacientes para a próxima semana é obtido pela soma dos pacientes que ainda permanecem internados, com base em uma probabilidade fixa de permanência, e o número de novos internamentos, que segue uma distribuição discreta de probabilidade.

Por substituições recursivas, podemos obter a distribuição marginal do processo em termos da inovação, isto é:

$$Y_t \stackrel{d}{=} \sum_{j=0}^{\infty} \phi^j \circ Z_{t-j}.$$

Para $\phi \in (0, 1)$ teremos um processo estacionário de segunda ordem, cuja média e a variância são dadas respectivamente por:

$$E(Y_t) = \frac{\mu}{1 - \phi} \text{ e } \text{Var}(Y_t) = \frac{\phi\mu + \sigma^2}{1 - \phi^2}.$$

Se $\phi = 1$, o processo se torna não estacionário, pois o processo mantém uma tendência de crescimento ao longo do tempo. Caso $\phi = 0$, o processo depende apenas da inovação, ou seja, os valores de Y_t não vão depender dos valores passados Y_{t-1} .

As funções de autocovariância e autocorrelação de defasagem h , são expressas respectivamente por:

$$\gamma(h) = \text{Cov}(Y_t, Y_{t+h}) = \sigma^2 \phi^h, \quad h = 0, \pm 1, \pm 2, \dots$$

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \quad h = 0, \pm 1, \pm 2, \dots$$

Podemos notar que a função de autocorrelação do processo não depende da distribuição da inovação. Além disso, apresenta um decaimento exponencial que depende apenas do deslocamento h . Em outras palavras, quanto maior for a defasagem, menor será o valor da função de autocorrelação, considerando $\phi \in (0, 1)$.

2.3 DISTRIBUIÇÃO POISSON INFLACIONADA DE ZEROS

Uma variável aleatória Z segue uma distribuição Poisson inflacionada de zeros, com parâmetros ρ e λ , denotada por $Z \sim \text{ZIP}(\rho, \lambda)$, se sua função de probabilidade, é dada por:

$$P(Z = z) = \begin{cases} \rho + (1 - \rho)e^{-\lambda}, & \text{if } z = 0 \\ (1 - \rho) \frac{e^{-\lambda} \lambda^z}{z!}, & \text{if } z = 1, 2, \dots \end{cases}$$

equivalentemente podemos escrever,

$$P(Z = z) = \rho I_{\{0\}}(z) + (1 - \rho) e^{-\lambda} \lambda^z / z!, \quad z = 0, 1, 2, \dots, \quad (2.3)$$

em que $\lambda > 0$, $0 < \rho < 1$ e, $I_{\{0\}}(z)$ é a função indicadora, definida por:

$$I_{\{0\}}(z) = \begin{cases} 1 & \text{se } z = 0, \\ 0 & \text{se } z \neq 0. \end{cases}$$

A média e a variância da variável aleatória Z são dadas respectivamente por:

$$E(Z) = \lambda(1 - \rho),$$

$$\text{Var}(Z) = \lambda(1 - \rho)(1 + \rho\lambda) = \mu_Z(1 + \rho\lambda)$$

Note que, se $\rho = 0$, Z segue uma distribuição de Poisson, com $\mu_Z = \sigma_Z^2$, se $\rho > 0$, teremos $\sigma^2 > \mu$, ou seja, nossa distribuição é sobredispersa.

A distribuição ZIP, faz parte de uma classe de distribuições inflacionadas em zeros, que tem o objetivo de modelar dados que apresentam uma quantidade maior de zeros que o esperado. Para mais detalhes, podemos consultar Garay et al. (2022), que apresentam uma classe de distribuições inflacionadas em zeros, que tem como casos particulares Poisson Inflacionada em Zeros (ZIP), Binomial Negativa Inflacionada em Zeros (ZINB) e Poisson Inversa Gaussiana Inflacionada em Zeros (ZIPIG). Além disso, também apresentam a extensão do modelo INAR(1), considerando a classe de distribuições inflacionadas em zeros.

2.4 MODELO ZINAR(1)

Al-Osh e Alzaid (1987) propuseram que a inovação segue uma a distribuição de Poisson. No entanto, em muitos casos, essa suposição pode não ser adequada para modelar os dados, como por exemplo, se os dados observados têm uma variância maior que a média (sobredispersão), a distribuição Poisson pode não ser adequada. Como resultado, foram desenvolvidos novos modelos que buscam atender as necessidades específicas na análise de certos conjuntos de dados. Em dados de contagem, é comum a incidência de zeros ser maior do que o esperado, assim Jazi, Jones e Lai (2012) sugerem que as inovações do modelo INAR(1) sigam a distribuição Poisson Inflacionada de Zeros (ZIP). Retomando o exemplo dos registros de entrada em uma UTI, mencionado anteriormente, é possível cogitar a possibilidade de períodos nos quais não há novas entradas de pacientes. Nessas circunstâncias, é interessante a escolha de uma distribuição que considere o excesso de zeros.

Como apresentado por Jazi, Jones e Lai (2012), dizemos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo autorregressivo para valores inteiros de primeira ordem, com inflação de zeros, denotado por ZINAR(1), se as inovações seguem a distribuição Poisson Inflacionada de Zeros, com parâmetros λ e ρ .

A média e a variância do processo ZINAR(1) são dadas respectivamente por:

$$\mu = \frac{\lambda(1 - \rho)}{1 - \phi} \quad \text{e} \quad \sigma^2 = \frac{\lambda(1 - \rho)(1 + \phi + \rho\lambda)}{1 - \phi^2}.$$

O parâmetro ρ determina a proporção de zeros na distribuição ZIP. Portanto, quanto mais próximo de 1, menor será a média do processo, assim como sua variabilidade.

2.5 MODELO RCINAR(1)

Em Seções anteriores, utilizamos o número de entradas em uma UTI como exemplo para ilustrar o funcionamento do modelo proposto. Ao retornar a esse exemplo, é possível considerar que a probabilidade de permanência pode não ser constante em um determinado período. Durante períodos que apresentam surtos epidêmicos de certas doenças, por exemplo, é possível que o número de entradas aumente consideravelmente. Por essa razão, é interessante propor um modelo que, além das suposições anteriores, considere um coeficiente variável como uma alternativa para capturar as variações do processo, ou seja, o coeficiente será aleatório e poderá se adaptar aos diferentes contextos.

Zheng, Basawa e Datta (2007) propõem um modelo autorregressivo de valor inteiro com coeficiente aleatório de primeira ordem, baseado na transformação logito. Eles estabelecem a ergodicidade do processo, tornando possível inferir sobre o comportamento do processo em longo prazo com base em observações de curto prazo. Além disso, apresentam estimadores de mínimos quadrados condicionais e quase verossimilhança, bem como propriedades assintóticas da distribuição desses estimadores.

Zheng, Basawa e Datta (2006), também apresentaram um modelo autorregressivo de valor inteiro, de ordem p com coeficiente aleatório para dados de contagem. Novamente, as propriedades de ergodicidade são abordadas e métodos de máxima verossimilhança, mínimos quadrados condicionais, quase verossimilhança modificada e método generalizado dos momentos são utilizados para estimar os parâmetros do modelo. É importante mencionar que para a análise de dados reais, foram propostos alguns modelos para capturar a tendência e a sazonalidade dos dados, mas foram observadas algumas problemáticas na estimação, como estimativas que não pertenciam ao espaço paramétrico.

2.5.1 Definição e Propriedades

Seja $\{Y_t\}_{t \in \mathbb{Z}}$ um processo autoregressivo para valores inteiros de primeira ordem, com coeficiente aleatório, denotado por RCINAR(1), definido por:

$$Y_t = \phi_t \circ Y_{t-1} + Z_t, \quad (2.4)$$

em que, $\{\phi_t\}_{t \in \mathbb{Z}}$ é uma sequência iid de variáveis aleatórias definida em $[0, 1)$; Z_t é uma sequência de variáveis iid com distribuição Poisson(λ), além disso, assumiremos que Z_t, ϕ_t são independentes entre si e independentes em relação a $Y_{t-1} \forall t \in \mathbb{Z}$, com isso, temos que as inovações e o coeficiente do processo não são influenciados pelos valores anteriores da série temporal.

Temos que, Y_t é um processo de Markov e, sua probabilidade de transição é dada por:

$$\begin{aligned}
 P_{ij} &= P(Y_t = i \mid Y_{t-1} = j, \phi_t = a) \\
 &= P(\phi_t \circ Y_{t-1} + Z_t = i \mid Y_{t-1} = j, \phi_t = a) \\
 &= \sum_{k=0}^{\min(i,j)} P(\phi_t \circ Y_{t-1} = k \mid Y_{t-1} = j, \phi_t = a) P(Z_t = i - k \mid Y_{t-1} = j, \phi_t = a) \\
 &= \sum_{k=0}^{\min(i,j)} P(\phi_t \circ Y_{t-1} = k \mid Y_{t-1} = j, \phi_t = a) P(Z_t = i - k) \\
 &= \sum_{k=0}^{\min(i,j)} \left[\binom{j}{k} \int_0^1 a^k (1-a)^{j-k} f_\phi(a) da \right] f_Z(i - k) \tag{2.5}
 \end{aligned}$$

Seja $E(\phi_t) = \phi$, $Var(\phi_t) = \sigma_\phi^2$, $E(Z_t) = \lambda$, $Var(Z_t) = \sigma_Z^2$ e, $\tau^2 = \phi^2 + \sigma_\phi^2$. Para $t \geq 2$

- (i) $E(Y_t | Y_{t-1}) = \phi Y_{t-1} + \lambda$
- (ii) $E(Y_t) = \lambda / (1 - \phi)$, se $E(Y_0) = \lambda / (1 - \phi)$
- (iii) $Var(Y_t | Y_{t-1}, \phi_t) = \phi_t (1 - \phi_t) Y_{t-1} + \sigma_z^2$
- (iv) $Var(Y_t | Y_{t-1}) = \sigma_\phi^2 Y_{t-1}^2 + (\phi(1 - \phi) - \sigma_\phi^2) Y_{t-1} + \sigma_z^2$
- (v) $Var(Y_t) = b / (1 - \tau^2)$
- (vi) $Cox(Y_{t+k}, Y_t) = \phi^k Var(Y_t)$, $k \geq 0$.

As provas das propriedades (i)-(iv) são dadas a seguir, já as propriedades (v) e (vi) podem ser vistas em Zheng, Basawa e Datta (2007).

$$\begin{aligned}
 (i) \ E(Y_t \mid Y_{t-1}) &= E(E(Y_t \mid Y_{t-1}, \phi_t) \mid Y_{t-1}) \\
 &= E(E(\phi_t \circ Y_{t-1} + Z_t \mid Y_{t-1}, \phi_t) \mid Y_{t-1}) \\
 &= E(E(\phi_t \circ Y_{t-1} \mid Y_{t-1}, \phi_t) \mid Y_{t-1}) + E(E(Z_t \mid Y_{t-1}; \phi_t) \mid Y_{t-1}) \\
 &= E(\phi_t Y_{t-1} \mid Y_{t-1}) + \lambda \\
 &= Y_{t-1} E(\phi_t) + \lambda \\
 &= \phi Y_{t-1} + \lambda
 \end{aligned}$$

$$\begin{aligned}
\text{(ii)} \quad E(Y_t) &= E(E(Y_t | Y_{t-1})) \\
&= E(E(Y_t | Y_{t-1}, \phi_t)) \\
&= E(E(\phi_t \circ Y_{t-1} + Z_t | Y_{t-1}, \phi_t)) \\
&= E(E(\phi_t \cdot Y_{t-1} | Y_{t-1}, \phi_t)) + E(E(Z_t)) \\
&= E(Y_{t-1}\phi_t) + \lambda \\
&= E(Y_{t-1}) + E(\phi_t) + \lambda \\
&= \phi E(Y_{t-1}) + \lambda \\
&= E(Y_t) - \phi E(Y_{t-1}) = \lambda \\
&= E(Y_t)(1 - \phi) = \lambda \\
E(Y_t) &= \frac{\lambda}{(1 - \phi)}
\end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad \text{Var}(Y_t | Y_{t-1}, \phi_t) &= \text{Var}(\phi_t \circ Y_{t-1} + Z_t | Y_{t-1}, \phi_t) \\
&= \text{Var}(\phi_t \circ Y_{t-1} | Y_{t-1}, \phi_t) + \text{Var}(Z_t) \\
&= \phi_t(1 - \phi_t)Y_{t-1} + \sigma_Z^2
\end{aligned}$$

$$\begin{aligned}
\text{(iv)} \quad \text{Var}(Y_t | Y_{t-1}) &= E[\text{Var}(Y_t | Y_{t-1}, \phi_t) | Y_{t-1}] + \text{Var}(E[Y_t | Y_{t-1}, \phi_t] | Y_{t-1}) \\
&= E[\phi_t(1 - \phi_t)Y_{t-1} + \sigma_Z^2 | Y_{t-1}] + \text{Var}(\phi_t Y_{t-1} | Y_{t-1}) \\
&= (\phi(1 - \phi) - \sigma_\phi^2)Y_{t-1} + \sigma_Z^2 + Y_{t-1}^2\sigma_\phi^2 \\
&= \sigma_\phi^2 Y_{t-1}^2 + (\phi(1 - \phi) - \sigma_\phi^2)Y_{t-1} + \sigma_Z^2
\end{aligned}$$

Zheng, Basawa e Datta (2007) provam que o processo $\{Y_t\}$ é uma cadeia de Markov irreduzível, aperiódica, e recorrente positiva (e portanto ergódica). Além disso propriedades como periodicidade, recorrência e ergodicidade são comentadas e podem ser vistas com mais detalhes em Ross (1996).

3 PROCESSO RCZINAR(1)

No Capítulo anterior, foram apresentados conceitos fundamentais para a compreensão deste e dos próximos capítulos. Um desses conceitos é o operador de *thinning* Binomial, que quantifica o número de observações remanescentes do tempo $t - 1$ no tempo t . Além disso, supomos que as inovações seguem uma distribuição Poisson inflacionada em zeros.

Além disso, foi apresentado o processo INAR(1) e suas extensões, como ZINAR(1), em que a inovação passa a seguir a distribuição Poisson inflacionada em zeros, e RCINAR(1), em que o coeficiente passa a ser aleatório, mas a inovação ainda segue distribuição Poisson.

O modelo RCINAR pode apresentar algumas limitações. Em particular, quando o conjunto de dados a ser analisado apresenta uma porcentagem significativa de zeros. A proposta deste capítulo é definir o modelo RCZINAR(1), em que a a distribuição da inovação segue uma distribuição Poisson inflacionada de zeros. Essa proposta é adequada para cenários que apresentam dados com mais zeros que o esperado. Neste capítulo apresentaremos, sob enfoque Bayesiano, a metodologia para a estimação dos parâmetros dos processos RCINAR e RCZINAR.

3.1 DEFINIÇÃO E PROPRIEDADES

Dizemos que o processo $\{Y_t\}_{t \in \mathbb{Z}}$ é um processo inteiro autoregressivo de primeira ordem, com coeficiente aleatório e inovação seguindo uma distribuição Poisson inflacionada em zeros, denotado por RCZINAR(1), definido por:

$$Y_t = \phi_t \circ Y_{t-1} + Z_t, \quad (3.1)$$

em que, $\{\phi_t\}_{t \in \mathbb{Z}}$ é uma sequência iid de variáveis aleatórias definidas em $[0, 1)$; $\{Z_t\}$ é uma sequência de variáveis i.i.d com distribuição Poisson inflacionada em zeros com parâmetros ρ e λ , além disso, assumiremos que Z_t e ϕ_t são independentes entre si e independentes a Y_{t-1} . Nesta dissertação, iremos utilizar a função de transformação logito para modelagem do coeficiente, O uso da função logito possibilita a modelagem com distribuições de probabilidade definidas no intervalo $(-\infty, +\infty)$ e garante que o coeficiente ϕ esteja no intervalo $[0, 1)$. Então, teremos W_t tal que $W_t = \log(\phi_t/1 - \phi_t)$. Assumiremos duas distribuições diferentes para W_t :

- (1) $W_t = \log\left(\frac{\phi_t}{1-\phi_t}\right) \sim N(\alpha, \sigma^2)$
- (2) $W_t = \log\left(\frac{\phi_t}{1-\phi_t}\right) \sim t_v(\alpha, \sigma^2)$

3.2 FUNÇÃO DE VEROSSIMILHANÇA

Seja $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ realizações do processo RCINAR(1). Denote o vetor de parâmetros desconhecidos por $\boldsymbol{\theta} = (\alpha, \sigma^2, \lambda)$ para $W_t \sim N(\alpha, \sigma^2)$ e $\boldsymbol{\theta} = (\alpha, \sigma^2, \nu, \lambda)$ para $W_t \sim t_\nu(\alpha, \sigma^2)$. Denotamos também a função de probabilidade conjunta de \mathbf{y} , por $f(\mathbf{y}|\boldsymbol{\theta})$. A função de verossimilhança é dada por:

$$L(y_1, y_2, \dots, y_n | \boldsymbol{\theta}) = f(y_1 | \boldsymbol{\theta}) \prod_{t=2}^n f(y_t | y_{t-1}; \boldsymbol{\theta}). \quad (3.2)$$

Para mais detalhes, sobre a função de máxima verossimilhança em contextos relacionados à análise de séries temporais, veja Hamilton (1994). Considerando que $Y_1 | \boldsymbol{\theta}$ segue uma distribuição Poisson, e utilizando a probabilidade de transição dada por (2.5), a função de verossimilhança pode ser reescrita como:

$$L(\mathbf{y} | \boldsymbol{\theta}) = f_Z(z_1) \prod_{t=2}^n \left\{ \sum_{k=0}^{\min(y_t, y_{t-1})} \left[\binom{y_{t-1}}{k} \int_0^t \phi_1^k (1 - \phi_1)^{y_{t-1}-k} dP_{\phi_1} \right] f_Z(y_t - k) \right\},$$

em que, $f_Z(\cdot)$ denota a função de probabilidade da distribuição de Poisson, e $\phi_1 = \exp(W_1)/(1 + \exp(W_1))$.

Por outro lado, para o processo RCZINAR(1), denotamos o vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\alpha, \sigma^2, \rho, \lambda)$ quando $W_t \sim N(\alpha, \sigma^2)$, $\boldsymbol{\theta} = (\alpha, \sigma^2, \nu, \rho, \lambda)$ para $W_t \sim t_\nu(\alpha, \sigma^2)$. A função de verossimilhança será dada pela Equação (3.2), em que Z segue uma distribuição Poisson inflacionada de zeros, dada por:

$$P(Z = z) = \begin{cases} \rho + (1 - \rho)e^{-\lambda}, & \text{if } z = 0 \\ (1 - \rho) \frac{e^{-\lambda} \lambda^z}{z!}, & \text{if } z = 1, 2, \dots \end{cases}$$

A seguir, apresentaremos a metodologia para estimar os parâmetros, para os processos RCINAR(1) e RCZINAR(1), sob abordagem Bayesiana.

3.2.1 Modelo RCINAR

Nesta seção apresentaremos o algoritmo para estimação dos parâmetros para o processo RCINAR(1), considerando $W_t \sim N(\alpha, \sigma^2)$ e $W_t \sim t_\nu(\alpha, \sigma^2)$.

O primeiro passo é a especificação de distribuições à Priori para os parâmetros. Como não temos conhecimento prévio sobre os parâmetros, escolhemos prioris fracamente informativas, que são escolhas realizadas de forma a não impor fortes suposições sobre os parâmetros.

Considere $W_t \sim N(\alpha, \sigma^2)$ e $W_t \sim t_\nu(\alpha, \sigma^2)$, com vetores de parâmetros $\theta = (\lambda, \alpha, \sigma^2)$ e $\theta = (\lambda, \alpha, \sigma^2, \nu)$, respectivamente. Assumindo independência entre os parâmetros das distribuições à priori:

(-) Para o caso Normal, temos que:

$$\pi(\theta) = \pi(\lambda, \alpha, \sigma^2) = \pi(\lambda)\pi(\alpha)\pi(\sigma^2).$$

A seguir, são apresentadas as seguintes distribuições à priori:

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

$$\alpha \sim N(\mu_\alpha, \sigma_\alpha^2),$$

$$\tau \sim \text{Gamma}(A_\tau, B_\tau).$$

Como sugerido por Garay et al. (2015), optamos por utilizar o parâmetro de precisão (τ) no lugar de σ^2 , tal que $\sigma^2 = 1/\tau$.

(-) Para o caso t, temos que:

$$\pi(\theta) = \pi(\lambda, \alpha, \tau, \nu) = \pi(\lambda)\pi(\alpha)\pi(\tau)\pi(\nu).$$

Além das distribuições à priori consideradas previamente, para o caso em que $W_t \sim t_\nu(\alpha, \sigma^2)$, consideramos também a priori hierárquica para o parâmetro ν , como sugerida por Cabral, Lachos e Madruga (2012), isto é:

$$\nu \mid \kappa \sim \text{Exp}(\kappa).$$

$$\kappa \sim U(0.01, 0.5).$$

em que, $N(\mu_\alpha, \sigma_\alpha^2)$ denota a distribuição Normal, com $-\infty < \mu < +\infty$, $\sigma^2 > 0$, $\text{Gamma}(A, B)$ denota a distribuição Gamma, com parâmetros $A > 0$, $B > 0$ e esperança A/B , $U(0.01, 0.5)$ denota a distribuição Uniforme definida no intervalo $(0.01, 0.5)$, $\text{Exp}(\kappa)$ denota a distribuição Exponencial, tal que, seu valor esperado é $1/\kappa$, com $\kappa > 0$. Utilizando a mesma abordagem que em Garay et al. (2020), os hiperparâmetros são fixos e conhecidos.

3.2.2 Algoritmo MCMC

De acordo com Gamerman e Lopes (2006, p. 142), o amostrador de Gibbs é um método MCMC, em que o kernel de transição é formado pelas distribuições condicionais completas. O amostrador de Gibbs é uma técnica MCMC muito utilizada em Inferência Bayesiana, quando as condicionais completas são possíveis de serem obtidas. Essa técnica permite a atualização sequencial dos parâmetros do modelo, baseada na geração de dados obtida pelas distribuições condicionais completas, resultando em amostras da distribuição a posteriori.

Considerando uma abordagem similar a Garay et al. (2020), com o uso de variáveis latentes, desenvolvemos um algoritmo do tipo MCMC.

Seja $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ realizações do processo RCINAR(1), definimos as variáveis latentes:

$$S_t = \phi_t \circ Y_{t-1}, \quad (3.3)$$

em que, pela equação (2.1), $S_t \mid \phi_t; y_{t-1} \sim \text{Bin}(y_{t-1}, \phi_t)$, quando $Y_{t-1} > 0$ e uma distribuição degenerada em zero quando $Y_{t-1} = 0$. E para $W_t \sim t_\nu(\alpha, \sigma^2)$, considerando sua representação hierárquica, temos que $U_t \sim \text{Gamma}(\nu/2, \nu/2)$ e:

$$W_t \mid U_t = u \sim N(\alpha, \sigma^2/u) \quad (3.4)$$

Pela definição do processo (2.4), temos que $Z_t = Y_t - S_t$ e, supondo independência entre $S_t \mid Y_{t-1}, \phi_t$, e $Z_t \mid \lambda$. Então:

$$\pi(z_t, s_t, w_t \mid y_{t-1}, \boldsymbol{\theta}) = \pi(z_t \mid \lambda) \pi(s_t \mid y_{t-1}; \phi_t) \pi(w_t \mid \alpha; \sigma^2) \quad (3.5)$$

Pelo teorema de Bayes, segue que:

$$\pi(z_t, s_t, w_t \mid y_t, y_{t-1}, \boldsymbol{\theta}) \propto \pi(y_t \mid z_t, s_t, w_t, y_{t-1}, \boldsymbol{\theta}) \pi(z_t, s_t, w_t \mid y_{t-1}, \boldsymbol{\theta}). \quad (3.6)$$

Da Equação (2.4), temos que

$$\pi(y_t \mid z_t, s_t, w_t, y_{t-1}, \boldsymbol{\theta}) = I_{\{z_t+s_t\}}(y_t) \quad (3.7)$$

então,

$$\pi(z_t, s_t, w_t | y_t, y_{(t-1)}, \boldsymbol{\theta}) \propto \pi(z_t, s_t, w_t | y_{(t-1)}, \boldsymbol{\theta}) I_{\mathcal{A}_t}(z_t, s_t, w_t) \quad (3.8)$$

em que,

$$\mathcal{A}_t = \{(z_t, s_t, w_t); z_t \in \{0, 1, \dots\}, 0 \leq s_t \leq y_t, z_t + s_t = y_t, w_t \in \{0, 1\}\}, \text{ tal que} \quad (3.9)$$

$$I_{\{\mathcal{A}_t\}}(z_t, s_t, w_t) = \begin{cases} 1 & \text{se } (z_t, s_t, w_t) \in \mathcal{A}_t, \\ 0 & \text{caso contrário.} \end{cases}$$

Utilizando as distribuições à priori dadas na Seção anterior, a função de verossimilhança completa de $\mathbf{s}, \mathbf{z}, \mathbf{w}, \boldsymbol{\theta}$ é dada por:

$$\begin{aligned} \pi(\mathbf{s}, \mathbf{z}, \mathbf{w}, \boldsymbol{\theta} | \mathbf{y}) &\propto \prod_{t=2}^n \{\pi(z_t, s_t, w_t | y_{t-1}, \boldsymbol{\theta})\} \pi(\boldsymbol{\theta}) \\ &\propto \prod_{t=2}^n \left\{ \left[\frac{\exp(-\lambda) \lambda^{z_t}}{z_t!} \right] \left[\binom{y_{t-1}}{s_t} \left(\frac{\exp(w_t)}{1 + \exp(w_t)} \right)^{s_t} \left(1 - \frac{\exp(w_t)}{1 + \exp(w_t)} \right)^{y_{t-1} - s_t} \right] \right\} \pi(\boldsymbol{\theta}) \end{aligned}$$

A partir da equação acima, podemos obter as condicionais completas.

Embora as condicionais completas de $\lambda, \alpha, \sigma^2$, tenham formas conhecidas, o mesmo não é válido para $\{s_t, w_t\}$. Nesses casos, utilizamos o algoritmo de Metropolis-Hastings, como proposto por Garay et al. (2020). A seguir, listamos os passos do algoritmo MCMC:

- *Passo 1:* Gere λ , a partir da distribuição condicional completa:

$$\lambda | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \alpha, \tau \sim \text{Gamma} \left(\sum_{t=2}^n (y_t - s_t) + A_\lambda; n - 1 + B_\lambda \right). \quad (3.10)$$

- *Passo 2:* Gere α , a partir da distribuição condicional completa:

(a) Para o caso Normal:

$$\alpha | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \lambda, \alpha, \tau \sim \text{N} \left(\left(\tau \sum_{t=2}^n w_t + \frac{\mu_\alpha}{\sigma_\alpha^2} \right) \left(\tau(n-1) + \frac{1}{\sigma_\alpha^2} \right)^{-1}; \left(\tau(n-1) + \frac{1}{\sigma_\alpha^2} \right)^{-1} \right). \quad (3.11)$$

(b) Para o caso t:

$$\alpha \mid \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \tau, \nu \sim \text{N} \left(\left(\tau \sum_{t=2}^n u_t w_t + \frac{\mu_\alpha}{\sigma_\alpha^2} \right) \left(\tau \sum_{t=2}^n u_t + \frac{1}{\sigma_\alpha^2} \right)^{-1}; \left(\tau \sum_{t=2}^n u_t + \frac{1}{\sigma_\alpha^2} \right)^{-1} \right). \quad (3.12)$$

▪ *Passo 3:* Gere τ , a partir da distribuição condicional completa:

(a) Para o caso Normal:

$$\tau \mid \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \lambda, \alpha \sim \text{Gamma} \left(\frac{n-1}{2} + A_\tau; \frac{\sum_{t=2}^n (w_t - \alpha)^2}{2} + B_\tau \right). \quad (3.13)$$

Faça $\sigma^2 = 1/\tau$.

(b) Para o caso t:

$$\tau \mid \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \nu \sim \text{Gamma} \left(\frac{n-1}{2} + A_\tau; \frac{\sum_{t=2}^n u_t (w_t - \alpha)^2}{2} + B_\tau \right) \quad (3.14)$$

Faça $\sigma^2 = 1/\tau$.

▪ *Passo 4:* (Apenas para o caso t) Gere U_t , a partir da distribuição condicional completa:

$$u_t \mid y_t, z_t, s_t, w_t, \lambda, \alpha, \tau, \nu \sim \text{Gamma} \left(\frac{\nu+1}{2}; \frac{\tau(w_t - \alpha)^2 \nu}{2} \right) \quad (3.15)$$

▪ *Passo 5:* Atualize os vetores $\mathbf{S} = (s_1, \dots, s_n)$ e $\mathbf{Z} = (z_1, \dots, z_n)$, por meio do algoritmo Metropolis-Hastings, seguindo os passos:

a) Se $y_{t-1} = 0$, então $s_t^* = 0$. Caso contrário, gere s_t^* pela distribuição Binomial(y_{t-1}, ϕ_t).

b) Faça $z_t^* = y_t - s_t^*$.

c) Gere um número aleatório k de uma distribuição Uniforme (0, 1).

d) Dada a amostra (s_t, z_t) obtida na iteração anterior, calcule probabilidade de aceitação da amostra proposta (s_t^*, z_t^*) :

$$a = \min \left\{ \frac{\pi(s_t^*, z_t^* \mid y_{t-1}, w_t^{-1}, \boldsymbol{\theta}) \pi(s_t \mid y_{t-1}, w_t^{-1})}{\pi(s_t, z_t \mid y_{t-1}, w_t^{-1}, \boldsymbol{\theta}) \pi(s_t^* \mid y_{t-1}, w_t^{-1})}, 1 \right\}, \quad (3.16)$$

em que, y_{t-1} indica a observação $t-1$, e $w_t^{-1} = \phi_t$.

e) Se $k < a$, não rejeite a amostra proposta e defina $(s_t, z_t) = (s_t^*, z_t^*)$, caso contrário, mantenha a amostra atual (s_t, z_t) .

▪ *Passo 6:*(Apenas para o caso Normal) Atualize o vetor \mathbf{W} , por meio do algoritmo de Metropolis-Hastings, seguindo passos:

- Gere a amostra w_t^* da distribuição proposta $N(w_t, 100)$
- Gere um número aleatório k de uma distribuição Uniforme de parâmetros $(0, 1)$.
- Dada a amostra w_t obtida na iteração anterior, calcule probabilidade de aceitação da amostra proposta w_t^* :

$$a = \min \left\{ \frac{\pi(w_t^* | y_t, z_t, s_t, w_t, \lambda, \alpha, \tau) f_w(w_t | w_t^*, 100)}{\pi(w_t | y_t, z_t, s_t, w_t, \lambda, \alpha, \tau) f_w(w_t^* | w_t, 100)}, 1 \right\},$$

em que, $f_w(\cdot)$ é a função de densidade de probabilidade da distribuição Normal, e

$\pi(w_t | y_t, z_t, s_t, w_t, \lambda, \alpha, \tau)$ pode ser obtida a seguir:

$$\pi(w_t | y_t, z_t, s_t, w_t, \lambda, \alpha, \sigma^2) \propto \left(\frac{\exp(w_t)}{1 + \exp(w_t)} \right)^{s_t} \left(\frac{1}{1 + \exp(w_t)} \right)^{y_t - 1 - s_t} \exp\left(\frac{-\tau(w_t - \alpha)^2}{2} \right).$$

- Se $k < a$, não rejeite a amostra proposta e defina $w_t = w_t^*$, caso contrário, mantenha a amostra atual w_t .

▪ *Passo 7:* (Apenas para o caso t) Atualize ν , por meio do algoritmo de Metropolis-Hastings, seguindo os passos:

- Gere um candidato ν_t^* da distribuição proposta $\text{LogNormal}(\log(\nu_t), 100)$.
- Gere um número aleatório k de uma distribuição $U(0, 1)$.
- Dada a amostra ν_t obtida na iteração anterior, calcule probabilidade de aceitação da amostra proposta ν_t^* :

$$a = \min \left\{ \frac{\pi(\nu^* | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \kappa) f(\nu | \nu^*, 100)}{\pi(\nu | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \kappa) f(\nu^* | \nu, 100)}, 1 \right\},$$

em que, $f(\nu)$ é uma função de densidade de probabilidade da distribuição log normal, e

$\pi(\nu_t | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \kappa)$ pode ser obtida a seguir:

$$\pi(\nu | \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{u}, \lambda, \alpha, \kappa) \propto \prod_{t=2}^n \left\{ \frac{\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} u_t^{\frac{\nu}{2}-1} \exp\left(\frac{\nu}{2} u_t - \kappa \nu \right) \right\}.$$

- Se $k < a$, não rejeite a amostra proposta e defina $\nu_t = \nu_t^*$, caso contrário, mantenha a amostra atual ν_t .

3.2.3 Modelo RCZINAR

Considere y_1, \dots, y_n realizações do processo RCZINAR(1), apresentado na Equação (3.1). Como para o processo RCINAR(1), vamos analisar dois casos, quando $W_t \sim N(\alpha, \sigma^2)$ e $W_t \sim t_\nu(\alpha, \sigma^2)$, com vetores de parâmetros $\theta = (\lambda, \rho, \alpha, \sigma^2)$ e $\theta = (\lambda, \rho, \alpha, \sigma^2, \nu)$, respectivamente.

A seguir, as respectivas Prioris para os parâmetros:

$$\pi(\lambda) \sim \text{Gamma}(A_\lambda, B_\lambda).$$

$$\pi(\rho) \sim U(0, 1).$$

$$\pi(\alpha) \sim N(\mu_\alpha, \sigma_\alpha^2).$$

$$\pi(\tau) \sim \text{Gamma}(A_\tau, B_\tau).$$

Para o caso t, também temos às seguintes Prioris:

$$\pi(\kappa) \sim U(0.01, 0.5).$$

$$\pi(\nu \mid \kappa) \sim \text{Exp}(\kappa).$$

3.2.4 Algoritmo MCMC

Considerando que Z_t é uma ZIP(ρ, λ), como apresentada na Equação (2.3). Além da variável latente, descrita na Equação (3.3), propomos a variável latente V_t , definida por:

$$V_t \mid \rho \sim \text{Ber}(\rho), \tag{3.17}$$

em que, $\text{Ber}(\rho)$, denota a distribuição Bernoulli com parâmetro ρ e,

$$Z_t \mid (V_t = 0, \lambda) \sim \text{Pois}(\lambda),$$

$$Z_t \mid V_t = 1 \text{ segue uma distribuição degenerada em zero.}$$

Para o caso t, as variáveis latentes utilizadas são S_t, U_t, V_t , já descritas em (3.3), (3.4), (3.17) respectivamente.

A seguir, apresentamos o algoritmo MCMC:

- *Passo 1:* Gere λ , a partir da distribuição condicional completa:

$$\lambda \mid \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{v}, \rho, \alpha, \tau \sim \text{Gamma} \left(\sum_{t=2}^n z_t (1 - v_t) + A_\lambda; \sum_{t=2}^n (1 - v_t) + B_\lambda \right).$$

- *Passo 2:* Gere ρ , a partir da distribuição condicional completa:

$$\rho \mid \mathbf{y}, \mathbf{s}, \mathbf{z}, \mathbf{w}, \mathbf{v}, \alpha, \tau, \lambda \sim \text{Beta} \left(\sum_{t=2}^n v_t + 1; \sum_{t=2}^n (1 - v_t) + 1 \right).$$

- *Passo 3:* Gere α , a partir da sua condicional completa dada em (3.11) para o caso normal, ou (3.12) para o caso t.

- *Passo 4:* Gere τ , a partir da sua condicional completa dada em (3.13) para o caso normal, ou (3.14) para o caso t.

- *Passo 5:* (Apenas para o caso t) Gere U_t , a partir da sua condicional completa dada em (3.15).

- *Passo 6:* Atualize o vetor latente \mathbf{S} , \mathbf{Z} , e \mathbf{V} , por meio do algoritmo Metropolis-Hastings, seguindo os passos:

a) Se $y_{t-1} = 0$, então $s_t^* = 0$. Caso contrário, gere s_t^* pela distribuição Binomial(y_{t-1}, ϕ_t).

b) Faça $z_t^* = y_t - s_t^*$.

c) Gere v_t^* da distribuição condicional completa $\pi(v_t \mid z_t, y_{t-1}, \alpha, \rho, \lambda)$, dada por:

$$V_t \mid Z_t = z_t, \rho, \lambda \sim \text{Bern} \left(\frac{\rho}{\rho + (1 - \rho) \exp(-\lambda)} \right), \text{ quando } z_t = 0,$$

uma degenerada em zero, quando $z_t > 0$.

d) Gere um número aleatório k , de uma distribuição $U(0, 1)$.

e) Dada a amostra (s_t, z_t, v_t) obtida na iteração anterior, calcule a probabilidade de aceitação da amostra proposta (s_t^*, z_t^*, v_t^*) :

$$a = \min \left\{ \frac{\pi(s_t^*, z_t^* \mid y_{t-1}, w_t^{-1}, \theta) \pi(s_t^{(j-1)} \mid y_{t-1}, w_t^{-1}) \pi(v_t^{(j-1)} \mid z_t^{(j-1)}, \rho, \lambda)}{\pi(s_t^{(j-1)}, z_t^{(j-1)} \mid y_{t-1}, w_t^{-1}, \theta) \pi(s_t^* \mid y_{t-1}, w_t^{-1}) \pi(v_t^* \mid z_t^*, \rho, \lambda)}, 1 \right\},$$

f) Se $k < a$ não rejeite a amostra proposta e defina $(s_t, z_t, v_t) = (s_t^*, z_t^*, v_t^*)$, caso contrário, $(s_t, z_t, v_t) = (s_t^{j-1}, z_t^{j-1}, v_t^{j-1})$.

- *Passo 7:* (Apenas para o caso Normal) Atualize o vetor \mathbf{W} , por meio do algoritmo de Metropolis-Hastings, seguindo os mesmos passos do modelo RCINAR, definido em (3.2.2).

- *Passo 8:* (Apenas para o caso t) Atualize ν , por meio do algoritmo de Metropolis-Hastings, seguindo os mesmos passos do modelo RCINAR, definido na Subseção 3.2.2.

3.3 PREVISÃO

Um dos interesses em estudar séries temporais é realizar previsões. No capítulo anterior, vimos algumas propriedades do modelo RCINAR(1), e nas Seções anteriores deste Capítulo, apresentamos o modelo com inflação em zeros, definido por RCZINAR(1) e também o algoritmo para estimação dos seus parâmetros. Nesta Seção propomos uma alternativa para o cálculo da previsão.

A esperança condicional, $E(Y_{t+h}|Y_t = y_t)$, é uma previsão pontual que minimiza o erro quadrático médio, como evidenciado por Jazi, Jones e Lai (2012).

Garay et al. (2022), apresentam o modelo INAR(1) com inovações inflacionadas em zeros, desenvolvem um algoritmo para estimação e, propõem duas alternativas Bootstrap para construção de intervalo de confiança para os parâmetros e previsão, sendo elas: Abordagem de bootstrap paramétrico e abordagem de bootstrap de blocos móveis. Bertail et al. (2024) realizam um estudo sobre o modelo INAR(p) com inovações inflacionadas em zeros, apresentam um método para estimação dos parâmetros e, desenvolvem um método de bootstrap regenerativo para construção de intervalos de confiança para os parâmetros e para estimar as distribuições preditivas.

Nesta dissertação usaremos o Bootstrap em Blocos para o cálculo da distribuição de predição. Essa abordagem preserva a estrutura de dependência temporal dos dados. A seguir o algoritmo para o Bootstrap em Blocos Móveis para previsão:

3.3.1 Abordagem Bootstrap em Blocos Móveis

O algoritmo de Bootstrap em Blocos Móveis consiste em dividir a série temporal original em blocos sobrepostos, preservando a dependência temporal dentro de cada bloco, e gerar novas amostras por meio de amostragem com reposição desses blocos. Para cada amostra bootstrap, os parâmetros do modelo são estimados utilizando o método MCMC, sendo posteriormente empregados para simular valores futuros de forma iterativa, em conformidade com a estrutura do modelo RCZINAR(1). Por fim, a distribuição empírica das simulações é utilizada como uma aproximação da distribuição preditiva. Os passos do algoritmo são descritos a seguir:

- *Passo 1:* Escolha um tamanho k para os blocos (usualmente $k = \sqrt{n}$, tal que k seja inteiro), e defina blocos sobrepostos do tipo: $B_1 = (y_1, \dots, y_k)$, $B_2 = (y_2, \dots, y_{k+1})$, ..., $B_{n-k+1} = (y_{n-k+1}, \dots, y_n)$

▪ *Passo 2:* Gere com reposição $(n/k) + 1$ blocos, que serão unidos e caso necessário truncados no final, para formarem uma nova série temporal de tamanho n .

▪ *Passo 3:* Estime os parâmetros de interesse θ , utilizando o algoritmo MCMC descrito na seção anterior para cada amostra bootstrap. Denotaremos os parâmetros por $\hat{\theta}^b = (\hat{\lambda}^b, \hat{\rho}^b, \hat{\alpha}^b, \hat{\sigma}^{2b})$, para $b = 1, 2, \dots, B$.

▪ *Passo 4:* Para $b = 1, 2, \dots, B$, repita M vezes os seguintes passos:

a) Gere $W_{n+h}^{(m)}$ e, depois faça $\phi_{n+h}^{(m)} = \exp\{W_{n+h}^{(m)}\}/(1 + \exp\{W_{n+h}^{(m)}\})$.

b) Gere $s_{n+h}^{(m)}$, tal que:

se $y_{n+h}^{(m)} = 0$ então $s_{n+h}^{(m)} = 0$. Caso contrário, gere $s_{n+h}^{(m)} \sim \text{Binomial}(y_{n+h-1}, \phi_{n+h}^{(m)})$.

c) Gere $z_{n+h}^{(m)} \sim \text{Pois}(\lambda)$.

d) E então calcule $y_{n+h}^{(m)} = s_{n+h}^{(m)} + z_{n+h}^{(m)}$.

▪ *Passo 5:* Para $b = 1, 2, \dots, B$, calcule:

$$\hat{p}_{n+h}^b(j) = \frac{\#\{m, y_{n+h}^{(m)} = j\}}{M}, \text{ para } h \geq 1, \text{ e } m = 1 \dots, M.$$

Utilize a distribuição empírica $\hat{p}_{n+h}^b(j)$, como a distribuição y_{n+h} , para $h \geq 1$.

3.4 FATOR DE BAYES

Nos Capítulos e Seções anteriores, abordamos os modelos RCINAR e RCZINAR, incluindo a estimativa de seus respectivos parâmetros através de inferência Bayesiana, além de discutir a predição utilizando o método Bootstrap. No entanto, também é importante determinar qual modelo é mais adequado para o conjunto de dados. Para isso, utilizaremos o critério de comparação do Fator de Bayes, a mesma proposta usada por Garay et al. (2020), que é uma técnica empregada na Inferência Bayesiana para essa finalidade. Assumindo que os dois modelos sejam igualmente prováveis, o Fator de Bayes é calculado como a razão das densidades marginais.

A densidade marginal para cada modelo M_k , pode ser definida como:

$$f(\mathbf{y} | M_k) = \int f(\mathbf{y} | \theta_k, M_k) \pi(\theta_k | M_k) d\theta_k, \text{ para } k = 1, 2.$$

em que θ_k são os parâmetros do modelo, $f(\mathbf{y} | \theta_k, M_k)$ é a função de verossimilhança, e $\pi(\theta_k | M_k)$ são as distribuições à priori dos parâmetros do modelo.

O Fator de Bayes maior que 1 indica que o modelo 1 modela melhor os dados, enquanto o Fator de Bayes menor que 1 sugere que o modelo 2 é o mais adequado.

4 ESTUDO DE SIMULAÇÃO

Neste capítulo, apresentamos um estudo de simulação que tem como objetivo analisar a precisão das estimativas realizadas via Inferência Bayesiana, utilizando o algoritmo Gibbs com auxílio do Metropolis Hastings (apresentado no Capítulo 3) para obter as estimativas dos parâmetros dos processos RCINAR(1) e RCZINAR(1).

4.1 CASO RCINAR(1)

A primeira parte do estudo visou estudar o comportamento do viés em dois cenários, analisando como os parâmetros se comportam quando estimados em conjunto e quando estimados individualmente. No primeiro caso, representado na Figura 1, estimamos um parâmetro enquanto mantinhamos os outros dois fixos. Especificamente, estimamos λ , e fixamos α e σ^2 , repetindo o processo para cada um dos outros parâmetros. No segundo caso, representado na Figura 2, todos os parâmetros foram estimados simultaneamente.

Ambos os cenários representam o modelo RCINAR(1), com $\lambda = 2, \alpha = -1.5, \sigma^2 = 5$. Realizamos um estudo de 300 réplicas Monte Carlo para tamanhos amostrais $n \in \{100, 200, 300, 500, 600, 800\}$.

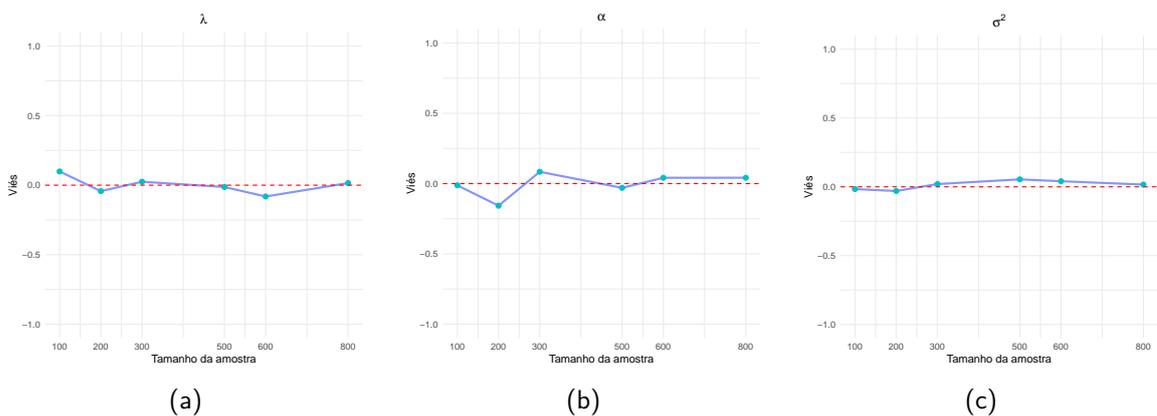


Figura 1 – Viés dos estimadores dos parâmetros do modelo RCINAR, para cada parâmetro estimado individualmente, para $\lambda = 2, \alpha = -1.5$ e $\sigma^2 = 5$.

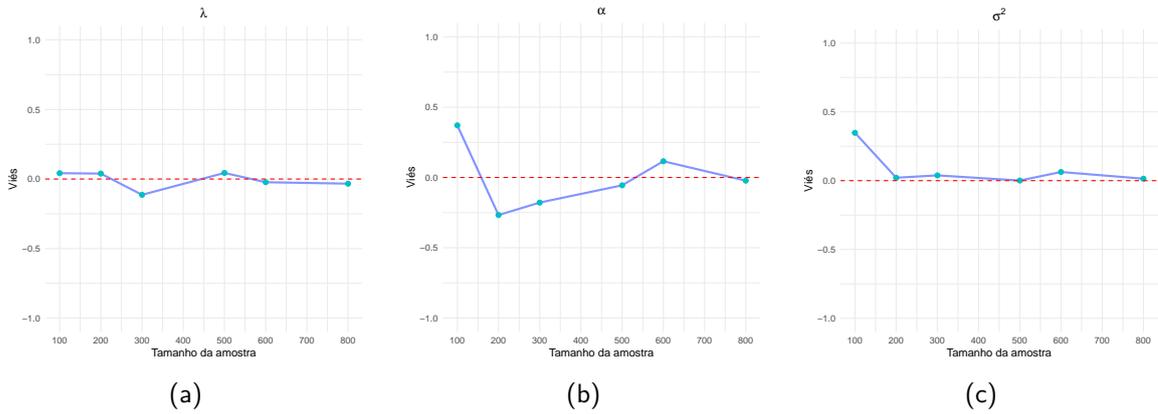


Figura 2 – Viés dos estimadores dos parâmetros do modelo RCINAR, estimados simultaneamente, para $\lambda = 2$, $\alpha = -1.5$ e $\sigma^2 = 5$.

Na Figura 1a - 1c, o viés permanece próximo de zero para a maioria dos tamanhos amostrais, indicando boa precisão na estimativa.

Em relação a Figura 2a ainda apresenta um bom comportamento, mantendo o viés próximo ao zero e, diferente do cenário anterior, 2b apresenta um viés mais acentuado, mas que tende a reduzir para tamanhos amostrais maiores. 2c também apresenta um bom comportamento, com viés apresentando valores próximos a zero.

A segunda parte do estudo tem como objetivo avaliar a precisão das estimativas dos parâmetros analisando o viés e o erro quadrático médio (EQM) para diferentes tamanhos amostrais. Foram consideradas amostras de tamanho $n \in \{100, 200, 300\}$. Fixamos $\alpha = \{-1.5, -0.75, 0.75, 1.5\}$, $\sigma^2 = \{2.5, 5\}$ e para o processo RCINAR(1) fixamos $\lambda = 2$.

Após as 300 réplicas serem geradas para cada cenário, foi analisada a precisão das estimativas dos parâmetros dos processos.

Tabela 1 – Viés e EQM das estimativas dos parâmetros do processo RCINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 5$

σ^2	α	n	$\hat{\lambda}$		$\hat{\alpha}$		$\hat{\sigma}^2$	
			Viés	EQM	Viés	EQM	Viés	EQM
5	-1.5	100	0.0420	0.0018	0.3709	0.1383	0.3475	0.1209
		200	0.0389	0.0797	-0.2674	0.2957	0.0212	0.0027
		300	-0.1136	0.0130	-0.1788	0.0329	0.0382	0.0015
	-0.75	100	0.2417	0.0605	-0.0239	0.0345	0.1556	0.0270
		200	0.0248	0.0162	-0.1488	0.2417	0.0961	0.0218
		300	0.0933	0.0144	-0.3748	0.1464	0.0656	0.0078
	0.75	100	0.1536	0.0331	-0.1319	0.3485	0.1924	0.0575
		200	0.1456	0.0811	-0.1397	0.1789	0.1129	0.0142
		300	0.1723	0.0304	-0.2541	0.1445	-0.1120	0.0127
	1.5	100	0.1741	0.0647	0.1118	0.0527	0.1510	0.0670
		200	0.0848	0.0700	-0.0019	0.0584	0.1546	0.0326
		300	0.2481	0.0715	-0.3016	0.1568	0.1738	0.0466

Tabela 2 – Viés e EQM das estimativas dos parâmetros do processo RCINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 2.5$

σ^2	α	n	$\hat{\lambda}$		$\hat{\alpha}$		$\hat{\sigma}^2$	
			Viés	EQM	Viés	EQM	Viés	EQM
2.5	-1.5	100	-0.2235	0.0512	0.9260	0.8654	-0.0546	0.0030
		200	0.3187	0.1017	-0.7621	0.5839	0.0579	0.0034
		300	0.1355	0.0185	-0.0740	0.0063	0.0595	0.0036
	-0.75	100	0.7786	0.6077	-0.9926	1.0012	0.0928	0.0087
		200	0.2580	0.0667	-0.2098	0.0446	0.0175	0.0003
		300	-0.0519	0.0027	-0.0101	0.0004	0.0595	0.0036
	0.75	100	-0.1573	0.0249	0.2573	0.0666	-0.0345	0.0012
		200	0.1746	0.0306	-0.2912	0.0849	0.0608	0.0037
		300	-0.1041	0.0108	0.1337	0.0179	0.0229	0.0005
	1.5	100	-0.1615	0.0261	-0.1840	0.0339	0.0746	0.0056
		200	0.3913	0.1532	-0.2019	0.0408	0.1785	0.0319
		300	-0.2836	0.08045	0.1675	0.0281	0.0368	0.0014

Como podemos observar nas Tabelas 4 e 2, o viés das estimativas dos parâmetros do modelo RCINAR(1) apresenta valores próximos de zero e tende a diminuir ainda mais conforme o tamanho da amostra aumenta.. Em relação ao erro quadrático médio (EQM), os valores são baixos e, na maioria dos cenários, o EQM tende a diminuir conforme o tamanho da amostra aumenta. Portanto, podemos concluir que as estimativas dos parâmetros se comportam conforme o esperado, ou seja, estão próximas aos valores reais dos parâmetros.

4.2 CASO RCZINAR(1)

Para o processo RCZINAR(1), analisamos a precisão das estimativas dos parâmetros por meio do viés e do erro quadrático médio (EQM) para diferentes tamanhos amostrais. Foram consideradas amostras de tamanho $n \in \{100, 200, 300\}$. Fixamos $\alpha = \{-1.5, -0.75, 0.75, 1.5\}$, $\sigma^2 = \{2.5, 5\}$, $\lambda = 2$ e $\rho = 0.7$. Após a geração de 300 réplicas para cada cenário, foi analisada a precisão das estimativas dos parâmetros do processo.

Tabela 3 – Viés e EQM das estimativas dos parâmetros do processo RCZINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 2.5$

σ^2	α	n	$\hat{\lambda}$		$\hat{\rho}$		$\hat{\alpha}$		$\hat{\sigma}^2$	
			Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
2.5	-1.5	100	0.2729	0.0746	0.0758	0.0058	0.5276	0.2888	-1.9361	3.7938
		200	0.1470	0.0221	0.0232	0.0006	0.1142	0.0486	0.6088	3.1434
		300	-0.2741	0.0751	-0.0140	0.0002	0.0951	0.0162	0.0035	0.6884
	-0.75	100	-0.2080	0.0433	0.0291	0.0009	-0.1408	0.0271	-2.0442	4.2057
		200	0.1154	0.0134	-0.0269	0.0007	0.0140	0.0009	3.6371	13.8595
		300	-0.0092	0.0002	-0.0040	0.0000	0.1705	0.0328	-1.3233	2.0025
	0.75	100	-0.2988	0.0893	-0.0969	0.0094	0.2137	0.0462	1.8496	3.4887
		200	-0.0181	0.0004	-0.0283	0.0008	-0.3553	0.1274	-0.6987	0.6561
		300	-0.2960	0.0877	-0.0247	0.0006	0.2316	0.0550	0.4345	0.2908
	1.5	100	-0.2711	0.0735	-0.1196	0.0143	0.3433	0.1182	3.1386	9.8912
		200	-0.1935	0.0376	-0.0250	0.0006	-0.3019	0.0927	-0.1191	0.0861
		300	0.0487	0.0025	-0.0019	0.0000	0.1052	0.0120	1.8448	3.4752

Tabela 4 – Viés e EQM das estimativas dos parâmetros do processo RCZINAR(1), para diferentes tamanhos de amostra, fixando $\lambda = 2$, $\alpha = (-1.5, -0.75, 0.75, 1.5)$, $\sigma^2 = 5$

σ^2	α	n	$\hat{\lambda}$		$\hat{\rho}$		$\hat{\alpha}$		$\hat{\sigma}^2$	
			Viés	EQM	Viés	EQM	Viés	EQM	Viés	EQM
5	-1.5	100	0.5739	0.3294	-0.0585	0.0034	0.9736	0.9499	-2.6259	7.0684
		200	0.4092	0.1676	0.0224	0.0005	0.6007	0.3618	-3.2363	10.5931
		300	-0.4576	0.2094	-0.0763	0.0058	0.2051	0.0456	3.1879	11.2714
	-0.75	100	-0.1927	0.0374	0.0694	0.0048	0.1534	0.0540	-4.4112	19.7081
		200	0.6114	0.3739	0.0102	0.0001	0.0762	0.0067	-3.9333	15.5206
		300	-0.4515	0.2039	-0.0816	0.0067	0.1306	0.0175	-2.2210	5.2006
	0.75	100	-0.2371	0.0562	0.0579	0.0034	0.2919	0.0857	2.6749	7.8990
		200	0.1848	0.0342	-0.0329	0.0011	-0.3342	0.1118	-2.8843	8.3377
		300	-0.1845	0.0340	-0.0073	0.0001	0.3907	0.1528	6.9065	47.8547
	1.5	100	0.1819	0.0331	-0.0229	0.0005	-0.3305	0.1096	1.0484	1.1645
		200	0.2057	0.0423	-0.0301	0.0009	-0.3110	0.0969	-0.4578	0.2217
		300	-0.0058	0.0001	-0.0418	0.0018	0.3909	0.1542	3.0346	9.4051

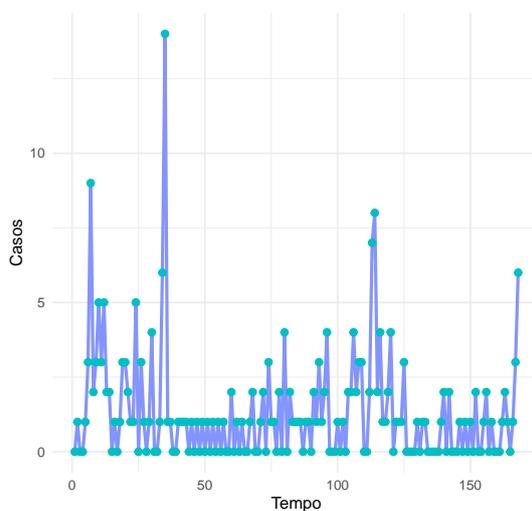
5 RESULTADOS E APLICAÇÕES

Para ilustrar as metodologias apresentadas nesta dissertação, aplicamos essas técnicas a um conjunto de dados reais. Consideramos o conjunto de dados sobre o número mensal de casos de poliomielite nos EUA, registrado entre 1970 e 1983. Originalmente, o banco de dados contém 168 observações, mas utilizamos apenas as primeiras 166 para estimar os parâmetros, reservando as duas últimas para fins de comparação no estudo de previsões. Esse conjunto foi analisado e disponibilizado inicialmente por Zeger (1988) e, conforme mencionado por Zheng, Basawa e Datta (2006), tornou-se um exemplo para modelagem de dados de contagem. Zheng, Basawa e Datta (2006) utilizam seis modelos diferentes, todos com abordagem clássica, incluindo o modelo RCINAR(1) como descrito no capítulo anterior.

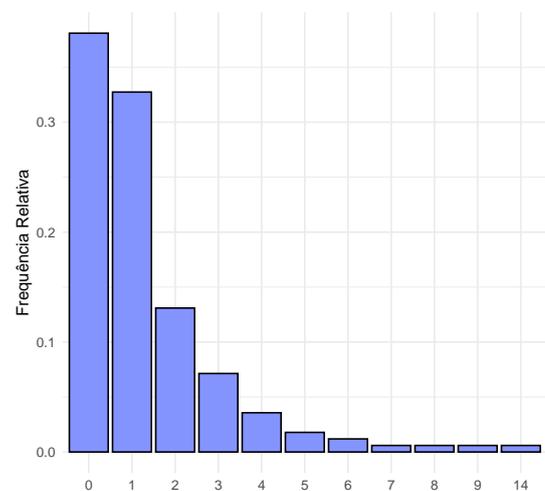
Na Tabela 5 são apresentadas algumas estatísticas descritivas da série. Podemos observar que o conjunto de dados apresenta uma porcentagem de 38% de zeros, com observações variando entre zero e quatorze. Os gráficos 3a e 3b auxiliam na visualização dos dados e podemos observar que as observações possuem uma grande quantidade de valores zeros e uns. Além disso, temos a Figura 4, que apresenta a função de autocorrelação (ACF) e a função de autocorrelação parcial (FACP), que sugerem um modelo autorregressivo de primeira ordem.

Tabela 5 – Estatísticas Descritivas

n	Mínimo	Média	Mediana	Máximo	Variância	% de zeros
168	0	1.33	1	14	3.50	38



(a)



(b)

Figura 3 – Número mensal de casos de Poliomielite nos EUA, registrados entre 1970 e 1983.

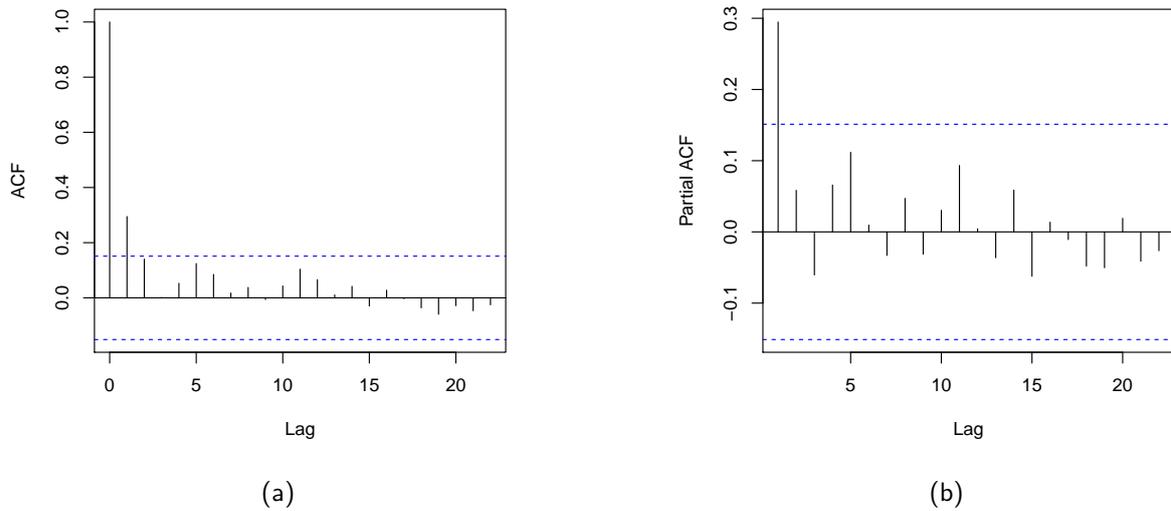


Figura 4 – ACF e PACF do conjunto de dados Poliomielite.

Na Tabela 6, são apresentadas as estimativas dos parâmetros utilizando as metodologias discutidas anteriormente, e estatísticas como o desvio padrão e o intervalo de credibilidade simétrico de 95% de confiança. Para cada modelo, foram realizadas 250000 iterações, com um período *burn-in* de 20% e *thin* de 2. Sendo assim, teremos 100000 realizações finais da distribuição a posteriori, e elas serão usadas para o cálculo da Inferência Bayesiana.

Tabela 6 – Tabela de resumo MCMC para os modelos INAR(1), ZINAR(1), RCINAR(1), RCZINAR(1) e RCINAR_T(1).

Modelo	Parâmetro	Média	DP	Mediana	$Q_{0.025}$	$Q_{0.975}$
INAR(1)	α	0.142	0.046	0.141	0.054	0.235
	λ	1.071	0.096	1.069	0.889	1.266
ZINAR(1)	α	0.133	0.049	0.132	0.040	0.231
	λ	1.440	0.165	1.436	1.128	1.775
	ρ	0.243	0.069	0.245	0.102	0.373
RCINAR _N (1)	α	-3.071	0.790	-3.096	-4.552	-1.444
	σ^2	18.868	6.619	18.344	7.306	34.000
	λ	1.018	0.090	1.015	0.849	1.199
RCINAR _T (1)	α	-2.6358	0.2372	-2.7806	-2.8890	-2.3025
	σ^2	0.0051	0.0049	0.0036	0.0005	0.0194
	λ	1.3420	0.0896	1.3398	1.1725	1.5227
	ν	3.0140	2.5519	2.3136	1.0369	9.1027
RCZINAR(1)	α	-3.172	0.829	-3.161	-4.776	-1.591
	σ^2	16.200	5.727	15.931	5.821	28.336
	λ	1.346	0.163	1.340	1.044	1.681
	ρ	0.229	0.073	0.232	0.078	0.367

Para esta aplicação, consideramos o modelo INAR(1) como M_1 , ZINAR(1) como M_2 , RCINAR_N(1) considerando $W_t \sim N(\alpha, \sigma^2)$ como M_3 , RCZINAR(1) considerando $W_t \sim N(\alpha, \sigma^2)$ como M_4 e, RCINAR_T(1) considerando $W_t \sim t_\nu(\alpha, \sigma^2)$ como M_5 . A Tabela 7 apresenta as densidades marginais estimadas para os respectivos modelos.

Tabela 7 – Densidade marginal estimada para os modelos estudados.

Modelo (M_k)	$f(y M_k)$
INAR(1)	$5.677908 \times e^{-129}$
ZINAR(1)	$7.314772 \times e^{-149}$
RCINAR _N (1)	$3.062886 \times e^{-122}$
RCZINAR _N (1)	$9.175483 \times e^{-129}$
RCINAR _T (1)	$2.5348450 \times e^{-139}$

O fator de Bayes do modelo M_3 , ou seja, RCINAR_N(1), em relação aos demais é: 5.39×10^6 em relação ao modelo M_1 , 4.19×10^{26} em relação ao modelo M_2 , 3.34×10^6 em relação ao modelo M_4 e 1.21×10^{17} em relação ao modelo M_5 . Isso indica que, em relação aos outros

modelos o RCINAR_N(1) é o mais indicado para a modelagem desse conjunto de dados.

As Figuras (5)-(7) apresentam o histórico das cadeias MCMC, bem como as densidades marginais a posteriori aproximadas dos parâmetros, considerando o modelo RCINAR(1).

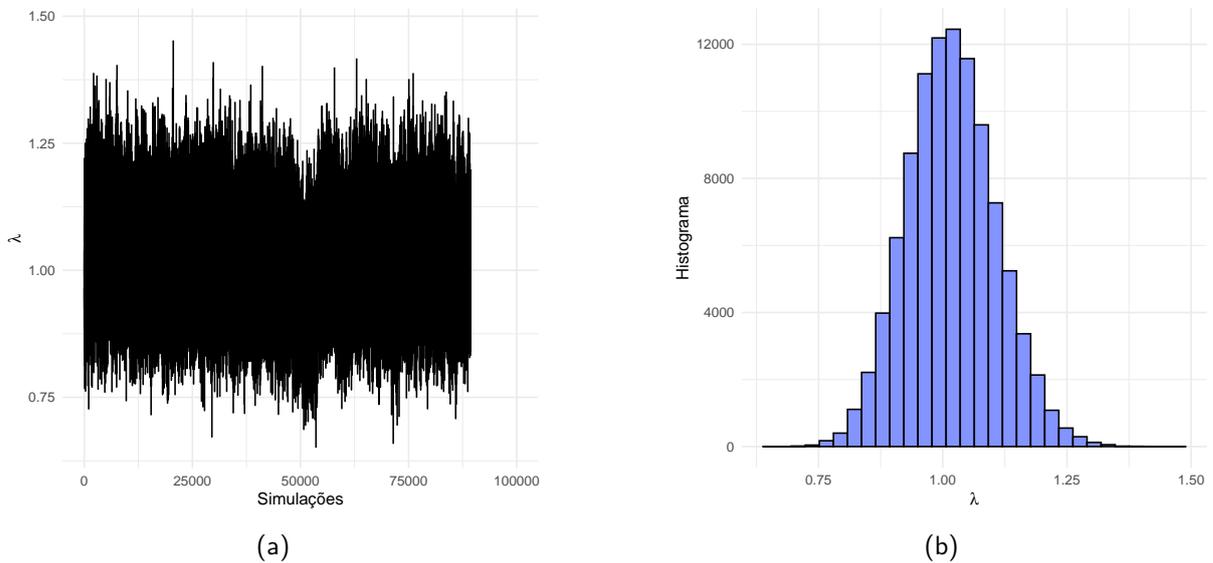


Figura 5 – Histórico das cadeias e densidade posterior marginal aproximada de λ , considerando o modelo RCINAR(1)

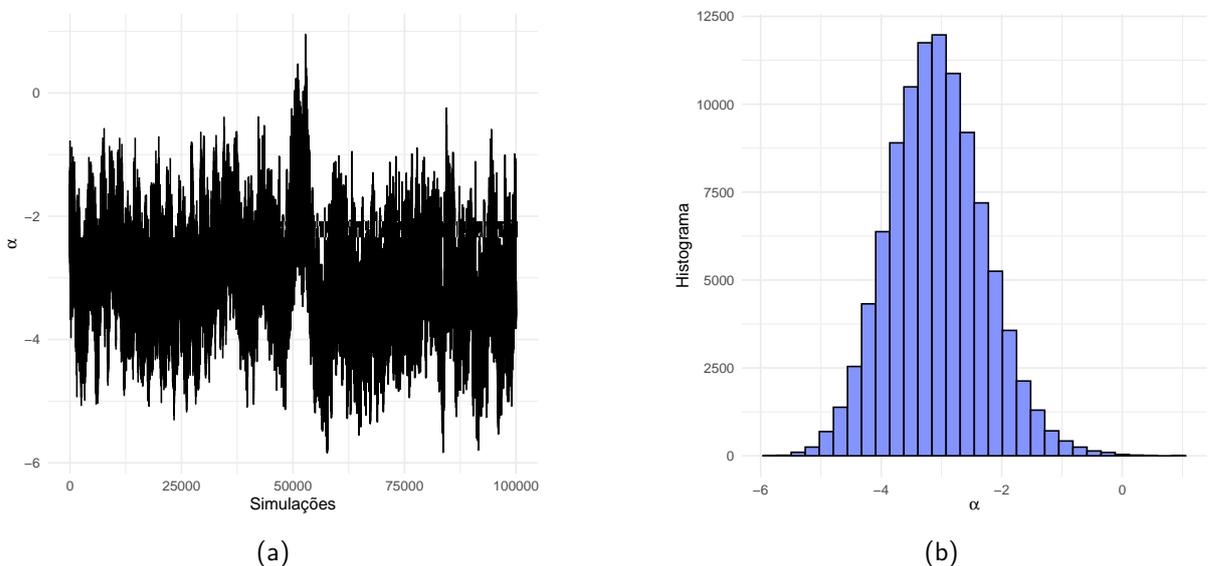


Figura 6 – Histórico das cadeias e densidade posterior marginal aproximada de α , considerando o modelo RCINAR(1)

De forma subsequente, foi obtida a distribuição preditiva, utilizando a estratégia apresentada na Seção 3.3 para as duas últimas observações do conjunto de dados Polio, utilizando o modelo RCINAR(1). A Figura 8 apresenta as distribuições preditivas, enquanto a Tabela 8

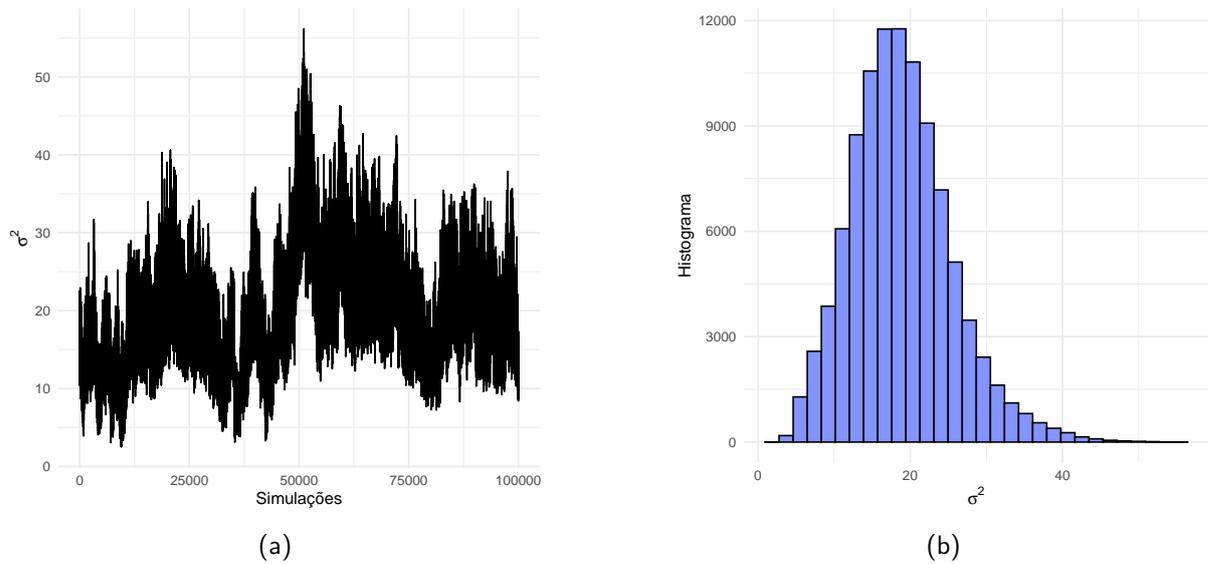


Figura 7 – Histórico das cadeias e densidade posterior marginal aproximada de σ^2 , considerando o modelo RCINAR(1)

fornece um resumo das estatísticas principais, incluindo a média, o desvio padrão, bem como os quantis $Q_{2.5\%}$ e $Q_{97.5\%}$.

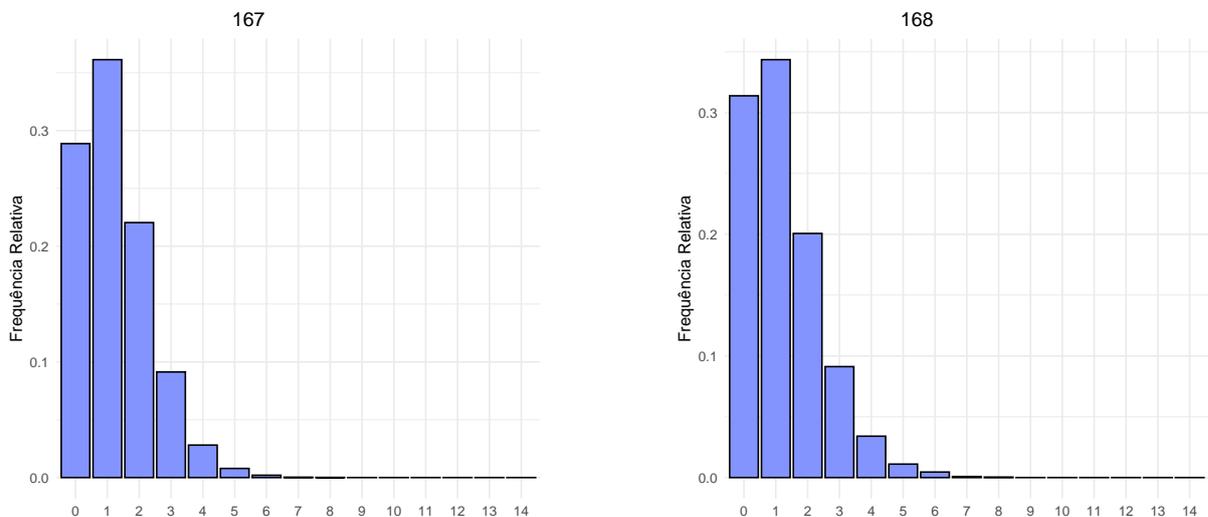


Figura 8 – Gráfico de barras da distribuição de previsão das observações 167 e 168, do modelo RCINAR(1)

Os valores reais das observações #167 e #168 são respectivamente 0 e 1. Na Figura (8), é possível observar que o valor mais provável para ambas as observações é 1. Para observação #167 a probabilidade de que ela assuma o valor 1 é de aproximadamente 36% enquanto a probabilidade de que o valor seja 0, é cerca de 28%. Embora o valor mais provável seja 1, a probabilidade do valor 0 também é considerável, e ao analisarmos o intervalo de confiança, a probabilidade da observação #167 ser zero, pode ser bem alta, isto é, com 95% de confiança,

Tabela 8 – Estatísticas da distribuição preditiva das observações #167 e #168, no modelo RCINAR. As linhas a partir de $j=9$ possuem todos os valores iguais a zero.

j	#167				#168			
	Média	SD	Q _{0.025}	Q _{0.975}	Média	SD	Q _{0.025}	Q _{0.075}
0	0.2887	0.0839	0.1452	0.4676	0.3138	0.0760	0.2008	0.4626
1	0.3613	0.0398	0.2886	0.5271	0.3436	0.0368	0.2737	0.4147
2	0.2204	0.0486	0.1254	0.4099	0.2007	0.0348	0.1357	0.2666
3	0.0913	0.0404	0.0225	0.2663	0.0912	0.0365	0.0273	0.1661
4	0.0281	0.0167	0.0050	0.1735	0.0340	0.0188	0.0050	0.0680
5	0.0079	0.0091	0.0000	0.1328	0.0111	0.0103	0.0000	0.0354
6	0.0020	0.0034	0.0000	0.1101	0.0045	0.0050	0.0000	0.0151
7	0.0003	0.0011	0.0000	0.1050	0.0008	0.0026	0.0000	0.0077
8	0.0001	0.0006	0.0000	0.1000	0.0003	0.0013	0.0000	0.0050

podemos afirmar que a média preditiva da observação #167 ser zero, está entre 14% e 46%. Por outro lado, para a observação #168, o valor mais provável coincide com o valor real, apresentando em média uma chance de 31%.

Com o intuito de verificar a metodologia proposta, implementamos também a distribuição preditiva para o modelo RCZINAR. Os resultados da tabela e do gráfico se encontram no Apêndice 6. Se observarmos os intervalos de confiança de 95%, podemos ver que são similares aos intervalos da distribuição preditiva do modelo RCINAR.

6 CONSIDERAÇÕES FINAIS

Nesta dissertação, estudamos o modelo RCINAR(1) com inovação Poisson e o modelo RCZINAR(1) com inovação Poisson inflacionada em zeros. Para ambos os modelos usamos a função de transformação *logito* para o coeficiente. Utilizamos uma abordagem Bayesiana para a estimação, propondo um algoritmo de Gibbs com auxílio do Metropolis-Hastings. Além disso, o método de Bootstrap em blocos para a distribuição aproximada da previsão também foi abordado.

Os resultados obtidos no estudo de simulação para o modelo RCINAR(1) indicam que o viés das estimativas dos parâmetros é pequeno e tende a se aproximar de zero conforme o tamanho da amostra aumenta. Além disso, na maioria dos cenários, o Erro Quadrático Médio diminuiu, sugerindo que a estimação foi eficaz. Já na aplicação a dados reais, de acordo com o fator de Bayes, o modelo RCINAR(1) apresentou um melhor ajuste.

Para trabalhos futuros, pretendemos, em primeiro lugar, desenvolver pacotes estatísticos na linguagem R Core Team (2024), possibilitando que outros pesquisadores utilizem as metodologias apresentadas nesta dissertação. Além disso, exploraremos modelos que incorporem inovações pertencentes à classe inflacionada em zeros, o que pode oferecer uma compreensão mais abrangente das características dos dados. Outra direção de pesquisa será a extensão deste estudo, com a investigação da abordagem Bayesiana para o processo RCINAR(p).

REFERÊNCIAS

- AL-OSH, M. A.; ALZAID, A. A. FIRST-ORDER INTEGER-VALUED AUTOREGRESSIVE (INAR(1)) PROCESS. *Journal of Time Series Analysis*, v. 8, n. 3, p. 261–275, maio 1987. ISSN 0143-9782, 1467-9892. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.1987.tb00438.x>>.
- AL-OSH, M. A.; ALZAID, A. A. Binomial autoregressive moving average models. *Communications in Statistics. Stochastic Models*, v. 7, n. 2, p. 261–282, jan. 1991. ISSN 0882-0287. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/15326349108807188>>.
- AWALE, M.; KASHIKAR, A. S. A random coefficient integer autoregressive model of order one with seasonal structure (RCINAR(1)s). *Model Assisted Statistics and Applications*, v. 18, n. 1, p. 1–12, mar. 2023. ISSN 15741699, 18759068. Disponível em: <<https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/MAS-211333>>.
- BERTAIL, P.; GARAY, A. M.; MEDINA, F. L.; JALES, I. C. A maximum likelihood and regenerative bootstrap approach for estimation and forecasting of INAR(p) processes with zero-inflated innovations. *Statistics*, v. 58, n. 2, p. 336–363, mar. 2024. ISSN 0233-1888, 1029-4910. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/02331888.2024.2344670>>.
- CABRAL, C. R. B.; LACHOS, V. H.; MADRUGA, M. R. Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *Journal of Statistical Planning and Inference*, v. 142, n. 1, p. 181–200, jan. 2012. ISSN 03783758. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378375811002771>>.
- COSTA, V. J.; FERREIRA, M.; CORDEIRO, M. T. A. Análise de séries temporais climáticas. *Revista de Ciências Agroveterinárias*, v. 14, n. 2, p. 169–177, 2015.
- GAMERMAN, D.; LOPES, H. F. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. 2nd ed. Boca Raton: Taylor & Francis, 2006, p. 142. (Texts in statistical science series, 68). ISBN 9781584885870.
- GARAY, A. M.; BOLFARINE, H.; LACHOS, V. H.; CABRAL, C. R. Bayesian analysis of censored linear regression models with scale mixtures of normal distributions. *Journal of Applied Statistics*, v. 42, n. 12, p. 2694–2714, dez. 2015. ISSN 0266-4763, 1360-0532. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/02664763.2015.1048671>>.
- GARAY, A. M.; MEDINA, F. L.; CABRAL, C. R. B.; LIN, T.-I. Bayesian analysis of the p -order integer-valued AR process with zero-inflated Poisson innovations. *Journal of Statistical Computation and Simulation*, v. 90, n. 11, p. 1943–1964, jul. 2020. ISSN 0094-9655, 1563-5163. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00949655.2020.1754819>>.
- GARAY, A. M.; MEDINA, F. L.; S., I. J. C.; BERTAIL, P. First-Order Integer Valued AR Processes with Zero-Inflated Innovations. In: CHAARI, F.; LESKOW, J.; WYLOMANSKA, A.; ZIMROZ, R.; NAPOLITANO, A. (Ed.). *Nonstationary Systems: Theory and Applications*. Cham: Springer International Publishing, 2022. v. 18, p. 19–40. ISBN 9783030821913 9783030821104. Disponível em: <https://link.springer.com/10.1007/978-3-030-82110-4_2>.

GONG, K.; PENG, Y.; DENG, H.; TORKZADEH, R. Time series analysis of consumer conversion behavior for an online retail store. In: . [S.l.: s.n.], 2015.

HAMILTON, J. D. Maximum Likelihood Estimation. In: *Time series analysis*. Princeton, N.J: Princeton University Press, 1994. p. 117–149. ISBN 9780691218632.

JAZI, M. A.; JONES, G.; LAI, C.-D. First-order integer valued AR processes with zero inflated poisson innovations: ZERO INFLATED INAR(1). *Journal of Time Series Analysis*, v. 33, n. 6, p. 954–963, nov. 2012. ISSN 01439782. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.2012.00809.x>>.

MCKENZIE, E. Some Simple models for discrete variate time series. *Journal of the American Water Resources Association*, v. 21, n. 4, p. 645–650, ago. 1985. ISSN 1093-474X, 1752-1688. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.1985.tb05379.x>>.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2024. Disponível em: <<https://www.R-project.org/>>.

RISTIC, M. M.; BAKOUCH, H. S.; NASTIC, A. S. A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference*, v. 139, n. 7, p. 2218–2226, jul. 2009. ISSN 03783758. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S037837580800400X>>.

ROSS, S. *Stochastic Processes*. Wiley, 1996. (Wiley series in probability and mathematical statistics). ISBN 9780471120629. Disponível em: <<https://books.google.com.br/books?id=ImUPAQAAAMAJ>>.

SCHNECKENREITHER, G.; HERRMANN, L.; REISENHOFER, R.; POPPER, N.; GROHS, P. Assessing the heterogeneity in the transmission of infectious diseases from time series of epidemiological data. *PLOS ONE*, v. 18, n. 5, p. e0286012, maio 2023. ISSN 1932-6203. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0286012>>.

SILVA, M. E. D.; OLIVEIRA, V. L. Difference Equations for the Higher-Order Moments and Cumulants of the INAR(1) Model. *Journal of Time Series Analysis*, v. 25, n. 3, p. 317–333, maio 2004. ISSN 0143-9782, 1467-9892. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.2004.01685.x>>.

STEUTEL, F. W.; Van Harn, K. Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability*, v. 7, n. 5, out. 1979. ISSN 0091-1798. Disponível em: <<https://projecteuclid.org/journals/annals-of-probability/volume-7/issue-5/Discrete-Analogues-of-Self-Decomposability-and-Stability/10.1214/aop/1176994950.full>>.

WEIß, C. H. Thinning operations for modeling time series of counts—a survey. *AStA Advances in Statistical Analysis*, v. 92, n. 3, p. 319–341, ago. 2008. ISSN 1863-8171, 1863-818X. Disponível em: <<http://link.springer.com/10.1007/s10182-008-0072-3>>.

ZEGER, S. L. A regression model for time series of counts. *Biometrika*, v. 75, n. 4, p. 621–629, 1988. ISSN 0006-3444, 1464-3510. Disponível em: <<https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/75.4.621>>.

ZHANG, J.; WANG, D.; YANG, K. A study of rcinar(1) process with generalized negative binomial marginals. *Communications in Statistics - Simulation and Computation*, Taylor Francis, v. 49, n. 6, p. 1487–1510, 2020.

ZHENG, H.; BASAWA, I. V.; DATTA, S. Inference for p th-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis*, v. 27, n. 3, p. 411–440, maio 2006. ISSN 0143-9782, 1467-9892. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9892.2006.00472.x>>.

ZHENG, H.; BASAWA, I. V.; DATTA, S. First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, v. 137, n. 1, p. 212–229, jan. 2007. ISSN 03783758. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378375806000267>>.

APÊNDICE A - PREVISÃO RCZINAR

Neste apêndice, apresentamos os resultados da análise preditiva para o modelo RCZINAR. A seguir, são exibidos a tabela com as estatísticas descritivas da distribuição preditiva e o gráfico correspondente, com destaque para os intervalos de confiança de 95%. Esses resultados complementam a análise principal, permitindo uma comparação direta com o desempenho do modelo RCINAR.

Tabela 9 – Estatísticas da distribuição preditiva das observações #167 e #168, no modelo RCZINAR. As linhas a partir de $j=9$ possuem todos os valores iguais a zero.

j	#167				#168			
	Média	SD	Q0.025	Q0.075	Média	SD	Q0.025	Q0.075
0	0.3265	0.0653	0.2255	0.4675	0.3437	0.0539	0.2558	0.4324
1	0.3397	0.0443	0.2693	0.4273	0.2970	0.0295	0.2612	0.3475
2	0.1992	0.0299	0.1505	0.2663	0.1943	0.0330	0.1446	0.2568
3	0.0920	0.0243	0.0501	0.1309	0.1033	0.0297	0.0603	0.1612
4	0.0322	0.0169	0.0101	0.0707	0.0389	0.0149	0.0198	0.0655
5	0.0083	0.0046	0.0000	0.0152	0.0148	0.0102	0.0000	0.0354
6	0.0018	0.0024	0.0000	0.0050	0.0053	0.0060	0.0000	0.0201
7	0.0004	0.0018	0.0000	0.0052	0.0020	0.0036	0.0000	0.0102
8	0.0000	0.0000	0.0000	0.0000	0.0005	0.0015	0.0000	0.0050
9	0.0000	0.0000	0.0000	0.0000	0.0001	0.0008	0.0000	0.0001

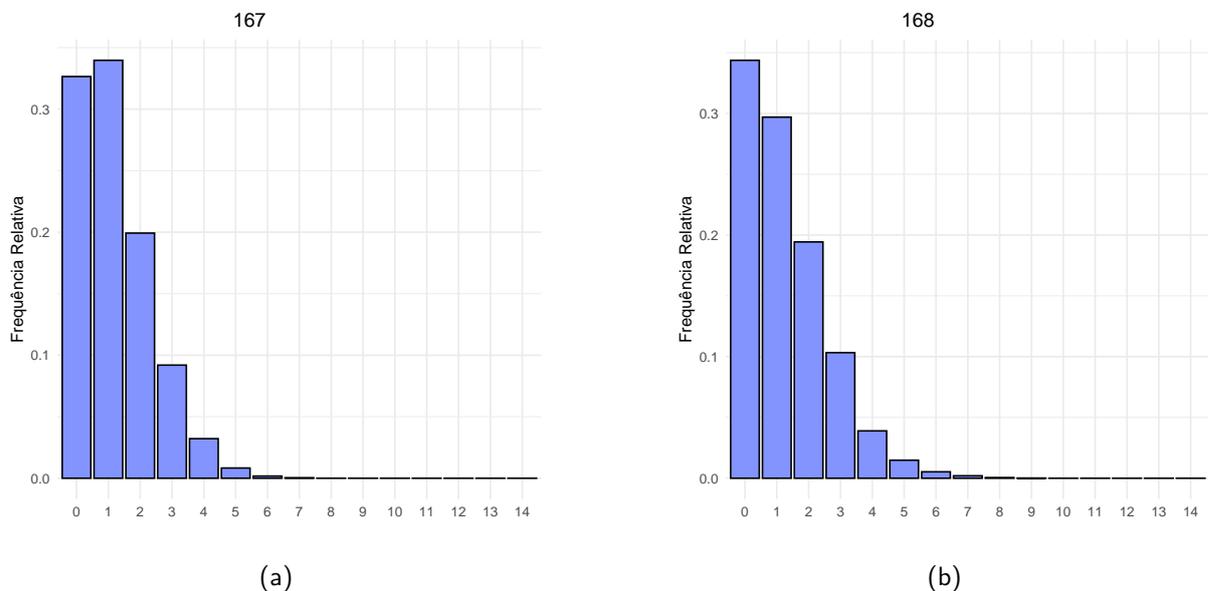


Figura 9 – Gráfico de barras da distribuição de previsão das observações 167 e 168, do modelo RCZINAR(1).