Especialização em *Data Science* e Estatística Aplicada

Módulo IV - Análise de Séries Temporais

Prof. Dr. Eder Angelo Milani

Goiânia, 2025

IME
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
FEN
FACULDADE DE ENFERMAGEM
UFG
UNIVERSIDADE FEDERAL DE GOIÁS

Conteúdo Programático

- Principais conceitos de séries temporais. (Aula 1)
- Funções de autocovariância e de autocorrelação. (Aula 1)
- Definição e estimação da tendência e da sazonalidade. (Aulas 2 e 3)
- Métodos de suavização. (Aulas 3 e 4)
- Modelagem Box-Jenkins: modelos AR, MA, ARMA, ARIMA e SARIMA.
- Análise de séries temporais interrompidas.

Conteúdo - Aula 5

1. Modelagem Box-Jenkins: ARIMA e SARIMA

- Introdução
- ARIMA
- SARIMA
  - Sazonalidade determinística
  - Sazonalidade estocástica

Introdução

Os modelos estudados até o momento são apropriados para descrever séries estacionárias, isto é, séries que se desenvolvem no tempo ao redor de uma média constante. Muitas séries encontradas na prática não são estacionárias.

O que fazer quando a série apresentar tendência?

Introdução

Voltando a Aula 1 ...

Obs.: um procedimento que é também utilizado para eliminar a tendência de um série é aquele de tomar diferenças, como foi definido anteriormente.

Mas o que é essa tal de diferença?

A primeira diferença de Z_t é definida por

$$\Delta Z_t = Z_t - Z_{t-1},$$

a segunda diferença é

$$\Delta^2 Z_t = \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}] = Z_t - 2Z_{t-1} + Z_{t-2}.$$

Modelo autorregressivo integrado de médias móveis

Se $W_t = \Delta Z_t$ for estacionária, podemos representar W_t por um modelo ARMA(p, q), ou seja,

$$\phi(B)W_t = \theta(B)a_t. \tag{1}$$

Se W_t for uma diferença de Z_t, então Z_t é uma *integral de* W_t, daí dizemos que Z_t segue um modelo autorregressivo, *integrado*, de médias móveis, ou modelo ARIMA,

$$\phi(B)\Delta^d Z_t = \theta(B)a_t, \tag{2}$$

de ordem (p, d, q) e escrevemos ARIMA(p, d, q), se p e q são as ordens de $\phi(B)$ e $\theta(B)$, respectivamente.

Modelo autorregressivo integrado de médias móveis

O modelo (2) supõe que a d-ésima diferença da série Z_t pode ser representada por um modelo ARMA, estacionário e invertível. Na maioria dos casos usuais, $d = 1$ ou $d = 2$, que corresponde a dois casos interessantes e comuns de não-estacionariedade homogênea:

(a) séries não-estacionárias quanto ao nível: oscilam ao redor de um nível médio durante algum tempo e depois saltam para outro nível temporário. Para torná-las estacionárias é suficiente tomar uma diferença; este é o caso típico de séries econômicas;

(b) séries não-estacionárias quanto à inclinação: oscilam numa direção por algum tempo e depois mudam para outra direção temporariamente. Para torná-las estacionárias é necessário tomar a segunda diferença.

Modelo autorregressivo integrado de médias móveis

Exemplo: Alguns casos particulares do modelo (2) são:

(i) ARIMA(0,1,1): $\Delta Z_t = (1 - \theta B)a_t$;

(ii) ARIMA(1,1,1): $(1 - \phi B)\Delta Z_t = (1 - \theta B)a_t$;

(iii) ARIMA(p,0,0) = AR(p)

(iv) ARIMA(0,0,q) = MA(q)

(v) ARIMA(p,0,q) = ARMA(p, q)

O modelo ARIMA(1,1,1) também pode ser escrito como

$$
\begin{aligned}
(1 - \phi B)\Delta Z_t &= (1 - \theta B)a_t \\
(1 - \phi B)(1 - B)Z_t &= (1 - \theta B)a_t,
\end{aligned}
$$

ou ainda por: $Z_t = (1 + \phi)Z_{t-1} - \phi Z_{t-2} + a_t - \theta a_{t-1}$.

Teste para raiz unitária

O teste Dickey-Fuller aumentado (do inglês *augmented Dickey-Fuller*) - ADF - é provavelmente o teste de hipóteses para raízes unitárias mais utilizado. A grosso modo, se tomarmos um processo da forma $Z_t = \theta_0 + \phi Z_{t-1} + a_t$, o objetivo do teste é, então, inferir se $\phi = 1$.

O problema é que estimar o modelo sob a hipótese de não estacionariedade nos faz incorrer no problema de regressão espúria (duas variáveis apresentam uma correlação forte e significativa, mas essa relação é falsa e não existe uma causa e efeito real entre elas).

Teste para raiz unitária

Dickey e Fuller (1979) propõem realizar o teste sob a série diferenciada e, três configurações que possibilitam obter as estatísticas de testes adequadas. Ainda que a série seja diferenciada, **sob a hipótese nula de existência de raiz unitária**, é possível que os resíduos das regressões estejam correlacionados. Para contornar esse problema, adicionam-se termos defasados da variável dependente, sendo que para determinar a quantidade geralmente se usam critérios de informação do tipo AIC e BIC.

As equações do teste são:

$$
\begin{aligned}
\Delta Z_t &= \gamma Z_{t-1} + \sum_{t=1}^{p} \beta_i \Delta Z_{t-1} + a_t, \\
\Delta Z_t &= \theta_0 + \gamma Z_{t-1} + \sum_{t=1}^{p} \beta_i \Delta Z_{t-1} + a_t, \\
\Delta Z_t &= \theta_0 + \beta t + \gamma Z_{t-1} + \sum_{t=1}^{p} \beta_i \Delta Z_{t-1} + a_t.
\end{aligned}
$$

Aplicação

Aplicação no *software* R.

SARIMA

Anteriormente, estudamos o problema da sazonalidade e os procedimentos de estimação e eliminação da componente sazonal determinística de uma série sazonal.

É possível que, mesmo após eliminar a componente sazonal determinística, ainda reste autocorrelação significativa em:

(i) "lags" de baixa ordem, indicando que os resíduos ainda são correlacionados, podendo-se ajustá-los através de um modelo ARIMA, por exemplo;

(ii) "lags" sazonais, isto é, múltiplos de período s. Isto significa que há necessidade de se considerar uma sazonalidade estocástica, ou seja, ajustar à série original um modelo **ARIMA sazonal (SARIMA)**.

Sazonalidade determinística

Consideremos, por simplicidade de exposição, dados observados mensalmente e sazonalidade de período $s = 12$. Tratemos, separadamente, os dois tipos de sazonalidade.

Quando $\{Z_t\}$ exibe um comportamento sazonal determinístico com período 12, um modelo que pode ser útil é:

$$Z_t = \mu_t + N_t,$$

onde μ_t é uma função determinística periódica, satisfazendo $\mu_t - \mu_{t-12} = 0$ e N_t é um processo estacionário que pode ser modelado por um ARMA(p, q).

Sazonalidade determinística

Dessa maneira, N_t satisfaz a equação

$$\phi(B)N_t = \theta(B)a_t,$$

onde a_t é ruído branco e μ_t é modelado utilizando variáveis "*dummies*" ou com pares de funções seno e cosseno, como já vimos anteriormente.

Sazonalidade determinística

Identificação

A identificação do modelos $Z_t = \mu_t + N_t$, é feita em dois passos:

1. Obtemos estimativas dos parâmetros que são utilizados para estimar a sazonalidade por meio de uma análise de regressão.
2. Calculamos os resíduos:
$$\hat{N}_t = Z_t - \hat{S}_t,$$
e examinamos as funções de autocorrelação e autocorrelação parcial para identificar um modelo ARMA(p, q) para N_t.

Sazonalidade determinística

Estimação

O processo de estimação é feito em partes, seguindo o roteiro:

- estimação de μ_t, sendo modelado por variáveis "*dummies*" ou com pares de funções seno e cosseno;
- obtenção de $\hat{N}_t = Z_t - \hat{S}_t$, e na sequência estimação do modelo ARMA(p, q) para N_t;
- por fim, estimação conjunta de μ_t e do modelo ARMA(p, q).

Aplicação

Aplicação no *software* R.

SARIMA

Do mesmo modo que podemos modelar uma série Z_t, observada mês a mês, por um modelo ARIMA, por exemplo:

$$Z_t = \phi Z_{t-1} + a_t,$$

poderíamos modelar uma associação ano a ano na série Z_t através de um modelo ARIMA sazonal, sendo dado por:

$$Z_t = \Phi Z_{t-12} + a_t.$$

De maneira similar, podemos considerar um modelo de médias móveis sazonais, como:

$$Z_t = a_t - \Theta a_{t-12}.$$

SARIMA

Portanto, a ideia é relacionar uma observação Z_t, correspondente a um determinado mês, janeiro por exemplo, com observações correspondentes a janeiros anteriores, através de um modelo ARIMA sazonal da forma:

$$\Phi(B^{12})\Delta_{12}^{D}Z_t = \Theta(B^{12})\alpha_t,$$

sendo que

- $\Phi(B^{12}) = 1 - \Phi_1 B^{12} - \cdots - \Phi_P B^{12P}$, é o operador autorregressivo sazonal de ordem P, estacionário;
- $\Theta(B^{12}) = 1 - \Theta_1 B^{12} - \cdots - \Theta_Q B^{12Q}$, é o operador de médias móveis sazonal de ordem Q, invertível;
- $\Delta_{12} = (1 - B^{12})$, é o operador diferença sazonal,
- $\Delta_{12}^{D} = (1 - B^{12})^{D}$, indicando o número de “diferenças sazonais”.

SARIMA

Uma diferença como o modelo ARIMA usual é que os $\alpha_t, \alpha_{t-1}, \ldots,$ não seriam ruído branco. Sem perda de generalidade, o valor observado de janeiro de um determinado ano, além de ser relacionada com janeiros anteriores, será também relacionada com dezembro, novembro do ano anterior, implicando que α_t, α_{t-1}, etc., sejam relacionados.

Para descrever esta relação introduzimos, para os α_t, o modelo ARIMA usual, dado por

$$\phi(B)\Delta^d \alpha_t = \theta(B) a_t,$$

sendo que agora a_t é ruído branco e

- $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, é o operador autorregressivo de ordem p, estacionário;
- $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, é o operador de médias móveis de ordem q, invertível;
- $\Delta^d = (1 - B)^d$, indica o número de diferenças.

SARIMA

Fazendo a junção das equações

$$\begin{aligned}\Phi(B^{12})\Delta_{12}^{D} Z_t &= \Theta(B^{12})\alpha_t,\\ \phi(B)\Delta^{d}\alpha_t &= \theta(B)a_t,\end{aligned}$$

obtemos

$$\phi(B)\Phi(B^{12})\Delta^{d}\Delta_{12}^{D} Z_t = \theta(B)\Theta(B^{12})a_t,$$

que é chamado de modelo **ARIMA sazonal multiplicativo** ou simplesmente de **SARIMA**, de ordem $(p, d, q) \times (P, D, Q)_{12}$.

O modelo SARIMA implica que devemos realizar d diferenças simples e D diferenças sazonais da série Z_t, de modo que o processo $W_t = \Delta^{d}\Delta_{12}^{D} Z_t$ seja estacionário.

SARIMA

Exemplo: Um modelo de médias móveis puro, SARIMA$(0,0,0) \times (0,0,Q)_{12}$ = SMA(Q), é da forma

$$Z_t = a_t - \Theta_1 a_{t-12} - \cdots - \Theta_Q a_{t-12Q}.$$

SARIMA

Exemplo: Um modelo de médias móveis puro, SARIMA$(0,0,0) \times (0,0,Q)_{12}$ = SMA(Q), é da forma

$$Z_t = a_t - \Theta_1 a_{t-12} - \cdots - \Theta_Q a_{t-12Q}.$$

Exemplo: Um modelo autorregressivo sazonal puro, SARIMA$(0,0,0)\times(P,0,0)_{12}$ = SAR(P), é da forma

$$Z_t = \Phi_1 Z_{t-12} + \cdots + \Phi_P Z_{t-12P} + a_t.$$

SARIMA

Exemplo: Um modelo SARIMA$(1,0,1) \times (1,0,1)_{12}$ tem a forma

$$(1-\phi B)(1-\Phi B^{12})Z_t = (1-\theta B)(1-\Theta B^{12})a_t,$$

ou

$$Z_t = \phi Z_{t-1} + \Phi Z_{t-12} - \phi\Phi Z_{t-13} + a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta\Theta a_{t-13}.$$

Aplicação

Aplicação no *software* R.

Especialização em *Data Science* e Estatística Aplicada

Módulo IV - Análise de Séries Temporais

Prof. Dr. Eder Angelo Milani

edermilani@ufg.br

IME
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

FEN
FACULDADE DE ENFERMAGEM

UFG
UNIVERSIDADE FEDERAL DE GOIÁS