Aula 2: Análise de séries temporais univariadas (ARIMA)

Marcus L. Nascimento

14 de novembro de 2025

Introdução

Introdução

- **Série temporal**: Sequência de variáveis aleatórias $Y_1, Y_2, \ldots, Y_n$ indexadas no tempo $(t = 1, 2, \ldots, n)$.
  - Por vezes, denotamos a sequência de variáveis aleatórias por $Y_{t_1}, Y_{t_2,\ldots,Y_{t_n}}$;
  - Frequentemente, tal sequência é considerada um subconjunto de uma sequência (possivelmente, infinita) $\{Y_t\}$ de variáveis aleatórias.
- Propósito em análise de séries temporais:
  - **Quantificar dependências** no tempo;
  - **Utilizar as correlações** para **explicar** as observações disponíveis e **inferir** sobre valores não observados.

Estacionariedade

- Um conceito fundamental em análise de séries temporais recai sobre a ideia de **estacionariedade**.
  - Medidas estatísticas (por exemplo, média, variância e correlação) são constantes ao longo do tempo;
  - Estabilidade no processo gerador subjacente facilita a identificação de padrões e relações nos dados;
  - Modelos estatísticos e de aprendizado de máquinas pressupõem estacionariedade.
- Estacionariedade **estrita** e **fraca**.

Estacionariedade estrita

DEFINIÇÃO: Uma série temporal $\{Y_t\}$ é dita **estritamente estacionária** se a distribuição conjunta de $(Y_{t_1}, \ldots, Y_{t_k})$ é idêntica à distribuição de $(Y_{t_1+s}, \ldots, Y_{t_k+s})$ para todo s, onde k é um número inteiro positivo arbitrário e $(t_1, \ldots, t_k)$ é uma coleção de k inteiros positivos.

OBSERVAÇÕES:

- A distribuição conjunta de $(Y_{t_1}, \ldots, Y_{t_k})$ é invariante sob translações temporais;
- Estacionariedade estrita é uma condição bastante forte;
- Difícil verificação empírica.

Estacionariedade fraca (segunda ordem)

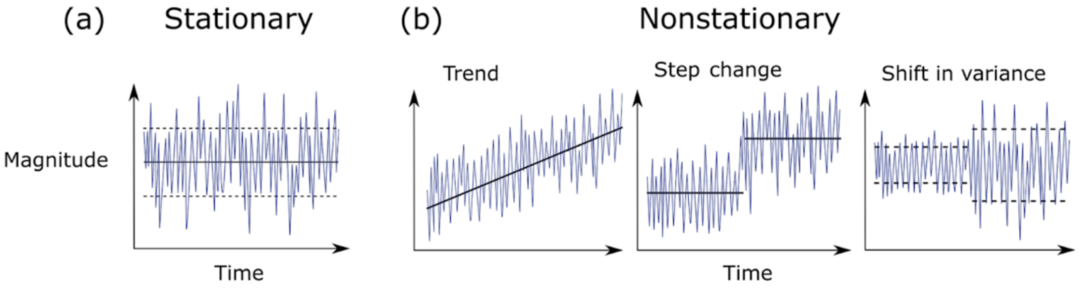
DEFINIÇÃO: Uma série temporal $\{Y_t\}$ é dita **fracamente estacionária** (ou estacionária de segunda ordem) se a média de Y_t e a covariância entre Y_t e Y_{t-l} são invariantes no tempo, onde l é um número inteiro arbitrário.

Mais especificamente, $\{Y_t\}$ é dita fracamente estacionária se:

1. $\mu_Y(t) = \mathrm{E}(Y_t) = \mu$, $|\mu| < \infty$;
2. $\sigma^2_Y(t) = E[(Y_t - \mu)^2] = \gamma_0$, $\gamma_0 < \infty$;
3. $\gamma_Y(t, t-l) = \mathrm{Cov}(Y_t, Y_{t-l}) = E[(Y_t - \mu)(Y_{t-l} - \mu)] = \gamma_l \ \forall l$ com $|\gamma_l| < \infty$.

Estacionariedade de segunda ordem implica que $Cov(Y_{t_1}, Y_{t_1-l}) = Cov(Y_{t_2}, Y_{t_2-l}) = \gamma_l$.

Estacionariedade Fraca



Função de autocorrelação (ACF)

- Sob estacionariedade fraca, a função de autocorrelação de defasagem l (lag l) é dada por:

$$\rho_l = \rho_Y(t, t-l) = \mathrm{Corr}(Y_t, Y_{t-l}) = \frac{\gamma_Y(t, t-l)}{\sqrt{\sigma_Y^2(t)\sigma_Y^2(t-l)}} = \frac{\gamma_l}{\gamma_0}$$

- A autocorrelação amostral de defasagem l (lag l), por sua vez, é definida como

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{n}(y_t - \bar{y})(y_{t-l} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}, \; 0 \leq l < n-1.$$

- A verificação quanto à presença ou não de autocorrelação em uma série temporal pode ser realizada através de testes de hipótese da seguinte forma:

$$
\begin{aligned}
\mathrm{H}_0 &: \rho_1 = \rho_2 = \ldots = \rho_m = 0 \\
\mathrm{H}_1 &: \rho_i \neq 0 \text{ para algum } i \in \{1, 2, \ldots, m\}.
\end{aligned}
$$

- Proposto por Box and Pierce (1970), o teste de Portmanteau aplica a seguinte estatística de teste:

$$
Q^*(m) = n \sum_{l=1}^{m} \hat{\rho}_l^2.
$$

- Assumindo que $\{Y_t\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com certas condições acerca dos momentos, $Q^*(m)$ é assintoticamente χ^2_m.

Função de autocorrelação (ACF) - Teste de Ljung-Box

- Com o intuito de aumentar o poder do teste em amostras finitas, Ljung and Box (1978) propuseram uma modificação na estatística $Q^*(m)$ da seguinte forma:

$$Q(m) = n(n+2)\sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{n-1}.$$

- Neste caso, a regra de decisão é rejeitar H_0 se $Q(m) > q_\alpha^2$, onde:
  - q_α^2 denota o $100(1-\alpha)$-ésimo percentil da distribuição χ_m^2;
  - α é um nível de significância fixado.

Função de autocorrelação parcial (PACF)

- Em geral, o conceito de correlação parcial refere-se a uma correlação condicional.

- No contexto de séries temporais, a autocorrelação parcial entre Y_t e Y_{t-l} é definida como a correlação entre Y_t e Y_{t-l} condicional em $Y_{t-l+1}, \ldots, Y_{t-1}$ (conjunto de observações entre os tempos t e $t-l$.

$$\phi_{l,l} = \frac{\mathrm{Cov}(Y_t, Y_{t-l} | Y_{t-l+1}, \ldots, Y_{t-1})}{\sqrt{\mathrm{Var}(Y_t | Y_{t-l+1} = y_{t-l+1}, \ldots, Y_{t-1} = y_{t-1}) \mathrm{Var}(Y_{t-l} | Y_{t-l+1} = y_{t-l+1}, \ldots, Y_{t-1} = y_{t-1})}}$$

EXEMPLOS:

- A autocorrelação parcial de primeira ordem é definida para ser igual à autocorrelação de primeira ordem;

- Autocorrelação de segunda ordem: $\phi_{2,2} = \frac{\mathrm{Cov}(Y_t, Y_{t-2} | Y_{t-1})}{\sqrt{\mathrm{Var}(Y_t | Y_{t-1} = y_{t-1}) \mathrm{Var}(Y_{t-2} | Y_{t-1} = y_{t-1})}}$;

Séries Temporais Lineares

Ruído branco

- Uma série temporal $\{W_t\}$ é denominada **ruído branco** se $\{W_t\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média 0 e variância constante τ^2 finita.

  - $\{W_t\}$ é estacionária;
  - Se $s \neq t$, $\text{Cov}(W_s, W_t) = \mathrm{E}[(W_s - \mu)(W_t - \mu)] = 0$;
  - $\rho_l = 0$ para todo l.

- Se W_t tem distribuição normal com média 0 e variância τ^2, a série temporal é denominada ruído branco gaussiano.

Série temporal linear

- Uma série temporal $\{Y_t\}$ é um processo linear se possui a seguinte representação:

$$y_t = \sum_{i=-\infty}^{+\infty} \psi_i w_{t-i} \tag{1}$$

para todo t e $\{\psi_i\}$ é uma sequência na qual $\sum_{i=-\infty}^{+\infty} |\psi_i| < \infty$.

- Um processo linear é denominado média móvel ou MA(∞) se $\psi_i = 0$ para todo $j < 0$, ou seja,

$$y_t = \sum_{i=0}^{\infty} \psi_i w_{t-i}. \tag{2}$$

- O ruído branco w_t pode ser interpretado como uma nova informação da série temporal no tempo t e é frequentemente chamado de inovação ou choque em t.

Série temporal linear

- Considerando um processo MA(∞), a covariância de defasagem l, $\gamma_Y(t, t-l)$, é dada por:

$$
\begin{aligned}
\gamma_Y(t, t-l) &= \mathrm{Cov}(Y_t, Y_{t-l}) \\
&= \mathrm{E}\left[\left(\sum_{i=0}^{\infty} \psi_i w_{t-i}\right)\left(\sum_{j=0}^{\infty} \psi_j w_{t-l-j}\right)\right] \\
&= \mathrm{E}\left(\sum_{i,j=0}^{\infty} \psi_i \psi_j w_{t-i} w_{t-l-j}\right) \\
&= \sum_{j=0}^{\infty} \psi_{j+l}\psi_j \mathrm{E}(w_{t-l-j}^2) = \tau^2 \sum_{j=0}^{\infty} \psi_{j+l}\psi_j .
\end{aligned}
$$

- Como $\mathrm{E}(Y_t) = 0$, o processo é estacionário por construção (resultado análogo pode ser obtido para o processo descrito na Equação (1)).

Série temporal linear

- O processo MA(∞) descrito na Equação (2) pode ser rescrito através da aplicação de um **operador de defasagem** (B):

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} \psi_i w_{t-i} \\ &= \sum_{i=0}^{\infty} \psi_i B^i w_t \\ &= \psi(B) w_t, \end{aligned}$$

onde $B^i w_t = w_{t-i}$ e $\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$.

- Além da notação B (*backshift*) que aplicaremos ao longo do curso, o operador de defasagem também é comumente denotado por L (*lag*).

Série temporal linear

- Substituindo o operador de desagem no polinômio $\psi(B)$ pela variável λ, pode ser verificado que todas as raízes do polinômio $\psi^{-1}(\lambda)$ ($\psi^{-1}(\lambda) = 0$) estarão fora do círculo unitário.

- Duas visões para estacionariedade:

  1. Convergência do processo MA(∞);

  2. Raízes do polínômio $\psi^{-1}(\lambda)$.

Processos Autorregressivos (AR)

Processo Autorregressivo de Ordem 1 - AR(1)

- Seja $\{W_t\}$ ruído branco, um processo AR(1) pode ser descrito como a seguir:

$$y_t = \phi_1 y_{t-1} + w_t. \tag{3}$$

- Um processo AR(1) pode ser descrito como um MA(∞):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + w_t \\ &= \phi_1(\phi_1 y_{t-2} + w_{t-1}) + w_t \\ &= \phi_1(\phi_1(\phi_1 y_{t-3} + w_{t-2}) + w_{t-1}) + w_t \\ &\vdots \\ &= \sum_{i=0}^{\infty} \phi_1^i w_{t-i}, \end{aligned}$$

onde $\sum_{i=0}^{\infty} |\phi_1^i| < \infty$, ou seja, $|\phi_1| < 1$ (estacionariedade).

Processo Autorregressivo de Ordem 1 - AR(1)

- Reescrevendo a Equação (3) em termos do operador de desagem (B), temos $y_t = (1 - \phi_1 B)^{-1} w_t$.
- No processo AR(1), $\psi^{-1}(\lambda) = \phi(\lambda) = (1 - \phi_1 \lambda)$, logo há apenas uma raiz $(\lambda = 1/\phi_1)$.
    - A condição $|1/\phi_1| > 1$ é análoga à condição $|\phi_1| < 1$, porém em termos das raízes do polinômio $\phi(\lambda)$.
- Sob estacionariedade, temos que:

$$
\begin{aligned}
\text{Var}(Y_t) &= \text{Var}(\phi_1 Y_{t-1} + W_t) \\
&= \text{Var}(\phi_1 Y_{t-1}) + \text{Var}(W_t) \\
&= \phi^2 \text{Var}(Y_{t-1}) + \tau^2 \\
&= \frac{\tau^2}{1 - \phi_1^2}.
\end{aligned}
$$

Processo Autorregressivo de Ordem 1 - AR(1)

- Ainda sob estacionariedade, temos que

$$
\begin{aligned}
\gamma_l &= \gamma_Y(t, t-l) = \mathrm{Cov}(Y_t, Y_{t-l}) = \mathrm{E}(Y_t Y_{t-l}) \\
&= \mathrm{E}((\phi_1 Y_{t-1} + W_t) Y_{t-l}) \\
&= \phi_1 \mathrm{E}(Y_{t-1} Y_{t-l}) + \mathrm{E}(W_t Y_{t-l}) \\
&= \phi_1 \mathrm{E}(Y_{t-1} Y_{t-l}) \\
&= \phi_1^2 \mathrm{E}(Y_{t-2} Y_{t-l}) \\
&\vdots \\
&= \phi_1^l \mathrm{E}(Y_{t-l} Y_{t-l}) \\
&= \phi_1^l \mathrm{E}(Y_t Y_t) \\
&= \phi_1^l \gamma_0,
\end{aligned}
$$

para $l = -1, 1, -2, 2, \ldots$.

Processo Autorregressivo de Ordem 1 - AR(1)

- Para encontrar γ_0, basta seguirmos de forma similar.

$$
\begin{aligned}
\gamma_0 &= \mathrm{E}(Y_t Y_t) \\
&= \mathrm{E}((\phi_1 Y_{t-1} + W_t) Y_t) \\
&= \phi_1 \mathrm{E}(Y_{t-1} Y_t) + \mathrm{E}(W_t Y_t) \\
&= \phi_1 \gamma_1 + \tau^2.
\end{aligned}
$$

- Como $\gamma_1 = \phi_1 \gamma_0$ e $\gamma_0 = \phi_1 \gamma_1 + \tau^2$, segue que $\gamma_0 = \frac{\tau^2}{1-\phi_1^2}$.

- A função de autocorrelação de um modelo AR(1), portanto, é dada por $\rho_l = \frac{\gamma_l}{\gamma_0} = \phi_l^l$, para $l = -1, 1, -2, 2, \ldots$.

Processo Autorregressivo de Ordem 2 - AR(2)

- Seja $\{W_t\}$ ruído branco, um processo AR(2) assume a seguinte forma:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t.$$

- Para encontrar a função de covariância do modelo AR(2), aplicamos ideias similares às aplicadas no modelo AR(1).

$$\begin{aligned}
\gamma_l &= E(Y_t Y_{t-l}) \\
&= E((\phi_1 Y_{t-1} + \phi_2 Y_{t-1} + W_t) Y_{t-l}) \\
&= \phi_1 E(Y_{t-1} Y_{t-l}) + \phi_2 E(Y_{t-2} Y_{t-l}) \\
&= \phi_1 \gamma_{l-1} + \phi_2 \gamma_{l-2} + \mathrm{E}(W_t Y_{t-l}).
\end{aligned}$$

Processo Autorregressivo de Ordem 2 - AR(2)

$$
\begin{aligned}
(l = 0) : \gamma_0 &= \phi_1 \gamma_{-1} + \phi_2 \gamma_{-2} + \mathrm{E}(W_t Y_{t-l}) \\
\gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \tau^2
\end{aligned}
$$

$$
\begin{aligned}
(l = 1) : \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \mathrm{E}(W_t Y_{t-l}) \\
&= \phi_1 \gamma_0 + \phi_2 \gamma_1 \\
&= \frac{\phi_1}{1 - \phi_2} \gamma_0
\end{aligned}
$$

$$
\begin{aligned}
(l = 2) : \gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0 + \mathrm{E}(W_t Y_{t-l}) \\
&= \phi_1 \gamma_1 + \phi_2 \gamma_0 \\
&= \left[\frac{\phi_1^2}{1 - \phi_2} + \phi_2 \right] \gamma_0
\end{aligned}
$$

Substituindo γ_1 e γ_2 em γ_0, temos:

$$\begin{aligned}
\gamma_0 &= \frac{\phi_1^2}{1-\phi_2}\gamma_0 + \left[\frac{\phi_2\phi_1^2}{1-\phi_2} + \phi_2^2\right]\gamma_0 + \tau^2 \\
&= \frac{(1-\phi_2)\tau^2}{(1-\phi_2) - \phi_1^2(1+\phi_2) + \phi_2^2(1-\phi_2)}.
\end{aligned}$$

A função de autocorrelação, por sua vez, é dada por:

$$\begin{aligned}
\rho_l &= \frac{\gamma_l}{\gamma_0} \\
&= \phi_1\frac{\gamma_{l-1}}{\gamma_0} + \phi_2\frac{\gamma_{l-2}}{\gamma_0} \\
&= \phi_1\rho_{l-1} + \phi_2\rho_{l-2}.
\end{aligned}$$

Processo Autorregressivo de Ordem 2 - AR(2)

- Note que, a partir dos resultados encontrados para as covariâncias, conhecendo os valores para as funções γ_0, γ_1 e γ_2, é possível um sistema com três equações e três parâmetros desconhecidos (ϕ_1, ϕ_2 e τ^2)

- Equações de Yule-Walker.

- Substituindo os valores teóricos de γ_0, γ_1, γ_2 por suas estimativas $\hat{\gamma}_0$, $\hat{\gamma}_1$, $\hat{\gamma}_2$, obtemos estimativas $\hat{\phi}_1$, $\hat{\phi}_2$ e $\hat{\tau}^2$.

- Tal método é de simples implementação e é utilizado no ajuste de modelos autorregressivos (uma vez que a ordem tenha sido fixada).

- Também aplicado em modelos autorregressivos mais gerais (ordem p). Neste caso, temos um sistema linear com $(p+1)$ equações e $(p+1)$ parâmetros.

Processo Autorregressivo de Ordem p - AR(p)

- Os resultados para os modelos AR(1) e AR(2) podem ser generalizados para um modelo AR(p):

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + w_t \\
&= (1 - \phi_1 B - \ldots - \phi_p B^p) w_t
\end{aligned}
$$

onde p é um inteiro não negativo e $\{W_t\}$ é um ruído branco.

- Processo estacionário caso as raízes de $\phi(\lambda) = (1 - \phi_1 \lambda - \ldots - \phi_p \lambda^p)$ estejam fora do círculo unitário.
- De forma mais geral, $\{Y_t\}$ é AR(p) se existe ϕ_0 tal que $\{(Y_t - \phi_0)\}$ é AR(p). Neste caso, sob estacionariedade, o valor esperado de Y_t é dado por:

$$
\mu = \mathrm{E}(Y_t) = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p}.
$$

Processo Autorregressivo de Ordem p - AR(p)

- Vimos que é possível estimar os parâmetros do modelo a partir de um sistema de equações uma vez fixada a ordem do processo autorregresivo.
- Uma pergunta relevante seria como definimos a ordem do processo autorregressivo neste caso.
- Considerando um modelo AR(p) estacionário gaussiano, temos os seguintes resultados para a função de autocorrelação parcial (PACF) amostral:
  - $\hat{\phi}_{p,p} \to \phi_p$, $n \to \infty$;
  - $\hat{\phi}_{l,l} \to 0$ para todo $l > p$;
  - $\text{Var}(\hat{\phi}_{l,l}) \to 1/n$ para todo $l > p$.
- Tais resultados indicam que o PACF de um modelo AP(p) zera na defasagem p.

Processo Autorregressivo de Ordem p - AR(p)

- Em análise de séries temporais, muito do nosso interesse está em fazer previsões a partir dos dados observados.

- Suponha que estejamos no tempo h (**origem**) e tenhamos interesse em prever Y_{h+l} onde $l \geq 1$ (**horizonte**).

  - $\mathcal{J}_h$ é a informação disponível em h;

  - $y_h(l)$ é a previsão l-passos à frente de Y_t com origem h.

Processo Autorregressivo de Ordem p - AR(p)

- Previsão 1 passo à frente:

$$y_h(1) = \mathrm{E}(Y_{h+1}|\mathcal{J}_n) = \phi_0 + \phi_1 y_h + \phi_2 y_{h-1} + \ldots + \phi_p y_{h+1-p}.$$

- Previsão 2 passos à frente:

$$\begin{aligned} y_h(2) &= \mathrm{E}(Y_{h+2}|\mathcal{J}_n) \\ &= \phi_0 + \phi_1 y_h(1) + \phi_2 y_h + \ldots + \phi_p y_{h+2-p} \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_h + \ldots + \phi_p y_{h+1-p}) + \phi_2 y_h + \ldots + \phi_p y_{t+2-p}. \end{aligned}$$

- Podemos aplicar o mesmo procedimento para computar $y_h(l)$ para qualquer l.
- As previsões são combinações lineares dos p valores observados mais recentes da série com coeficientes computados indutivamente a partir dos coeficientes do modelo.
- Previsões convergem rapidamente para a média da série $(\phi_0/(1 - \phi_1 - \ldots - \phi_p))$.

Processos de Médias Móveis (MA)

Processo de Média Móvel de Ordem 1 - MA(1)

- Seja $\{W_t\}$ ruído branco, um processo MA(1) pode ser descrito como a seguir:

$$y_t = c_0 + \theta_1 w_{t-1} + w_t.$$

- Por conveniência, assumimos $c_0 = 0$ para computar as autocovariâncias.

$$\begin{aligned}
\gamma_l &= \mathrm{E}(Y_t Y_{t-l}) \\
&= \mathrm{E}((\theta_1 W_{t-1} + W_t)(\theta_1 W_{t-l-1} + W_{t-l}) \\
&= \mathrm{E}(\theta_1^2 W_{t-1} W_{t-l-1}) + \mathrm{E}(\theta_1 W_{t-1} W_{t-l}) + \mathrm{E}(\theta_1 W_t W_{t-l-1}) + \mathrm{E}(W_t W_{t-l}).
\end{aligned}$$

- Para $l = 0$, temos:

$$\begin{aligned}
\gamma_0 &= \mathrm{E}(\theta_1^2 W_{t-1} W_{t-1}) + \mathrm{E}(\theta_1 W_{t-1} W_t) + \mathrm{E}(\theta_1 W_t W_{t-1}) + \mathrm{E}(W_t W_t) \\
&= \theta_1^2 \tau^2 + \tau_2 \\
&= (1 + \theta_1^2)\tau^2.
\end{aligned}$$

Processo Média Móvel de Ordem 1 - MA(1)

- Para $l = 1$, temos:

$$\begin{aligned}\gamma_1 &= \mathrm{E}(\theta_1^2 W_{t-1} W_{t-2}) + \mathrm{E}(\theta_1 W_{t-1} W_{t-1}) + \mathrm{E}(\theta_1 W_t W_{t-2}) + \mathrm{E}(W_t W_{t-1}) \\ &= \theta_1 \tau^2.\end{aligned}$$

- Para $l > 1$, temos $\gamma_l = 0$.

- Note que a função de autocovariância será zero para toda defasagem maior que 1 em um modelo MA(1).

- $\rho_0 = 1$, $\rho_1 = \theta_1/(1 + \theta_1^2)$, $\rho_l = 0$ para $l > 1$.

Processo Média Móvel de Ordem 2 - MA(2)

- Seja $\{W_t\}$ ruído branco, um processo MA(2) é dado pela seguinte equação:

$$y_t = c_0 + \theta_1 w_{t-1} + \theta_2 w_{t-2} + w_t.$$

- Assumindo $c_0 = 0$, teremos as autocovariâncias como a seguir:

$$\begin{aligned}
\gamma_0 &= (1+\theta_1^2+\theta_2^2)\tau^2 \\
\gamma_1 &= (\theta_1+\theta_1\theta_2)\tau^2 \\
\gamma_2 &= \theta_2\tau^2 \\
\gamma_l &= 0. \text{ para } l > 2.
\end{aligned}$$

- Consequentemente, temos:

$$\begin{aligned}
\rho_0 &= 1 \\
\rho_1 &= (\theta_1+\theta_1\theta_2)/(1+\theta_1^2+\theta_2^2) \\
\rho_2 &= \theta_2/(1+\theta_1^2+\theta_2^2) \\
\rho_l &= 0. \text{ para } l > 2.
\end{aligned}$$

Processo Média Móvel de Ordem q - MA(q)

- Seja $\{W_t\}$ ruído branco, generalizando os modelos MA(1) e MA(2), descrevemos um modelo MA(q):

$$y_t = c_0 + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q} + w_t.$$

- Processos de médias móveis são fracamente estacionários.
  - $\mathrm{E}(Y_t) = c_0$;
  - $\mathrm{Var}(Y_t) = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\tau^2$.
- Assumindo $c_0 = 0$, a função de autocovariância γ_l e, consequentemente, a função de autocorrelação ρ_l serão iguais a 0 para $l > q$.
- Tal resultado dá uma indicação acerca da escolha da ordem de um processo de médias móveis.

Processo Média Móvel de Ordem q - MA(q)

- Assim como verificamos que um processo AR(1) é um processo MA(∞), também por recursão, podemos verificar que um processo MA(1) é um processo AR(∞).

$$\begin{aligned}
y_t &= \theta_1 w_{t-1} + w_t \\
&= \theta_1(x_{t-1} - \theta_1 w_{t-2}) + w_t \\
&= \theta_1 x_{t-1} - \theta_1^2 w_{t-2} + w_t \\
&= \theta_1 x_{t-1} - \theta_1^2(x_{t-2} - \theta_1 w_{t-3}) + w_t \\
&= \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 w_{t-3} + w_t \\
&\vdots \\
&= \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 w_{t-3} - \ldots + w_t
\end{aligned}$$

- Ressalta-se que o resultado vale para $|\theta_1| < 1$

Processos Autorregressivos de Médias Móveis (ARMA)

Processos Autorregressivos de Médias Móveis

- Seja $\{W_t\}$ ruído branco, um processo autorregressivo de médias móveis de ordem p e q, ARMA(p,q), é descrito da seguinte forma:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j} + w_t.$$

- Uma série temporal é dita por um processo ARIMA se, quando tomada a diferença um número finito de vezes, a mesma se torna um processo ARMA.
- $Y \sim \text{ARIMA}(p, d, q)$ se $Y \sim \text{ARMA}(p, q)$ após d diferenças serem tomadas.
    - $Y \sim \text{ARIMA}(p, d, q) \iff \nabla^d Y \sim \text{ARMA}(p, q)$.

Processos Autorregressivos de Médias móveis

- Vimos anteriormente que podemos utilizar o ACF e o PACF no ajuste de modelos MA(q) e AR(p).
- No caso dos modelos ARIMA(p,d,q), podemos proceder da seguinte forma:
    1. Caso necessário, transforme a série temporal em estacionária via diferenças;
    2. Verifique se a mesma é um ruído branco. Caso positivo, não há a necessidade de ajustar um modelo;
    3. Ajuste um modelo AR(p) aos dados e calcule os resíduos;
    4. Ajuste um modelo MA(q) aos resíduos ou aos dados caso o modelo AR(p) não seja considerado adequado;
    5. Utilizando as ordens p e q determinadas nos passos anteriores, ajuste um modelo ARMA(p,q) aplicando máxima verossimilhança;
    6. Analise os resíduos e teste para ruído branco.

Processos Autorregressivos de Médias móveis

- Uma alternativa à identificação da ordem de um processo autorregressivo via PACF ou de um processo de médias móveis via ACF ou mesmo de um processo ARIMA, seguindo os passos descritos anteriormente, se dá através de critérios de seleção de modelos.

- Dentre os critérios de seleção de modelos, destacamos o AIC (Akaike Information Criterion) e o BIC (Bayesian information criterion).

- O AIC (Akaike Information Criterion) é dado pela seguinte equação;

$$\text{AIC} = \underbrace{-2 \times \log(\text{verossimilhança})}_{\text{bondade do ajuste}} + \underbrace{2 \times (\text{número de parâmetros})}_{\text{penalização}},$$

onde a verossimilhança é avaliada considerando o modelo estimado.

- Considerando um modelo AR(p) estimado utilizando Yule-Walker, ao invés de avaliar a função de verossimilhança, podemos aplicar:

$$\text{AIC} = n \times (\log(\tilde{\tau}^2) + 1) + 2(p + 1),$$

onde n é o tamanho da amostra e $\tilde{\tau}^2$ é o valor estimado para a variância do ruído branco (τ^2).

- Para identificação da ordem de um processo AR via critérios de seleção de modelos, pode, por exemplo, realizar os seguintes passos:

1. Ajuste de modelos AR com diferentes ordens;
2. Calcule o critério para cada um dos modelos;
3. Selecione o modelo com a melhor métrica.

Processos Autorregressivos de Médias móveis

- Há ainda uma versão "corrigida" do *Akaike Information Criterion*, AIC_c.

- Considerando um modelo ARIMA, o AIC_c é dado pela seguinte equação:

$$\mathrm{AIC}_c = \mathrm{AIC} + 2 \times \frac{(p+q+k+1)(p+q+k+2)}{n-p-q-k-2},$$

  onde $k = 1$ se $\mu \neq 0$ e $k = 0$ se $\mu = 0$.

- O BIC (Bayesian information criterion), por sua vez, é dado pela seguinte equação;

$$\mathrm{BIC} = \underbrace{-2 \times \log(\text{verossimilhança})}_{\text{bondade do ajuste}} + \underbrace{(\text{número de parâmetros}) \times \log(\text{tamanho da amostra})}_{\text{penalização}},$$

  onde a verossimilhança é avaliada considerando o modelo estimado.

Processos Autorregressivos de Médias móveis

- As penalizações para o AIC e o BIC são 2 e $\log(n)$ respectivamente.
  - BIC tende a selecionar modelos AR de ordens menores quando o tamanho da amostra é moderado ou grande.
- A ordem dos processos é identificada através dos modelos que produzem os menores valores para o AIC ou para o BIC.
- Vale destacar que critérios de seleção de modelos tendem a ser úteis na seleção dos valores de p e q e tendem a não ser tão eficazes na seleção de d.
  - Tomar diferenças altera o dado no qual a verossimilhança é computada;
  - Critérios de seleção de modelos não são comparáveis para diferentes valores de d.

Processos Autorregressivos de Médias móveis

- Em R, modelos ARIMA são comumente ajustados através da função `auto.arima()`.
    - Aplica uma variação do algoritmo descrito em Hyndman and Khandakar (2008);
    - Combina testes de raiz unitária, AIC_c e estimação por máxima verossimilhança.
- Os argumentos da função `auto.arima()` possibilitam a especificação de diferentes variações do algoritmo.
- A seguir, descrevemos a versão padrão do algoritmo.

Processos Autorregressivos de Médias móveis

1. O número de diferenças $0 \leq d \leq 2$ é determinado aplicando repetidos testes KPSS.
2. Os valores de p e q são escolhidos pela minimização do AIC$_c$. O algoritmo aplica uma busca *stepwise* em vez de considerar todas as combinações possíveis de p e q.
   a. Quatro modelos iniciais são ajustados: ARIMA(0,d,0), ARIMA(2,d,2), ARIMA(1,d,0), ARIMA(0,d,1).
   μ é incluído a menos que $d = 2$; se $d \leq 1$, um modelo adicional também é ajustado: ARIMA(0,d,0) com $\mu = 0$.
   b. O melhor modelo ajustado em (a) de acordo com o AIC$_c$ é fixado como o "modelo atual".
   c. Variações do modelo atual são consideradas:
      - Varie p e q do modelo atual adicionando e subtraindo 1 unidade;
      - Inclua/exclua μ do modelo atual.
   d. Repita o passo 2(c) até que não seja encontrado um AIC$_c$ menor.

Processos Autorregressivos de Médias móveis

- À medida que vamos combinando diferentes componentes com o intuito de formar modelos mais complexos, aplicar o **operador de defasagem** torna-se conveniente.

- Além de "retroceder" variáveis, o operador também é conveniente na descrição de processos de diferenças.

  - Diferença de primeira ordem:

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t.$$

  - Diferença de segunda ordem:

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t+2} = (1 - B)^2 y_t.$$

Processos Autorregressivos de Médias móveis

- De forma geral, diferenças de ordem d podem, portanto, ser escritas como $(1-B)^d y_t$.
- O operador de defasagem pode ser aplicado na combinação de diferenças. Em particular, termos envolvendo B podem ser multiplicados. Por exemplo, uma diferença sazonal seguida por uma primeira diferença:

$$(1-B)(1-B^m)y_t = (1-B-B^m+B^{m+1})y_t = y_t - y_{t-1} - y_{t-m} + y_{t-m+1}.$$

- Utilizando o operador de defasagem, o modelo ARIMA(p,d,q) é escrito da seguinte forma:

$$\underbrace{(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)}_{\text{AR}(p)} \underbrace{(1-B)^d}_{d \text{ diferenças}} (y_t - \mu) = \underbrace{(1+\theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)}_{\text{MA}(q)} w_t.$$

Processos Autorregressivos de Médias móveis

- Através desta notação, podemos escrever um guia (passo a passo) sobre como previsões em modelos ARIMA(p,d,q) podem ser construídas:

  1. Expanda o modelo ARIMA de tal forma que y_t esteja na parcela esquerda da equação, enquanto os demais termos estejam na parcela direita da equação;
  2. Reescreva a equação substituindo h (origem) por $h + l$;
  3. No lado direito da equação, substitua as futuras observações por suas previsões, os erros futuros por zero e os erros passados por seus respectivos resíduos.

- Começando por $l = 1$, os passos acima podem ser repetidos para $l = 2, 3, \ldots$ até que todas as previsões sejam calculadas.

EXEMPLO: ARIMA(3,1,1).

- Primeiro passo:

$$
\begin{aligned}
&(1-\hat{\phi}_1 B-\hat{\phi}_2 B^2-\hat{\phi}_3 B^3)(1-B)y_t=(1+\hat{\theta}_1)w_t \Rightarrow\\
&\left[1-(1+\hat{\phi}_1)B+(\hat{\phi}_1-\hat{\phi}_2)B^2+(\hat{\phi}_2-\hat{\phi}_3)B^3+\hat{\phi}_3 B^4\right]y_t=(1+\hat{\theta}_1)w_t \Rightarrow\\
&y_t-(1+\hat{\phi}_1)y_{t-1}+(\hat{\phi}_1-\hat{\phi}_2)y_{t-2}+(\hat{\phi}_2-\hat{\phi}_3)y_{t-3}+\hat{\phi}_3 y_{t-4}=w_t+\hat{\theta}_1 w_{t-1} \Rightarrow\\
&y_t=(1+\hat{\phi}_1)y_{t-1}-(\hat{\phi}_1-\hat{\phi}_2)y_{t-2}-(\hat{\phi}_2-\hat{\phi}_3)y_{t-3}-\hat{\phi}_3 y_{t-4}+w_t+\hat{\theta}_1 w_{t-1}.
\end{aligned}
$$

- Segundo passo:

$$
\hat{y}_h(1)=(1+\hat{\phi}_1)y_h-(\hat{\phi}_1-\hat{\phi}_2)y_{h-1}-(\hat{\phi}_2-\hat{\phi}_3)y_{h-2}-\hat{\phi}_3 y_{h-3}+w_h(1)+\hat{\theta}_1 w_h.
$$

- Terceiro passo ($h = n$):

$$\hat{y}_n(1) = (1 + \hat{\phi}_1)y_n - (\hat{\phi}_1 - \hat{\phi}_2)y_{n-1} - (\hat{\phi}_2 - \hat{\phi}_3)y_{n-2} - \hat{\phi}_3 y_{n-3} + \hat{\theta}_1 \tilde{w}_n.$$

  - Podemos escrever a previsão dois passos à frente:

$$\hat{y}_n(2) = (1 + \hat{\phi}_1)\hat{y}_n(1) - (\hat{\phi}_1 - \hat{\phi}_2)y_n - (\hat{\phi}_2 - \hat{\phi}_3)y_{n-1} - \hat{\phi}_3 y_{n-2}.$$

  - O processo é repetido de tal forma que as previsões de todos os períodos futuros sejam obtidas.

Teste da raiz unitária

Teste da raiz unitária

- Vimos que o processo AR(p) é estacionário quando todas as raízes do polinômio $\phi(\lambda)$ estão fora do círculo unitário.
- Quando uma das raízes é igual a 1, dizemos que o processo possui **raiz unitária**.
  - O processo é não estacionário.
- Considerando, por exemplo, um processo AR(1) tal que $y_t = \phi_0 + \phi_1 y_{t-1} + w_t$ é possível pensar no seguinte teste de hipóteses:

$$\begin{aligned} H_0 : \phi_1 &= 1 \text{ (processo não-estacionário)} \\ H_1 : \phi_1 &< 1 \text{ (processo estacionário).} \end{aligned}$$

- Neste caso, utilizar um teste t parece natural para avaliar H_0; contudo, a estatística de teste t não possui as distribuições usuais.

Teste da raiz unitária

- Tomando a diferença entre y_t e y_{t-1}, o processo AR(1) pode ser reparametrizado da seguinte forma:

$$\begin{aligned} \Delta y_t &= \phi_0 + (\phi_1 - 1)y_{t-1} + w_t \\ &= \phi_0 + \alpha_0 y_{t-1} + w_t. \end{aligned} \tag{4}$$

- Com base na reparametrização, o teste pode ser reescrito:

$$\begin{aligned} H_0 : \alpha_0 &= \phi_1 - 1 = 0 \\ H_1 : \alpha_0 &< 0. \end{aligned}$$

- Testar $\phi_1 = 1$ é equivalente a testar $\alpha_0 = 0$
- Testes da raiz unitária são comumente computados utilizando a reparametrização em (4).

Teste da raiz unitária

- Modelo de regressão linear simples.
  - Estimação via Mínimos Quadrados Ordinários (MQO).
- O teste procede considerando uma estatística de teste t:

$$t_{\phi_1=1} = \frac{\hat{\phi}_1 - 1}{\text{se}(\hat{\phi}_1)}.$$

  - Teste t conhecido como teste de Dickey-Fuller (DF);
  - A distribuição limite de $t_{\phi_1=1}$ é chamada de distribuição Dickey-Fuller (DF) e não possui representação em forma fechada (quantis são calculados via aproximação numérica ou simulação).

Teste da raiz unitária

- Considerando um processo AR(p), o processo pode ser reescrito aplicando uma ideia similar de reparametrização:

$$\Delta y_t = \phi_0 + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \ldots + \alpha_{p-1} \Delta y_{t-(p-1)} + w_t,$$

onde

  - Os parâmetros α são funções dos parâmetros ϕ;
  - $\alpha_0 = \phi_1 + \phi_2 + \ldots + \phi_p - 1$ Estimação via Mínimos Quadrados Ordinários (MQO).

- O processo será não-estacionário quando $\alpha_0 = 0$ $(\phi_1 + \phi_2 + \ldots + \phi_p = 1)$.

Teste da raiz unitária

- Novamente, o teste pode ser reescrito:

$$
\begin{aligned}
H_0 : \alpha_0 &= 0 \\
H_1 : \alpha_0 &< 0.
\end{aligned}
$$

- Assim como no teste de Dickey-Fuller, temos um modelo de regressão linear simples.
  - Estimação via Mínimos Quadrados Ordinários (MQO).
- Teste de Dickey-Fuller aumentado (ADF).

Material suplementar

Modelos ARIMA sazonais (SARIMA)

- Sazonalidade é um comportamento comumente apresentado ao analisarmos séries temporais, ou seja, os dados frequentemente apresentam ciclos ou comportamentos periódicos.

- Em algumas aplicações, a sazonalidade possui importância secundária, sendo removida dos dados.

  - O procedimento de remoção da sazonalidade de uma série temporal é denominado **ajuste sazonal**;

  - O procedimento de inferência é realizado com base na série temporal ajustada sazonalmente.

- Muitos dados econômicos publicados pelo IBGE e pelo Banco Central são ajustados sazonalmente (por exemplo, IPCA e IBC-Br).

Modelos ARIMA sazonais (SARIMA)

- Em muitas aplicações, no entanto, a sazonalidade é uma característica tão relevante quanto as demais e deve ser modelada.

- Nos casos em que desejamos modelar a sazonalidade, uma abordagem possível se dá através de modelos ARIMA sazonais (SARIMA).

- Um modelo ARIMA sazonal consiste em incluir termos sazonais adicionais ao modelo ARIMA:

$$\text{ARIMA}\ \underbrace{(p, d, q)}_{\substack{\text{componente} \\ \text{não sazonal}}}\ \underbrace{(P, D, Q)_m}_{\substack{\text{componente} \\ \text{sazonal}}},$$

onde m é o número de observações por ano.

Modelos ARIMA sazonais (SARIMA)

- A componente sazonal do modelo consiste de termos similares à componente não sazonal, porém envolve defasagens do período sazonal.

- Como exemplo, podemos considerar um modelo ARIMA(1,1,1)(1,1,1)$_4$ (com $\mu = 0$) para uma série temporal trimestral:

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)(1 + \Theta_1 B^4)w_t.$$

Observação: Note que os termos referentes à componente sazonal são simplesmente multiplicados pelos termos da componente não sazonal.

Modelos ARIMA para d fracionário (ARFIMA)

- Ao lidarmos com modelos ARIMA, precisamos fixar um número inteiro d referente ao número de diferenças que devem ser tomadas a fim de que a série temporal $\{(1-B)^d Y_t\}$ seja um processo estacionário.
  - O procedimento padrão para determinar se uma série temporal é estacionária ou não se dá através da aplicação de testes de raiz unitária (Aula 05);
  - Testes de raiz unitária (Aula 05) tendem a apresentar problemas para distinguir entre séries de fato não estacionárias e séries estacionárias com quebra estrutural ou mudança de regime;
  - Em casos nos quais tais testes não conseguem diferenciar claramente, costuma-se aplicar uma diferença de primeira ordem.
- Modelos ARFIMA são desenhados para representar séries temporais que exibem uma **dependência de longo prazo** tal que não sejam classificadas como estacionárias, mas não podem ser ditas não estacionárias.

Modelos ARIMA para d fracionário (ARFIMA)

- Modelos ARFIMA permitem que a série temporal seja integrada de forma fracionária.
  - Generaliza o modelo ARIMA de ordem de integração inteira d ao permitir que d assuma valores entre $-0,5 < d < 0,5$.
- O conceito de integração fracionária usualmente refere-se à definição de uma série temporal com dependência de longo prazo ou memória longa.
- Processos ARIMA estacionários podem ser considerados séries temporais de memória curta.
  - Modelos AR(p) possuem memória infinita (MA(∞)), mas os efeitos dos valores passados seguem uma forma geométrica, tendendo a valores próximos de zero rapidamente;
  - Modelos MA(q) possuem memória de exatamente q períodos, decaindo a zero rapidamente.

- Um modelo ARFIMA(p,d,q) é caracterizado pela seguinte equação:

  $(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(1 - B)^d y_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q) w_t,$

  onde

  $$(1 - B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d) B^k}{\Gamma(-d)\Gamma(k+1)},$$

  na qual $\Gamma(\cdot)$ denota a função gamma e d pode assumir qualquer valor real.

- O processo é não estacionário para $d \geq 0,5$.

Modelos ARIMA para d fracionário (ARFIMA)

- Assumindo que $d \in [0, 0.5)$, Hosking (1981) demonstrou que a função de autocorrelação de um processo ARFIMA é proporcional a k^{2d-1} à medida que $k \to \infty$.

    - A autocorrelação de um processo ARFIMA decai hiperbolicamente para zero ($k \to \infty$) em contraste com o decaimento geométrico (mais rápido) do processo estacionário ARMA.

- Para $d \in (0, 0.5)$, o processo ARFIMA é dito de memória longa ou de dependência de longo prazo positiva; para $d \in (-0.5, 0)$, o processo é dito de dependência de longo prazo negativa.

Referências

Box, G. E. P., and David A. Pierce. 1970. "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association* 65 (332): 1509–1526.

Hosking, J. R. M. 1981. "Fractional differencing." *Biometrika* 68 (1): 165–176.

Hyndman, R. J., and Y. Khandakar. 2008. "Automatic time series forecasting: The forecast package for R." *Journal of Statistical Software* 27 (1): 1–22.

Ljung, G. M., and G. E. P. Box. 1978. "On a measure of lack of fit in time series models." *Biometrika* 65 (2): 297–303.