UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Instituto de Física
Bacharelado em Engenharia Física
Tópicos especiais de engenharia física

Implementação de um modelo autoregressivo (AR)

Introdução

Modelos autoregressivos são modelos para a previsão do comportamento de séries temporais estacionárias que são construídos pela combinação linear dos valores da série no passado. Apenas os pontos no passado da série que tem mais influência nos pontos futuros são utilizados para a combinação. O modelo desenvolvido aplicado neste trabalho tem como objetivo o estudo da série temporal do número de passageiros mensais das linhas aéreas dos EUA de 1949 até 1960. [1]

1. Teste de estacionariedade

1.1 Verificação visual da estacionariedade

Para o modelo autoregressivo ser aplicado, primeiramente é necessário verificar se a série é estacionária, ou seja, se os parâmetros da série são constantes no tempo. Nesse caso, a média (μ) e a variância (σ^2).

Primeiramente, pode ser feita uma análise visual da série para a verificação desses parâmetros (figura 1).

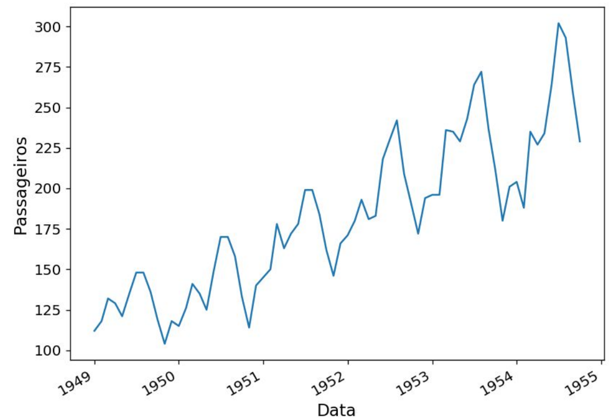

Figura 1. Série temporal do número mensal de passageiros nas linhas aéreas do EUA entre 1949 e 1955.

Pode-se perceber pela imagem que a série tem uma tendência de aumentar no tempo e que a variância também não parece ser constante. Assim não sendo uma série estacionária. Entretanto, para testar quantitativamente a estacionariedade, utiliza-se o teste de Dickey Fuller (DF).

1.2 Teste de Dickey Fuller

O teste de dickey fuller pode ser entendido como um teste de hipótese para verificar se os coeficientes que multiplicam os pontos da série temporal nos tempos anteriores são maiores ou iguais a 1, ou seja, que e evolução da série temporal não terá uma tendência na média ou na variância. O teste se dá pela seguinte forma:

Dado um modelo autoregressivo de ordem 1 - AR(1):

$$x_t = \varphi x_{t-1} + \varepsilon_t$$

Então fazendo-se a diferença com o ponto anterior (x_{t-1}):

$$\Delta x_t = \beta x_{t-1} + \varepsilon_t$$

Onde $\beta = \varphi - 1$

Se $\beta \geq 1$, a série é não estacionária. Se

$\beta < 1$ a série é estacionária.

Então, aplicando-se a fórmula anterior na série temporal e fazendo um ajuste linear, encontramos $\beta = -0.041$. O número é muito próximo de zero e, aliado à incerteza, pode-se aproximar $\beta = 0$ e afirmar que a série é não estacionária, como visto no teste visual.

Para transformar a série em estacionária, aplica-se o logaritmo natural, toma-se a média com uma janela de 2 pontos e, por último, deriva-se a série.

Realizando-se o teste de dickey fuller novamente, tem-se $\beta = -0.46$. Assim a série se torna estacionária. Após a transformação anterior, tem-se:

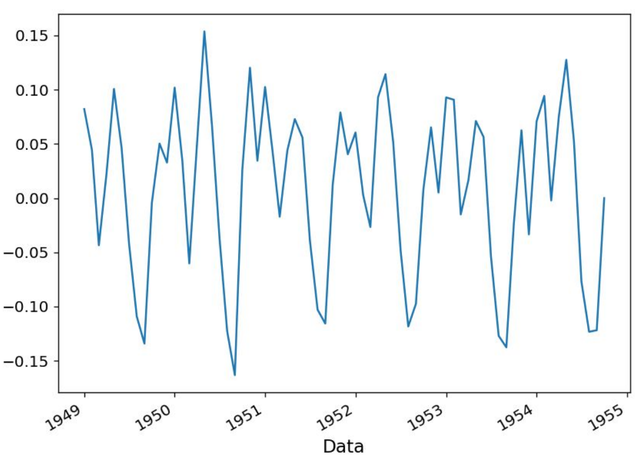

Figura 2. Série temporal estacionária após a transformação da série de passageiros mensais nas linhas aéreas do EUA entre 1949 e 1955.

2. Cálculo da autocorrelação parcial (PACF)

Após se ter a série estacionária, é necessário escolher a ordem do modelo autoregressivo a ser aplicado, ou seja, identificar os pontos da série no passado que mais influenciam a série no futuro. Para isso, calcula-se a função de autocorrelação parcial (PACF). Basicamente, é modelado o ponto da série no tempo ‘t’ com os pontos anteriores. Os coeficientes da série nos pontos anteriores geram as dependências que serão incluídas no gráfico da PACF. Matematicamente, tem-se:

$$x_t = c + \sum_{i=1}^{k} \varphi_{t-i} + \varepsilon_t$$

Para obter os coeficientes dos pontos anteriores, utilizou-se as equações de Yule-Walker:

$$\rho_m = \sum_{t=1}^{k} \varphi_t \rho_{m-t} + \varepsilon^2 \delta_{m,t}$$

Onde $m = 0, 1, ..., k$ em que ρ_m é a autocovariância e ε^2 é a variância da série.

Somando para todos os t onde tem-se m+1 equações. Pode-se montar um sistema de equações expandindo t até k. Tem-se então na forma matricial:

$$\begin{pmatrix} \rho(0) & \rho(1) & \cdots & \rho(k-1) \\ \rho(1) & \rho(0) & \cdots & \rho(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}$$

Calculando os termos acima para a série temporal e invertendo a matriz para encontrar os coeficientes φ, tem-se o seguinte gráfico para a PACF:

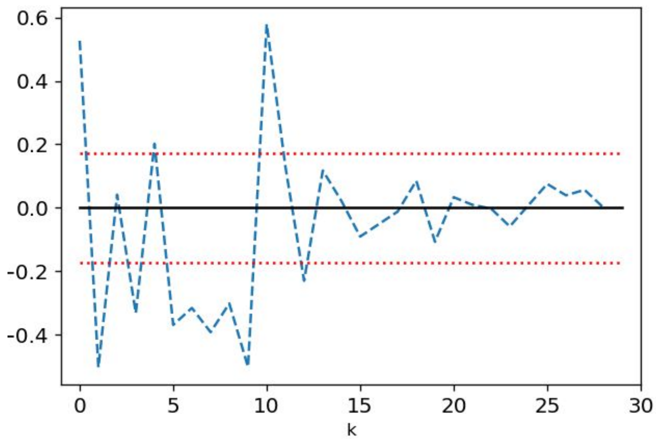

Figura 3. Função de autocorrelação parcial calculada com o limite para a escolha da ordem do modelo AR.

Para escolher a ordem do modelo, escolhemos os coeficientes que estão dentro do limite $\pm 1.96\sqrt{N}$ sendo N o número de pontos da série temporal [2]. Desta forma escolhemos k = 13.

Com os coeficientes da PACF pode-se aplicar o modelo de autoregressão de ordem 13. Antes, entretanto, é necessário calcular a constante da combinação linear que, através das equações de Yule-Walker é dada por:

$$C = \mu - \sum_{i=1}^{k} \varphi_i \mu$$

Aplicando-se a equação acima, obtém-se $C = 0.026$.

3. Implementação do modelo AR

Substituindo-se os coeficientes calculados pelo método de Yule-Walker na equação do modelo autoregressivo e calculando-se a predição de 13 valores futuros, tem-se:

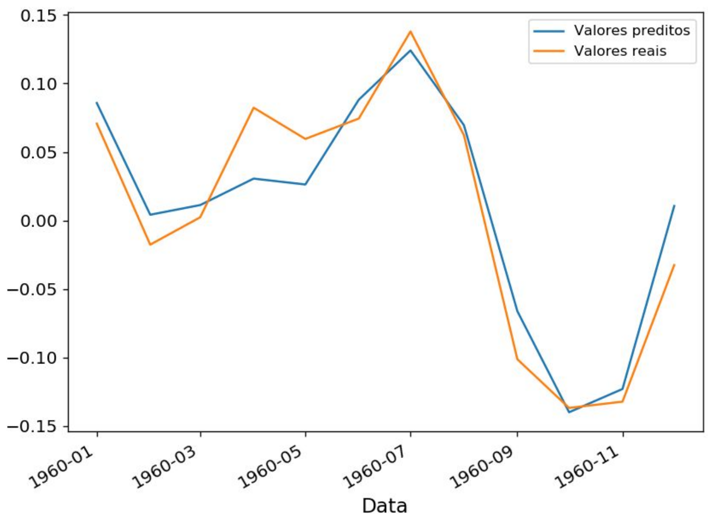

Figura 4. Predição dos valores para a série temporal com a aplicação de um modelo autoregressivo de ordem k = 13.

Conclusão

Os valores preditos, como observados na figura acima acompanham a mesma tendência dos valores reais da série temporal para o modelo autoregressivo de ordem 13. Foi possível aplicar um modelo autoregressivo para uma série com tendência temporal e sazonalidade após a devida transformação para uma série estacionária.

Referências:

1 -
https://www.kaggle.com/chirag19/air-pass

engers

2 - Wikipedia. Disponível em: https://en.wikipedia.org/wiki/Partial_autocorrelation_function. Acesso em: 03 de maio de 2020.