

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Lucas de Oliveira Ferreira de Sales

**Análise de desempenho do gráfico de Shewhart
para o processo POMINAR(1)**

Natal - RN

Dezembro de 2017

Lucas de Oliveira Ferreira de Sales

Análise de desempenho do gráfico de Shewhart para o processo POMINAR(1)

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Orientador: Prof. Dr. André Luís Santos de Pinho

Natal - RN

Dezembro de 2017

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Sales, Lucas de Oliveira Ferreira de.

Análise de desempenho do gráfico de Shewhart para o processo POMINAR(1) / Lucas de Oliveira Ferreira de Sales. - 2017.
57 f.: il.

Monografia (Bacharelado em Estatística) - Universidade Federal do Rio Grande do Norte. Centro de Ciências Exatas e da Terra. Departamento de Estatística. Natal, RN, 2017.

Orientador: André Luís Santos de Pinho.

1. Estatística - Monografia. 2. Controle estatístico de processos - Monografia. 3. Gráfico X - Monografia. 4. Processos autocorrelacionados - Monografia. 5. Séries temporais de valores inteiros - Monografia. I. Pinho, André Luís Santos de. II. Título.

RN/UF/CCET


CDU 519.2

Lucas de Oliveira Ferreira de Sales

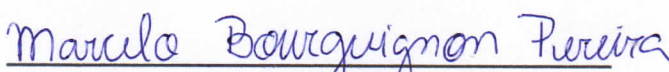
Análise de desempenho do gráfico de Shewhart para o processo POMINAR(1)

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

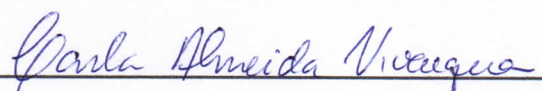
Trabalho aprovado. Natal - RN, Dezembro de 2017:



Prof. Dr. André Luís Santos de Pinho
Orientador



Professor Dr.
Marcelo Bourguignon Pereira



Professor Dra.
Carla Almeida Vivacqua

Natal - RN
8 de dezembro de 2017

À minha mãe, por todo amor e ensinamentos.

Agradecimentos

A Deus, pelo dom da vida.

A toda minha família, por todo apoio e amor incondicional, em especial: minha mãe Fransuilma, que sempre demonstra em cada segundo de sua vida a pessoa espetacular que ela é, e meu pai Mozart por todo apoio que tem me dado mesmo morando longe.

Ao professor André por te aceitado ser meu orientador neste trabalho e por toda sua disponibilidade em me ajudar. Aos professores do departamento de estatística, em especial: Pledson Guedes, por toda sua motivação e determinação para fazer o curso cada vez melhor, à Professora Dione por todos os ensinamentos durante toda minha jornada no PET, ao professor Damião Nóbrega por toda ajuda, conselhos e ensinamentos, aos professores Carla Vivacqua, Marcus Nunes, Marcelo Bourguignon, Francisco Moisés, Paulo Roberto, Iloneide Ramos, Luz Milena e Bruno Monte por toda ajuda e apoio. Aos professores: Paulo Roberto, Viviane Klein e Ailton Rodrigues do departamento de matemática, por toda ajuda nas disciplinas de cálculo e análise matemática.

Aos meus grandes amigos Erika Rayanne e Rodrigo Matheus que me acompanham desde os primeiros dias da graduação e estão sempre comigo nos momentos mais divertidos e nos mais complicados.

Aos meus amigos ex PETianos e atuais PETianos, Francimário, Inara, Elias, Fidel, Joyce, Kalil, Felipe, Pedro, Antony Iury, Kaio, Igor, Daniel Nóbrega, Diêgo costa, Félix, Carla, Rayland, Fabrício, Wesley, Fillipe, Ivonaldo, Thiago e Wanderlan, por todas as excelentes tardes e noites na sala do PET.

Ao Programa de Educação Tutorial- PET, pelo suporte financeiro e estrutura oferecida a mim durante três anos e meio da minha graduação.

Aos meus amigos da turma de 2014.1 e a todos os colegas que fiz nestes quatro anos de curso, em especial: Isabelle, Antony Maciel, Taynná, José Edson, Ruanderson, Rodrigo Amorim, Adryan, Marylaine, Arine, Mariana Lopes, Igor Maurício, Gustavo Fontoura, Luana Pontes, Ronnevon e Waldemar, por compartilharem boa parte da caminhada da graduação junto a mim .

“Seja forte! Não como uma onda que tudo destrói, mas como uma rocha que tudo suporta.”

Autor desconhecido.

Resumo

Desde a revolução industrial até os dias atuais é de interesse da indústria o monitoramento da qualidade de seus produtos. Nesse contexto é que se adequam as técnicas do controle estatístico de processos (CEP), em especial, os gráficos de controle, que têm como finalidade o monitoramento de alguma determinada característica de qualidade. Nos dias atuais, por causa da natureza de muitos processos é comum a modelagem desses dados através de modelos que captem a autocorrelação entre as observações e que se adequem à natureza da variável de interesse. Deste modo, modelos de séries temporais de valores inteiros juntamente com gráficos de controle modificados estão sendo muito utilizados na perspectiva de melhor monitorar processos com presença de autocorrelação e que a variável monitorada possua uma medida expressa por números inteiros. Com base nisso, este trabalho tem como finalidade principal construir o gráfico de Shewhart, também conhecido como gráfico \bar{X} , adequado para monitorar dados de um processo POMINAR(1) e avaliar seu desempenho em alguns cenários. Tanto a construção do gráfico, quanto a avaliação do desempenho serão realizadas através de estudos empíricos utilizando o método de Monte Carlo na geração do limite de controle e na avaliação de desempenho do gráfico. O gráfico de controle proposto foi avaliado em relação ao número médio de amostras até um alarme falso (NMAF) e o número médio de amostras até a detecção de um alarme verdadeiro (NMA) em 50 cenários distintos e apresentou um desempenho satisfatório, ou seja, os NMAF's foram altos, em torno de 240 e 400 amostras até detectar um alarme falso dependendo sempre da combinação dos parâmetros e dos tamanhos amostrais e os NMA's em torno de 3 e 20 amostras. Em particular, à medida que o deslocamento na média se aproxima de três desvios padrão, o NMA se aproxima de 3, ou seja, necessitando apenas de 3 amostras para detectar tal alteração.

Palavras-chave: Controle estatístico de Processos. Gráfico \bar{X} . Processos autocorrelacionados. Séries temporais de valores inteiros.

Abstract

Since the industrial revolution to the present day it is in the industry's interest to monitor the quality of their products. That way, the techniques of statistical process control (SPC), especially control charts, are used to monitor certain quality characteristics. Nowadays, because of the nature of many processes, it is common to model such data through models that capture the autocorrelation among. Thus, time series models of integer values together with modified control charts are being widely used in order to better monitor processes with presence of autocorrelation and that the monitored variable has a measure expressed by integers values. Based on all these features, this study has the main purpose of constructing the appropriate Shewhart chart, or \bar{X} Chart, to monitor data from a POMINAR(1) process and to evaluate its performance in some scenarios. Both the construction of the chart and the performance evaluation will be performed through computational studies using the Monte Carlo method in the generation of the control limit and in the performance evaluation of the chart. We will analyze the proposed control chart based on the in-control average run length (ARL_0) and out-control average run length (ARL_1) in 50 different scenarios. The modified Shewhart chart presented a good performance in relation to ARL_0 which was around 240 and 400 samples, in relation to detection of mean deviations the chart present ARL_1 around 20 and 3 samples. But as the deviation from the mean increases the ARL_1 decreases to 3 samples.

Keywords: Autocorrelation Process. Statistical Process Control. Time Series of Integer Values. \bar{X} Chart.

Lista de ilustrações

Figura 1 – Gráfico de Shewhart de um processo de produção de eixos.	17
-----------------------------------------------------------------------------	----

Lista de tabelas

Tabela 1 – Combinação de parâmetros que ilustram a sobredispersão do processo.	32
Tabela 2 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,3$, $\beta=0,3$, $\lambda=2$ e $p=0,3$.	37
Tabela 3 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,4$, $\beta=0,6$, $\lambda=3$ e $p=0,4$.	38
Tabela 4 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,4$, $\beta=0,5$, $\lambda=5$ e $p=0,5$.	38
Tabela 5 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,6$, $\beta=0,9$, $\lambda=7$ e $p=0,6$.	39
Tabela 6 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,7$, $\beta=0,9$, $\lambda=9$ e $p=0,4$.	40

Sumário

	Lista de ilustrações	9
	Lista de tabelas	10
1	INTRODUÇÃO	13
2	O GRÁFICO \bar{X} PARA MONITORAMENTO DA MÉDIA	16
2.1	Interpretação do gráfico \bar{X}	17
2.2	Desempenho dos gráficos de controle	18
3	DEFINIÇÕES BÁSICAS E CONSTRUÇÃO DOS PROCESSOS POIS- SON INAR(1) E INARCH(1)	20
3.1	Definições básicas	20
3.2	Construção do processo INAR(1)	20
3.2.1	Processo Poisson INAR(1)	22
3.3	Construção do processo INARCH(1)	23
4	CONSTRUÇÃO DO PROCESSO POMINAR(1)	26
4.1	Definição e propriedades probabilísticas	26
4.2	Estimação dos parâmetros	29
5	GRÁFICO DE SHEWHART PARA O PROCESSO POMINAR(1)	31
5.1	Construção dos limites de controle	31
5.2	Obtenção do limite superior de controle	33
5.3	Obtendo os limites em uma situação prática	35
6	ANÁLISE DE DESEMPENHO DO GRÁFICO PARA O PROCESSO POMINAR(1)	36
7	CONSIDERAÇÕES FINAIS	41
7.1	Sugestões de trabalhos futuros	42
	APÊNDICE A – FUNÇÕES UTILIZADAS NA ESTIMAÇÃO DO LSC	43
	APÊNDICE B – ROTINAS COMPUTACIONAIS PARA GERAÇÃO DO NMAF	46

APÊNDICE C – ROTINAS COMPUTACIONAIS PARA GERAÇÃO DO NMA	50
REFERÊNCIAS	56

1 Introdução

Costa, Epprecht e Carpinetti (2005) comentam que desde a revolução industrial até os dias atuais existe uma enorme preocupação com a qualidade, funcionamento e produção de vários itens de consumo, tais como automóveis, roupas e geladeiras, por exemplo. Tendo em vista a necessidade do monitoramento da qualidade dos processos, industriais ou não, o CEP (controle estatístico de qualidade) é uma área de estudo que tem como objetivo maior estabelecer informações sobre o comportamento dos processos utilizando algumas ferramentas para conseguir verificar e obter a estabilidade dos processos.

A principal ferramenta do controle estatístico de processo é o gráfico de controle, que baseia-se no monitoramento periódico de amostras de uma determinada característica de qualidade de interesse. O monitoramento é realizado através da plotagem de medidas da característica de qualidade *versus* o tempo, ou número de amostras, e verificando se estas medidas se distribuem aleatoriamente em torno de um linha central, denominada limite médio ou linha média, e dentro de duas faixas de confiança denominadas limites de controle, caso isso ocorra o processo é dito sob controle. Se existir alguma causa de variabilidade de fontes não-usuais, ou não aleatórias, os pontos plotados tendem a sair dos limites de controle e o processo é dito como fora de controle. As principais vantagens do uso do gráfico de controle são: ser um método visual de monitoramento, ou diagnóstico, facilitando assim a compreensão para indivíduos menos treinados. Os gráficos de controle evitam ajustes desnecessários no processo e auxiliam na identificação do instante em que o processo foi alterado, permitindo assim buscar soluções para a causa do desajuste do processo. Para mais detalhes sobre o contexto de CEP ver Montgomery (2000).

Os gráficos de controle, em geral, baseiam-se na suposição que as amostras coletadas no processo monitorado são independentes entre si ao longo do tempo, porém em muitas situações, pela própria natureza do processo, essa suposição de independência é violada. Por exemplo, em indústrias que tratam com processos químicos, que envolvem temperaturas, ou misturas, é natural que sequências de ensaios químicos tenham relações entre si, caso o tempo entre a retirada de cada amostra não seja grande o suficiente para dissipar esta relação. Outra situação corriqueira é quando se monitora acessos em uma determinada página *online*, pois esses acessos podem estar diretamente ligados aos horários de coleta da amostra, de forma que amostras coletadas em horários próximos tenham uma relação mais forte do que amostras coletadas em horários mais distantes. Estes exemplos servem para ilustrar a quebra na suposição de independência entre as amostras.

Processos de naturezas similares aos exemplificados são conhecidos como processos autocorrelacionados. Este mecanismo da autocorrelação influencia no desempenho e na

precisão de detecção dos gráficos de controle, para mais detalhes ver Wiel (1996).

Dada essa situação de quebra de suposições nos gráficos usuais, uma área que vem em grande crescente no contexto do controle estatístico de processos é a que tem como finalidade propor, baseado em modelos de séries temporais, mudanças nos gráficos de controle de forma que possam levar em conta essa autocorrelação natural do processo, porém sem interferir no seu desempenho. Claro, Costa e Machado (2007) propuseram uma alteração no gráfico de Shewhart, também conhecido gráfico \bar{X} , e Média Móvel Ponderada Exponencialmente (*Exponentially Weighted Moving Average-EWMA*), de forma tal que dados provenientes de um modelo auto-regressivo de ordem 1 (AR(1)) pudesse ser bem monitorado pelos gráficos citados apesar de existir a presença da autocorrelação. Porém esse trabalho contemplou apenas processos cuja natureza da variável seja contínua e que os dados se adequem ao modelo AR(1). No contexto de processos os quais a natureza da variável monitorada seja discreta, os modelos de séries temporais utilizados contemplam a natureza da variável, ou seja, são modelos para dados discretos e em muitos casos usam-se modelos para dados discretos inflados de zeros devido a natureza da variável monitorada. Weiß e Testik (2009), Weiß (2011) e Rakitzis, Castagliola e Maravelakis (2016) abordam e reforçam a ideia da importância do procedimento correto na análise de um processo autocorrelacionado, principalmente quando tratamos com valores inteiros. Além disso, esses trabalhos trazem alterações nos gráficos de somas acumuladas (*Cumulative Sum-CUSUM*) para processos cuja natureza da variável seja discreta.

Neste contexto, este trabalho tem como principal objetivo propor alterações no gráfico \bar{X} usual, com a finalidade de monitorar, de forma satisfatória, dados provenientes do processo autorregressivo misto para valores inteiros com inovações Poisson POMINAR(1) (*Poisson Mixed Interger Autoregressive*), proposto por Orozco (2017). Deste modo, o objetivo específico deste trabalho consiste na verificação, através de simulações, do número médio de amostra até a detecção de um alarme, para diferentes combinações de parâmetros, ou seja, avaliar o desempenho do gráfico proposto ao monitorar dados modelados pelo processo POMINAR(1).

O trabalho está organizado da seguinte forma, no Capítulo 2, apresenta-se a construção, interpretação e métodos de avaliação do gráfico de Shewhart. Posteriormente, no Capítulo 3, são definidos os processo INAR(1) (*Interger Value First-Order Autoregressive*), Poisson INAR(1) e INARCH(1) (*Integer-valued first-order Autoregressive Conditional Heteroskedasticity*), juntamente com os operadores *thinning* binomial e *thinning* Poisson, servindo como base principal para o Capítulo 4 que consta da construção, definição, propriedades, métodos de estimação e características do processo POMINAR(1). No Capítulo 5, está descrito todo o procedimento para a obtenção dos limites de controle para o monitoramento de processos modelados pelo POMINAR(1), incluindo a obtenção dos limites em uma situação prática. O capítulo 6 apresenta as análises do desempenho do

gráfico proposto. Por fim, no capítulo 7 temos as conclusões obtidas através deste estudo e sugestões de trabalhos futuros.

2 O gráfico \bar{X} para monitoramento da média

Uma gama de variáveis, ou características de qualidade podem ser expressas em termos de uma medida numérica. Segundo Montgomery (2000) ao lidarmos com uma característica de qualidade que possa ser mensurada através de medidas numéricas é usualmente necessário o monitoramento tanto do valor médio da característica como de sua variabilidade. Neste contexto de monitoramento é que são inseridos os gráficos de controle.

Conforme comentado no Capítulo 1, os gráficos de controle baseiam-se no monitoramento periódico de amostras para averiguar o comportamento de determinada característica de interesse. Quando essa característica monitorada é a média amostral, e deseja-se detectar grandes desvios no processo na média do processo (1,5 desvios padrão ou mais na média), utiliza-se o gráfico de médias \bar{X} para o monitoramento.

O gráfico de controle \bar{X} , ou gráfico de Shewhart, foi proposto pelo engenheiro e estatístico Walter A. Shewhart em meados dos anos 30 e baseia-se em duas suposições: são retiradas m amostras de tamanho n , em que cada uma das m amostras são independentes entre si. Assumimos normalidade para a variável monitorada, sendo a suposição de normalidade um pouco mais flexível dado que se a variável de interesse não seguir distribuição normal, porém a amostra for suficientemente grande, pelo teorema central do limite podemos utilizar a aproximação normal. Assim como todo gráfico de controle, o gráfico \bar{X} de Shewhart possui três limites de controle, sendo eles: limite médio, ou linha média (LM), limite superior de controle (LSC) e limite inferior de controle (LIC).

Seja X uma característica de qualidade monitorada, μ_0 e σ_0 a média e desvio padrão, respectivamente, desta característica quando o processo esta sob controle, n o tamanho de cada subamostra das m amostras independentes, então os limites de controle para o gráfico \bar{X} são dados de forma geral por:

$$\begin{aligned} LSC &= \mu_0 + k \frac{\sigma_0}{\sqrt{n}}, \\ LM &= \mu_0 \text{ e} \\ LIC &= \mu_0 - k \frac{\sigma_0}{\sqrt{n}}. \end{aligned}$$

Ainda nos anos 30, Shewhart propôs a utilização de $k = 3$ baseando-se no lema: “Se o processo estiver em controle, evite ajustes desnecessários. Com a abertura de 3 desvios-padrão, enquanto o processo estiver sob controle, raramente um ponto cairá na região de ação do gráfico”. Com base nesse lema, Shewhart afirmava que isso iria prevenir intervenções desnecessárias na produção, tornando o gráfico útil e eficiente para as condições

de uma indústria de produção. Outra justificativa mais formal é dada em Costa, Epprecht e Carpinetti (2005) que diz que os intervalos considerando $k = 3$ em torno da média sob controle (μ_0) engloba cerca de 99,73% dos valores das médias amostrais \bar{X} , desta forma se o processo permanecer insento de causas especiais, a média amostral \bar{X} terá uma baixa probabilidade de cair foras do limites de controle, e conseqüentemente é mais verossímil que um valor da média amostral fora dos limites realmente seja consequência de alguma alteração no processo.

A Figura 1 ilustra o gráfico \bar{X} com $k = 3$, ou de Shewhart, para 30 amostras de tamanho $n=5$ de um processo de produção de eixos. Os dados desse processo de produção estão presente em Costa, Epprecht e Carpinetti (2005).

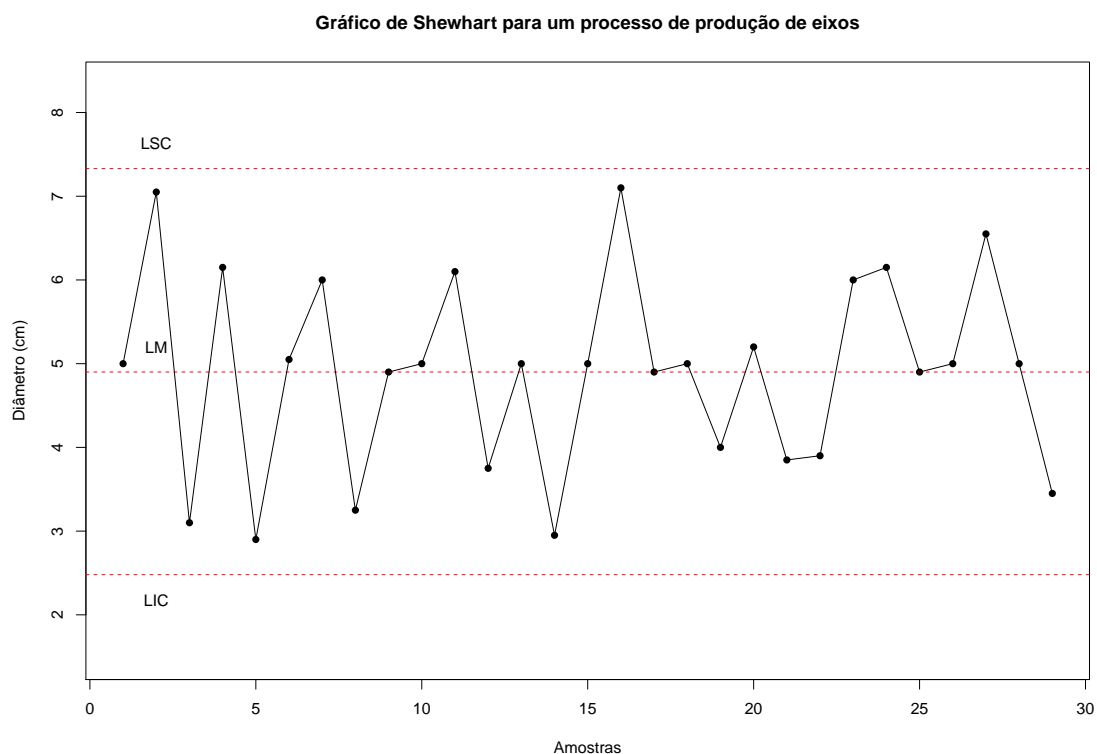


Figura 1 – Gráfico de Shewhart de um processo de produção de eixos.

A Figura 1 apresenta a estrutura básica presente em um gráfico \bar{X} . Essa estrutura basicamente é composta por pontos (médias amostrais) plotados seguindo determinados padrões em torno dos limites de controles. A seguir, veremos como interpretar esses padrões presentes no gráfico de controle \bar{X} .

2.1 Interpretação do gráfico \bar{X}

Como já visto, os gráficos de controle auxiliam no monitoramento de uma característica de qualidade que estamos interessados, e com base nesse monitoramento podemos

concluir se o processo está ou não sob controle. Para compreensão de como interpretar o gráfico e verificarmos se o processo está sob controle ou não, considere a seguinte situação: foram retiradas m amostras, independentes e espaçadas no tempo, de tamanho n e com base no processo sob controle foram construídos os limites de controle em seguida foram plotadas no gráfico as médias de cada uma das m amostras independentes. Com base na situação descrita temos que:

- O processo estará sob controle se os pontos plotados no gráfico \bar{X} estiverem flutuando aleatoriamente, sem nenhum padrão evidente, em torno da linha média;
- O processo estará fora de controle caso algumas das situações abaixo ocorram:
 - 1- Um ou mais pontos acima do LSC são indícios que houve algum aumento na média do processo para aquela característica observada;
 - 2- Um ou mais pontos abaixo do LIC são indícios que houve alguma diminuição na média do processo para aquela característica observada;
 - 3- Uma sequência de oito pontos consecutivos de um mesmo lado (acima ou abaixo) da linha central, ou seis pontos em uma sequência crescente ou decrescente, podem representar uma tendência ao aumento ou diminuição da média do processo para a característica que está sendo monitorada;
 - 4- Uma grande sequência de pontos alternadamente para cima e para baixo, ou um padrão não usual, ou não aleatório, podem ser indícios que a suposição de independência foi violada, ou que a média está sofrendo alguma tendência cíclica;
 - 5- Um ou mais pontos muito próximos dos limites de controle pode ser um alerta de possível alteração.

Vale ressaltar que estas sugestões de interpretações tem como finalidade serem um referencial para tomadas de decisões com base no gráfico \bar{X} , e que para alguns processos específicos os critérios de paradas podem ser alterados, porém sempre coerente com a natureza do processo.

2.2 Desempenho dos gráficos de controle

Assim como a construção de limites de controle e interpretações, outro ponto de extrema importância quando tratamos de gráficos de controle é o seu desempenho, ou seja, o quão rápido e preciso o gráfico de controle está sendo para determinada situação ou processo.

Para um melhor entendimento e compreensão de como se comporta um gráfico de controle, podemos imaginá-lo como um teste de hipótese, em que para cada uma das m amostras verificamos as seguintes hipóteses: H_0 : o processo está sob controle e H_1 : o

processo está fora de controle. Consideremos μ_x a média verdadeira da nossa característica de qualidade e μ_0 a média da característica quando o processo está sob controle, então podemos definir hipóteses anteriores da seguinte forma:

$$H_0 : \mu_x = \mu_0 \text{ vs } H_1 : \mu_x \neq \mu_0.$$

Para o gráfico \bar{X} , quando o valor da média amostral está dentro dos limites de controle é um sinal que existem evidências estatísticas para não rejeitarmos H_0 , em contrapartida se o valor da média amostral cair fora dos limites de controle é um sinal que existem evidências estatísticas para rejeitarmos a hipótese H_0 , o que equivale a “aceitarmos” H_1 .

Assim como os testes de hipóteses, os gráficos de controle não são infalíveis, ou seja, é passível que com alguma probabilidade ao rejeitarmos ou não H_0 estejamos cometendo um erro. Desta forma definimos α_0 como a probabilidade de rejeitarmos H_0 quando ela é verdadeira, β_0 a probabilidade de não rejeitarmos H_0 quando deveríamos rejeitá-la e $p_0 = 1 - \beta_0$ a probabilidade de rejeitarmos H_0 quando realmente deveríamos rejeitar, ou seja, detectar a mudança. Com base nas probabilidades α_0 , β_0 e p_0 , definidas anteriormente, o desempenho dos gráficos de controle é dado através das seguintes quantidades:

- Número médio de amostras até a detecção do alarme verdadeiro (NMA): O NMA representa o número esperado de amostras, após a mudança na média, necessários para detectar a alteração na média do processo;
- Número médio de amostras até um alarme falso (NMAF): O NMAF representa o número esperado de amostras para que a sinalização do gráfico (um ponto fora dos limites de controle) seja um alarme falso, ou seja, o ponto caiu fora dos limites, porém não houve alteração no processo.

Montgomery (2000) comenta que quando as amostras da característica de qualidade são independentes, o número de observações necessárias até um ponto exceder os limites de controle será uma variável aleatória que segue uma distribuição geométrica de parâmetro p e valor esperado $1/p$, sendo assim o NMA e NMAF são dados por:

$$NMA = \frac{1}{p_0} \text{ e}$$

$$NMAF = \frac{1}{\alpha_0},$$

ou seja, o NMA é o valor esperado de uma variável geométrica cujo parâmetro é p_0 e NMAF é o valor esperado de uma variável geométrica cujo parâmetro é α_0 .

3 Definições básicas e construção dos processos Poisson INAR(1) e INARCH(1)

O presente capítulo consta de algumas definições preliminares para melhor compreensão do trabalho. Estes conceitos principais tratam sobre: processos estocásticos, séries temporais, definições dos operadores *thinning* binomial e *thinning* Poisson e a construção dos processos INAR(1), Poisson INAR(1) e INARCH(1) que são a base fundamental do processo POMINAR(1).

3.1 Definições básicas

A seguir serão postuladas algumas definições básicas para uma compreensão mais clara do estudo de séries temporais.

Definição 3.1.1. Um processo estocástico é uma coleção de variáveis aleatórias $\{X_t\}_{t \in T}$ definida sobre um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, T é o conjunto de índices do processo ou espaço paramétrico.

Definição 3.1.2. Uma série temporal é uma sequência de observações $\{X_t\}_{t \geq 0}$ sobre uma variável de interesse feitas em um determinado período de tempo t .

Definição 3.1.3. Um processo $\{X_t\}$ é estacionário de segunda ordem, se os dois primeiros momentos de X_t existem e não dependem do tempo, isto é, $E[X_t] = \mu$ e $E[X_t^2] = \delta$.

Estas definições são de extrema importância, visto que uma série temporal, definida em 3.1.2, é um processo estocástico. Em muitos modelos de séries temporais o conceito de estacionariedade fraca, definido em 3.1.3, é utilizado como uma estrutura simplificadora para a redução do número de parâmetros a serem estimados em uma única realização de uma série temporal.

3.2 Construção do processo INAR(1)

Nesta seção abordaremos o modelo de série temporais para valores inteiros INAR(1) apresentado por Al-Osh e Alzaid (1987) que baseia-se no operador proposto por Steutel e Harn (1979), pertencente à classe dos operadores *thinning*, chamado de operador *thinning* binomial. Desta forma, iremos apresentar a definição do operador *thinning* binomial, tais como suas propriedades e definiremos o modelo INAR(1).

Definição 3.2.1. Seja X uma variável aleatória que assume valores inteiros não negativos, α um número real tal que $\alpha \in [0,1]$. Definimos o operador *thinning* binomial, denotado por “ \circ ”, como

$$\alpha \circ X = \sum_{i=1}^X B_i(\alpha),$$

em que $\{B_i(\alpha)\}_{i=1}^X$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), chamada de série de contagem, com distribuição Bernoulli de parâmetro α , independentes de X e com função de probabilidade dada por

$$\mathbb{P}(B_i(\alpha) = 1) = 1 - \mathbb{P}(B_i(\alpha) = 0) = \alpha.$$

Silva e Oliveira (2004) apresentam e demonstram as propriedades aqui abordadas referentes ao operador dado na Definição 3.2.1.

Lema 3.2.1. *Sejam X e Y variáveis aleatórias identicamente distribuídas que assumem valores inteiros não-negativos, α_1 e α_2 constantes reais pertencentes ao intervalo $[0,1]$, suponha que a série de contagem de $\alpha_i \circ X$, é independente da série de contagem de $\alpha_i \circ Y$, em que $i = 1,2$. Então o operador thinning binomial apresenta as seguintes propriedades:*¹

1. $0 \circ X = 0$.
2. $1 \circ X = X$.
3. $\alpha_1 \circ (\alpha_2 \circ X) \stackrel{d}{=} (\alpha_1 \alpha_2) \circ X$.
4. Se X e Y são independentes, $\alpha_1 \circ (X + Y) \stackrel{d}{=} \alpha_1 \circ X + \alpha_1 \circ Y$.
5. $E[\alpha_1 \circ X] = \alpha_1 E[X]$.
6. $E[\alpha_1 \circ X | X] = \alpha_1 X$.
7. $Var(\alpha_1 \circ X | X) = \alpha_1(1 - \alpha_1)X$.
8. $Var(\alpha_1 \circ X) = \alpha_1^2 Var(X) + \alpha_1(1 - \alpha_1)E[X]$.
9. $E[(\alpha_1 \circ X)^2] = \alpha_1^2 E[X^2] + \alpha_1(1 - \alpha_1)E[X]$.
10. $E[X(\alpha_1 \circ Y)] = \alpha_1 E[XY]$.
11. $E[X(\alpha_1 \circ Y)^2] = \alpha_1^2 E[XY^2] + \alpha_1(1 - \alpha_1)E[XY]$.
12. $E[(\alpha_1 \circ X)(\alpha_2 \circ Y)] = \alpha_1 \alpha_2 E[XY]$.

¹ O símbolo $\stackrel{d}{=}$ representa, ou denota, que as variáveis aleatórias da igualdade possuem a mesma distribuição.

13. $E[(\alpha_1 \circ X)^2(\alpha_2 \circ Y)] = \alpha_1^2 \alpha_2 E[X^2 Y] + \alpha_1(1 - \alpha_1)\alpha_2 E[XY]$.
14. $Cov(\alpha_1 \circ X, \alpha_2 \circ Y) = \alpha_1 \alpha_2 Cov(X, Y)$.

Com a definição do operador *thinning* binomial, dada em 3.2.1, podemos definir o processo proposto por, (??), e Al-Osh e Alzaid (1987) da seguinte forma:

Definição 3.2.2. Um processo discreto para valores inteiros não negativos $\{X_t\}_{t \geq 1}$, $X_t \in \mathbb{N}_0$, é um processo INAR(1) se satisfaz a recursão

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \quad \text{para } t \geq 1, \quad (3.1)$$

em que $\alpha \in [0,1]$, $\mathbb{N}_0 = \{0,1,\dots\}$, “ \circ ” é o operador *thinning* binomial, $\{\varepsilon_t\}_{t \geq 1}$, $\varepsilon_t \in \mathbb{N}_0$, é uma sequência de variáveis aleatórias i.i.d. para valores inteiros não negativos com média μ_ε e variância σ_ε^2 , e a série de contagem de $\alpha \circ X_{t-1}$ é independente de ε_t .

Para $\alpha \in (0,1)$, um processo $\{X_t\}_{t \geq 1}$ que satisfaz a Equação (3.1) é um processo estacionário de segunda ordem. Então o processo INAR(1) possui média e variância constante tal que a covariância entre X_t e X_{t+k} depende só de k (Al-Osh e Alzaid (1987); Jin-Guan e Yuan (1991)).

3.2.1 Processo Poisson INAR(1)

O processo Poisson INAR(1) é um caso particular do processo INAR(1), definido na Seção 3.2, que tem como característica a suposição que os ε_t seguem uma distribuição Poisson. Esse processo é parte essencial na construção do POMINAR(1), proposto por Orozco (2017).

Definição 3.2.3. Seja $\{X_t\}_{t \geq 1}$ um processo INAR(1), que satisfaz (3.1), se ε_t tem distribuição Poisson com parâmetro λ ($\varepsilon_t \sim Po(\lambda)$), então $\{X_t\}_{t \geq 1}$ é um processo Poisson INAR(1), de parâmetros α e λ , em que $E[\varepsilon_t] = \mu_\varepsilon = \lambda$, $Var(\varepsilon_t) = \sigma_\varepsilon^2 = \lambda$.

Utilizando as propriedades do operador *thinning* binomial e a estacionariedade dos processos INAR(1) e Poisson INAR(1), Al-Osh e Alzaid (1987) propuseram as seguintes propriedades, referentes ao processo Poisson INAR(1):

A média e variância marginal do modelo Poisson INAR(1) são dadas por

$$E[X_t] = Var(X_t) = \frac{\lambda}{1 - \alpha}.$$

Podemos escrever a média e variância condicional do processo Poisson INAR(1) da seguinte forma:

$$E[X_t|X_{t-1}] = \alpha X_{t-1} + \lambda$$

e

$$Var(X_t|X_{t-1}) = \alpha(1 - \alpha)X_{t-1} + \lambda.$$

As funções de autocovariância, $\gamma(k)$, e autocorrelação, $\rho(k)$, referentes ao modelo Poisson INAR(1) de defasagem k podem ser expressas como

$$\gamma(k) = Cov(X_{t-k}, X_t) = \alpha^k \gamma(0), \quad \text{para } k = 0, 1, \dots,$$

e

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \alpha^k \quad \text{para } k = 0, 1, \dots$$

Observe que a função de autocorrelação $\rho(k)$ decai exponencialmente com a defasagem k , pois assumimos $\alpha \in (0, 1)$.

Uma interpretação prática do processo Poisson INAR(1) é dada em Freeland (1998). A interpretação sugere vermos o processo Poisson INAR(1) como um processo de nascimento e morte. Processos de nascimento e morte são processos em que os indivíduos do tempo $t - 1$ tem um probabilidade α de permanecerem vivos até o tempo t e com relação a cada tempo t , supomos que o número de nascimentos segue um distribuição Poisson de parâmetro λ . A equação a seguir representa o processo de nascimento e morte descrito,

$$\underbrace{X_t}_{\text{População no tempo } t} = \underbrace{\alpha \circ X_{t-1}}_{\text{Sobreviventes no tempo } t} + \underbrace{\epsilon_t}_{\text{nascimentos}}.$$

3.3 Construção do processo INARCH(1)

Na presente seção, iremos construir o processo INARCH(1), definido de uma forma alternativa por Fernández (2013) baseado no operador proposto por Ferland, Latour e Oraichi (2006), pertencente à classe dos operadores *thinning*, chamado de operador *thinning* Poisson. Desta forma, iremos apresentar a definição do operador *thinning* Poisson, tais como suas propriedades e definiremos o modelo INARCH(1) da forma proposta por Fernández (2013).

Definição 3.3.1. Seja X uma variável aleatória assumindo valores inteiros não negativos, β um número real tal que $\beta \geq 0$. O operador *thinning* Poisson “ $*$ ” é definido como

$$\beta * X = \sum_{i=1}^X P_i(\beta),$$

em que $\{P_i(\beta)\}_{i=1}^X$ é uma sequência de variáveis aleatórias i.i.d. com distribuição Poisson de parâmetro β e independentes de X .

Usualmente a sequência $\{P_i(\beta)\}_{i=1}^X$ é chamada série de contagem de $\beta * X$. Note que, pela forma que definimos o operador *thinning* Poisson, $\beta * X|X = x$, tem distribuição Poisson de parâmetro βx ($\beta * X|X = x \sim \text{Poisson}(\beta x)$).

Perceba que a diferença básica dos operadores dados nas Definições 3.2.1 e 3.3.1 está na distribuição de probabilidade associada à série de contagem.

Fernández (2013) demonstrou as propriedades aqui apresentadas referentes ao operador *thinning* Poisson dado na Definição 3.3.1.

Lema 3.3.1. *Sejam X_i , $i = 1, \dots, m$, uma sequência de variáveis aleatórias identicamente distribuídas de valores inteiros, β_i , $i = 1, \dots, m$, uma sequência de constantes reais não negativas, suponha que as séries de contagem de $\beta_i * X_i|X_i = x_i$, são independentes e identicamente distribuídas de acordo com uma distribuição Poisson de parâmetro $\beta_i x_i$ ($\text{Poisson}(\beta_i x_i)$), e independentes dos X_i . Então,*

1. $0 * X_1 = 0$.
2. $\beta_1 * (X_1 + X_2) \stackrel{d}{=} \beta_1 * X_1 + \beta_1 * X_2$, se X_1, X_2 são independentes.
3. $E[\beta_1 * X_1] = \beta_1 E[X_1]$.
4. $\text{Var}(\beta_1 * X_1) = \beta_1^2 \text{Var}(X_1) + \beta_1 E[X_1]$.
5. $E[\beta_1 * X_1|X_1] = \beta_1 X_1$.
6. $\text{Var}(\beta_1 * X_1|X_1) = \beta_1 X_1$.
7. $\text{Cov}(\beta_1 * X_1, \beta_2 * X_2) = \beta_1 \beta_2 \text{Cov}(X_1, X_2)$.
8. $E[(\beta_1 * X_1)^2] = \beta_1 E[X_1] + \beta_1^2 E[X_1^2]$.
9. $E[(\beta_1 * X_1)^r] = \sum_{k=1}^r S(r, k) \beta_1^k E[X_1^k]$, em que $S(r, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^r$ é o número de Stirling de segunda espécie que pode ser usado para contar o número de formas de particionar um conjunto de r objetos em k subconjuntos não vazios.
10. $E[(\beta_1 * X_1)X_2] = \beta_1 E[X_1 X_2]$.
11. $E[(\beta_1 * X_1)^2 X_2] = \beta_1 E[X_1 X_2] + \beta_1^2 E[X_1^2 X_2]$.
12. $E[(\beta_1 * X_1)^r X_2] = \sum_{k=1}^r S(r, k) \beta_1^k E[X_1^k X_2]$.
13. $E\left[(\beta_1 * X_1) \prod_{i=2}^m X_i\right] = \beta_1 E\left[\prod_{i=1}^m X_i\right]$.
14. $E\left[(\beta_1 * X_1)^r \prod_{i=2}^m X_i\right] = \sum_{k=1}^r S(r, k) \beta_1^k E\left[X_1^k \prod_{i=1}^m X_i\right]$.
15. $E[(\beta_1 * X_1)(\beta_2 * X_2)] = \beta_1 \beta_2 E[X_1 X_2]$.

16. $E \left[\prod_{i=1}^m (\beta_i * X_i) \right] = \left[\prod_{i=1}^m \beta_i \right] E \left[\prod_{i=1}^m X_i \right]$.
17. $E[(\beta_1 * X_1)^2 (\beta_2 * X_2)] = \beta_1^2 \beta_2 E[X_1^2 X_2] + \beta_1 \beta_2 E[X_1 X_2]$.
18. $E[(\beta_1 * X_1)^r (\beta_2 * X_2)] = \beta_2 \sum_{k=1}^r S(r, k) \beta_1^k E[X_1^k X_2]$.
19. $E[\beta_1 * \dots (\beta_1 * (\beta_1 * X_1))] = \beta_1^r E[X_1]$.

Com o operador *thinning* Poisson definido em 3.3.1, podemos definir o processo INARCH(1) da forma proposta por Fernández (2013).

Definição 3.3.2. Um processo discreto para valores inteiros não negativos $\{X_t\}_{t \geq 1}$, $X_t \in \mathbb{N}_0$, é um processo INARCH(1) se satisfaz a seguinte recursão

$$X_t = \beta * X_{t-1} + \varepsilon_t, \quad \text{para } t \geq 1, \quad (3.2)$$

em que $\beta \geq 0$, “ $*$ ” é o operador *thinning* Poisson, $\{\varepsilon_t\}_{t \geq 1}$, $\varepsilon_t \in \mathbb{N}_0$, é uma sequência de variáveis aleatórias i.i.d. com distribuição Poisson de parâmetro λ ($\varepsilon_t \sim Po(\lambda)$), assume-se que a série de contagem de $\beta * X_{t-1}$ é independente de ε_t .

Para $\beta \in (0,1)$, um processo $\{X_t\}_{t \geq 1}$ que satisfaz a Equação (3.2) é um processo estacionário de segunda ordem. Heinen (2003) e Ferland, Latour e Oraichi (2006) mostraram que o processo INARCH(1) possui média e variância constante tal que a covariância entre X_t e X_{t+k} depende só de k , conforme pode ser visto nas equações a seguir.

$$E[X_t] = \frac{\lambda}{1 - \beta},$$

$$Var(X_t) = \frac{\lambda}{(1 - \beta)(1 - \beta^2)}$$

e

$$\rho(k) = Corr(X_{t-k}, X_t) = \beta^k \quad \text{para } k = 0, 1, \dots$$

As propriedades probabilísticas acima decorrem das propriedades do operador *thinning* Poisson, dadas no Lema 3.3.1, e pela estacionariedade do processo.

O processo INARCH(1) pode ser interpretado como uma população no tempo $t-1$, a qual alguns indivíduos morrem, outros geram descendentes e outros sobrevivem, porém não gera descendentes até o tempo t e existe um determinado número de novos membros para a população no tempo t . A equação a seguir ilustra esta interpretação.

$$\underbrace{X_t}_{\text{População no tempo } t} = \underbrace{\beta * X_{t-1}}_{\text{População no tempo } t-1} + \underbrace{\varepsilon_t}_{\text{Novos membros}} .$$

4 Construção do processo POMINAR(1)

No Capítulo 3 foram definidos os processos Poisson INAR(1) e INARCH(1) que serão a base para a compreensão do processo que aqui será apresentado. Desta forma, o presente capítulo apresentará o processo POMINAR(1), proposto por Orozco (2017) e que consiste em um processo de séries temporais autoregressivo misto com inovações Poisson para valores inteiros.

4.1 Definição e propriedades probabilísticas

Definição 4.1.1. O processo POMINAR(1) é definido da seguinte forma:

$$X_t = \begin{cases} \alpha \circ X_{t-1} + \varepsilon_t, & \text{com probabilidade } p, \\ \beta * X_{t-1} + \varepsilon_t, & \text{com probabilidade } 1 - p, \end{cases} \quad (4.1)$$

em que $\alpha, \beta \in [0,1]$, $p \in [0,1]$, $\{\varepsilon_t\}$ é uma sequência de variáveis aleatórias i.i.d. com distribuição Poisson de parâmetro λ , ε_t é independente de todo $\{B_i(\alpha)\}$ e de todo $\{P_i(\beta)\}$, $\{B_i(\alpha)\}$ e $\{P_i(\beta)\}$ também são independentes entre si, em que $\{B_i(\alpha)\}$ e $\{P_i(\beta)\}$ foram definidos em (3.2.1) e (3.3.1), respectivamente. Os operadores *thinnings* no tempo t são realizados independentemente um do outro.

Em linhas gerais, podemos resumir o processo dado na Definição 4.1.1 como um processo que com uma certa probabilidade p será o processo Poisson INAR(1) e com probabilidade $1 - p$ será o processo INARCH(1).

Orozco (2017) propõe uma ideia geral e uma possível interpretação que permita uma melhor compreensão da forma com que se comporta este processo misto. Considere a seguinte situação: suponha que X_t descreve o número total de casais num determinado país dentro de um período de tempo t , seja ε_t o número total da população no tempo t independente de X_{t-1} . Neste país existem incentivos financeiros para que as pessoas adultas tenham filhos. Portanto, um casal, vai ter a possibilidade de ter ou não um filho. Caso não tenha um filho não receberá nenhum incentivo, se tiver um filho terá o incentivo. Nesse caso, $(\alpha \circ X_{t-1})$ será o total de sobreviventes ainda existem no período $[t-1, t)$, e essa situação seguirá o modelo INAR(1) baseado no operador *thinning* binomial. Entretanto, também é possível que o casal queira ter, *nenhum* filho ou, *um* filho ou, *2* ou, *mais* filhos, em cujo caso, receberá incentivos além de outros benefícios. Nesta última situação o uso do modelo INARCH(1) baseado no operador *thinning* Poisson pode ser apropriado, e $(\beta * X_{t-1})$ significará o número total de casais no tempo $t - 1$.

O processo POMINAR(1) pode ser reduzido a alguns casos dependendo das combinações dos parâmetros α, β e p , estes casos são:

1. Se $p = 1$, então $X_t = \alpha \circ X_{t-1} + \varepsilon_t$.

Note que o processo se reduz ao Poisson INAR(1), e no contexto do POMINAR(1) é chamado POMINAR(1)-I.

2. Se $p = 0$, então $X_t = \beta * X_{t-1} + \varepsilon_t$.

Note que o processo se reduz ao INARCH(1), e no contexto do POMINAR(1) é chamado POMINAR(1)-II.

3. Se $\alpha = 0$, então $X_t = \begin{cases} \varepsilon_t, & \text{com probabilidade } p, \\ \beta * X_{t-1} + \varepsilon_t, & \text{com probabilidade } 1 - p. \end{cases}$

O processo reduzido é chamado POMINAR(1)-III.

4. Se $\beta = 0$, então $X_t = \begin{cases} \alpha \circ X_{t-1} + \varepsilon_t, & \text{com probabilidade } p, \\ \varepsilon_t, & \text{com probabilidade } 1 - p. \end{cases}$

O processo reduzido é chamado POMINAR(1)-IV.

5. Se $\alpha = 0$ e $\beta = 0$, ou $p = 1$ e $\alpha = 0$, ou $p = 0$ e $\beta = 0$, então $X_t = \varepsilon_t, \forall t$. Assim, $\{X_t\}$ é uma sequência de variáveis aleatórias i.i.d. com uma distribuição Poisson de parâmetro λ .

O processo reduzido é chamado POMINAR(1)-V.

Baseado na forma como Li, Wang e Zhang (2015) definiram o processo ZIMINAR(1), o processo POMINAR(1), dado na Definição 4.1.1, pode ser reescrito como:

$$X_t = S_{1,t}(\alpha \circ X_{t-1} + \varepsilon_t) + S_{2,t}(\beta * X_{t-1} + \varepsilon_t), \quad (4.2)$$

em que $\mathbf{S}_t = (S_{1,t}, S_{2,t})^\top$, $t \geq 1$ é um vetor aleatório com função de probabilidade dada por

$$\begin{aligned} \mathbb{P}(\mathbf{S}_t = (1,0)^\top) &= p, \\ \mathbb{P}(\mathbf{S}_t = (0,1)^\top) &= 1 - p, \end{aligned} \quad (4.3)$$

em que \mathbf{S}_t é uma sequência de variáveis aleatórias Bernoulli independentes entre si, \mathbf{S}_t é independente de todo $\{B_i(\alpha)\}$, $\{P_i(\beta)\}$, ε_t e X_{t-k} para $k \geq 1$.

O objetivo de reescrevermos o processo segundo a Equação (4.2) é tornar algebricamente mais simples as demonstrações dos teoremas que garantirão a existência e unicidade do processo e as propriedades probabilísticas referentes ao POMINAR(1). Orozco (2017), em seu trabalho, demonstrou dois teoremas de fundamental importância na construção do processo POMINAR(1). Estes teoremas garantem a estacionariedade de segunda ordem, unicidade e ergodicidade do processo, a existência de esperança e variância condicional, e

a existência teórica das funções de autocovariância e autocorrelação. Estes teoremas são dados a seguir:

Teorema 4.1.1. *Existe uma única série para valores inteiros aleatórios estritamente estacionária $\{X_t\}$ que satisfaz a equação (4.2) e $Cov(X_k, \varepsilon_t) = 0$, para $k < t$. Os momentos incondicionais de primeira e segunda ordem da série estritamente estacionária $\{X_t\}$ existem, além disso o processo é ergódico.*

Como consequência direta da demonstração do Teorema 4.1.1, temos as seguintes expressões para os dois primeiros momentos de X_t :

$$E[X_t] = \frac{\lambda}{1 - C_1}, \quad (4.4)$$

$$E[X_t^2] = \frac{C_4 E[X_{t-1}] + \lambda^2 + \lambda}{1 - C_2}, \quad (4.5)$$

e utilizando as expressões (4.4) e (4.5) obtemos a seguinte expressão para a variância de X_t ,

$$Var(X_t) = \frac{\lambda^2[(1 - C_1)^2 - (1 - C_2)] + \lambda(1 - C_1)[C_4 + (1 - C_1)]}{(1 - C_2)(1 - C_1)^2},$$

em que, $C_1 = p\alpha + (1 - p)\beta$, $C_2 = p\alpha^2 + (1 - p)\beta^2$, $C_3 = p\alpha(1 - \alpha) + (1 - p)\beta$, $C_4 = C_3 + 2E[\varepsilon_t]C_1$.

Teorema 4.1.2. *Suponha que $\{X_t\}$ segue o processo POMINAR(1), então a média condicional k passos à frente e a variância condicional k passos à frente são,*

$$E[X_{t+k}|X_t] = C_1^k X_t + \frac{1 - C_1^k}{1 - C_1} E[\varepsilon_t]$$

e

$$Var(X_{t+k}|X_t) = (C_2^k - C_1^{2k})X_t^2 + (1 - C_2^k)E[X_t^2] + (C_1^{2k} - 1)(E[X_t])^2 + \left(\frac{C_4(C_1^k - C_2^k)}{C_1 - C_2} - 2C_1^k(1 - C_1^k)E[X_t] \right) (X_t - E[X_t]), \quad t \geq 1.$$

Com base nos Teoremas 4.1.1 e 4.1.2 podemos associar a esperança e variância condicional k passos à frente com a esperança e variância no instante de tempo t do seguinte modo:

$$\lim_{k \rightarrow \infty} E[X_{t+k}|X_t] = E[\varepsilon_t]/(1 - C_1) = E[X_t],$$

que é a média do processo no instante t , tal como

$$\lim_{k \rightarrow \infty} Var(X_{t+k}|X_t) = E[X_t^2] - (E[X_t])^2 = Var(X_t),$$

que é a variância do processo no instante t . Vale ressaltar que este resultado é válido quando $C_1, C_2 \in [0,1)$.

Após definidas e conhecidas $E[X_t]$, $E[X_t^2]$ e a $Var(X_t)$, tem-se que as funções de autocovariância e autocorrelação são dadas pelo Teorema a seguir.

Teorema 4.1.3. *Suponha que $\{X_t\}$ segue o processo POMINAR(1), definido pela equação (4.2), então a função de autocovariância e função de autocorrelação, respectivamente são,*

$$Cov(X_{t+k}, X_t) = p^2 \alpha^k \frac{\lambda}{1-\alpha} + (1-p)^2 \beta^k \frac{\lambda}{(1-\beta)(1-\beta^2)}$$

e

$$Corr(X_{t+k}, X_t) = \frac{p^2 \alpha^k \frac{\lambda}{1-\alpha} + (1-p)^2 \beta^k \frac{\lambda}{(1-\beta)(1-\beta^2)}}{\lambda^2 \frac{[(1-C_1)^2 - (1-C_2)] + \lambda(1-C_1)[C_4 + (1-C_1)]}{(1-C_2)(1-C_1)^2}}.$$

4.2 Estimação dos parâmetros

Quando tratamos da estimação de parâmetros de um modelo de séries temporais é usual a utilização de três métodos: máxima verossimilhança, máxima verossimilhança condicional e Yule-Walker. Porém, no contexto de séries temporais de valores inteiros, em geral, o estimador via máxima verossimilhança condicional (EMVC) e estimadores de Yule-Walker (EYW) são os mais utilizados. Deste modo, Orozco (2017) em seu trabalho comparou estes dois estimadores para avaliar suas propriedades em relação aos parâmetros α , β , λ e p presentes no processo POMINAR(1), comparação esta realizada através de estimativas pontuais, erro quadrático médio e viés dos estimadores. Estas quantidades comparativas foram obtidas via simulações de Monte Carlo para diversos cenários, os quais, contaram com diversas combinações de parâmetros e tamanhos amostrais.

Após a comparação dos estimadores de máxima verossimilhança condicional e Yule-Walker, Orozco (2017) sugeriu a utilização do EMVC para estimação dos parâmetros do processo POMINAR(1), pois em suas simulações apresentou um excelente desempenho em relação às estimativas pontuais e ao erro quadrático médio, sendo muito superior ao EYW.

O método de máxima verossimilhança condicional é baseado na função de verossimilhança condicionando a observação do tempo t (instante atual) à observação no tempo $t - k$ (o instante k passos atrás), em particular, quando $k = 1$, temos que a função de verossimilhança é condicionada ao tempo imediatamente anterior ao instante atual. Quando tratamos especificamente de um processo de séries temporais, a ideia geral do estimador de máxima verossimilhança condicional é obter as estimativas dos parâmetros utilizando a distribuição conjunta do processo $\{X_t\}$.

Deste modo, o método da máxima verossimilhança condicional pode ser expresso em função das probabilidades de transição do seguinte modo:

$$L(\alpha, \beta, \lambda, p | X) = \prod_{t=2}^T \mathbb{P}(X_t = j | X_{t-1} = i),$$

em que, $\mathbb{P}(X_t = j | X_{t-1} = i)$ são as probabilidades de transição. Para o processo POMINAR(1), a probabilidade de transição possui forma analítica, e pode ser escrita da seguinte forma:

$$\mathbb{P}(X_t = j | X_{t-1} = i) = p \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1 - \alpha)^{i-k} \left[\frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \right] + (1-p) \frac{e^{-(\beta i + \lambda)} (\beta i + \lambda)^j}{j!}. \quad (4.6)$$

A probabilidade dada em (4.6) pode ser interpretada como:

$$\mathbb{P}(X_t = j | X_{t-1} = i) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1 - \alpha)^{i-k} \left[\frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \right]$$

com uma determinada probabilidade p , pois com esta probabilidade o processo será INAR(1) e deste modo a probabilidade de transição resume-se a esta parcela exclusivamente, ou,

$$\mathbb{P}(X_t = j | X_{t-1} = i) = \frac{e^{-(\beta i + \lambda)} (\beta i + \lambda)^j}{j!},$$

com probabilidade $1 - p$, pois o processo será INARCH(1) e deste modo a probabilidade de transição será expressa apenas por esta parcela.

Esta interpretação decorre pela forma a qual o processo POMINAR(1) é construído, pois no instante de tempo t o processo será INAR(1) com probabilidade p , ou INARCH(1) com probabilidade $1 - p$.

A função log-verossimilhança condicional $\ell(\alpha, \beta, \lambda, p)$, para o processo POMINAR(1) é dada por:

$$\begin{aligned} \ell(\alpha, \beta, \lambda, p) &= \log \left[\prod_{t=2}^T \mathbb{P}(X_t = j | X_{t-1} = i) \right] \\ &= \sum_{t=2}^T \log [\mathbb{P}(X_t = j | X_{t-1} = i)] \\ &= \sum_{t=2}^T \log \left[p \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1 - \alpha)^{i-k} \left[\frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \right] + (1-p) \frac{e^{-(\beta i + \lambda)} (\beta i + \lambda)^j}{j!} \right]. \end{aligned}$$

Ao tentarmos maximizar a função log-verossimilhança condicional para o POMINAR(1), não encontramos solução analítica, deste modo devemos recorrer a métodos computacionais para obtermos $\hat{\alpha}$, $\hat{\beta}$, $\hat{\lambda}$ e \hat{p} que são os valores que maximizam a função log-verossimilhança e por sua vez serão as estimativas de máxima verossimilhança condicional para os parâmetros $\alpha, \beta, \lambda, p$.

5 Gráfico de Shewhart para o processo PO-MINAR(1)

Conforme comentado no Capítulo 1, os gráficos usuais, em particular o de Shewhart, apresentam baixo desempenho na detecção de alterações na média quando o processo apresenta estrutura de autocorrelação, ou seja, para processos autocorrelacionados. Sendo assim, o presente capítulo tem como objetivo apresentar um método computacional, baseado no método proposto por Gandy e Kvaløy (2013), de obtenção dos limites de controle para o gráfico de Shewhart para o processo POMINAR(1) definido no Capítulo 4.

5.1 Construção dos limites de controle

Os limites de controle possuem uma relação direta com o número médio de amostras até o alarme falso (NMAF) e com o número médio de amostras até a detecção de um alarme verdadeiro (NMA), visto que, conforme descrito na Seção 2.2, eles são expressos pela seguinte expressão

$$NMA = \frac{1}{p_0} \text{ e}$$

$$NMAF = \frac{1}{\alpha_0},$$

em que as probabilidades α_0 e p_0 dependem dos limites de controle. Visto isso, é de interesse encontrar limites de controle que forneçam um NMAF satisfatoriamente alto, ou seja, uma probabilidade α_0 baixa, e um NMA satisfatoriamente baixo, ou seja, uma probabilidade p_0 alta, em que à medida que o desvio em relação à média aumente o NMA diminua.

Usualmente as probabilidades α_0 e p_0 são dadas pela seguinte expressão:

$$\alpha_0 = P(\bar{X} > LSC | \mu = \mu_0) + P(\bar{X} < LIC | \mu = \mu_0) \quad (5.1)$$

$$p_0 = P(\bar{X} > LSC | \mu = \mu_0 + \delta\sigma_0) + P(\bar{X} < LIC | \mu = \mu_0 + \delta\sigma_0),$$

em que μ é a média do processo no instante atual, μ_0 e σ_0 a média e o desvio padrão do processo sob controle, respectivamente, e δ indica quantos desvios padrão a média do processo sob controle foi afetada por alguma mudança não aleatória.

Percebe-se pela expressão 5.1 que os limites de controle podem ser vistos como quantis da distribuição da média que forneçam uma determinada probabilidade (α_0 ou p_0)

e desde modo fixando esta probabilidade podemos obter os limites como sendo os quantis da distribuição que forneçam essa probabilidade e obtendo assim um certo controle em relação ao NMA e NMAF já que eles dependem diretamente dessas probabilidade que são funções do LSC e LIC.

Conforme visto no Capítulo 2 usualmente o gráfico de Shewhart considera a probabilidade $\alpha_0 = 0,0027$, obtendo assim um NMAF=371, o que proporciona os famosos limites 3σ . Porém, no contexto de processo autocorrelacionados, em particular, para processos de valores inteiros não negativos, como o caso do POMINAR(1), é comum adotar o LIC igual a zero, visto que nenhum valor da média será inferior a zero, e focar geralmente em desvio positivos da média do processos como exemplifica e comenta Rakitzis e Castagliola (2016) em seu trabalho. Desde modo, para fins deste trabalho iremos considerar apenas desvio positivos na média do processo e o LIC=0.

Outro fato para considerar o LIC=0 para o contexto do pominar é a sobredispersão do processo para uma série de combinação de parâmetros, pois dados provenientes do processo POMINAR(1) possuem, em muitos casos, a média menor que a variância, como podemos perceber na Tabela 1. Este fato proporcionaria para muitas combinações de parâmetros um LIC que gerasse um grande número de alarmes falsos devido a variabilidade dos dados, por isso é de maior interesse neste trabalho os desvio positivos em relação a média.

Tabela 1 – Combinação de parâmetros que ilustram a sobredispersão do processo.

α	β	λ	p	μ_X	σ_X^2	σ_X^2/μ_X
0,3	0,3	2	0,3	2,86	3,05	1,06
0,4	0,6	3	0,4	6,25	8,65	1,38
0,4	0,5	5	0,5	9,09	10,78	1,18
0,6	0,9	7	0,6	25,00	71,96	2,87
0,7	0,9	9	0,4	50,00	201,87	4,03

Conforme comentado, podemos ver os limites de controle como quantis que deixam determinada probabilidade a esquerda, assim para o processo POMINAR(1) a ideia geral para obtenção do LSC baseia-se na ideia proposta por Gandy e Kvaløy (2013) que tem como princípio as seguintes etapas:

- O limite será baseado em um quantil o qual nos fornecerá uma probabilidade de 0,0027 a direita, conforme o gráfico usual de Shewhart;
- Com base em uma amostra do processo POMINAR(1) de tamanho m , com subamostras de tamanho n , estimar via Método de Monte Carlo o quantil amostral que nos forneça a probabilidade desejada (0,0027);

Por fim, avaliaremos o gráfico de controle com base no LSC, estimado, através do NMA e NMAF que serão obtidos via simulação de Monte Carlo.

5.2 Obtenção do limite superior de controle

A obtenção do limite superior de controle para o gráfico de Shewhart do processo POMINAR(1) foi feita com base no método de Monte Carlo, com número de réplicas igual a 1000, e seguiu o seguinte algoritmo:

1. Define-se os parâmetros do processo e o tamanho de cada subamostra, denotado por n . Fixou-se o tamanho amostral $m = 500$, em que m denota o tamanho da amostra;
2. Gera-se uma amostra do processo POMINAR(1) de tamanho $500 \times n$, em que n e os parâmetros do processo foram definidos no passo 1.
3. Agrupam-se as observações, sequencialmente, em 500 grupos de tamanho n e obtêm-se as médias de cada um dos 500 grupos, formando assim uma nova amostra de tamanho 500, formada agora pelas médias de cada grupo.
4. Estima-se o quantil da distribuição das 500 médias amostrais, obtidas no passo 3, que acumula uma probabilidade de 0,9973 à esquerda;
5. Repetem-se, de forma independente, 1000 vezes os passos 2,3 e 4, e utiliza-se o LSC como a média dos 1000 quantis estimados.

O esquema abaixo ilustra o algoritmo utilizado para obtenção do LSC.

Réplica 1	Réplica 2	...	Réplica 1000
\bar{x}_1	\bar{x}_1	...	\bar{x}_1
\vdots	\vdots	\vdots	\vdots
\bar{x}_{500}	\bar{x}_{500}	...	\bar{x}_{500}
\hat{q}_1	\hat{q}_2	...	\hat{q}_{1000}

$$\widehat{LSC} = \frac{\sum_{i=1}^{1000} \hat{q}_i}{1000},$$

em que \bar{x}_j é a média da j -ésima amostra após o agrupamento, $j = 1, \dots, 500$, e \hat{q}_i é o quantil amostral da i -ésima réplica $i = 1, \dots, 1000$.

A rotina computacional, programada utilizando o *software* R Core Team (2017), versão 3.4.1 para a plataforma Windows, das funções que geram o LSC pode ser vista no Apêndice A.

Após a obtenção do LSC, baseado no algoritmo descrito anteriormente, avaliou-se o desempenho do gráfico baseado no NMA e no NMAF. Essas quantidades foram obtidas através de 10000 réplicas de Monte Carlo e com o auxílio do *software* R Core Team (2017), em que o código pode ser visto nos Apêndices B e C. O algoritmo para obtenção do NMA e NMAF é dado a seguir:

1. Gera-se uma amostra do processo POMINAR(1) de tamanho $5000 \times n$, em que n e os parâmetros do processo são os mesmo utilizados na obtenção do LSC.
2. Defini-se um valor δ que representa a alteração do processo ponderada pelo desvio padrão da média, σ_0/\sqrt{n} , em que σ_0 é o desvio padrão do processo sob controle ;
3. Agrupam-se as observações, sequencialmente, em 5000 grupos de tamanho n e obtêm-se as médias de cada um dos 5000 grupos, formando assim uma nova amostra de tamanho 5000, formada agora pelas médias de cada grupo;
4. Compara-se o LSC com as 5000 médias acrescidas de $\delta \times \sigma_0/\sqrt{n}$, e armazena-se a posição da primeira amostra a qual o valor da média amostral, acrescida, foi superior ao LSC ;
5. Repete-se, de forma independente, 10000 vezes os passos 3 e 4, e obtêm-se o NMA ou NMAF com base na média das posições obtidas no passo 4.

Quando deseja-se obter o NMAF, utiliza-se o algoritmo anterior com $\delta = 0$, ou seja, sem alteração na média do processo.

O esquema a seguir ilustra o algoritmo de obtenção do NMAF e NMA.

Réplica 1	Réplica 2	...	Réplica 10000
$\bar{x}_1 + \delta \frac{\sigma_0}{n}$	$\bar{x}_1 + \delta \frac{\sigma_0}{n}$...	$\bar{x}_1 + \delta \frac{\sigma_0}{n}$
\vdots	\vdots	\vdots	\vdots
$\bar{x}_{5000} + \delta \frac{\sigma_0}{n}$	$\bar{x}_{5000} + \delta \frac{\sigma_0}{n}$...	$\bar{x}_{5000} + \delta \frac{\sigma_0}{n}$
a_1	a_2	...	a_{10000}

$$\widehat{NMA} \text{ ou } \widehat{NMAF} = \frac{\sum_{i=1}^{10000} a_i}{10000},$$

em que os a_i , $i = 1, 2, \dots, 10000$, representa o número da primeira amostra a qual o gráfico detectou a alteração na média na i -ésima réplica de Monte Carlo.

5.3 Obtendo os limites em uma situação prática

Até o presente momento vimos como obter os limites de controle para uma situação teórica, ou seja, assumindo que o processo POMINAR(1) é o melhor modelo e que os valores dos parâmetros são conhecidos, pois trata-se de simulações. Porém, em uma situação prática não se tem o conhecimento de fato dos parâmetros do processo. Deste modo, para obtenção dos limites de controle em uma situação real devemos seguir o seguinte algoritmo:

1. Obter uma amostra do processo quando ele estiver sob controle, de tamanho $m \geq 100$, com subamostras de tamanho $n = 1$;
2. Ajustar alguns modelos, baseados na amostra, e verificar através de critérios de comparação, como por exemplo AIC (Critério de Informação Akaike) e BIC (Critério de Informação Bayesiano), qual modelo melhor se adequa ao dados;
3. Caso o processo POMINAR(1) seja um bom modelo para os dados, utiliza-se a amostra coletada no passo 1 para estimar os parâmetros do processo via máxima verossimilhança condicional, conforme recomendado na Seção 4.2;
4. Com base nos parâmetros estimados no passo 3 segue-se o mesmo procedimento de obtenção do LSC descrito na Seção 5.2.

Após a obtenção do LSC, têm-se a ferramenta para poder monitorar as observações posteriores as que se fazem presente na amostra sob controle.

6 Análise de desempenho do gráfico para o processo POMINAR(1)

No presente capítulo serão apresentados e analisados os resultados das simulações, que seguiram os procedimentos descritos no Capítulo 5. Este estudo de simulação tem como finalidade verificar o desempenho, ou eficiência, do gráfico de Shewhart para o processo POMINAR(1), pelo método proposto, através do NMA e NMAF para 50 cenários distintos. Estes cenários são compostos pelas combinações de parâmetros presentes na Tabela 1 e uma variação do tamanho de cada subamostra, $n = 1, \dots, 10$.

As combinações de parâmetros apresentadas na Tabela 1 foram escolhidas com base em Orozco (2017), trabalho que propôs o processo POMINAR(1), pois além destas combinações apresentarem distintos valores da média, são combinações que proporcionam a estacionariedade do processo em questão. Em relação às propriedades de detecção do gráfico, foram considerados 5 diferentes desvios na média, ou seja, $\delta = 0; 1,5; 2; 2,5; 3$, em que quando $\delta = 0$ representa o NMAF, pois não houve alteração na média.

Neste estudo de simulação obtivemos o LSC com base nos procedimentos descritos no Capítulo 5 e utilizando os valores verdadeiros dos parâmetros na obtenção do LSC. Como este estudo apresenta a proposta de monitorar amostras com subgrupos de tamanho $n \geq 1$, o LSC não foi expresso de forma inteira, visto que as médias de observações do POMINAR(1) poderão assumir valores em \mathbb{R}^+ (reais positivos incluindo o zero).

A análise do desempenho do gráfico foi feita com base nos NMA's e NMAF's, expressos em valores inteiros, pois indicam o número esperado de amostras até a detecção de algum alarme no gráfico, seja ele verdadeiro ou falso, sendo assim expressando estas quantidades em valores inteiros facilita a interpretação prática e as comparações entre os cenários. O valor inteiro atribuído ao NMA ou NMAF é o menor inteiro, o qual seja maior que o número real obtido na simulação. Por exemplo, se obtivermos na simulação um NMA simulado de 15,03 amostras o seu valor será expresso por 16 amostras.

A primeira combinação de parâmetros, $\alpha=0,3$, $\beta=0,3$, $\lambda=2$ e $p=0,3$, considerada neste estudo é a que proporciona um menor valor da média e menor desvio padrão. Deste modo é esperado que desvio menores possam ser detectados mais rapidamente que em situações com médias e desvios padrão maiores, pois a variação dos dados neste contexto, apesar da sobredispersão do processo, não é tão grande em relação a média. Na Tabela 2 encontram-se os resultados das simulações para a primeira combinação de parâmetros. Pode-se perceber que para esta combinação de parâmetros à medida que os desvios em relação a média, sob controle, aumenta o gráfico detecta tal alteração no processo de forma

mais rápida, se aproximando de 3 amostras à medida que a alteração se aproxima de três desvios padrão. Nesta configuração de parâmetros, nota-se que NMA's para os diferentes tamanho de subamostras não diferem muito. Sendo assim, em uma situação prática na qual o processo possua uma configuração de parâmetros que forneça uma média e desvio padrão relativamente baixos, similares a esta configuração, o tamanho das subamostras não influenciam diretamente na detecção de alterações na média, deste modo recomenda-se a utilização de subamostras de tamanho $n = 2$, pois obteve-se bons resultados em relação à detecção de alterações e apresentou o segundo melhor NMAF para esta combinação.

Tabela 2 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,3$, $\beta=0,3$, $\lambda=2$ e $p=0,3$.

n	LSC	δ				
		0	1,5	2	2,5	3
1	8,52	260	16	15	8	4
2	6,69	306	17	9	6	3
3	5,92	233	20	8	5	3
4	5,47	220	15	7	5	3
5	5,18	225	19	9	5	3
6	4,96	248	16	8	4	3
7	4,78	277	14	8	4	3
8	4,65	325	13	7	4	3
9	4,53	231	16	7	3	3
10	4,45	287	14	7	4	2

Na Tabela 3 estão dispostos os resultados referentes à seguinte combinação de parâmetros: $\alpha=0,4$, $\beta=0,6$, $\lambda=3$ e $p=0,4$. Esta combinação proporciona uma média e desvio padrão não tão altos, porém maiores que na combinação 1. Nesta configuração de parâmetros pode-se perceber o início da influência do tamanho das subamostras na detecção de desvios na média, apesar de não ser tão chamativa existe uma diferença, até considerável, quando se trata do NMA passando-se de $\delta = 1,5$ para $\delta = 2$. Outro fato que salta aos olhos é a eficiência do gráfico, neste cenário, quando $n = 1$. Este fato mostra que em situações práticas, nas quais a média e desvio padrão sejam similares ao desta configuração, realizar o monitoramento da média utilizando $n = 1$ proporciona um excelente desempenho e um ótimo custo benefício por utilizar-se de subamostras de tamanho 1.

Tabela 3 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,4$, $\beta=0,6$, $\lambda=3$ e $p=0,4$.

n	LSC	δ				
		0	1,5	2	2,5	3
1	16,31	348	16	16	8	4
2	12,88	274	25	12	7	4
3	11,54	299	20	12	6	4
4	10,69	264	15	7	5	3
5	10,18	263	21	9	5	3
6	9,83	267	20	9	5	3
7	9,53	281	16	10	5	3
8	9,29	305	13	7	4	3
9	9,11	351	17	9	4	3
10	8,96	304	17	8	5	3

Os resultados da terceira combinação de parâmetros, $\alpha=0,4$, $\beta=0,5$, $\lambda=5$ e $p=0,5$, estão presentes na Tabela 4. Esta combinação de parâmetros apresenta, similar a primeira combinação, um índice de dispersão próximo de 1,1, porém com média e variância maiores que na primeira combinação. Similar a primeira configuração de parâmetros, este cenário não apresenta uma evidente influência do tamanho das subamostras na detecção de alterações na média. Em contra partida, destaca-se os tamanhos de subamostras $n = 2,3$ que fornecem excelentes NMAF. Deste modo, recomenda-se que em situações similares a esta utilize-se tamanhos de subamostras iguais a 2 ou 3 pois fornecem um excelente NMAF e uma rápida detecção da alteração na média, principalmente quando δ aproxima-se de 3.

Tabela 4 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,4$, $\beta=0,5$, $\lambda=5$ e $p=0,5$.

n	LSC	δ				
		0	1,5	2	2,5	3
1	19,48	293	23	10	7	4
2	16,11	344	21	8	5	3
3	14,72	348	17	8	4	3
4	13,86	289	16	8	4	3
5	13,34	273	15	9	4	3
6	12,96	274	15	7	4	3
7	12,64	295	15	8	4	3
8	12,38	319	15	7	4	3
9	12,19	270	13	8	4	3
10	12,39	320	14	7	4	2

A quarta configuração de parâmetros, $\alpha=0,6$, $\beta=0,9$, $\lambda=7$ e $p=0,6$, diferentes das outras três configurações apresentadas até agora, possui uma variância que chega a ser quase

três vezes o valor da média. Neste cenário é importante verificar qual o comportamento do gráfico, principalmente na detecção de desvios na média, pois ele ilustra situações com coeficientes de variação grandes e média maior que dos cenários anteriores.

Os resultados das simulações para a quarta configuração são apresentados na Tabela 5. Nota-se que por causa da grande variabilidade dos dados em relação à média, a detecção de desvios com $\delta \leq 2$ não apresentou resultados tão bons quanto os das três configurações anteriores, que possuíam médias e variâncias menores. Em contra partida, o procedimento demonstrou eficiência na detecção de grandes desvios, como para $\delta = 2,5$ e 3, em que os valores dos NMA's estão em torno de 10 a 4 amostras para detectar tais alterações. Outro ponto importante foi o fato de para este cenário os NMAF's serem todos superiores a 345 amostras. Para combinações de parâmetros que sejam similares à esta configuração, aconselha-se usar o gráfico proposto quando desejar monitorar grandes alterações na média, visto que ele apresentou desempenho satisfatório para $\delta = 2,5$ e 3 e ótimos valores de NMAF. Em relação a influência, na detecção, dos tamanhos das subamostras percebe-se que à medida que n aumenta o gráfico tende a detectar alterações mais rápido, e o NMAF diminui um pouco, porém continua em uma faixa de segurança muito boa o que possibilita a utilização do gráfico.

Tabela 5 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,6$, $\beta=0,9$, $\lambda=7$ e $p=0,6$.

		δ				
n	LSC	0	1,5	2	2,5	3
1	57,06	424	66	33	9	10
2	45,49	424	48	25	13	7
3	41,11	379	39	20	11	6
4	38,64	375	34	18	9	5
5	36,90	355	31	16	8	5
6	35,65	365	31	15	8	4
7	34,82	347	30	15	8	4
8	34,18	372	30	15	8	4
9	33,51	356	30	15	8	4
10	34,15	380	27	14	7	4

A última configuração de parâmetros presente neste estudo de simulação consta com $\alpha=0,7$, $\beta=0,9$, $\lambda=9$ e $p=0,4$, e apresenta a maior média e o maior coeficiente de variação. O objetivo geral de avaliar o gráfico neste cenário é verificar seu poder de detecção em situações que possuam um valor da média alto, e um grande coeficiente de variação, coeficiente este que para esta situação específica é 4.

Na Tabela 6 encontram-se os resultados das simulações para a quinta configuração dos parâmetros. Percebe-se que os resultados presentes nesta tabela, assemelham-se com

os resultados obtidos ao analisarmos a Tabela 5 para a quarta combinação, ou seja, para desvios com $\delta \leq 2$ o gráfico proposto não apresentou resultados tão bons quanto nos três cenários iniciais. Porém, esta configuração apresentou excelentes NMAF's todos maiores que 395 amostras. Com base nesta simulação, similar aos resultados para a quarta combinação, é recomendado a utilização deste gráfico, para cenários similares ao desta configuração de parâmetros, apenas quando deseja-se detectar desvio de alta magnitude, ou seja, $\delta \geq 2,5$. Em relação ao número de subamostras, eles influenciam de forma positiva na detecção de desvios na média, porém, como já é característico do gráfico, esta influência diminui à medida que o desvio aumenta, estabilizando o NMA próximo a 4 amostras.

Tabela 6 – NMA's e NMAF's do gráfico proposto para a combinação de parâmetros: $\alpha=0,7$, $\beta=0,9$, $\lambda=9$ e $p=0,4$.

n	LSC	δ				
		0	1,5	2	2,5	3
1	97,68	407	56	31	16	9
2	81,57	429	46	23	12	6
3	74,97	396	39	19	10	6
4	71,43	422	37	18	10	5
5	68,70	398	33	14	9	5
6	67,04	428	33	17	8	5
7	65,60	421	31	15	8	4
8	64,57	419	32	16	8	4
9	63,58	416	30	16	8	4
10	64,59	424	30	15	8	4

Baseado nas análises feitas sobre os resultados das simulações, nota-se o bom desempenho do gráfico proposto para o processo POMINAR(1) nos mais distintos cenários. Especialmente quando estamos interessados em detectar desvios com $\delta \geq 2$ o gráfico apresentou excelentes NMA's, ou seja, se aproximando de 3 amostras, e ótimos NMAF's (todos acima de 220 amostras), para todos os cenários. Outro fator importante, que ressalta nas análises, é a necessidade de tamanho de subamostras pequenas para detectar os desvios, o que influência de forma positiva na utilização prática do gráfico, visto que subamostras menores, em geral, geram situações com menos custos e possibilitam em muitos casos o incentivo da utilização do gráfico de controle para monitoramento do processo.

7 Considerações Finais

Neste trabalho foi proposto e avaliado um método de monitoramento para a média do processo POMINAR(1), proposto por Orozco (2017), através do contexto do controle estatístico de processos, mais especificamente gráficos de controle. Deste modo, apresentamos uma revisão bibliográfica dos conceitos iniciais do controle estatístico de processo, do impacto da autocorrelação nos gráficos usuais e foram definidos os processos INAR(1), Poisson INAR(1) e INARCH(1) processos estes que compõem o processo misto POMINAR(1).

Desenvolvemos rotinas computacionais que permitem gerar valores e estimar os parâmetros do processo POMINAR(1), obtivemos, baseado nas ideias de Gandy e Kvaløy (2013) e através de simulações, os limites superiores de controle para o processo e propusemos também como obter estes limites de controle em situações práticas, em que o POMINAR(1) seja um modelo adequado ao processo.

Com base nas simulações realizadas, pode-se perceber uma boa eficiência do gráfico proposto, visto que para combinações de parâmetros distintas ele apresentou o número médio de amostras até um alarme falso em torno de 240 a 400 amostras, os quais variam de acordo com as combinações dos parâmetros. Porém, este número médio, em geral, é bem satisfatório, visto que estamos no contexto de processos autocorrelacionados. Quando refere-se à detecção de desvios na média, o gráfico pelo método proposto apresentou que para situações com médias sob controle menores que 10 o número médio de amostras até detectar um alarme verdadeiro fica em torno 20 e 3 amostras. Em contra partida para médias mais altas, superiores a 25 o gráfico mostrou ser eficaz apenas na detecção de desvios com $\delta \geq 2$. Este fato pode ser explicado devido à sobredispersão dos dados, ou seja, a média é menor que a variância, implicando que quanto maior a média do processo maior será sua variância tendo assim impacto direto no poder de detecção do gráfico. Outro ponto de destaque é que independente do valor da média do processo, à medida que o valor δ , relacionado ao desvio na média do processo, se aproxima de 3 o número de amostras até sinalizar a alteração se aproxima de 3.

As simulações apresentaram que o gráfico de Shewhart proposto tem uma excelente aplicabilidade, visto que na maioria dos casos, tamanhos de subamostras pequenos, como 1, 2 e 3, são suficientes para fornecer um bom poder de detecção e um número médio de amostras para sinalizar um alarme falso alto, o que em situações práticas é ideal pois gera um custo menor para o produtor, porém, não trás perdas em relação ao monitoramento do processo, ou seja, com as subamostras exemplificadas o gráfico detecta bem os desvios na média e demorará a sinalizar algo que não seja uma alteração verdadeira na média.

Deste modo concluímos que o gráfico de Shewhart proposto para o processo misto de séries temporais para valores inteiros POMINAR(1), apresentou um ótimo desempenho, principalmente na detecção de grandes desvios na média e mostrou-se uma excelente opção de monitoramento de processos em situações práticas.

7.1 Sugestões de trabalhos futuros

Algumas possíveis continuações do que foi feito neste trabalho são: propor os limites de controle para o processo POMINAR(1) de forma analítica, considerar cenários em que ocorram alterações na média e no desvio padrão do processo e ajustar o método para um gráfico que permita detectar pequenos desvios na média como os gráficos CUSUM e EWMA.

APÊNDICE A – Funções utilizadas na estimação do LSC

```
##### Código MONOGRAFIA: LUCAS SALES #####

##### FUNÇÕES #####
#####

#Função que gera os valores do POMINAR(1)
rpominar=function(theta,n){
  alpha=theta[1]
  betha=theta[2]
  lambda=theta[3]
  p=theta[4]
  c1=p*alpha+betha*(1-p)
  y=epsilon=NULL
#Valor inicial da série temporal sendo a parte inteira da média
  y[1]= round(lambda/(1-c1),0)
  for(i in 2:(n+300)){
    epsilon[i]= rpois(1,lambda) #Erro com distribuição poisson lambda
    u=runif(1) #Se a uniforme < P ele é inar(1), c.c. é inarch(1)
    if(u<p) {
      y[i]=rbinom(1,y[i-1],alpha) + epsilon[i] #INAR(1)
    } else{
      y[i]=rpois(1,y[i-1]*betha) + epsilon[i] #INARCH(1)
    }
  }
  return(y[301:(300+n)])
}

#Função logverssomilhaça do POMINAR(1)
lik=function(theta,x){
  #Theta é o vetor de parâmetro alpha,betha,lambda e p
  alpha=theta[1]
  betha=theta[2]
```

```

lambda=theta[3]
p=theta[4]
n=length(x) #Tamanho amostral
##### OS SOMATÓRIOS DO LOG EM FORMA DE FUNÇÃO#####
probtransi=function(theta,x,t){
  soma=0 #Começando a variável soma como 0
  aux=NULL
  for( k in 0:min(x[t-1],x[t])){
    aux[k+1]=rbinom(1,1,p)
    if(aux[k+1]==1){
      soma=soma+ (dbinom(k,x[t-1],alpha) * dpois(x[t]-k,lambda) )
    }else{
      soma=soma+(dpois(x[t],betha*x[t-1]+lambda))
    }
  }
  return(soma)
}

##### O SOMATÓRIO DO T (A LOG VEROSSIMILHANÇA FINAL!) #####
loglik=function(theta,x){
  -sum(sapply(2:n,function(t){
    log(probtransi(theta,x,t)) })))
}
est = try(optim(theta,
                function(theta){loglik(theta,x)},
                control=list(fnscale=-1),lower=c(0.01,0.01,0,0.01)
                ,upper=c(.99,.99,Inf,1) , method="L-BFGS-B"),T)est$par
}

#Função que gera os quantis empíricos do POMINAR(1)
qmeanpominar=function(b,theta,m,n){
  q1=matrix(rep(0,b*2),ncol=2)
  for(i in 1:b){
    aux=matrix(rpominar(theta,n*m),nrow=m)
    dados=rowMeans(aux)
    q1[i,]=quantile(dados,probs=c(0,0.9973))
    #q1[i,]=quantile(dados,probs=c(0.025,0.975))
  }
}

```

```
(q=colMeans(q1))

}

#Função que encontra os ARLS para o POMINAR(1)
ARL=function(theta,q,delta,m,n){
  alpha=theta[1]
  betha=theta[2]
  lambda=theta[3]
  p=theta[4]
  c1=p*alpha+betha*(1-p)
  c2=p*alpha^2+betha^2*(1-p)
  c3=p*alpha*(1-alpha)+betha*(1-p)
  c4=c3+2*lambda*c1
  #Média teórica e variâncias teóricas
  mu=lambda/(1-c1)
  sig2hat=(lambda^2*((1-c1)^2 -(1-c2)) + ( lambda*(1-c1)*(c4+1-c1))/
  (( 1-c2)*(1-c1)^2) )
  lsc=q[2];lic=q[1]
  aux=matrix(rpominar(theta,n*m),nrow=m)
  dados=rowMeans(aux)+delta*sqrt(sig2hat/n)
  ( nma=which(dados>lsc)[1] )

  # if(nma==NA){
  #   nma=m
  # }
  return(nma)
}
```

APÊNDICE B – Rotinas computacionais para geração do NMAF

```
##### SIMULAÇÕES #####
#####
source(Funções.R)
#####
##### COMBINAÇÃO 1 #####

#Parâmetros do processo
alpha=.3;betha=0.3;p=.3;lambda=2;m=100;n=1;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)

#NMAF
nmaf=NULL
m=5000;
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
delta=0 #Variação da média
for(i in 1:r){
  print(i)
  nmaf[i]=ARL(theta,q,delta,m,n)
}
mean(nmaf)

#####
##### COMBINAÇÃO 2 #####

#Parâmetros do processo
alpha=.4;betha=0.6;p=.4;lambda=3;m=100;n=1;b=1000;
```

```

r=10000;theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)

#NMAF
nmaf=NULL
m=5000
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
delta=0 #Variação da média
for(i in 1:r){
  print(i)
  nmaf[i]=ARL(theta,q,delta,m,n)
}
mean(nmaf)

#####
##### COMBINAÇÃO 3 #####

#Parâmetros do processo
alpha=.4;betha=0.5;p=.5;lambda=5;m=100;n=1;b=1000;r=10000
;theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)

#NMAF
nmaf=NULL
m=5000
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
delta=0 #Variação da média
for(i in 1:r){
  print(i)
  nmaf[i]=ARL(theta,q,delta,m,n)
}
mean(nmaf)

#####

```

```
##### COMBINAÇÃO 4 #####
```

```
#Parâmetros do processo
alpha=.6;betha=0.9;p=.6;lambda=7;m=100;n=1;b=1000;r=10000
;theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)

#NMAF
nmaf=NULL
m=5000
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
delta=0 #Variação da média
for(i in 1:r){
  print(i)
  nmaf[i]=ARL(theta,q,delta,m,n)
}
mean(nmaf)
```

```
#####
##### COMBINAÇÃO 5 #####
```

```
#Parâmetros do processo
alpha=.7;betha=0.9;p=.4;lambda=9;m=100;n=1;b=1000;r=10000
;theta=c(alpha,betha,lambda,p)
##Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)

#NMAF
nmaf=NULL
m=5000
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
delta=0 #Variação da média
for(i in 1:r){
  print(i)
  nmaf[i]=ARL(theta,q,delta,m,n)
```

```
}  
mean(nmaf)
```

APÊNDICE C – Rotinas computacionais para geração do NMA

```
##### SIMULAÇÕES NMA #####
#####
source(Funções.R)
#####
##### COMBINAÇÃO 1 #####

#Parâmetros do processo
alpha=.3;betha=0.3;p=.3;lambda=2;m=100;n=7;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado

#NMA
nma1=NULL
m=1000;
delta=1.5
for(i in 1:r){
  print(i)
  nma1[i]=ARL(theta,q,delta,m,n)
}
nma2=NULL
m=1000;
delta=2
for(i in 1:r){
  print(i)
  nma2[i]=ARL(theta,q,delta,m,n)
}
nma3=NULL
m=1000;
```

```
delta=2.5
for(i in 1:r){
  print(i)
  nma3[i]=ARL(theta,q,delta,m,n)
}
nma4=NULL
m=1000;
delta=3
for(i in 1:r){
  print(i)
  nma4[i]=ARL(theta,q,delta,m,n)
}

#####
##### COMBINAÇÃO 2 #####

#Parâmetros do processo
alpha=.4;betha=0.6;p=.4;lambda=3;m=100;n=7;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado

#NMA
nma1=NULL
m=1000;
delta=1.5
for(i in 1:r){
  print(i)
  nma1[i]=ARL(theta,q,delta,m,n)
}
nma2=NULL
m=1000;
delta=2
for(i in 1:r){
  print(i)
```

```
    nma2[i]=ARL(theta,q,delta,m,n)
}
nma3=NULL
m=1000;
delta=2.5
for(i in 1:r){
  print(i)
  nma3[i]=ARL(theta,q,delta,m,n)
}
nma4=NULL
m=1000;
delta=3
for(i in 1:r){
  print(i)
  nma4[i]=ARL(theta,q,delta,m,n)
}
```

```
#####
##### COMBINAÇÃO 3 #####
```

```
#Parâmetros do processo
alpha=.4;betha=0.5;p=.5;lambda=5;m=100;n=7;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
```

```
#NMA
nma1=NULL
m=1000;
delta=1.5
for(i in 1:r){
  print(i)
```

```

    nma1[i]=ARL(theta,q,delta,m,n)
}
nma2=NULL
m=1000;
delta=2
for(i in 1:r){
  print(i)
  nma2[i]=ARL(theta,q,delta,m,n)
}
nma3=NULL
m=1000;
delta=2.5
for(i in 1:r){
  print(i)
  nma3[i]=ARL(theta,q,delta,m,n)
}
nma4=NULL
m=1000;
delta=3
for(i in 1:r){
  print(i)
  nma4[i]=ARL(theta,q,delta,m,n)
}

#####
##### COMBINAÇÃO 4 #####

#Parâmetros do processo
alpha=.6;betha=0.9;p=.6;lambda=7;m=100;n=7;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
#Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado

#NMA
nma1=NULL
m=1000;

```

```
delta=1.5
for(i in 1:r){
  print(i)
  nma1[i]=ARL(theta,q,delta,m,n)
}
nma2=NULL
m=1000;
delta=2
for(i in 1:r){
  print(i)
  nma2[i]=ARL(theta,q,delta,m,n)
}
nma3=NULL
m=1000;
delta=2.5
for(i in 1:r){
  print(i)
  nma3[i]=ARL(theta,q,delta,m,n)
}
nma4=NULL
m=1000;
delta=3
for(i in 1:r){
  print(i)
  nma4[i]=ARL(theta,q,delta,m,n)
}

#####
##### COMBINAÇÃO 5 #####

#Parâmetros do processo
alpha=.7;betha=0.9;p=.4;lambda=9;m=100;n=7;b=1000;r=10000;
theta=c(alpha,betha,lambda,p)
##Amostra sob controle
aux=matrix(rpominar(theta,n*m),nrow=m)
amostra0=rowMeans(aux)
q=qmeanpominar(b,theta,500,n) #Quantil com parâmetro estimado
```

```
#NMA
nma1=NULL
m=1000;
delta=1.5
for(i in 1:r){
  print(i)
  nma1[i]=ARL(theta,q,delta,m,n)
}
nma2=NULL
m=1000;
delta=2
for(i in 1:r){
  print(i)
  nma2[i]=ARL(theta,q,delta,m,n)
}
nma3=NULL
m=1000;
delta=2.5
for(i in 1:r){
  print(i)
  nma3[i]=ARL(theta,q,delta,m,n)
}
nma4=NULL
m=1000;
delta=3
for(i in 1:r){
  print(i)
  nma4[i]=ARL(theta,q,delta,m,n)
}
```

Referências

- AL-OSH, M.; ALZAID, A. A. First-order interger-valued autoregressive (inar(1)) process. *Journal of Time Series Analysis*, Wiley Online Library, v. 8, n. 3, p. 261–275, 1987.
- CLARO, F. A. E.; COSTA, A. F. B.; MACHADO, M. A. G. Gráficos de controle de ewma e de x-barra para monitoramento de processos autocorrelacionados. *Produção*, Associação Brasileira de Engenharia de Produção, p. 536–546, 2007.
- COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. *Controle estatístico de qualidade*. [S.l.]: Atlas São Paulo, 2005.
- FERLAND, R.; LATOUR, A.; ORAICHI, D. Integer-valued garch process. *Journal of Time Series Analysis*, Wiley Online Library, v. 27, n. 6, p. 923–942, 2006.
- FERNÁNDEZ, L. M. Z. Forecasting time series with integer values. Universidade Federal de Pernambuco, 2013.
- FREELAND, R. K. *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. Tese (Doutorado) — University of British Columbia, 1998.
- GANDY, A.; KVALØY, J. T. Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 40, n. 4, p. 647–668, 2013.
- HEINEN, A. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.
- JIN-GUAN, D.; YUAN, L. The integer-valued autoregressive (inar (p)) model. *Journal of time series analysis*, Wiley Online Library, v. 12, n. 2, p. 129–142, 1991.
- LI, C.; WANG, D.; ZHANG, H. First-order mixed integer-valued autoregressive processes with zero-inflated generalized power series innovations. *Journal of the Korean Statistical Society*, Elsevier, v. 44, n. 2, p. 232–246, 2015.
- MONTGOMERY, D. C. *Introdução Ao Controle Estatístico Da Qualidade*. [S.l.]: Grupo Gen-LTC, 2000.
- OROZCO, D. L. R. *Um novo processo autorregressivo misto para séries temporais de valores inteiros de primeira ordem com inovações Poisson (POMINAR(1))*. Dissertação (Mestrado) — Brasil, 2017.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>.
- RAKITZIS, A. C.; CASTAGLIOLA, P. The effect of parameter estimation on the performance of one-sided shewhart control charts for zero-inflated processes. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 45, n. 14, p. 4194–4214, 2016.

- RAKITZIS, A. C.; CASTAGLIOLA, P.; MARAVELAKIS, P. E. Cumulative sum control charts for monitoring geometrically inflated poisson processes: An application to infectious disease counts data. *Statistical methods in medical research*, SAGE Publications Sage UK: London, England, p. 0962280216641985, 2016.
- 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*, Wiley Online Library, v. 25, n. 3, p. 317–333, 2004.
- STEUTEL, F.; HARN, K. V. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, JSTOR, p. 893–899, 1979.
- WEISS, C. H. Detecting mean increases in poisson inar (1) processes with ewma control charts. *Journal of Applied Statistics*, Taylor & Francis, v. 38, n. 2, p. 383–398, 2011.
- WEISS, C. H.; TESTIK, M. C. Cusum monitoring of first-order integer-valued autoregressive processes of poisson counts. *Journal of Quality Technology*, American Society for Quality, v. 41, n. 4, p. 389, 2009.
- WIEL, S. V. Monitoring processes that wander using integrated moving average models. *Technometrics*, Taylor & Francis, v. 38, n. 2, p. 139–151, 1996.