Modelos autorregressivos de valores inteiros (INAR)

Chrstian do Espirito Santo

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

8 de dezembro de 2023

Introdução

- Séries com natureza discreta de números inteiros não negativos.
- Baseados nos modelos AR adpatados para aplicações com distribuições discretas como Poisson, Binomial, geométrica e outros.
- Proposto por Al-Osh& Alzaid em 1987 para modelar séries de contagem, em seguida foi generalizado de INAR(1) para ordem p e incluída a condição de estacionariedade.

Equação

- o processo INAR(1) é descrito pela equação $y_t = p \circ y_{t-1} + \epsilon_t$,
- Sendo p um numero real tal que p $\in [0, 1]$.

- ϵ_t chamada de fator de inovação do processo, e uma seqeência de variáveis aleatórias i.i.d de valores inteiros não negativos com media μ e variância σ^2
- Se p=1 o processo será um passeio aleatório, se p=0 será uma sequência $\{\epsilon_t\}_{t \in \mathbb{Z}}$.
- e $\circ$ é um operador chamado operador thinning.

Operador Thinning

- O Operador thinig foi cirada como um substituto do multiplicador no modelo AR, com a alteração para ser aplicada um multiplicador inteiro denotado por $\circ$.
- $\alpha \circ X = \sum_{j=1}^{X} Y_j$;
- X é v.a inteira e não negativa;
- Y_j é i.i.d e seue uma distribuição Bernoulli(α);
- $\alpha \in [0, 1]$

Interpretação

- $y_t = p \circ y_{t-1} + \epsilon_t,$
- Podemos interpretar o modelo INAR(1) como uma soma de duas distribuições discretas.
- avaliando $W_t = p \circ y_{t-1}$ temos $W_t | y_{t-1} \sim Binomial(p, y_{t-1})$;
- ϵ_t segue uma distribuição discreta como Poisson, Geométrica, Binomial negativa entre outras;
- Com base nisso podemos ver que o processo INAR(1) possui um forte fator de interpretabilidade;
- $$\underbrace{y_t}_{\text{N alunos no tempo } t} = \underbrace{p \circ y_{t-1}}_{\text{sobreviventes no tempo } t-1} + \underbrace{\epsilon_t}_{\text{Novos alunos}}$$

INAR(p)

- Assim como o INAR(1), sua generalização segue similar a um AR(p), com a diferenca do operador thinning, mais uma simiaridade é i comportamnto de acf.
- $y_t = p_1 \circ y_{t-1} + p_2 \circ y_{t-2} + \ldots + p_p \circ y_{t-p} + \epsilon_t;$
- $y_t = \phi_1 y_{t-1} + \phi y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t;$

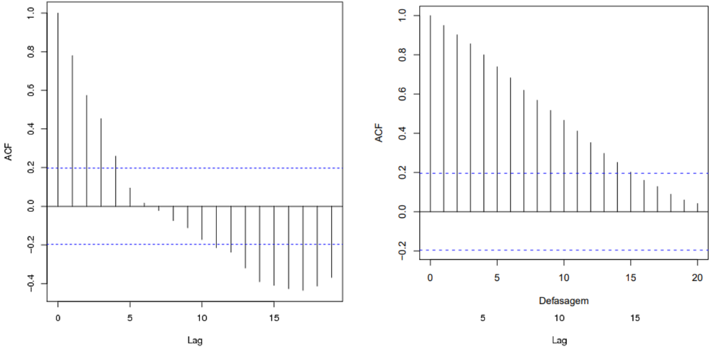


Figura: Função de autocorrelação parcial dos processos AR(1) e INAR(1)

Estimação dos parâmetros

- A estimação dos parâmetros do processo INAR (1) é mais complexo do que para um AR(1), ja que as duistribuição condicional de y_t dado y_{t-1} e uma comvolução da distribuição de $\{\epsilon_t\}_{t\in\mathbb{Z}}$ e de uma binomial(p,$y_t - 1$).
- O modelo INAR possui 3 tipos de estimação mais usuais;
- Estimador de Yule-Walker;
- Estimador dos minimos quadrados condicionais;
- Estimador de maxima verossimilhança condicional.

Estimador de Yule-Walker

- A partir de uma amostra de tamanho n os parâmetros são estimados usando a função de autocorrelação, autocovariância amostral e a mpédia aritimética .
- $\hat{p} = \hat{\rho}(1) = \frac{\gamma(\hat{1})}{\gamma(\hat{0})} = \frac{\sum_{t=1}^{n-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$;
- e uma estimativa baseada no primeiro momento do processo, dada por $\mathbf{E}(y_t) = \frac{\lambda}{(1-p)}$;
- $\hat{\lambda} = (1 - \hat{p})\frac{1}{n}\sum_{t-1}^{n} y_t$;

Estimador por minimos quadrados condicional

- Nesse estimador, sendo $\{y_t\}_{t\in\mathbb{Z}}$ processo INAR(1) com $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{Po}(\lambda)$ temos $\mathbb{E}[y_t|y_{t-1}] = py_{t-1} + \lambda = g(\theta, y_{t-1})$, onde θ é o vetor de parâmetros a serem estimados po uma mostra de tamanho n.
- Minimizando $Q_n(\theta) = \sum_{t=2}^{n}[y_t - g(\theta, y_{t-1})]^2$ derivando e igualando a 0 obtemos:
- $\hat{p} = \frac{\sum_{t=2}^{n} y_t y_{t-1} - \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}}{\sum_{t=2}^{n} y_{t-1}^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n} y_{t-1}\right)^2}$;
- $\hat{\lambda} = \frac{1}{n-1}\left(\sum_{t=2}^{n} y_t - \hat{p}\sum_{t=2}^{n} y_{t-1}\right)$;

Estimador de máxima verossimilhança condicional

- Dado $\{y_t\}_{t\in\mathbb{Z}}$ o processo INAR(1) com $\{\epsilon_t\}_{t\in\mathbb{Z}} \sim \text{Po}(\lambda)$ temos a distribuição condicional $f(y_t|y_{t-1})$ que é uma convolução de uma distribuição binomial do operador thinning $p \circ y_{t-1}$ e uma Poisson de ϵ_t que pode ser descrita como
- $f(y_t|y_{t-1}) = e^{-\lambda} \sum_{i=0}^{\min(y_t, y_{t-1})} \frac{\lambda^{(y_t)-i}}{(y_t - i)!} \binom{y_{t-1}}{i} p^i (1-p)^{(y_{t-1})-i}$;
- A verossimilhança incondiciona é dada por $L(p, \lambda; y) = \frac{[\lambda/(1-p)]^{y_0}}{y_0!} \exp(\frac{\lambda}{1-p}) \prod_{t=1}^{n} f(y_t \mid y_{t-1})$;
- $\frac{\partial l(p,\lambda; Y|y_0)}{\partial p} = \sum_{t=1}^{n} \frac{y_{t-1}}{1-p} \frac{[f(y_{t-1}-1|y_{t-1}-1) - f(y_t|y_{t-1})]}{f(y_t|y_{t-1})}$;
- $\frac{\partial l(p,\lambda; y|y_0)}{\partial \lambda} = \sum_{t=1}^{n} \frac{[f(y_{t-1}-1|y_{t-1}-1) - f(y_t|y_{t-1})]}{f(y_t|y_{t-1})}$.

INAR X AR

- Modelos de séries de contagem ou eventos discretos x Modelos de séries continuas como temperatura, financeira e etc,
- $\epsilon_t \sim Poisson(\lambda)$ x $\epsilon_t \sim Normal(0, \sigma^2)$;
- distribuição assintotica onde Poisson(λ), $\lambda \to \infty$, Normal
- Monte Carlo para aprximar a estatística de teste ;

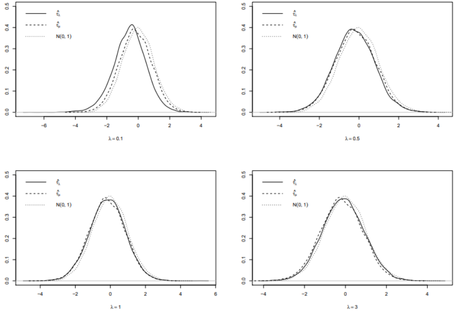


Figura: Teste Empirico de Marcelo Bouruignon Pereira

Qualidade do ar em Cariacica em 2007

- Dados disponibilizados pelo IEMA calculando o indice de qualidade do ar(IQA) entre 01/01/2007 e 21/03/2007.
- A serie conta com 80 observações no qual 41% corespodem ao valor 1 e com limite superior igual 5, caracteristicas que dificultariam o uso de qualquer modelo baseado na distribuição gaussiana pela magnetude pequena e não assumir valores negativos.;

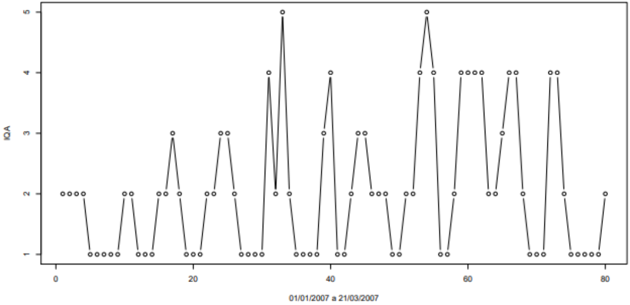


Figura: Estudo e aplicação de Marcelo Bouruignon Pereira

Estimação dos parâmetros

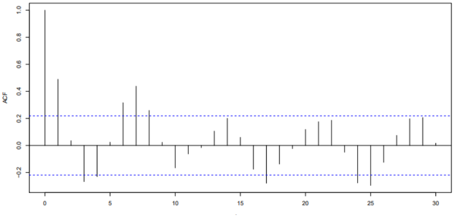


Figura: ACF

- O modelo estimado foi o $\text{SINAR}(1)_7$;

Tabela: Estimativas dos parâmetros do modelo.

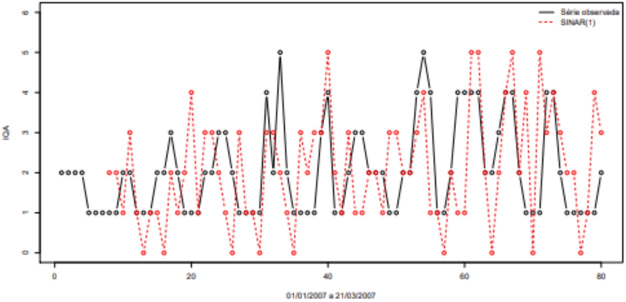
Método	$\hat{P}$	$\hat{\lambda}$
YW	0.4526	1.1507
MQC	0.4367	1.1849
MVC	0.6416	0.7336

Conclusões

- $y_t = 0.6416 \circ y_{t-7} + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{Po}(0.7336).$

Tabela: previsão 3 passos a frente

h	y_{n+h}	$\hat{y}_{n+h}$	$\hat{e}_{n+h}$
1	2	2.0168	-0.0168
3	2	1.3752	0.6248
5	3	1.3752	1.6248



Agradecimentos

Quero agradecer em primeiro lugar a Deus, ser supremo, fonte de força inesgotável por ter me dado força, perseverança e paciência, qualidades mais do que necessárias para concluir o mestrado.

Agradeço muito especialmente aos meus pais pelo interesse em me ver concluir o curso e por serem minha eterna inspiração para continuar na luta por um objetivo na vida.

Ao professor Dr. Valdério Anselmo Reisen por ter me conduzido até aqui, ter confiado em meu potencial, pela dedicação na minha formação acadêmica e por todo o apoio dos últimos cinco anos.

Ao meu orientador Professor Dr. Klaus Leite Pinto Vasconcellos pela orientação, paciência, dedicação, sugestões e valiosas recomendações que tornaram possível este trabalho.

Ao professor Dr. Domingos do departamento de matemática da Universidade Federal do Espiríto Santo pela confiança depositada em mim.

Aos meus amigos de Pernambuco, com destaque para os amigos Agrinaldo, Josivandro, Dona Dada, Jeremias, Iván, Lutemberg, Manoel, Natasha, Poema, Josimar e Helton, que se tornaram a minha família nesses dois anos de mestrado e com certeza foram fundamentais para a conclusão do mestrado, gostaria de expressar minha profunda gratidão.

Aos demais colegas da pós-graduação e classe que de alguma forma contribuíram para o desfecho meritório deste mestrado: Davis, Cícero, Laércio, Diego, Silvio, Francisco e Marcela.

A Valéria Bittencourt, secretária da pós-graduação em estatística, pela competência, carinho, amizade e atenção.

A minha namorada Gleicielle pela paciência, apoio e incentivo constantes.

Aos amigos de infância pelo apoio enquanto estava em outro estado.

Aos amigos Alessandro e Giovanni, pessoas nas quais sempre estiveram dispostas a me ajudar desde a época de graduação.

Ao amigo Fabio Fajardo pela troca de conhecimento e apoio.