

Uma aplicação do modelo Normal Generalizada-ARMA

Eder Angelo Milani¹²

Marinho G. Andrade³

Carlos Alberto Ribeiro Diniz¹

1 Introdução

O modelo ARMA é uma abordagem padrão em estudos de séries temporais, mas esta abordagem impõe a suposição de que os ruídos seguem uma distribuição gaussiana. Entretanto, existem séries que violam esta suposição, como por exemplo, séries que apresentam caudas mais pesadas ou mais leves que as da distribuição gaussiana, séries de dados de contagem, séries assimétricas. Isto vem motivando muitos autores a estudar processos não gaussianos, podemos citar por exemplo Zeger (1988) que propôs um modelo *parameter-driven* com a distribuição poisson, Davis(2003) e Davis e Wu (2009) que apresentaram modelos utilizando as distribuições poisson e binomial negativa, respectivamente, para dados de contagem, Jose *et al.* (2008) que propuseram o modelo autorregressivo de primeira ordem com distribuição marginal normal-Laplace.

A distribuição normal generalizada, que pode assumir as distribuições normal e Laplace como casos particulares, foi proposta por Nadarajah (2005). Uma entre várias características desta distribuição é o fato de que suas caudas podem ser mais leves ou mais pesadas do que as da distribuição normal. Motivados a melhor explicar séries que apresentam simetria, mas com curtosi diferente da distribuição gaussiana, criamos o modelo Normal Generalizada - ARMA (NG-ARMA) que é um modelo *observation-driven* que se mostra como uma alternativa para modelar a média de séries temporais que apresentam simetria, mas com caudas mais leves ou mais pesadas do que as da distribuição gaussiana, tendo ainda o modelo ARMA como um caso particular.

2 Material e métodos

Considere $\{Y_t, t \in \mathbb{N}\}$ um processo estocástico e seja \mathfrak{F}_{t-1} o conjunto de informações das observações até o instante $(t-1)$, definido como $\mathfrak{F}_{t-1} = \{x_t, \dots, x_1, y_{t-1}, \dots, y_1, \mu_{t-1}, \dots, \mu_1\}$. Admita que y_t tem distribuição condicional dada por

$$f_{Y_t|\mathfrak{F}_{t-1}}(y_t|\mathfrak{F}_{t-1}) = \frac{s}{2\sigma\Gamma(1/s)} \exp\left(-\left|\frac{y_t - \mu_t}{\sigma}\right|^s\right), \quad (1)$$

¹Des-UFSCar. e-mail: edinhomilani@hotmail.com

²Agradecimento à Capes pelo apoio financeiro.

³ICMC/USP.

ou seja, $Y_t | \mathfrak{F}_{t-1}$ tem distribuição Normal Generalizada com os parâmetros μ_t , σ e s em que $E[Y_t | \mathfrak{F}_{t-1}] = \mu_t$.

Como no modelo GARMA, proposto por Benjamin *et al.* (2003), o nosso interesse aqui é modelar a média do processo estocástico, logo μ_t é relacionada com o preditor linear, η_t , por uma função de ligação, g , monótona, bijetiva e duas vezes diferenciável. A relação é expressa por

$$g(\mu_t) = \eta_t = \mathbf{x}'_t \boldsymbol{\beta} + \tau_t$$

em que τ_t representa os termos autorregressivo e de médias móveis, sendo adotado como $\tau_t = \sum_{j=1}^p \phi_j \{g(y_{t-j}) - \mathbf{x}'_{t-j} \boldsymbol{\beta}\} + \sum_{j=1}^q \theta_j \{g(y_{t-j}) - \eta_{t-j}\}$, e \mathbf{x} são as covariáveis.

Quando adotamos como função de ligação a função identidade o preditor linear pode ser escrito como

$$\mu_t = \mathbf{x}'_t \boldsymbol{\beta} + \sum_{j=1}^p \phi_j (y_{t-j} - \mathbf{x}'_{t-j} \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}), \quad (2)$$

assumindo que não haja covariáveis, a equação (2) pode ser reduzida ao modelo ARMA(p, q)

$$\mu_t = \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}). \quad (3)$$

O modelo NG-ARMA que é definido pelo par de equações (1) e (3) ou por (1) e (2), apresenta uma maior flexibilidade em relação ao modelo ARMA por assumir a distribuição condicional normal generalizada em vez da distribuição normal, sendo que o segundo é um caso particular do primeiro. Portanto o modelo NG-ARMA é uma alternativa para modelar séries temporais que apresentam simetria e curtosi diferente da distribuição normal.

2.1 Estimação de máxima verossimilhança

Denote os parâmetros do modelo a serem estimados por $\boldsymbol{\lambda} = (\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s)$ sendo que $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$ e $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$. Seja $\{y_1, \dots, y_n\}$ a série temporal, μ_t como descrito na equação (3), $r = \max(p, q)$ e $\underline{y}_r = \{y_1, \dots, y_r\}$ as r primeiras observações. Utilizando a função de verossimilhança condicional introduzida por Cox (1975), que é baseada inteiramente na distribuição condicional da resposta atual, dados seus valores passados, e nas informações passadas das covariáveis, estas últimas se existirem, a função de verossimilhança condicionada nas r primeiras observações é dada por

$$L(\beta_0, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma, s | \underline{y}_r) = \prod_{t=r+1}^n f(y_t | \mathfrak{F}_{t-1})$$

$$= \left(\frac{s}{2\sigma\Gamma(1/s)} \right)^{n-r} \exp \left\{ -\frac{1}{\sigma^s} \sum_{t=r+1}^n \left| y_t - \beta_0 + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \theta_j (y_{t-j} - \mu_{t-j}) \right|^s \right\}$$

As estimativas de máxima verossimilhança são encontradas utilizando métodos iterativos, como por exemplo, o método “L-BFGS-B” que está disponível no *software* R: *A Language and Environment for Statistical Computing* (2013). Os intervalos de confiança assintóticos são obtidos considerando as estimativas pontuais e a inversa da matriz de informação observada de Fisher.

3 Resultados e discussões

Consideramos a série de vazões médias mensais de Furnas observada de janeiro de 1931 até dezembro de 1997. Por se tratar de uma série que apresenta apenas valores positivos e sazonalidade, utilizamos a padronização dada por

$$Z_t = \frac{\log(Y_t) - \mu_m}{\sigma_m}$$

sendo que m representa observações dentro de um período r , $m = 1, \dots, S$. Como a série foi observada por um número inteiro de períodos, $T = nS$, introduzimos o índice r para representar um ciclo qualquer, logo, $r = 1, 2, \dots, n$. Disto, podemos assumir $t = t(r, m) = S(r - 1) + m$ e finalmente μ_m e σ_m^2 são estimados por: $\hat{\mu}_m = \frac{1}{n} \sum_{r=1}^n \log(Y_{t(r,m)})$ e $\hat{\sigma}_m^2 = \frac{1}{n} \sum_{r=1}^n [\log(Y_{t(r,m)}) - \hat{\mu}_m]^2$, com $m = 1, \dots, S$.

Adotamos $S = 12$ por ser uma série de vazão média mensal que tem um ciclo hidrológico de 12 meses e $n = 67$. Na Figura 1 em (a) é apresentada a série padronizada.

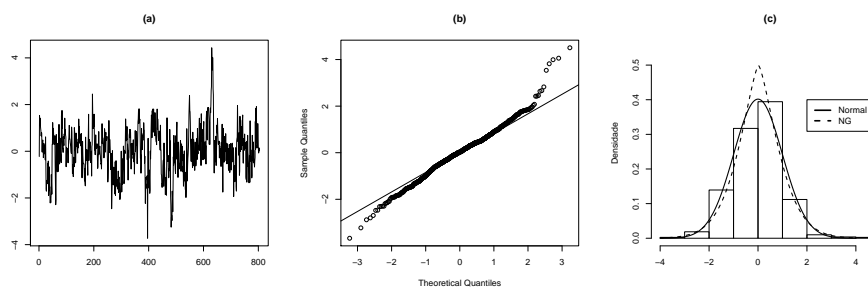


Figura 1: Série padronizada em (a), qqnorm em (b) e histograma em (c)

Para a verificação de normalidade nos dados utilizamos o histograma e o gráfico qqnorm, apresentados na Figura 1. Notamos pelo histograma que a distribuição normal generalizada se adequa melhor aos dados do que a distribuição normal, e através do gráfico qqnorm verificamos que as caudas da série se diferem das caudas da distribuição normal. Com esses resultados observamos que a distribuição normal generalizada se apresenta como uma alternativa para modelar a série.

Os gráficos das funções de autocorrelação e autocorrelação parcial são apresentados na Figura 2, indicando que um provável modelo a se ajustar à série será o modelo NG-ARMA(3,0).

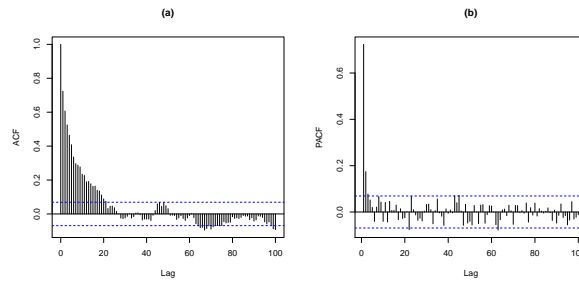


Figura 2: Acf em (a) e pacf em (b)

Para a escolha do melhor modelo que represente a série de vazão média mensal ajustamos algumas diferentes ordens utilizando os modelos ARMA e NG-ARMA sem intercepto, os resultados estão apresentados na Tabela 1. Notamos pela mesma que o modelo que apresenta menor AIC e menor BIC foi o modelo NG-ARMA(1,1).

Tabela 1: Seleção de modelo utilizando os critérios AIC e BIC

Modelo	Log-Verossimilhança	AIC	BIC
NG-ARMA(3,0)	-875,6	1.761,3	1.785,0
NG-ARMA(4,0)	-876,1	1.764,1	1.792,6
NG-ARMA(1,1)	-874,6	1.757,2	1.776,2
ARMA(3,0)	-977,4	1.962,9	1.981,9
ARMA(1,1)	-980,5	1.967,0	1.981,2

As estimativas de máxima verossimilhança e intervalo de confiança de 95% do modelo NG-ARMA(1,1) estão apresentados na Tabela 2. Podemos observar que todos os parâmetros são significativos e que a série apresenta caudas mais pesadas do que as da distribuição normal, pois o valor do parâmetro s é menor que 2.

Tabela 2: Estimativas do modelo NG-ARMA(1,1)

Parâmetro	EMV	IC
ϕ_1	0,872	(0,825; 0,919)
θ_1	-0,324	(-0,416; -0,231)
σ	0,788	(0,693; 0,883)
s	1,483	(1,284; 1,682)

Através dos gráficos de autocorrelação e qqnorm dos resíduos que são apresentados na Figura 3, notamos que o ajuste do modelo NG-ARMA(1,1) eliminou a correlação dos resíduos, mas as caudas dos resíduos apresentam uma divergência em relação às caudas da distribuição normal.

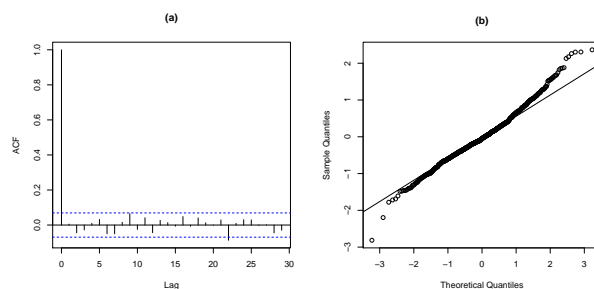


Figura 3: Acf em (a) e qqnorm em (b)

4 Conclusões

Motivados por séries que apresentam caudas mais leves ou mais pesadas do que as da distribuição normal, propomos uma nova classe de modelos para séries temporais que flexibiliza o modelo normal ARMA e que ainda o tem como caso particular. Através da aplicação com a série de vazões apresentada, mostramos que o modelo NG-ARMA é uma ótima alternativa ao modelo ARMA com distribuição normal quando os dados apresentam curtosi diferente da distribuição normal.

Referências

- [1] Benjamin, M. A.; Rigby, R. A.; Stasinopoulos, D. M. Generalized Autoregressive Moving Average Models. **Journal of the American Statistical Association**. v. 98, p. 214-223, 2003.
- [2] Cox, D. R. Partial likelihood. **Biometrika**. v. 62, p. 69-76, 1975.
- [3] Davis, R. A.; Dunsmuir, W. T. M.; Streett, S. B. Observation-driven Models for Poisson Counts. **Biometrika**. v. 90, p. 777-790, 2003.
- [4] Davis, R. A.; Wu, R. A negative Binomial Model for Time Series of Counts. **Biometrika**. v. 96, p. 735-749, 2009.
- [5] Jose, K. K.; Tomy, L.; Sreekumar, J. Autoregressive Processes with Normal-Laplace Marginals. **Statistics and Probability Letters**. v. 78, p. 2456-2462, 2008.
- [6] Nadarajah, S. A Generalized Normal Distribution. **Journal of Applied Statistics**. v. 32, p. 685-694, 2005.
- [7] Zeger, S. L. A Regression Model for Time Series of Counts. **Biometrika**. v. 75, p. 621-629, 1988.