

Séries Temporais

James D Santos

2025-04-09

Table of contents

*Prefácio

Estas são as notas de aula estão sendo produzidas para a utilização na disciplina Séries Temporais, no Bacharelado de Estatística.

Considere-as um rascunho, e portanto, sujeita a erros.

Dúvidas e sugestões podem ser enviadas para o e-mail james@ufam.edu.br

Summary

In summary, this book has no content whatsoever.

1 + 1

[1] 2

Introdução

1 Notações

Serão utilizadas letras minúsculas para designar tanto variáveis aleatórias quanto seus respectivos valores observados, entando a diferença clara no contexto. Exemplo: em

$$x_t \sim \text{Normal}(0, 1),$$

x_t representa uma variável aleatória, enquanto que em $x_t = 0$ é um valor observado.

Vetores serão denotados por negritos e sempre serão vetores-coluna. Exemplo

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}.$$

O vetor \mathbf{x}' é o transposto de \mathbf{x} .

Para $\mathcal{T} = \{1, 2, \dots\}$,

- Se $A \subset \mathcal{T}$. Então $x_A = \{x_t, t \in A\}$.
- $x_{a:b} = x_a, x_{a+1}, \dots, x_{b-1}, x_b$.
- Um vetor de dimensão q observado no tempo t é escrito como

$$\mathbf{x}_t = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}_t.$$


2 O que é uma análise de séries temporais?

Considera-se que uma série temporal é uma coleção de observações realizadas ao longo do tempo. Será utilizada a notação x_t para designar o valor registrado no tempo t e $\mathcal{D}_t == \{x_1, \dots, x_t\}$ representará a série observada até o tempo t .

Existem três objetivos principais no estudo de séries temporais


- *Previsão*: Dado \mathcal{D}_t a previsão trata do problema de realizar inferências sobre x_{t+h} , com $h > 0$.
- *Suavização (ou alisamento)*: Dado \mathcal{D}_t a suavização trata do problema de realizar inferências baseadas x_{t-h} , com $h > 0$
- *Monitoramento*: detectar em tempo real as mudanças ou discrepâncias no comportamento do processo.

Note que tais objetivos só fazem sentido se há alguma estrutura de dependência entre as variáveis que compõe a série temporal. Para ilustrar, considere a figura abaixo representa o gráfico a série temporal com o número anual de embarques e desembarques de passageiros em vôos domésticos no aeroporto Eduardo Gomes.



intro_files/figure-pdf/unnamed-chunk-1-1.pdf

Ainda considerando a série acima, seja x_t o número de embarques e desembarques registrado no ano t . A figura abaixo mostra o diagrama de dispersão entre x_t e x_{t-1} , de onde é possível observar a correlação positiva, estimada em 0,86.



intro_files/figure-pdf/unnamed-chunk-2-1.pdf

De posse desses resultados, pode-se imaginar um primeiro modelo, no qual a relação entre o presente e o passado imediato é ditado por uma regressão linear simples, gerando a equação


$$\hat{x}_t = 7,589 \times 10^5 + 0,7109x_{t-1}.$$

Sabendo que $x_{2017} = 2.376.505$, uma previsão para 2018 seria $\hat{x}_{2018} = 2.448.357$. O valor observado em 2018 foi 2.572.159, gerando um erro de previsão igual a $x_{2018} - \hat{x}_{2018} = 195.654$ embarques e desembarques domésticos.

3 Exemplos de séries temporais

3.1 Eletrocardiograma

```
ts.plot(ECG)
```



intro_files/figure-pdf/unnamed-chunk-4-1.pdf

Figure 1: 1800 medidas da taxa cardíaca instantânea, em batidas por minuto, de um indivíduo.

3.2 Produto Interno Bruto Brasileiro

```
ts.plot(PIB)
```

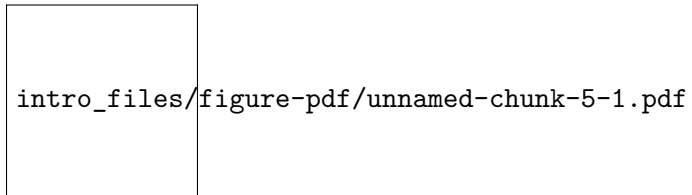


Figure 2: PIB entre 1967 e 2014 corrigidos pelo valor do dólar em 4/2015.

3.3 Mortes por doenças pulmonares no Reino Unido

```
ts.plot(ldeaths)
```

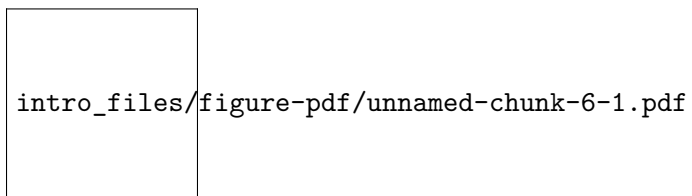


Figure 3: PIB entre 1967 e 2014 corrigidos pelo valor do dólar em 4/2015.

Criando séries no R

Esta seção tem por objetivo mostrar algumas funções em R para a criação e análise exploratória de séries temporais.

4 A classe ts

Para todos os efeitos, uma série temporal é um vetor numérico. O vetor abaixo armazena o número de nascidos vivos por mês na cidade de Manaus em 2021, sendo `x[1]` o mês de janeiro e assim sucessivamente.

```
x <- c( 3043, 2902, 3166, 3014, 3095, 2955, 3087, 3141,  
       3129, 3096, 3191, 3222)
```

Por sua vez, o gráfico da série temporal pode ser construído utilizando a função `plot`, com o argumento `type='l'`.

```
plot(x, type = 'l')
```

ts_window_date_files/figure-pdf/unnamed-chunk-2-1.pdf

Contudo, é útil construir a série temporal como um objeto da classe `ts`. Tal função possui dois argumentos importantes:

- **frequency**: representa o número de observações por unidade de tempo. Por exemplo, se tempo está sendo contado em anos, mas o dados são mensais, então **frequency=12**; se os dados forem trimestrais, **frequency=4** e assim por diante.
- **start**: representa o tempo da primeira observação. Pode ser representado por um único número ou por um vetor de comprimento dois. Esse último caso só é utilizado quando **frequency** é diferente de 1 e representa a ordem, em relação à frequência, da primeira observação. Por exemplo, com **frequency=12**, o vetor **start=c(1996,2)** implica que a primeira observação data de fevereiro de 1996.

No código abaixo, o vetor criado anteriormente é colocado com um objeto `ts`

```
x <- ts( x, start = c(2021,1), frequency = 12)
x
```

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 2021 | 3043 | 2902 | 3166 | 3014 | 3095 | 2955 | 3087 | 3141 | 3129 | 3096 | 3191 | 3222 |

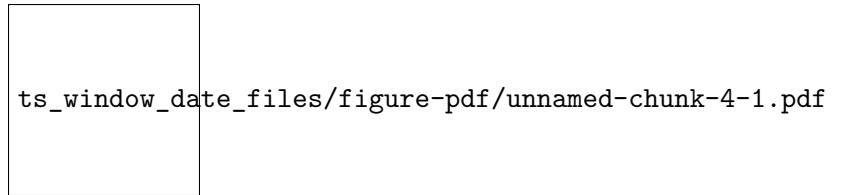
```
ts.plot(x)
```

ts_window_date_files/figure-pdf/unnamed-chunk-3-1.pdf

No gráfico acima, a parte decimal no eixo x representa a fração do tempo entre de um ano (começando em 0 e acumulando 1/12 para cada mês subsequente).

O gráfico pode ser customizado do mesmo modo que um `plot`. Abaixo segue um exemplo.

```
plot(x, ylab = 'No. nascidos vivos mensal', lwd = 2, col = 'seagreen')
```



A função `start` retorna o início da série, `end` seu fim e `frequency` o número de observações por unidade de tempo. Observe o exemplo abaixo.

```
start(x)
```

```
[1] 2021    1
```

```
end(x)
```

```
[1] 2021   12
```

```
frequency(x)
```

```
[1] 12
```

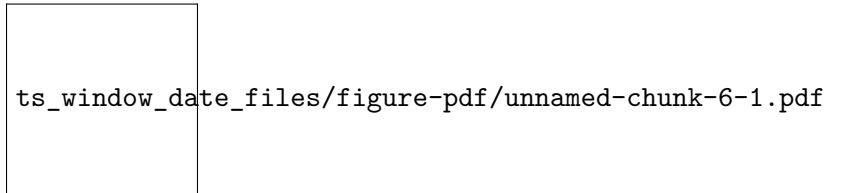
A partir das informações acima, sabe-se a série `x` é mensal (`frequency=12`), que sua primeira observação data de janeiro de 2021 e a última de dezembro de 2021.

5 A função `window`

A função `window` seleciona um subconjunto da série temporal. Abaixo foram selecionados apenas os nascimentos entre Junho e Agosto e estes valores foram registrados no gráfico.

```
z <- window(x, start=c(2021,6), end = c(2021,8))

plot(x, ylab = 'No. nascidos vivos mensal', lwd = 2, col = 'seagreen')
lines(z, col = 'brown', lwd= 2)
```



6 O pacote data.table

Assim como números e textos possuem classes específicas, as datas no ambiente R também possuem sua própria classe, denominada **Date**.

```
# 3 de agosto de 1998 (formato americano)
x <- '1998/8/3'
as.Date(x)
```

```
[1] "1998-08-03"
```

Para que o R entenda uma data fora do padrão americano, é necessário passar o formado para o argumento **format**. Seguem alguns exemplos:

```
# 3 de agosto de 1998 (formato nacional)
x <- '3/8/1998'
as.Date(x, format = '%d/%m/%Y')
```

```
[1] "1998-08-03"
```

```
x <- '3-8-1998'
as.Date(x, format = '%d-%m-%Y')
```

```
[1] "1998-08-03"
```

```
# agosto de 1998
x <- '8/1998'
as.Date(x, format = '%m/%Y')
```

```
[1] NA
```

Ao se trabalhar com fontes originais, é comum ter como unidade amostral um evento com sua data registrada. Em geral, nosso objetivo é determinar a quantidade de eventos dentro de dias, semanas, meses ou anos. O pacote `data.table` permite lidar com esse problema de modo rápido, criando um objeto deste tipo utilizando a função `fread`.

Para ilustrar, será utilizada a base de dados de acidentes com aeronaves, mantida pela Força Aérea Brasileira, que registra diariamente o número de acidentes com aeronaves.

```
library(data.table)
url <- 'https://drive.google.com/uc?authuser=0&id=1iYrnwXgmLK07x8b330aD73sc0VruZEuz&export=csv'

aereo <- fread(url, encoding = 'Latin-1')
aereo$ocorrencia_dia <- as.Date(aereo$ocorrencia_dia, '%d/%m/%Y')
```

Um objeto do tipo `data.table` permite uma série de consultas. Em geral, pode-se fazer `aereo[a,b,c]`, onde `a` é uma consulta/função nas linhas, `b` nas colunas e `c` é um agrupador. Uma excelente introdução pode ser vista em [Introduction to data.table](#).

Abaixo, foi selecionada a coluna de interesse `ocorrencia_dia`.

```
fab_dia <- aereo[, 'ocorrencia_dia', ]
head(fab_dia)
```

```
      ocorrencia_dia
1:      2023-04-05
2:      2023-06-24
3:      2023-06-27
4:      2023-06-30
5:      2023-06-25
6:      2023-06-23
```

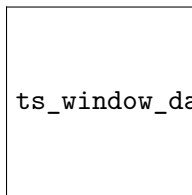
Ao utilizar o operador `.N` em `[, .N, c]`, é retornado o número de linhas que possuem o agrupamento em `c`. Abaixo, as datas do banco são agrupadas por ano.


```
fab_ano <- fab_dia[, .N, by=.(year(ocorrencia_dia))]  
fab_ano <-fab_ano[ order(year) ]  
head(fab_ano)
```

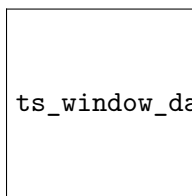
```
   year    N  
1: 2013 654  
2: 2014 569  
3: 2015 471  
4: 2016 403  
5: 2017 432  
6: 2018 444
```

Os comandos a seguir criam dois objetos do tipo `ts`, sendo um para o número anual de acidentes e outro para o mensal

```
fab_ano <- ts( fab_ano, start = 2013)  
plot(fab_ano[,2], lwd = 2, ylab = 'No. acidentes/ano', xlab = 'Ano')
```



```
fab_mes <- fab_dia[, .N, by=.(year(ocorrencia_dia), month(ocorrencia_dia))]  
  
fab_mes <-fab_mes[ order(year, month) ]  
fab_mes <- ts( fab_mes[,3], start = c(2013, 1), frequency = 12)  
plot(fab_mes, lwd = 2, ylab = 'No. acidentes/mês', xlab = 'Ano')
```



7 Exercício

Exercício 1

A série abaixo contém a data dos óbitos maternos no Brasil a partir de 2010.

```
url <- 'https://drive.google.com/uc?authuser=0&id=1tYFFT9L2iopKmBDUI3P8qNIRa0nMYj7d&export=
```

Crie uma série temporal com o número de óbitos mensal e faça um gráfico. Crie uma janela para colocar no gráfico o período da pandemia de COVID-19.

Séries Estacionárias

Uma coleção do tipo $\{x(t), t \in \mathcal{T}\}$, $\mathcal{T} \subseteq \mathbb{R}$, onde $x(t)$ é uma variável aleatória para cada t fixado, é denominada processo estocástico.

Um processo estocástico é dito ser fortemente estacionário se sua distribuição é invariante ao índice. Portanto, para qualquer t_1, \dots, t_k , a distribuição de $x(t_1), \dots, x(t_k)$ é a mesma de $x(t_1 + h), \dots, x(t_k + h)$.

Example .1. Se $x(t) \sim \text{Normal}(0, 1)$ e $x(t)$ é independente de $x(s)$ para todo $t \neq s$, então, para qualquer t_1, \dots, t_k ,

$$\begin{aligned} P(x(t_1) < x_1, \dots, x(t_k) < x_k) &= \prod_{i=1}^k P(x(t_i) < x_i) = \prod_{i=1}^k P(x(t_i + h) < x_i) \\ &= P(x(t_1 + h) < x_1, \dots, x(t_k + h) < x_k) \end{aligned}$$

logo, $\{x(t), t \in \mathbb{R}\}$ é um processo fortemente estacionário. ■

Theorem .1. Seja $\{x(t), t \in \mathbb{N}\}$ um processo estocástico com

$$x(t) = x_{(t-1)} + \varepsilon_t, \quad \varepsilon_t \sim \text{Normal}(0, 1),$$

para $t = 1, 2, \dots$ com a condição de que $x_0 \sim \text{Normal}(0, 1)$ e que $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0 \quad \forall s \neq t$. Então,

$$x_t = x_{t-1} + \varepsilon_t = \dots = x_0 + \sum_{j=1}^t \varepsilon_j \sim \text{Normal}(0, t + 1).$$

Como $x_t \sim \text{Normal}(0, t + 1)$ e, para qualquer $h > 0$, $x_{t+h} \sim \text{Normal}(0, t + h + 1)$, temos que este processo não é fortemente estacionário. ■

Definition .1. Um processo estocástico $\{y_t\}$ é dito ser fracamente estacionário (ou de segunda ordem) se

$$\begin{aligned} E(y_t) &= \mu, \\ Var(y_t) &= \nu, \\ Cov(y_t, y_s) &= E(y_t y_s) - E(y_t)E(y_s) = \gamma(t - s) \end{aligned}$$

onde μ e ν são constantes independentes de t e $\gamma(t - s)$ depende de t e s somente através da diferença $|t - s|$. ■

Example .2. Considere o processo estocástico $\{x(t), t \in \mathbb{N}\}$, onde

$$x(t) = \varepsilon(t) + \frac{1}{2}\varepsilon(t - 1)$$

onde $\varepsilon(t) \sim \text{Normal}(0, \nu)$ para $t = 1, \dots$, $\varepsilon(t)$ é independente de $\varepsilon(s)$ para todo $s \neq t$ e $\varepsilon(0) = 0$. Então:

$$\begin{aligned} E(x(t)) &= E(\varepsilon(t)) + \frac{1}{2}E(\varepsilon(t - 1)) = 0 \\ Var(x(t)) &= Var(\varepsilon(t)) + \frac{1}{4}Var(\varepsilon(t - 1)) = \frac{5}{4}\nu \end{aligned}$$

e

$$\begin{aligned} Cov(x(t), x(t + h)) &= Cov\left(\varepsilon(t) + \frac{1}{2}\varepsilon(t - 1), \varepsilon(t + h) + \frac{1}{2}\varepsilon(t + h - 1)\right) \\ &= Cov(\varepsilon(t), \varepsilon(t + h)) + \frac{1}{2}Cov(\varepsilon(t), \varepsilon(t + h - 1)) \\ &\quad + \frac{1}{2}Cov(\varepsilon(t - 1), \varepsilon(t + h)) + \frac{1}{4}Cov(\varepsilon(t - 1), \varepsilon(t + h - 1)) \\ &= \begin{cases} \frac{4}{5}\nu, & h = 0 \\ \frac{1}{2}\nu, & |h| = 1, \\ 0, & \text{caso contrário.} \end{cases} \end{aligned}$$

Portanto, o processo é fracamente estacionário. ■

Quando $\mathcal{T} = \{t \in D \subseteq \mathbb{Z}\}$, utiliza-se a notação $x(t) = x_t$. Além disso, se t pode ser interpretado como tempo, x_t é uma série temporal. Uma série temporal é dita ser estacionária se ela é fracamente estacionária e o mesmo princípio se aplica nessas notas de aula.

8 Processo estacionário ergódico

Seja x_1, x_2, \dots, x_n uma série temporal estacionária. Então, a média μ pode ser estimada por \bar{x}_n , uma vez que $E(\bar{x}) = \mu$. A variância dessa estatística é

$$\begin{aligned}
Var(\bar{x}) &= Cov(\bar{x}_n, \bar{x}_n) = Cov\left(\sum_{i=1}^n \frac{x_i}{n}, \sum_{j=1}^n \frac{x_j}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Cov(x_i, x_j) \\
&= \frac{1}{n^2} \left[\sum_{i=1}^n Cov(x_i, x_i) + 2 \sum_{i=1}^n \sum_{j \neq i}^n Cov(x_i, x_j) \right] \\
&= \frac{1}{n^2} \left[n\nu + 2 \sum_{h=1}^{n-1} (n-h) \gamma(h) \right] = \frac{\nu}{n} + \frac{2}{n} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \gamma(h)
\end{aligned}$$

Note que, diferente do caso independente e identicamente distribuído, \bar{x} não é necessariamente um estimador adequado, conforme pode ser constatado no exemplo abaixo.

Example .3. Seja x_t um processo onde $x_0 \sim \text{Normal}(0, \nu)$ e $x_t = x_0$ para todo $t > 0$. Como

$$\begin{aligned}
E(x_t) &= E(E(x_t|x_0)) = E(x_0) = 0 \\
Var(x_t) &= E(Var(x_t|x_0)) + Var(E(x_t|x_0)) = E(0) + Var(x_0) = \nu \\
Cov(x_t, x_{t-h}) &= E(Cov(x_t, x_{t-h}|x_0)) + Cov(E(x_t|x_0), E(x_{t-h}|x_0)) \\
&= E(0) + Cov(x_0, x_0) = Var(x_0) = \nu
\end{aligned}$$

e, portanto, o processo é fracamente estacionário. Contudo,

$$Var(\bar{x}) = \frac{\nu}{n} + \frac{2}{n} \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \nu = \nu,$$

portanto, o erro padrão não decai com o aumento do tamanho da amostra.

■

A partir do exemplo acima, fica claro que \bar{x} nem sempre será um estimador adequado para uma série estacionária.

Definition .2. Uma série temporal estacionária é dita ser ergódica para a média se

$$\sum_{i=1}^n \frac{x_i}{n} \xrightarrow{p} \mu,$$

quando $n \rightarrow \infty$.

A partir deste momento será considerado que toda série temporal estacionária é ergódica e, portanto \bar{x} é um estimador para μ .

Example .4. Considere novamente o processo no (**estacionario_nao_ergodico?**). Como $\bar{x}_n = x_0$, tem-se que, para $\varepsilon > 0$ arbitrário, $\bar{x} \sim \text{Normal}(0, \nu)$ e

$$P(|\bar{x}_n - 0| > \varepsilon) = 2P(x_0 > \varepsilon) = 2 \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{y^2}{2\nu}} d\nu > \frac{1}{2}$$

logo, \bar{x} não converge em probabilidade para 0 e, portanto, o processo não é ergódico na média. ■

9 Ruído branco

Definition .3. A série estacionária x_t é dita ser um ruído branco se $E(x_t) = 0$, $Var(x_t) = \nu$ e

$$Cov(x_t, x_s) = 0, \quad (1)$$

para todo $t \neq s$. ■

É imediato que o ruído branco é uma série temporal estacionária. Além disso, pela Desigualdade de Chebyshev, para qualquer $\varepsilon > 0$,

$$P(|\bar{x}_n| \geq \varepsilon) \leq \frac{E(\bar{x}_n^2)}{\varepsilon^2} = \frac{Var(\bar{x}_n)}{\varepsilon^2} = \frac{\nu}{n\varepsilon^2}$$

logo $\lim_n P(|\bar{x}_n| \leq \varepsilon) = 0$ e $\bar{x}_n \xrightarrow{p} 0$. Portanto, o ruído branco é ergódico.

Considere agora a série temporal $y_t = \mu + x_t$, onde x_t é um ruído branco. Então

$$\bar{y}_n = \mu + \bar{x}_n \xrightarrow{p} \mu$$

e \bar{y}_n é um estimador para μ .

Em certos momentos, será considerado que x_t e x_s , para todo $t \neq s$ são independentes (essa é uma condição mais forte, pois implica em $Cov(x_t, x_s) = 0$). Esse processo é denominado ruído branco independente.

Por último, também será considerado a possibilidade de que $x_t \sim \text{Normal}(0, \nu)$, com x_t e x_s independentes para todo $t \neq s$. Esse processo será denominado ruído branco gaussiano.

Revisão sobre o modelo linear

10 Definição

Para $i = 1, \dots, n$, considere o modelo linear abaixo:

$$y_i = \beta_0 + \sum_{j=1}^{p-1} x_{i,j} + \varepsilon_i = \underbrace{(1 \ x_{i,1} \ \cdots \ x_{i,p-1})}_{\mathbf{f}'_i} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}}_{\boldsymbol{\beta}} + \varepsilon_i = \mathbf{f}'_i \boldsymbol{\beta} + \varepsilon_i,$$

onde x_i é fixado e ε é um ruído branco gaussiano. Pela independência entre y_i e y_j , pode-se fazer a seguinte representação estocástica de \mathbf{y} :

$$\mathbf{y} = \mathbf{F}' \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

onde $\boldsymbol{\varepsilon} \sim \text{Normal}(\mathbf{0}, \nu \mathbf{I}_n)$ e \mathbf{F} é uma matriz $p \times n$ conhecida com i -ésima coluna dada por \mathbf{f}_i :

$$\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_n).$$

A função de verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\beta}, \nu) &\propto \left(\frac{1}{\nu}\right)^{\frac{T}{2}} \exp \left\{ -\frac{1}{2\nu} (\mathbf{y} - \mathbf{F}' \boldsymbol{\beta})' (\mathbf{y} - \mathbf{F}' \boldsymbol{\beta}) \right\} \\ &\propto \left(\frac{1}{\nu}\right)^{\frac{T}{2}} \exp \left\{ -\frac{1}{2\nu} [(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{F} \mathbf{F}' (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + R] \right\} \end{aligned}$$

onde

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{F} \mathbf{F}')^{-1} \mathbf{F} \mathbf{y}, \\ R &= (\mathbf{y} - \mathbf{F}' \hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{F}' \hat{\boldsymbol{\beta}}) \end{aligned}$$

É sabido que:

- $\hat{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\beta}$
- R é conhecido como *soma de quadrados de resíduos*
- $\hat{\nu} = R/(n - p)$ é um estimador não viciado para ν .

Além disso, tem-se que

$$\begin{aligned}\hat{\beta} &\sim \text{Normal}_p(\beta, (\mathbf{F}\mathbf{F}'_n)^{-1}\nu) \\ \frac{R}{\nu} &\sim \chi^2_{n-p} \\ \sqrt{n-p} \frac{\hat{\beta} - \beta}{\sqrt{R}} &\sim t_{n-p}(\mathbf{0}_p, (\mathbf{F}'\mathbf{F})^{-1})\end{aligned}$$

11 Resíduos e valores ajustados

O valor ajustado da i -ésima observação é dado por

$$\hat{y}_i = \mathbf{f}'_i \hat{\beta} \quad (3)$$

e, concluímos que

$$\hat{\mathbf{y}} \sim \text{Normal}(\mathbf{F}'\beta, \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}\nu)$$

O respectivo resíduo é dado por

$$e_i = y_i - \hat{y}_i,$$

e, como o vetor de resíduos é dado por

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}},$$

tem-se que

$$\mathbf{e} \sim \text{Normal}(\mathbf{0}_n, \nu(\mathbf{I}_n - \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}))$$

Denotando $\mathbf{H} = \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}\mathbf{F}$, de h_i como sendo o i -ésimo elemento na diagonal de \mathbf{H} , defini-se o resíduo studentizado por

$$\tilde{e}_i = \frac{e_i}{\sqrt{\hat{\nu}(1 - h_i)}}$$

Acontece que, se a suposição de ruído branco gaussiano for verdadeira, \tilde{e}_i tende a se comportar como um ruído branco.

12 Seleção de modelos

Modelos lineares podem ser comparados através do critério de informação de Akaike (AIC):

$$-2 \log L(\hat{\beta}, \hat{\nu}) - 2(p + 1).$$

Considera-se como mais parcimonioso o modelo com menor valor do AIC.

13 Robustez do modelo linear

Embora tenhamos utilizado o ruído branco gaussiano, estes resultados ainda podem ser aplicados quando ε_n é um ruído branco qualquer.

De fato, pode-se mostrar que os estimadores são os mesmos obtidos pelo método dos mínimos quadrados. Neste caso, as distribuições dos estimadores podem ser utilizadas como aproximações.

Tendência

14 O que é tendência?

Diz-se que uma série temporal observada possui tendência quando ela exibe um padrão de crescimento ou decrescimento em médio/longo prazo.

A Figure ?? mostra a série do número mensal de acidentes com aeronaves, construída através dos dados diários mantidos pela Força Aérea Brasileira. Note uma tendência de decrescimento na série até meados de 2016, substituída então por uma tendência de crescimento.

```
url <- 'https://www.dropbox.com/scl/fi/kq4jwbovu94u857238sus/N-mensal-de-acidentes-com-aer  
x <- read.csv(url, h = T)  
acidentesFAB <- ts( x, start = c(2013,1), frequency=12)  
ts.plot(acidentesFAB, lwd = 2, xlab = 'Ano', ylab = 'No. acidentes')
```

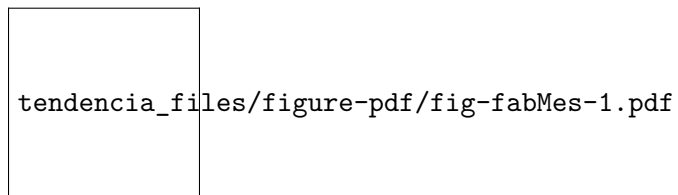


Figure 4: Número mensal de acidentes envolvendo aeronaves. Fonte: FAB

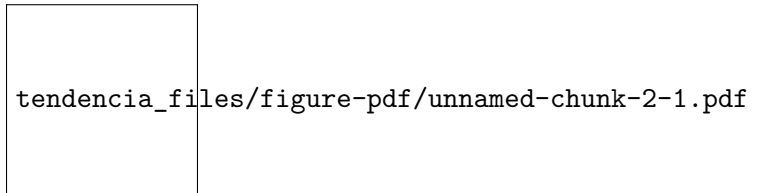
15 Tendência aleatória e tendência determinística

A tendência pode ser duas naturezas: determinística ou aleatória.

A tendência aleatória é construída ao acaso. Considere, por exemplo, o passeio aleatório definido por $x_0 = 0$ e $x_t = x_{t-1} + \varepsilon_t$, onde ε_t é um ruído branco gaussiano com $\nu = 1$. Já foi

mostrado que $E(x_t) = 0$ e $Var(x_t) = t$. A figura abaixo apresenta uma série simulada desse processo.

```
set.seed(1)
x = 0
for( t in 2:100) x[t] = x[t - 1] + rnorm(1,0,1)
ts.plot(x, lwd = 2)
```



Observe que a série exibe uma tendência, mas não há qualquer explicação para a sua existência, uma vez que este comportamento é fruto do acaso. Ainda, teremos que $E(x_t) = 0$, o que torna o padrão observado irrelevante.

Na tendência determinística, há uma função $T(\cdot)$ que determina seu comportamento. Nesse caso, é assumido que

$$y_t = T(t) + \varepsilon_t,$$

onde ε_t é uma série estacionária com média 0 e variância ν . Deste modo, $E(y_t) = T(t)$, o que implica que $T(\cdot)$ representa o comportamento médio da série. O problema de estimar $T(\cdot)$ é denominado suavização.

Na prática, é impossível determinar se uma tendência é aleatória ou determinística, cabendo ao estatístico procurar se há motivos para acreditar que está analisando o segundo tipo. A partir deste momento, toda tendência será considerada determinística.

16 O modelo de tendência polinomial

Considere que a série temporal foi observada até o tempo s . Então, a tendência é definida como uma função $T : (0, t] \rightarrow \mathbb{R}$. O Teorema de Weierstrass afirma que, se T é contínua, então para qualquer $\delta > 0$, existe um polinômio $u(\cdot)$ tal que

$$|T(t) - u(t)| < \delta.$$

Isto quer dizer que $T(\cdot)$ sempre pode ser aproximada por um polinômio. Assim, para determinada ordem p , é correto afirmar que

$$y_t = \beta_0 + \sum_{j=1}^p \beta_j t^j + \varepsilon_t \quad (4)$$

onde ε_t é uma série estacionária, é um modelo razoável para uma série temporal com tendência. Assumindo que ε_t é um ruído branco gaussiano, tem-se o modelo de tendência polinomial de grau p .

Fazendo $\mathbf{f}'_t = (1, t, \dots, t^p)$, o modelo de tendência polinomial é reescrito como

$$\mathbf{y} = \mathbf{F}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

e inferências sobre $\boldsymbol{\beta}$ e $\boldsymbol{\nu}$ são feitas utilizando o modelo linear tradicional.

Example .5. Considere o número anual de nascidos vivos no estado do Amazonas entre os anos 2000 e 20013:

```
x <- c( 67646 , 70252 , 70671 , 70751 , 71345 ,
       73488 , 75584 , 73469 , 75030 , 75729 ,
       74188 , 76202 , 77434 , 79041)

nascidos <- ts(x, start =2000)

ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
stats::acf(nascidos)
```

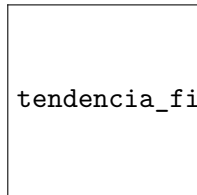


Figure 5: Número de nascimentos anual no estado do Amazonas (Fonte: SINASC/SUS)

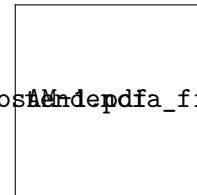


Figure 6: Correlograma da série (Fonte: SINASC/SUS).

Vamos ajustar um modelo de tendência polinomial de ordem 1, ou seja

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

onde $t = 1, \dots, 14$ representa os tempos 2000, \dots , 2013.

```
tempo <- 1:14
mod <- lm( nascidos ~ poly(tempo, 1, raw = TRUE))
```

As estimativas de máxima verossimilhança para β_0 e β_1 são:

```
mod$coefficients
```

```
(Intercept) poly(tempo, 1, raw = TRUE)
68266.879 715.178
```

ou seja,

$$\hat{T}(t) = \hat{\beta} + \hat{\beta}_1 t = 68.267 + 715t$$

Os resíduos do modelo linear `mod` podem ser obtidos via função `residuals`. Abaixo, verificamos que a série dos resíduos oscila em torno de zero e que nenhuma autocorrelação parece ser relevante, o que dão indícios de que os erros são um ruído branco.

```
res <- residuals(mod)
ts.plot( res, main = '')
acf(res, main = '')
```

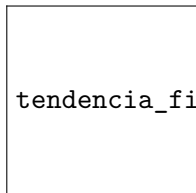


Figure 7: Série dos resíduos

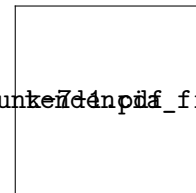


Figure 8: Correlograma dos resíduos

Abaixo, o teste de Shapiro-Wilks não gera evidências contra a suposição de normalidade e o teste de Box-Pierce não gera evidências contra a hipótese de ruído branco.

```
shapiro.test(res)
```

```
Shapiro-Wilk normality test
```

```
data:  res
W = 0.96982, p-value = 0.8743
```

```
Box.test(res)
```

Box-Pierce test

```
data: res  
X-squared = 0.02156, df = 1, p-value = 0.8833
```

É interessante notar que, para $t = 1, \dots, 14$,

$$\hat{T}(t) = \hat{\beta}_0 + \hat{\beta}_1 t = \hat{y}_t,$$

logo, os valores preditos do modelo são uma estimativa para a tendência nos pontos observados.

```
ts.plot( cbind( nascidos, fitted(mod)), col = 1:2, lwd = 2)
```

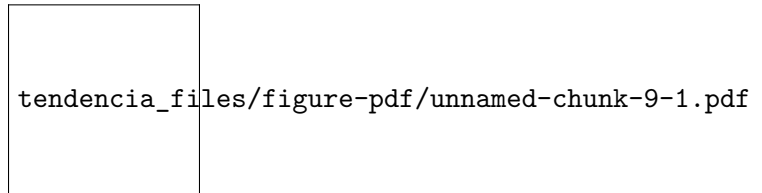


Figure 9: Linha preta: série original. Linha vermelha: tendência estimada

16.1 Previsão

A previsão é realizada utilizando o modelo ajustado, extrapolando para um tempo não observado. Por exemplo a estimativa para 2014 ($t = 15$) é

$$\hat{T}(15) = \hat{\beta} + \hat{\beta}_1 15 = 78.992$$

(o valor real foi 81.145).

É importante ressaltar que esse tipo de modelo é interessante para fazer inferências sobre a tendência, mas pode ser inadequado para previsões, uma vez que o polinômio é uma aproximação apenas para o intervalo observado.

16.2 Seleção de modelos lineares

O valor do Critério de Informação de Akaike (AIC) é dado por $-2L(\hat{\theta}) + 2k$ onde L é a função de verossimilhança e $\hat{\theta}$ e k são o estimador de máxima verossimilhança para θ e sua dimensão, respectivamente. O modelo com menor AIC é considerado mais adequado.

Considere o nível anual, em pés, do Lago Huron. Essa série já vem carregada no R sob o nome `LakeHuron`.

```
ts.plot(LakeHuron, lwd = 2, ylab = 'Nível (pés)')
stats::acf(LakeHuron)
```

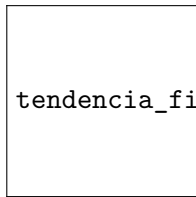


Figure 10: Nível anual do Lago Huron, entre 1875 e 1972

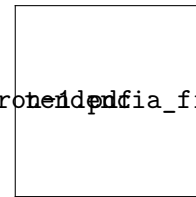


Figure 11: Correlograma da série

Vamos ajustar alguns modelos para tentar explica a tendêndia dessa série.

```
tempo <- 1 : length(LakeHuron)
mod1 <- lm( LakeHuron ~ poly(tempo, 1, raw = T))
mod2 <- lm( LakeHuron ~ poly(tempo, 2, raw = T))
mod3 <- lm( LakeHuron ~ poly(tempo, 3, raw = T))
mod4 <- lm( LakeHuron ~ poly(tempo, 4, raw = T))
mod5 <- lm( LakeHuron ~ poly(tempo, 5, raw = T))
mod6 <- lm( LakeHuron ~ poly(tempo, 6, raw = T))
```

```
AIC(mod1)
```

```
[1] 306.0957
```

```
AIC(mod2)
```

```
[1] 287.8407
```

```
AIC(mod3)
```

[1] 289.8391

```
AIC(mod4)
```

[1] 291.7127

```
AIC(mod5)
```

[1] 293.475

```
AIC(mod6)
```

[1] 291.7054

Entre os modelos ajustados, o de ordem 2 foi aquele com o menor valor do AIC. Sua tendência estimada é

O polinômio ajustado foi

$$\hat{T}(t) = 581 - 0,091t + 0,001t^2$$

Abaixo, apresentamos a análise de resíduos desse modelo.

```
res <- residuals(mod2)
ts.plot( res, main = '')
abline(h = 0, lty = 2)
acf(res, main = '')
```

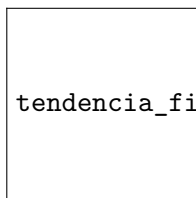


Figure 12: Série dos resíduos

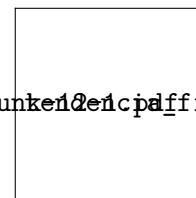


Figure 13: Correlograma dos resíduos

Os resíduos parecem oscilar em torno de zero com uma variância constante, mas o correlograma sugere que não temos um ruído branco. O teste de Box-Pierce, dado abaixo, confirma a nossa suspeita. Deste modo, este modelo não é adequado.

```
Box.test(res)
```

```
Box-Pierce test
```

```
data: res  
X-squared = 50.973, df = 1, p-value = 9.365e-13
```

17 Métodos não paramétricos para estimação da tendência

O modelo de tendência polinomial é robusto quando relaxamos a necessidade do ruído branco ser gaussiano. Nesse sentido, as estimativas ainda são válidas, mas perdemos todos os testes de hipóteses.

Os métodos não paramétricos independem da distribuição do ruído, sendo úteis para a análise exploratória.

17.1 Médias Móveis

O método das médias móveis consiste em obter $\hat{T}(t)$ através da média da série considerando os valores vizinhos à y_t . Para o tempo t e $m = 2h + 1$, com $h = 1, 2, \dots$, considere o conjunto $\mathcal{V}(m)_t = \{t - h, \dots, t + h\}$. Defini-se a média móvel de ordem m (notação m -MM) como

$$\hat{T}_h(t) = \frac{1}{m} \sum_{i=t-h}^{t+h} y_i, \quad h < t < n - h$$

.

Para compreender melhor esse estimador, considere que a relação entre pontos vizinhos é aproximadamente linear, ou seja, para qualquer $t \in \mathcal{V}(m)_t$ existem $a_{\mathcal{V}}$ e $b_{\mathcal{V}}$ tais que

$$y_t \approx a_{\mathcal{V}} + b_{\mathcal{V}}t + \varepsilon_t,$$

onde ε_t é considerado uma série temporal estacionária e ergódica. Então

$$\begin{aligned} E(\hat{T}(t)) &= \frac{a_{\mathcal{V}} + b_{\mathcal{V}}(t - h) + \dots + a_{\mathcal{V}} + b_{\mathcal{V}}t + \dots + a_{\mathcal{V}} + b_{\mathcal{V}}(t + h)}{m} \\ &= a_{\mathcal{V}} + b_{\mathcal{V}}t \end{aligned}$$

$$Var(\hat{T}(t)) = \frac{\nu}{m} + \frac{2}{m} \sum_{j=1}^{2h} j\gamma(j)$$

Observe que, como $T(\cdot)$ é determinística e os ruídos são estacionários e ergódicos, então $Var(\hat{T}(t))$ converge para zero quando $m \rightarrow \infty$. Contudo, $T(\cdot)$ é localmente linear, logo \hat{T} é um estimador razoável para valores baixos de h . Este é um exemplo típico de *trade off* entre vies e variância, onde não é possível minimizar os dois simultaneamente.

Utilizaremos a função `ma(x,m)`, do pacote `forecast` para encontrar m -MM para a série 'x

Example .6. Abaixo, apresentamos a série anual histórica de nascidos vivos no Amazonas desde 1994 até 2021.

```
x <- c(47780, 47966, 49112, 56070, 57180, 62037,
67646, 70252, 70671, 70751, 71345, 73488, 75584,
73469, 75030, 75729, 74188, 76202, 77434, 79041,
81145, 80097, 76703, 78066, 78087, 77622, 75635,
78454)

nascidos <- ts(x, start =1994)
ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
```

tendencia_files/figure-pdf/unnamed-chunk-14-1.pdf

```
require(forecast)
```

Carregando pacotes exigidos: forecast

Warning: package 'forecast' was built under R version 4.3.1

Registered S3 method overwritten by 'quantmod':

```
method      from
as.zoo.data.frame zoo
```

```
oo <- par( mfrow = c(2,2), mar = c(2,2,1,1))
ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
lines( ma(nascidos,3) , col =2, lwd = 2)
legend('bottomright', legend = c('Série original','3-MM'), fill = c(1,2,3), bty='n')
```



```

ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
lines( ma(nascidos,5) , col =2, lwd = 2)
legend('bottomright', legend = c('Série original','5-MM'), fill = c(1,2,3), bty='n')
ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
lines( ma(nascidos,7) , col =2, lwd = 2)
legend('bottomright', legend = c('Série original','7-MM'), fill = c(1,2,3), bty='n')
ts.plot(nascidos, lwd = 2, ylab = 'No. nascidos vivos')
lines( ma(nascidos,9) , col =2, lwd = 2)
legend('bottomright', legend = c('Série original','9-MM'), fill = c(1,2,3), bty='n')

```

tendencia_files/figure-pdf/unnamed-chunk-15-1.pdf

```
par(oo)
```

Considere a estimativa obtida pela média móvel de ordem 3.

```

tendencia <- ma(nascidos, 3)
tendencia

```

Time Series:

Start = 1994

End = 2021

Frequency = 1

```

[1]      NA 48286.00 51049.33 54120.67 58429.00 62287.67 66645.00 69523.00
[9] 70558.00 70922.33 71861.33 73472.33 74180.33 74694.33 74742.67 74982.33
[17] 75373.00 75941.33 77559.00 79206.67 80094.33 79315.00 78288.67 77618.67
[25] 77925.00 77114.67 77237.00      NA

```

Vamos estimar o ruído da série (e eliminar as coordenadas vazias)

```

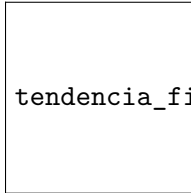
ruído <- nascidos - tendencia
ruído <- ruído[is.na(ruído) == F]

```

Abaixo, apresentamos as principais estatísticas sobre os resíduos. A série histórica dos resíduos oscila em torno de zero e não há motivos para suspeitar de que sua variância é constante. O correlograma apresenta autocorrelações baixas, como o esperado em um ruído branco. O teste

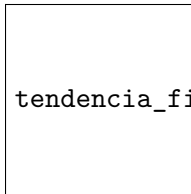
de Shapiro-Wilks não dá evidências contra normalidade, o que suporta a hipótese de ruído branco gaussiano. O teste de Box-Pierce apresenta um p-valor de 0,04 e, em conjunto com as demais evidências, vamos considerá-lo significativo ao nível de 4%.

```
ts.plot(ruido)
abline( h = 0, lty = 2)
abline( h = 2*sd(ruido), lty=2)
abline( h = -2*sd(ruido), lty=2)
```



tendencia_files/figure-pdf/unnamed-chunk-18-1.pdf

```
acf(ruido)
```



tendencia_files/figure-pdf/unnamed-chunk-18-2.pdf

```
shapiro.test(ruido)
```

Shapiro-Wilk normality test

data: ruido

W = 0.9711, p-value = 0.6521

```
Box.test(ruido)
```

Box-Pierce test

data: ruido

X-squared = 4.1804, df = 1, p-value = 0.04089

■

Até o momento, foi considerado que a ordem da média móvel é escrita como $m = 2h + 1$, ou seja, a ordem é sempre ímpar. Sem Para definir a média móvel para uma ordem par, perda de generalidade, assuma que $m = 4$. Como não é possível escolher um número igual de vizinhos à y_t , tem-se duas possibilidades para $\mathcal{V}(4)_t$:

$$\mathcal{V}'(4)_t = \{y_{t-1}, y_t, y_{t+1}, y_{t+2}\}$$

e

$$\mathcal{V}''(4)_t = \{y_{t-2}, y_{t-1}, y_t, y_{t+1}\}.$$

Para cada possibilidade, tem-se

$$\hat{T}'(t) = \frac{1}{4} \sum_{i=t-1}^{t+2} y_i$$

e

$$\hat{T}''(t) = \frac{1}{4} \sum_{i=t-2}^{t+1} y_i.$$

A média móvel 2-MM será definida por

$$\hat{T}(t) = \frac{T'(t) + T''(t)}{2}$$

Observe que a média agora é ponderada, uma vez que

$$\hat{T}(t) = \frac{y_{t-2}}{8} + \frac{y_{t-1}}{4} + \frac{y_t}{4} + \frac{y_{t+1}}{4} + \frac{y_{t+2}}{8}.$$

Para o caso de $m = 2h$ com $h = 1, 2, \dots$, defini-se m -MM por

$$\begin{aligned} \hat{T}(t) &= \frac{1}{2} \left[\frac{1}{m} \sum_{i=t-h}^{t+h-1} y_i + \frac{1}{m} \sum_{i=t-h+1}^{t+h} y_i \right] \\ &= \frac{y_{t-h}}{2m} + \frac{1}{m} \sum_{i=t-h+1}^{t+h-1} y_i + \frac{y_{t+h}}{2m}. \end{aligned}$$

Note que o argumento de $\hat{T}(t)$ é aproximadamente não viciado para m pequeno não se altera, uma vez que os pesos para os tempos $t - j$ e $t + j$ são simétricos

O estimador para tendência conhecido como média móvel ponderada de ordem m é dado por

$$\hat{T}(t) = \sum_{i=t-h}^{t+h} w_i y_i,$$

com $w_i > 0$, $w_{t-j} = w_{t+j}$ e $\sum_{i=t-h}^{t+h} w_i = 1$. O método tradicional é obtido fazendo $w_i = 1/m$.

Por último, como as primeiras e últimas observações são removidas, esses métodos não são os mais recomendados.

Outras médias móveis

É importante ressaltar que os métodos estatísticos são ferramentas universais e que geralmente sofrem modificações ao serem aplicados em outras áreas. Deste modo, existem outras definições de médias móveis que podem causar confusão.

Em epidemiologia por exemplo, a média móvel de ordem m é definida por

$$\hat{T}(t) = \frac{1}{m} \sum_{j=1}^m y_{t-j+1}$$

ou seja, a soma dos valores mais recentes em relação à t . Observe que o contexto é diferente: em uma epidemia por exemplo, deseja-se estimar $T(t)$ onde t é o tempo mais recente e, em geral, se utiliza o 7-MM tirando a média dos últimos 7 dias. A mesma lógica vale para o mercado financeiro, que tira a média dos últimos 5 dias de preço de fechamento.

Ainda no mercado financeiro, o importante é captar a mudança da tendência o mais rápido possível. Deste modo, utiliza-se uma média móvel ponderada definida por

$$\hat{T}(t) = \frac{2}{m(m+1)} \sum_{j=1}^m (m-j+1)y_{t-j+1}.$$

Na definição acima, o último preço da ação, dado por y_t , é o valor mais importante e por isso recebe o maior peso. Note que os pesos não são simétricos.

18 Suavização do gráfico de dispersão estimada localmente (loess)

No método de suavização do gráfico de dispersão, deseja-se estimar $f(x) = E(y|x)$, através da coleção $(y_1, x_1), \dots, (y_n, x_n)$, para um valor qualquer de x .

A estimativa \hat{f} para o ponto x' é calculada considerando os seguintes passos:

1. Fixe um valor inteiro positivo $q \leq n$.
2. Dentro do conjunto x_1, \dots, x_n , encontre os q valores mais próximos de x' (via distância euclidiana). Denote este conjunto por \mathcal{V} e denote por d a maior distância encontrada.

3. Para cada x_1, \dots, x_n seja

$$v_j(x') = \begin{cases} \left(1 - \left|\frac{x_j - x'}{d}\right|\right)^3, & \text{se } |x_j - x| \leq d \\ 0, & \text{caso contrário} \end{cases}$$

o peso associado à x_j (valores próximos de x' receberão o peso máximo e valores afastados recebem menor peso)

4. Ajuste o modelo de regressão ponderado, minimizando

$$\sum_{i=1}^n v_i(x') \left(y_i - \sum_{j=0}^p \beta_j x^j \right)^2$$

5. Estime $f(x')$ por

$$\hat{f}(x') = \sum_{j=0}^p \hat{\beta}_j x^j$$

Oberve que este método pode ser utilizado para estimar a tendência da série. Abaixo, vamos analisar a série de taxa de desemprego mensal, entre março de 2002 e dezembro de 2015.

```
url <- 'https://www.dropbox.com/s/rmgymzsic99qawd/desemprego.csv?dl=1'

banco <- read.csv(url, sep = ';', h = F)

desemprego <- ts( banco$V2, start = c(2002,3), frequency=12)

ts.plot(desemprego, ylab = 'Taxa de desemprego')
```

tendencia_files/figure-pdf/unnamed-chunk-19-1.pdf

```
acf(desemprego, lag = 30)
```

tendencia_files/figure-pdf/unnamed-chunk-19-2.pdf

Vamos estimar a tendência

```
# criando a variável regressora
tempo <- 1 : length(desemprego)

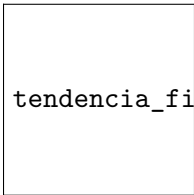
# aplicando o loess
lw <- loess( desemprego ~ tempo)

# transformando o valor predito em uma série temporal

fit <- ts(lw$fitted, start = start(desemprego), frequency = frequency(desemprego) )

# gráfico da tendência estimada

ts.plot( desemprego, ylab = 'Taxa de desemprego' , lwd = 2)
lines(fit, lwd = 2, col = 'tomato')
legend('topright', c('Observado','Ajustado'),fill=c(1,'tomato'), bty='n')
```

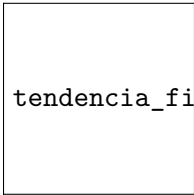


tendencia_files/figure-pdf/unnamed-chunk-20-1.pdf

Vamos eliminar a tendência estimada e avaliar o restante.

```
yt <- desemprego - fit

ts.plot(yt)
```



tendencia_files/figure-pdf/unnamed-chunk-21-1.pdf

```
acf(yt)
```

tendencia_files/figure-pdf/unnamed-chunk-21-2.pdf

Fica claro o comportamento sazonal, o que implica que o restante não é uma série estacionária.

Sazonalidade

19 Padrão sazonal

Padrões que surgem sistematicamente ao longo do tempo são denominados sazonais. Exemplos: flutuações de temperatura entre estações, início e fim do semestre letivo, Natal, dias úteis, feriados flutuantes como a Páscoa e o Carnaval.

Um padrão sazonal pode ser modelado através de uma função periódica.

Definition .4. Dizemos que $s(\cdot)$ é uma função periódica de período p se

$$s(t) = s(t + kp), \forall k = 1, 2, \dots$$

Example .7. A função

$$s(t) = \begin{cases} -1, & t = 1, 4, 7, \dots \\ 0, & t = 2, 5, 8, \dots \\ 1, & t = 3, 6, 9, \dots \end{cases}$$

é periódica e possui período $p = 3$.

Example .8. A função

$$s(t) = \cos\left(2\pi\frac{t}{4}\right)$$

é periódica (com $p = 4$). De fato, para $k = 1, 2, 3, \dots$,

$$\begin{aligned} g(t + 4k) &= \cos\left(2\pi\frac{t + 4k}{4}\right) = \cos\left(\pi\frac{t}{4} + 2\pi k\right) \\ &= \cos\left(2\pi\frac{t}{4}\right) \underbrace{\cos(2\pi k)}_1 - \sin\left(\pi\frac{t}{4}\right) \underbrace{\sin(2\pi k)}_0 \\ &= \cos\left(2\pi\frac{t}{4}\right) = g(t). \end{aligned}$$

Seja $s(\cdot)$ uma função periódica. Então, uma série temporal sazonal aditiva é descrita como

$$x_t = s(t) + \varepsilon_t,$$

onde ε_t é um ruído estacionário.

Atenção. Ao criar um objeto do tipo `ts` no `R`, o argumento `frequency` é considerado o período do padrão sazonal. A maioria das funções voltadas para padrões sazonais acessam essa informação no objeto.

20 Gráficos sazonal de subséries

Seja x_t uma série sazonal de período p . Para construir um gráfico sazonal de subséries:

- Faça p subséries:

$$x_1, x_{1+p}, x_{1+2p}, \dots$$

$$x_2, x_{2+p}, x_{2+2p}, \dots$$

$$\dots$$

$$x_p, x_{2p}, x_{3p}, \dots$$

- Calcule a média de cada subsérie.
- Faça um gráfico de cada subsérie, cada um com uma linha horizontal com o valor de sua respectiva média.

Observe que, para um padrão sazonal, a subsérie

$$x_i, x_{i+p}, x_{i+2p}, \dots$$

é equivalente à

$$s(i) + \varepsilon_i, s(i) + \varepsilon_{i+p}, s(i) + \varepsilon_{i+2p}, \dots$$

e, supondo que ε_t é um ruído branco ergódico, teremos que

$$\hat{s}(i) = \frac{1}{m+1} \sum_{j=0}^m x_{i+jp}$$

onde $m+1$ é o tamanho da amostra correspondente à subsérie.

No `R`, o comando `monthplot(x)` faz o gráfico de subséries. Considerando um padrão sazonal, cada subsérie deve oscilar em torno de uma média constante. Vejamos alguns exemplos.

Example .9. A série abaixo (gráfico à esquerda) apresenta as temperaturas médias mensais, em Fahrenheits, no Castelo de Nottingham entre 1920 e 1939. O gráfico à direita apresenta o gráfico de subséries. Como a série possui período 12, este último gráfico já identifica cada subsérie com a inicial do mês correspondente. Note que a maioria das subséries oscila em torno da média

```
ts.plot(nottem)
monthplot(nottem)
```

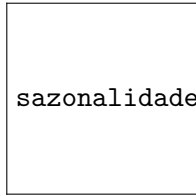


Figure 14: Gráfico da série de temperaturas no Castelo de Nottingham

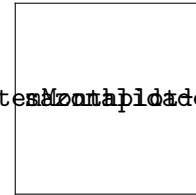


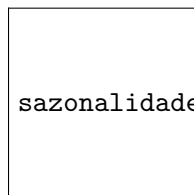
Figure 15: Gráfico de subséries

■

É possível que a tendência construa uma falsa impressão sobre o comportamento sazonal. Portanto, é interessante remover a tendência da série antes de fazer o gráfico das subséries.


Example .10. Abaixo apresentamos a série mensal da concentração de CO₂, em partes por milhão, no Mauna Loa. Observe a clara tendência de crescimento e o padrão sazonal.

```
ts.plot(co2)
```



O gráfico das subséries é dado abaixo. Observe que há uma tendência crescente em cada subsérie.

```
monthplot(co2)
```




sazonalidade_files/figure-pdf/unnamed-chunk-3-1.pdf

A tendência crescente nas subséries não nos permitia trabalhar com uma função periódica. Contudo o gráfico está sendo influenciado pela tendência. Vamos removê-la e fazer o gráfico novamente.


```
# estimação da tendência via loess
tempo <- 1 : length(co2)
model <- loess( co2 ~tempo)
tend <- fitted(model)

# série sem tendência e o gráfico de subséries
plot( co2-tend)
```



sazonalidade_files/figure-pdf/unnamed-chunk-4-1.pdf

```
monthplot(co2-tend)
```



sazonalidade_files/figure-pdf/unnamed-chunk-4-2.pdf

■

Também é possível que o padrão sazonal não seja bem representado por uma função periódica. Isto ocorre quando o gráfico de subséries possui tendência.

Example .11. A `fig-AirPassengersMonthplot` apresenta o gráfico da série `AirPassengers` e de suas subséries. Observe que todas as subséries possuem um padrão de crescimento, que deve estar sendo governado pela tendência crescente da série.

```
ts.plot(AirPassengers)
monthplot(AirPassengers)
```

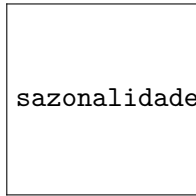


Figure 16: Gráfico da série `AirPassengers`

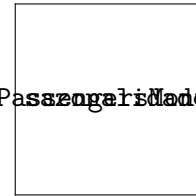
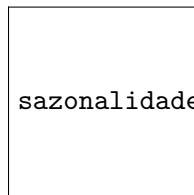


Figure 17: Gráfico de subséries da série `AirPassengers`. Observe as tendências monotonicamente crescente em cada subgráfico

Abaixo, vamos estimar a tendência da série e removê-la. Note que o padrão das subséries se altera, mostrando o real comportamento do padrão sazonal: é uma tendência de crescimento nos meses de verão americano (junho, julho e agosto) e um padrão de queda começando em novembro e pegando so três subsequentes meses do inverno.

```
# estimação da tendência via loess
tempo <- 1 : length(AirPassengers)
model <- loess( AirPassengers ~tempo)
tend <- fitted(model)

# série sem tendência e o gráfico de subséries
plot( AirPassengers-tend)
```



```
monthplot(AirPassengers-tend)
```



21 Dessazonalização

Existem séries que apresentam tanto tendência quanto sazonalidade. Isso pode dificultar o estudo da tendência, uma vez que a sazonalidade interfere no movimento de subida e descida da série. Nesses casos, é interessante obter uma estimativa da parte sazonal para removê-la da série.

Considere a série

$$y_t = T^*(t) + s^*(t) + \varepsilon_t,$$

onde $T^*(t)$ é a tendência e $s^*(t)$ uma função periódica, com período igual a p . É sempre verdade que existe uma constante real c tal que

$$\sum_{j=1}^p s^*(t+j) = c$$

para qualquer $t = 0, 1, \dots$. Contudo, note que c é uma constante (que, por sua vez, faz parte da tendência da série). Fazendo $T(t) = T^*(t) + c$ e $s(t) = s^*(t) - c$, teremos

$$y_t = T^*(t) + c + s^*(t) - c + \varepsilon_t = T(t) + s(t) + \varepsilon_t,$$

e sempre podemos assumir que

$$\sum_{t=1}^p s(t) = 0.$$

Isto é equivalente a afirmar que o efeito sazonal desaparece quando as p observações são agregadas. Deste modo, o estimador para tendência p -MM estima a tendência sem o efeito sazonal.

Considere novamente a série com o número de óbitos por doenças pulmonares no Reino Unido. A figura abaixo mostra a série original e a tendência, estimada pela 12-MM. Observe o padrão claro de tendência decrescente.

22 Decomposições de séries

Em geral, assume-se que a série temporal admite a seguinte decomposição:

$$x_t = T(t) + s(t) + \varepsilon_t$$

A estimação não paramétrica dos componentes da decomposição constitui-se de ferramenta exploratória essencial para a análise de séries temporais. A decomposição clássica é realizada através dos seguintes passos:

- Estime a tendência da série, utilizando a p -MM, obtendo \hat{T}
 - Remova a tendência da série: $\tilde{x}_t = x_t - \hat{T}$
 - Encontre as p subséries de \tilde{x} e estime os valores de s através de suas médias.
 - Estime os resíduos: $\hat{\varepsilon}_t = x_t - \hat{T}(t) - \hat{s}(t)$
- A Figure ?? apresenta a decomposição clássica da série de óbitos por doenças pulmonares no Reino Unido.

```
plot(decompose(ldeaths))
```

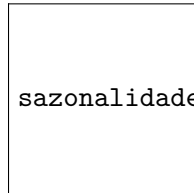
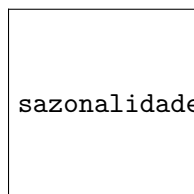


Figure 18: Decomposição da série de óbitos por doenças pulmonares no Reino Unido

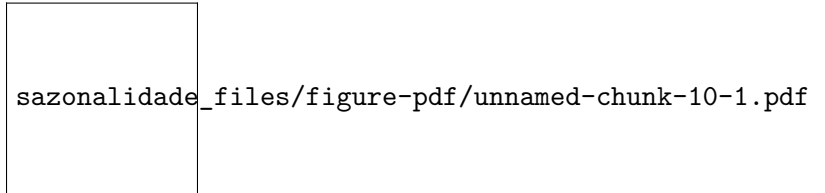
Como é utilizada uma média móvel, a estimativa da tendência e do ruído é prejudicada no início e fim da série. Outro problema que este método não consegue estimar padrões sazonais que não sejam representados por uma função periódica. Para ilustrar, apresentamos abaixo a decomposição clássica da série **AirPassengers**. Como a estimação da parte sazonal utilizou a média das subséries, as tendências sazonais vistas no Example ?? foram para os resíduos (identificados co)

```
plot(decompose(AirPassengers))
```



Uma solução mais robusta é conhecida como STL, que realiza uma série de estimativas de tendência, tanto para a geral quanto para a sazonal, utilizando o loess. Os detalhes podem ser vistos no paper original [STL](#). Abaixo, apresentamos o STL para a série `AirPassengers`. Observe que a sazonalidade foi melhor estimada, embora ainda exista um padrão sazonal nos resíduos.

```
# além da série, devemos colocar o período desejado na função
plot(stl(AirPassengers, 12))
```



23 Modelo de Forma livre - ou fatores sazonais

Considere uma série temporal com sazonalidade dada por uma função periódica $s(\cdot)$ de período p . Sejam

$$\beta_j = s(t + jp).$$

Os parâmetros β_1, \dots, β_p são denominados fatores sazonais. Note que é necessário colocar a restrição

$$\beta_p = -\beta_1 - \dots - \beta_{p-1},$$

pois

$$\sum_{j=1}^p \beta_j = 0$$

(ou seja, existem na prática $p - 1$ fatores sazonais para serem estimados).

Example .12. Seja x_t uma série sazonal de período $p = 4$ (por, exemplo, em dados trimestrais). Então

$$\begin{aligned} E(x_t) &= s(t) = \beta_1, t = 1, 5, 9, \dots, \\ E(x_t) &= s(t) = \beta_2, t = 2, 6, 10, \dots, \\ E(x_t) &= s(t) = \beta_3, t = 3, 7, 11, \dots, \\ E(x_t) &= s(t) = -\beta_1 - \beta_2 - \beta_3, t = 4, 8, 12, \dots, \end{aligned}$$

Seja $\mathbf{E}_{j,m-1}$ o vetor coluna de comprimento $m - 1$ que possui a j -ésima entrada igual a 1 e as demais iguais a zero. Por exemplo

$$\mathbf{E}_{2,3} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Para um período p e para $j = 1, \dots, p - 1$ faça

$$\mathbf{f}_{j+kp} = \mathbf{E}_{j,p-1}$$

com $k = 0, 1, 2, \dots$. Para $s = 1, 2, \dots$

$$\mathbf{f}_{sp} = -\mathbf{1}_{p-1} \quad (5)$$

Então,

$$x_t = \mathbf{f}'_t \boldsymbol{\beta} + \varepsilon_t,$$

onde $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_{p-1})$ representa os efeitos sazonais. Disto, teremos

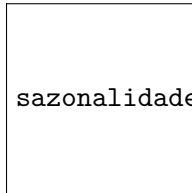
$$\mathbf{x} = \mathbf{F}'_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde a matriz \mathbf{F}_n é formada pelas colunas $\mathbf{f}_1, \dots, \mathbf{f}_p$.

Portanto, a estimação de $\boldsymbol{\beta}$ é feita através da teoria de modelos lineares.

Example .13. Considere novamente a série `nottem`, representada na figura abaixo.

```
ts.plot(nottem)
abline(h = mean(nottem))
```



sazonalidade_files/figure-pdf/unnamed-chunk-11-1.pdf

Observe que a série pode ser vista como uma função periódica com um nível constante (conhecido em regressão como intercepto):

$$x_t = \mu + s(t) + \varepsilon_t$$

Considerando que os erros são ergódicos, podemos estimar a média μ pela média amostral. Deste modo, a série sem tendência é:

$$\tilde{x}_t = x_t - \bar{x} = s(t) + \varepsilon_t$$

Vamos ajustar um modelo linear para \tilde{x} . Para construir a matriz de regressão \mathbf{F}'_n vamos utilizar a função `cycle` que identifica o índice do padrão sazonal associado a cada observação.

```
xtil <- nottem - mean(nottem)
mes <- cycle(nottem)
mes <- as.factor(mes)
mod <- lm( xtil ~ mes - 1)
mod
```

Call:

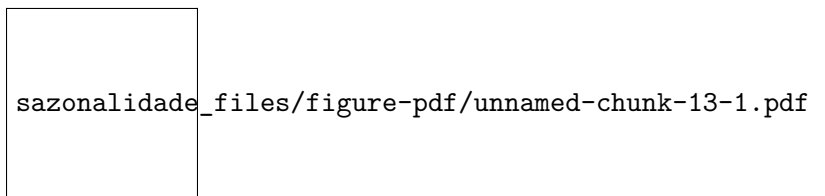
```
lm(formula = xtil ~ mes - 1)
```

Coefficients:

| mes1 | mes2 | mes3 | mes4 | mes5 | mes6 | mes7 | mes8 |
|---------|---------|---------|---------|--------|--------|---------|---------|
| -9.3446 | -9.8496 | -6.8446 | -2.7496 | 3.5204 | 9.0004 | 12.8604 | 11.4804 |
| mes9 | mes10 | mes11 | mes12 | | | | |
| 7.4404 | 0.4554 | -6.4596 | -9.5096 | | | | |

Nos resultados acima podemos notar que os meses de Novembro até Abril possuem temperaturas menores que a média. O gráfico dos efeitos sazonais é dado a seguir, mostrando que a passagem dos meses possui um padrão de onda, como um cosseno.

```
plot.new()
plot.window(xlim = c(0,13), ylim= c(-11,11))
points(coefficients(mod), pch = 16)
axis(1, at = 1:12, labels = c('J','F','M','A','M','J','J','A','S','O','N','D'))
axis(2)
title(ylab = 'Valor do fator sazonal',xlab='Fator sazonal')
```



Abaixo analisamos os resíduos do modelo. Os resíduos parecem flutuar e torno de zero com variância constante. O correlograma apresenta algumas leves autocorrelações nas defasagens 1 e 2. O teste de Shapiro-Wilks não rejeita a normalidade, mas o teste de Box-Pierce rejeita a hipótese de ruído branco. Voltaremos a este problema posteriormente.

```
res <- residuals(mod)
ts.plot(res)
acf(res)
```

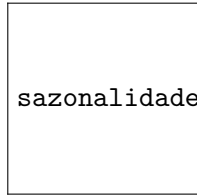


Figure 19: Gráfico dos resíduos

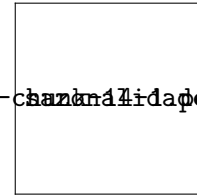


Figure 20: Correlograma dos resíduos

```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data:  res
W = 0.99125, p-value = 0.1612
```

```
Box.test(res)
```

Box-Pierce test

```
data:  res
X-squared = 13.109, df = 1, p-value = 0.0002939
```

■

24 Regressão harmônica simples

24.1 A função harmônica

A função

$$s(t) = A \cos(2\pi\omega t_i + \phi), \quad (6)$$

é denominada harmônico, onde A e $\phi \in (0, 2\pi)$ são denominados amplitude e fase. Essa função é periódica, com período igual a $p = 1/\omega$ onde ω é denominado frequência (angular).

```
curve( cos( 2*pi/12 *x), 0,24, lwd = 2, ylab = '')
curve( .5*cos( 2*pi/12 *x), add = T, lty = 2,lwd = 2)
curve( cos( 2*pi/12 *x+2), 0,24, lwd = 2, ylab = '')
curve( cos( 2*pi/12 *x + 4), add = T, lty = 2,lwd = 2)
curve( cos( 2*pi/12 *x), 0,24, lwd = 2, ylab = '')
curve( cos( 2*pi/6 *x), add = T, lty = 2,lwd = 2)
```

Cuidado. Período é o tempo necessário para que o padrão sazonal se repita, dando uma volta completa. Já a frequência é o número de vezes que o padrão se repete por unidade de tempo. Por exemplo, para $p = 12$ (meses), $\omega = 1/12$ implica que cada unidade do tempo (mês) representa um doze ávos de uma volta.

Entretanto, quando criamos um objeto do tipo **ts** alimentamos o argumento **frequency** com o valor do período. A aparente confusão ocorre porque o argumento **frequency** se refere ao número de observações (ou seja, frequência) por unidade de tempo.

Também é possível escrever um objeto **ts** atribuindo o valor ω através do argumento **deltat**. Por exemplo, para um período de 12 meses, usamos **deltat=1/12**.

Pela definição, como $\omega = 1/p$, teremos que $\omega \in (0, 1)$. Assuma que $a = \omega + 1/2$, onde $\omega \in (0, 1/2)$. Então, podemos notar que

$$\begin{aligned} A \cos(2\pi\omega t + \phi) &= A \cos\left(2\pi\left(a - \frac{1}{2}\right)t + \phi\right) \\ &= A \cos(2\pi at + \phi) \cos(\pi t) \\ &= A^* \cos(2\pi at + \phi). \end{aligned}$$

onde $A^* = A \cos(\pi t)$. Na prática, é impossível distinguir entre (A, ω) e (A^*, a) . Para permitir a existência da identificabilidade, vamos assumir que $A > 0$ e $\omega \in (0, 1/2)$. Isto implica que o menor período detectável é $p = 2$.

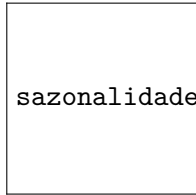


Figure 21: A amplitude representa o valor máximo e mínimo do harmônico. No gráfico estão representados harmônicos de período 12 com diferentes amplitudes. Linha cheia: $A=1$. Linha tracejada: $A=0,5$

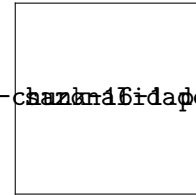


Figure 22: A fase tem papel de locação, deslocando da onda mas mantendo sua forma. No gráfico estão representados harmônicos de período 12 com diferentes fases. Linha cheia: $\phi = 2$. Linha tracejada: $\phi = 4$

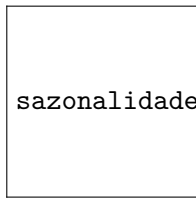


Figure 23: A frequência representa o número de voltas por unidade de tempo. No gráfico estão representados harmônicos com frequências diferentes. Linha cheia: $\omega = 1/12$ (padrão anual). Linha tracejada: $\omega = 1/6$ (padrão semestral)

24.2 O modelo de regressão harmônica simples

O modelo de regressão harmônica simples é dado por

$$y_t = A \cos(2\pi\omega t_i + \phi) + \varepsilon_t,$$

onde ε_t é um ruído branco. Como ω é considerado conhecido, este modelo pode ser linearizado:

$$A \cos(2\pi\omega t_i + \phi) = A [\cos(2\pi\omega t_i) \cos(\phi) - \sin(2\pi\omega t_i) \sin(\phi)]$$

e, fazendo $\beta_1 = A \cos(\phi)$ e $\beta_2 = -A \sin(\phi)$, teremos que

$$y_t = \beta_1 \cos(2\pi\omega t_i) + \beta_2 \sin(2\pi\omega) + \varepsilon_t$$

Note que sempre é possível recuperar os parâmetros originais:

$$\begin{cases} \beta_1 = A \cos(\phi) \\ \beta_2 = -A \sin(\phi) \end{cases} \Rightarrow \begin{cases} \beta_1^2 = A^2 \cos^2(\phi) \\ \beta_2^2 = A^2 \sin^2(\phi) \end{cases} \Rightarrow \begin{cases} A = \sqrt{\beta_1^2 + \beta_2^2} \\ \phi = \cos^{-1}\left(\frac{\beta_1}{A}\right) \end{cases}$$

Fazendo

$$\boldsymbol{\beta}' = (\beta_1, \beta_2)$$

$$\mathbf{f}'_t = (\cos(\omega t_i), \sin(\omega t_i))$$

teremos

$$\mathbf{f}'_t \boldsymbol{\beta} = A \cos(\omega t_i + \phi).$$

Portanto, podemos escrever:

$$y_t = \mathbf{f}'_t \boldsymbol{\beta} + \varepsilon_t \tag{7}$$

Considerando $\varepsilon_t \sim \text{Normal}(0, \nu)$ teremos

$$\mathbf{y} | \boldsymbol{\beta}, \nu \sim \text{Normal}(\mathbf{F}'_n \boldsymbol{\beta}, \nu \mathbf{I}_T).$$

Example .14. Considere novamente a série `nottem`. A função `harmonic` do pacote `TSA` constrói a matriz necessária para compor o modelo de regressão harmônica simples.

```
require(TSA)
```

Carregando pacotes exigidos: TSA

Warning: package 'TSA' was built under R version 4.3.2

Attaching package: 'TSA'

The following objects are masked from 'package:stats':

acf, arima

The following object is masked from 'package:utils':

tar

```
har <- harmonic(nottem)
mod <- lm( nottem ~ har)
mod
```

Call:

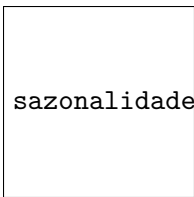
lm(formula = nottem ~ har)

Coefficients:

| (Intercept) | harcos(2*pi*t) | harsin(2*pi*t) |
|-------------|----------------|----------------|
| 49.040 | -11.473 | -1.391 |

Note que não removemos a média da série, logo a mesma está sendo estimada pelo intercepto μ . O gráfico do harmônico gerado é mostrado abaixo.

```
param <- coefficients(mod)
curve( param[1]+ param[2]*cos(2*pi*x/12) + param[3]*sin(2*pi*x/12),1,12, ylab = 'Harmônico
```



sazonalidade_files/figure-pdf/unnamed-chunk-18-1.pdf

Em relação aos resíduos do modelo, temos

```
res <- residuals(mod)
ts.plot(res)
```

```
stats::acf(res)
```

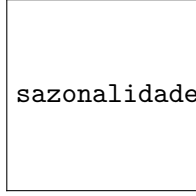


Figure 24: Resíduos da regressão harmônica simples

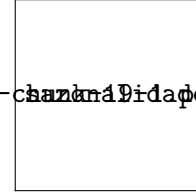


Figure 25: Correlograma dos resíduos

No gráfico dos resíduos podemos ver uma série oscilando em torno de zero com o que parecer ser uma variância constante. Já no correlograma podemos notar autocorrelações significativas na defasagem um em algumas outras, o que elimina a hipótese de ruído branco (o teste de Box-Pierce tem p-valor $< 10^{-5}$, confirmando a nossa suspeita). ■

25 Periodograma

Considere o modelo de regressão harmônica simples com frequência desconhecida. Então, considerando os parâmetros $\theta = \{\beta_1, \beta_2, \nu, \omega\}$ e um ruído branco gaussiano, sabemos que a função de verossimilhança pode ser reescrita como

$$L(\theta) \propto \left(\frac{1}{\nu}\right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\nu} \left[(\beta - \hat{\beta}(\omega))' \mathbf{F} \mathbf{F}' (\beta - \hat{\beta}(\omega)) + R(\omega) \right] \right\}$$

obs: como \mathbf{F} depende de ω , tanto $\hat{\beta}$ quanto R dependem de ω e, por isso, estão escritos como função da frequência.

Como

$$\begin{aligned} \mathbf{F}_n \mathbf{F}_n' &= \begin{pmatrix} \sum_{t=1}^n \cos(2\pi\omega t)^2 & \sum_{t=1}^n \cos(2\pi\omega t) \sin(2\pi\omega t) \\ \sum_{t=1}^n \cos(2\pi\omega t) \sin(2\pi\omega t) & \sum_{t=1}^n \sin(2\pi\omega t)^2 \end{pmatrix} \\ &= \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix} = \frac{n}{2} \mathbf{I}_2 \end{aligned}$$

e

$$\begin{aligned} R(\omega) &= (\mathbf{y} - \mathbf{F}_n' \hat{\beta}(\omega))' (\mathbf{y} - \mathbf{F}_n' \hat{\beta}(\omega)) = \mathbf{y}' \mathbf{y} - \hat{\beta}(\omega)' (\mathbf{F}_n \mathbf{F}_n') \hat{\beta}(\omega) \\ &= \mathbf{y}' \mathbf{y} - \frac{n}{2} \hat{\beta}(\omega)' \hat{\beta}(\omega) \end{aligned}$$

Considere a função de verossimilhança perfilada de ω :

$$L(\omega | \hat{\beta}(\omega)) \propto \exp \left\{ \frac{1}{2\nu} R(\omega) \right\} \propto \exp \left\{ \frac{n}{4\nu} \hat{\beta}(\omega)' \hat{\beta}(\omega) \right\}$$

Faça $I(\omega) = \frac{n}{2} \hat{\beta}(\omega)' \hat{\beta}(\omega)$. Então,

$$L(\omega | \hat{\beta}(\omega)) \propto \exp \left\{ \frac{1}{2\nu} I(\omega) \right\}$$

É fácil notar que, quanto maior for o valor de $I(\omega)$, maior será a função de verossimilhança perfilada.

Definition .5. O gráfico $(\omega, I(\omega))$ é denominado periodograma e serve para nos auxiliar a encontrar as frequências mais importantes da série. $I(\omega)$ também é conhecido como espectro (ou densidade espectral).

Como

$$\hat{\beta}' \hat{\beta} = \hat{\beta}_1^2 + \hat{\beta}_2^2,$$

logo,

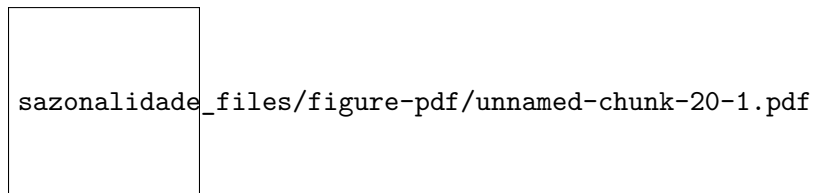
$$I(\omega) = \frac{n}{2} [\hat{\beta}_1^2 + \hat{\beta}_2^2] = \frac{n}{2} \hat{A}^2$$

No periodograma, restringimos a busca nos valores $\omega_k = k/n$, com $1 \leq k < n/2$. Se tivermos um pico na frequência ω_k , então podemos estimar o período como sendo:

$$p = \frac{1}{\omega_k} = \frac{n}{k}.$$

Utilizamos a função `periodogram` do pacote `TSA` para fazer o periodograma. Abaixo, apresentamos o periodograma da série `nottem`.

```
p <- periodogram(nottem)
```



Note que o periodograma possui um pico dominante. Abaixo, mostramos que este pico é o período 12.

```
# encontrando em qual coordenada ocorre o maior valor do periodograma
k <- which(p$spec == max(p$spec))

# identificando a frequência correspondente
```

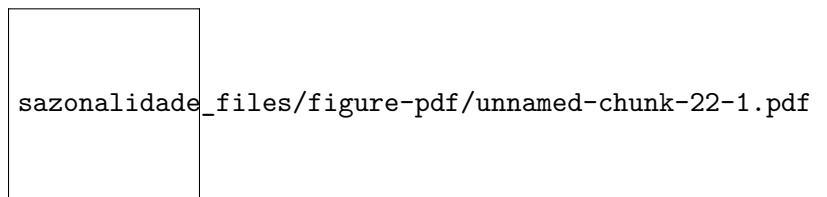
```
omega <- p$freq[k]

# identificando o período da série
1/omega
```

```
[1] 12
```

No Example ?? aplicamos uma regressão harmônica simples e descobrimos que o ruído resultante não era branco e gaussiano. Abaixo, mostramos o periodograma dos resíduos obtidos nesse exemplo.

```
p2 <- periodogram(res)
```



Observe que uma nova frequência dominante surge. Vamos identificá-la.

```
# encontrando em qual coordenada ocorre o maior valor do periodograma
k <- which(p2$spec == max(p2$spec))

# identificando a frequência correspondente
omega <- p2$freq[k]

# identificando o período da série
1/omega
```

```
[1] 6
```

Portanto, ainda existe um padrão sazonal, de período 6, para ser explicado nessa série.

26 Regressão Harmônica

Na série `nottem`, vimos que o periodograma apontou que o período deveria ser 12. Este primeiro período encontrado será denominado ‘fundamental’. Após ajustarmos uma regressão harmônica simples, com $p = 12$, os resíduos mostraram outra componente sazonal com período $p' = 6$.

Na verdade, é comum termos períodos que são frações do período fundamental. Após identificar a frequência (ou período) fundamental, teremos as seguintes frequências para explorar:

$$\omega_k = \frac{k}{p},$$

com $1 \leq k < \lfloor p/2 \rfloor$, onde $\lfloor a \rfloor$ é a parte inteira de a .

O harmônico gerado por ω_k , dado por

$$A_k \cos(2\pi\omega_k t + \phi_k) \quad (8)$$

é denominado harmônico de ordem k .

- O harmônico de ordem 1 completa um ciclo em p unidades de tempo.
- O harmônico de ordem 2 completa um ciclo em $p/2$ unidades de tempo.
- O harmônico de ordem k completa um ciclo em p/k unidades de tempo.

Podemos lidar com os diversos harmônicos de modo bastante direto:

$$y_t = \sum_{k=1}^m A_k \cos(\omega_k t + \phi_k) + \varepsilon_t \quad (9)$$

ou ainda, utilizando os resultados já discutidos,

$$\begin{aligned} y_t &= \sum_{k=1}^m A_k [\cos(\omega_k t) \cos(\phi_k) - \sin(\omega_k t) \sin(\phi_k)] + \varepsilon_t \\ &= \sum_{k=1}^m [\beta_{k,1} \cos(\omega_k t) + \beta_{k,2} \sin(\omega_k t)] + \varepsilon_t, \end{aligned}$$

onde

$$m = \begin{cases} \lfloor p/2 \rfloor, & \text{se } p \text{ é ímpar} \\ p/2 - 1, & \text{se } p \text{ é par} \end{cases}$$

Fazendo

$$\mathbf{f}'_t = (\cos(w_1 t), \sin(w_1 t), \dots, \cos(w_m t), \sin(w_m t)) \quad (10)$$

e

$$\boldsymbol{\beta}' = (\beta_{1,1}, \beta_{1,2}, \dots, \beta_{m,1}, \beta_{m,2}) \quad (11)$$

teremos

$$y_t = \mathbf{f}_t' \boldsymbol{\beta} + \varepsilon_t,$$

e o tradicional modelo linear:

$$\mathbf{y} = \mathbf{F}_t' \boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

:::{#exm-nottemRegressaoHarmonica} Já vimos que a função `harm(x)` gera a matriz necessária para ajusta a regressão harmônica simples para a série temporal x . Ela possui um argumento, denotado por `m` e cujo *default* é um, que gera uma regressão harmônica com `m` harmônicos.

Abaixo, mostramos os resultados para a regressão harmônica simples e com os dois primeiros harmônicos para a série `nottem`. Note que o AIC deste último sugere que os dois primeiros harmônicos deve ser utilizados

```
simples <- lm( nottem ~harmonic(nottem) )
doisHarm <- lm( nottem ~harmonic(nottem, 2))

AIC(simples)
```

```
[1] 1134.345
```

```
AIC(doisHarm)
```

```
[1] 1091.866
```

Os resíduos do modelo com dois harmônicos são explorados abaixo. A série dos resíduos parece flutuar em torno de zero com variância constante, enquanto que o correlograma exibe uma pequena autocorrelação na defasagem um. O teste de Shapiro-Wilks aceita a normalidade, mas o teste de Ljung-Box rejeita a hipótese de ruído branco.

```
res2 <- residuals(doisHarm)
ts.plot(res2)
```

sazonalidade_files/figure-pdf/unnamed-chunk-25-1.pdf

```
acf(res2)
```

```
shapiro.test(res2)
```

Shapiro-Wilk normality test

```
data:  res2
W = 0.98919, p-value = 0.06944
```

```
Box.test(res2, type = 'Ljung-Box')
```

Box-Ljung test

```
data:  res2
X-squared = 10.905, df = 1, p-value = 0.0009591
```

Previsão para modelos lineares

27 O valor ajustado como previsão

Seja $\mathcal{D}_t = \{y_1, \dots, y_t\}$ a série temporal observada. Portanto, para qualquer $h > 0$, o valor y_{t+h} é desconhecido. Inferências pontuais sobre esse valor são denominadas previsões, sendo denotadas por $\hat{y}_t(h)$, onde o valor de h é denominado horizonte (de previsão).

Considere que a série temporal y_t pode ser escrita como um modelo linear. Então, pelo princípio da substituição, o valor ajustado $\hat{y}_{t+h} = \mathbf{f}'_{t+h} \hat{\boldsymbol{\beta}}$ é um estimador para y_{t+h} e, portanto, $\hat{y}_t(h) = \hat{y}_{t+h}$ uma previsão para o valor da série no horizonte h .

Como $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{F}_n \mathbf{F}_n')^{-1} \nu)$, teremos que

$$\hat{y}_{t+h} \sim N(\mathbf{f}'_{t+h} \boldsymbol{\beta}, \nu \mathbf{f}'_{t+h} (\mathbf{F}_n \mathbf{F}_n')^{-1} \mathbf{f}_{t+h})$$

Como

$$\frac{\hat{y}_{t+h} - \mathbf{f}'_{t+h}\boldsymbol{\beta}}{\sqrt{\nu \mathbf{f}'_{t+h}(\mathbf{F}_n \mathbf{F}'_n)^{-1} \mathbf{f}_{t+h}}} \sim N(0, 1)$$

Um intervalo aproximado, de previsão $\gamma 100\%$, pode ser dado por

$$\left(\mathbf{f}'_{t+h} \hat{\boldsymbol{\beta}} + z_{\frac{1-\gamma}{2}} \sqrt{\hat{\nu} \mathbf{f}'_{t+h}(\mathbf{F}_n \mathbf{F}'_n)^{-1} \mathbf{f}_{t+h}}, \mathbf{f}'_{t+h} \hat{\boldsymbol{\beta}} + z_{\frac{1+\gamma}{2}} \sqrt{\hat{\nu} \mathbf{f}'_{t+h}(\mathbf{F}_n \mathbf{F}'_n)^{-1} \mathbf{f}_{t+h}} \right)$$

Em geral, utiliza-se γ igual a 0,8 ou 0,9.

28 Aplicação na série co2

A série co2 apresenta a média de concentração de carbono, em partes por milhão, em Mauna Loa.

```
ts.plot(co2)
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-1-1.p

É possível observar uma tendência crescente e um padrão sazonal. A série vai até 1997. Vamos remover os anos de 1996 e 1997 para utilizá-los na previsão.

```
co2_1995 <- window( co2, end = c(1995,12))
```

Vamos primeiramente eliminar a tendência, utilizando o loees, para estudar o padrão sazonal. Abaixo, mostramos a tendência estimada.

```
require(TSA)
```

Carregando pacotes exigidos: TSA

Warning: package 'TSA' was built under R version 4.3.2

Attaching package: 'TSA'

The following objects are masked from 'package:stats':

acf, arima

The following object is masked from 'package:utils':

tar

```
require(forecast)
```

Carregando pacotes exigidos: forecast

Warning: package 'forecast' was built under R version 4.3.1

Registered S3 method overwritten by 'quantmod':

| method | from |
|-------------------|------|
| as.zoo.data.frame | zoo |

Registered S3 methods overwritten by 'forecast':

| method | from |
|--------------|------|
| fitted.Arima | TSA |
| plot.Arima | TSA |

```
tempo <- 1:length(co2_1995)
lw <- loess( co2_1995 ~ tempo)
tendLoees <- lw$fitted
ts.plot(tendLoees)
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-3-1.p

```
semTend <- co2_1995 - tendLoees
```

```
ts.plot(semTend)
```

```
lines( ma(semTend, 12))
monthplot(semTend)
```

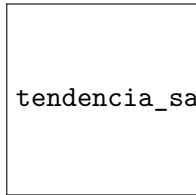


Figure 26: Gráfico da série sem tendência

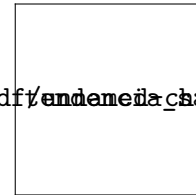
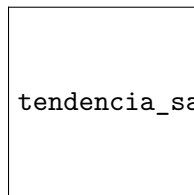


Figure 27: Gráfico de subséries

O sinal sazonal tem um efeito de ± 4 somados à tendência. O primeiro gráfico acima mostra a série subtraída da estimativa via loees, junto com uma média móvel de ordem 12, que oscila em torno de zero, o que são indícios de que a tendência foi removida. O gráfico de subséries apresenta comportamento estacionário para alguns meses. Outro parecem ter uma tendência, como abril, por exemplo. Contudo, o valor desse efeito é baixo se comparado com a tendência geral, o que nos permite assumir uma função periódica para a sazonalidade.

Abaixo apresentamos o periodograma. A frequência fundamental representa um período de 12 meses e a segunda frequência relevante mostra a necessidade do harmônico de ordem 2.

```
per <- periodogram(semTend)
```



```
tail( 1/per$freq[ order(per$spec)] , 3)
```

```
[1] 6.00000 12.16216 11.84211
```

Agora, vamos considerar apenas a tendência estimada, procurando por um polinômio de ordem adequada.

```
aic <- NULL
for(i in 1:15){
  mod <- lm( tendLooes ~ poly( tempo, i ,raw = T))
  aic[i] <- AIC(mod)
```

```
}
ts.plot(aic, type = 'o')
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-6-1.p

Vamos construir o modelo final. Para poder utilizar esse modelo para fazer previsões, precisamos construir a matriz de regressão, utilizando o comando `model.frame`, antes de construir o objeto `lm`. Essa matriz foi denominada por `X` abaixo.

```
ordemP <- 7
X <- model.matrix(~poly(tempo, ordemP, raw = T)+ harmonic(co2_1995, 2))
modFinal <- lm( co2_1995 ~ X)
ts.plot(co2_1995)
lines(ts(modFinal$fitted.values, start = start(co2), frequency = frequency(co2)), lwd = 2,
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-7-1.p

Abaixo, criamos a matriz de regressão com com os tempos correspondentes aos anos de 1996 e 1997. Note que vamos utilizar o nome `X` novamente.

```
# criando a matriz para previsão
n <- length(tempo)
tempoPrev <- (tempo[n]+1):(tempo[n]+24)
tempoPrev <- ts(tempoPrev, frequency = 12)
X <- model.matrix( ~ poly( tempoPrev,ordemP, raw = T) + harmonic(tempoPrev,2))
```

Agora, vamos utilizar a função `predict` e conjunto coma matriz criada anteriormente, para obter os valores previstos. Também vamos obter o intervalo de previsão de 95%.

```
pred <- predict(modFinal, data.frame(X), interval = 'prediction', level = .95)
pred <- ts(pred, start = c(1996,1), frequency = 12)
```

```
head(pred)
```

| | | fit | lwr | upr |
|-----|------|----------|----------|----------|
| Jan | 1996 | 361.6882 | 360.7432 | 362.6331 |
| Feb | 1996 | 362.5388 | 361.5834 | 363.4943 |
| Mar | 1996 | 363.5099 | 362.5435 | 364.4763 |
| Apr | 1996 | 364.6676 | 363.6897 | 365.6456 |
| May | 1996 | 365.4696 | 364.4787 | 366.4605 |
| Jun | 1996 | 365.1953 | 364.1894 | 366.2011 |

Abaixo, mostramos os valores previstos e os observados.

```
ts.plot(co2, xlim = c(1994,1998), ylim = c(350,370))  
lines(pred[,1], lwd =2, col =2)
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-10-1.

Abaixo, o mesmo gráfico mas com intervalo de previsão de 90%.

```
require(scales)
```

Carregando pacotes exigidos: scales

Warning: package 'scales' was built under R version 4.3.3

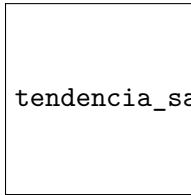
```
ts.plot(co2, xlim = c(1994,1998), ylim = c(350,370), type = 'p')  
polygon( 1996+c(0:23,23:0)/12, c(pred[,2],pred[24:1,3]), col = alpha('lightpink',.3), bord  
lines(pred[,1], lwd =2, col =2)
```

tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-11-1.

Note que o modelo conseguiu prever o ano de 1996 de modo satisfatório e os cinco primeiros meses de 1997. Este modelo parece ser adequado para previsões com o horizonte de doze meses.

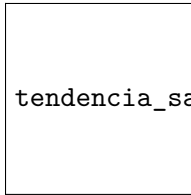
Vale ressaltar que este modelo não satisfaz a hipótese de ruído branco. Os gráficos dos resíduos revelam ainda características típicas de séries estacionárias.

```
res <- rstudent(modFinal)
ts.plot(res)
```



tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-12-1.

```
acf(res)
```



tendencia_sazonalidade_files/figure-pdf/unnamed-chunk-12-2.

```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data:  res
W = 0.99684, p-value = 0.5439
```

```
Box.test( res, type = 'Ljung-Box')
```

Box-Ljung test

```
data:  res
X-squared = 233.31, df = 1, p-value < 2.2e-16
```

29 Avaliando a qualidade da previsão

Considere que o objetivo principal da análise é a previsão. Um modelo pode falhar em alguma suposição, como normalidade dos erros, mas ainda sim produzir boas previsões. Por isso, é importante conseguir medir o quão bom é o modelo, o que implica estudar a diferença entre o previsto e o realizado. Vamos definir o erro de previsão por

$$u_t = \hat{y}_{t-1}(1) - y_t$$

Fixamos um valor J para separar as últimas J observações

$$y_{t-J+1}, \dots, y_t,$$

e, partir destas observações, calculamos a performance de previsão segundo alguma métrica a ser minimizada. As métricas mais comuns são:

- MAD (desvio médio absoluto):

$$MAD = \frac{1}{J} \sum_{i=t-J+1}^t |u_i|$$

- EQM (erro quadrático médio):

$$EQM = \frac{1}{J} \sum_{i=t-J+1}^t (u_i)^2$$

- MAPE (erro percentual médio)

$$MAPE = \frac{1}{J} \sum_{i=t-J+1}^t \frac{|u_i|}{y_i} \times 100\%$$

- SMAPE (erro simétrico percentual médio)

$$SMAPE = \frac{1}{J} \sum_{i=t-J+1}^t 2 \frac{|u_i|}{y_i + \hat{y}_{i-1}(1)} \times 100\%$$

- MedAPE (erro percentual mediano)

$$MedAPE = mediana \left(\frac{|u_i|}{y_i} \right) \times 100\%, \quad i = t - J + 1, \dots, t.$$

- MASE (erro escalonado médio)

$$MASE = 100\% \times \frac{1}{J} \frac{\sum_{i=t-J+1}^t |u_i|}{\frac{1}{J-1} \sum_{i=t-J+2}^t |y_i - y_{i-1}|}.$$

O MASE tem uma interpretação muito interessante: se $MASE > 100\%$, então o modelo é pior do que simplesmente fazer

$$y_{t-1}(1) = y_{t-1},$$

ou seja, prever y_t como sendo igual a y_{t-1} . Isto é considerado um “modelo ingênuo” (*naïve model*), sendo considerado o modelo de previsão mais básico.

Métodos de suavização exponencial

30 Introdução

Considere uma série temporal com a decomposição

$$y_t = \text{sinal}_t + \text{ruído}_t$$

Seja \mathcal{D}_t a série observada. Para qualquer $0 < k \leq t$ o valor suavizado $\tilde{\text{sinal}}_k$ corresponde a uma estimativa do sinal considerando a amostra \mathcal{D}_t .

Os métodos de suavização (ou alisamento) que serão estudados aqui são formulados considerando a seguinte lógica:

1. Obtenha uma previsão para o sinal no tempo $t - 1$
2. Obtenha uma estimativa para o sinal no tempo t
3. Calcule o valor suavizado através de uma combinação linear convexa dos resultados obtidos em 1. e 2.

Considere que função de previsão para y_t é

$$\hat{y}_{t-1}(1) = F(\theta_{t-1})$$

onde θ_t é o valor suavizado dos parâmetros envolvidos no tempo t . Assuma que o erro de previsão é aditivo, ou seja,

$$e_t = y_t - \hat{y}_{t-1}(1) = y_t - F(\theta_{t-1}).$$

Como a suavização de θ_t é obtida utilizando \hat{y}_{t-1} e θ_{t-1} , os modelos clássicos de suavização exponencial podem ser escritos da forma de equações de espaço estado com inovações, ou seja,

$$\begin{aligned} y_t &= F(\theta_{t-1}) + \Phi(\theta_{t-1})e_t \\ \theta_t &= G(\theta_{t-1}) + \psi(\theta_{t-1})e_t \end{aligned}$$

onde θ_t é denominado *estado*, $\psi(\cdot)$ e $\Phi(\cdot)$ são funções escalares, $F(\cdot)$ e $G(\cdot)$ são funções vetoriais.

Ao assumir que e_t é um ruído branco gaussiano, a função de verossimilhança para θ_0 e os demais parâmetros fixos é

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\nu}} \exp \left\{ -\frac{1}{2\nu} \left(\frac{y_i - F(\theta_{t-1})}{\Psi(\theta_{t-1})} \right)^2 \right\}$$

e suas estimativas, de máxima verossimilhança, são obtidas por maximização numérica de $\log L$.

Considerando a amostra \mathcal{D}_t , note que o valor θ_t já foi obtido. Podemos escrever

$$y_{t+h} = F(\theta_{t+h-1}) + \Psi(\theta_{t+h-1})e_{t+h}$$

e, utilizando diversas vezes a relação recursiva entre os estados, podemos escrever y_{t+h} como função de θ_t e dos erros e_{t+h}, \dots, e_{t+1} . Com isso, podemos encontrar a distribuição de y_{t+h} .

Um caso particular importante é dado quando $F(\cdot)$, $G(\cdot)$, são lineares, $\Psi(\cdot) = 1$ e $\psi(\cdot)$ não depende de θ_t . Nesse caso, o modelo de espaço estado é dado por:

$$\begin{aligned} y_t &= F'\theta_{t-1} + e_t \\ \theta_t &= G\theta_{t-1} + \psi e_t \end{aligned}$$

Nesse caso

$$\begin{aligned} y_{t+h} &= F'\theta_{t+h-1} + e_{t+h} = F'G\theta_{t+h-2} + F'\psi e_{t+h-1} + e_{t+h} \\ &= \dots \\ &= F'G^{h-1}\theta_t + \sum_{j=1}^{h-1} F'G^{j-1}\psi e_{t+h-j} + e_{t+h} \end{aligned}$$

Como y_{t+h} pode ser escrito como combinação linear de ruídos gaussianos, teremos que a distribuição para a previsão de horizonte h tem distribuição normal com média e variância dadas por

$$\begin{aligned} E(y_{t+h}|\mathcal{D}_t) &= F'G^{h-1}\theta_t \\ Var(y_{t+h}|\mathcal{D}_t) &= \nu \left[1 + \sum_{j=1}^{h-1} F'G^{j-1}\psi\psi'G'^{j-1}F \right] \end{aligned}$$

31 Suavização exponencial simples

31.1 Definição do método

Considere uma série temporal

$$y_t = \mu_t + \varepsilon_t$$

onde o nível μ_t é o sinal e ε_t um ruído branco. Como

$$E(y_t|\mathcal{D}_{t-1}) = \mu_t,$$

pelo método dos momentos, y_t é um estimador para μ_t (e, **depois observado**, y_t se torna estimativa para o sinal no tempo t).

Vamos adicionar a restrição de que $\mu_t \approx \mu_{t+1}$, ou seja, a série flutua em torno de pequenas oscilações no nível. Com isso, dada a amostra \mathcal{D}_t , teremos

$$\hat{y}_{t-1}(1) = E(y_t|\mathcal{D}_{t-1}) = \mu_t \approx \mu_{t-1}$$

A vantagem da aproximação é que podemos estimar μ_{t-1} , uma vez que temos a amostra \mathcal{D}_{t-1} . Deste modo, **antes de observar** y_t ,

$$\hat{y}_{t-1}(1) = \tilde{\mu}_{t-1}$$

é um estimador para μ_t , com $t > 1$.

Deste modo, temos duas fontes de informação sobre o nível: a observação y_t e o valor suavizado $\tilde{\mu}_{t-1}$, que representa a previsão $\hat{y}_{t-1}(1)$. O método de suavização exponencial simples consiste em ponderar essas fontes, criando a seguinte estimativa ponderada para o sinal:

$$\tilde{\mu}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1}(1),$$

onde $\alpha \in (0, 1)$. Também pode-se escrever o método da seguinte forma:

$$\tilde{\mu}_t = \alpha y_t + (1 - \alpha)\tilde{\mu}_{t-1}.$$

Tal forma nos permite entender o nome exponencial, uma vez que

$$\begin{aligned} \tilde{\mu}_t &= \alpha y_t + (1 - \alpha)\tilde{\mu}_{t-1} \\ &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + (1 - \alpha)^2\tilde{\mu}_{t-2} \\ &= \sum_{j=1}^{t-1} \alpha(1 - \alpha)^j y_{t-j} + (1 - \alpha)^t \tilde{\mu}_0 \end{aligned}$$

Esse modelo pode ser escrito na seguinte forma de espaço estado:

$$\begin{aligned} y_t &= \tilde{\mu}_{t-1} + e_t \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \alpha e_t \end{aligned}$$

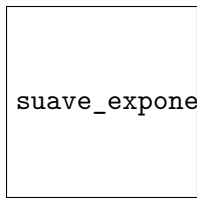
logo, identificando $F = 1, G = 1, \psi = \alpha$, teremos que

$$y_{t+h}|\mathcal{D}_t \sim \text{Normal}(\tilde{\mu}_t, \nu[1 + \alpha^2(h-1)])$$

31.2 Aplicação: nível do Nilo

Considere novamente a série Nile, cujos valores representam o fluxo anual do rio Nilo entre 1871 e 1970. Note que a série aparenta oscilar em torno de um nível constante após 1898.

```
ts.plot(Nile, ylab = expression(Fluxo~em~10^8~m^3) , lwd = 2)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-1-1.pdf

Vamos utilizar a função `ets(y, model)`, do pacote `forecast`, onde `y` é a série temporal e `model='ANN'` representa o modelo de suavização exponencial.

```
require(forecast)
```

Carregando pacotes exigidos: forecast

Warning: package 'forecast' was built under R version 4.3.1

Registered S3 method overwritten by 'quantmod':

```
method      from  
as.zoo.data.frame zoo
```

```
mod <- ets(Nile, 'ANN')  
mod
```

ETS(A,N,N)

Call:

```
ets(y = Nile, model = "ANN")
```

Smoothing parameters:

```
alpha = 0.2455
```

Initial states:

```
l = 1110.6869
```

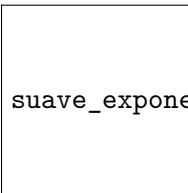
```
sigma: 144.2318
```

| | AIC | AICc | BIC |
|--|----------|----------|----------|
| | 1458.781 | 1459.031 | 1466.597 |

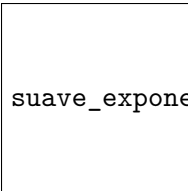
Acima podemos ver as estimativas dos parâmetros: $\hat{\alpha} = 0.2455$, $\tilde{\mu}_0 = 1110.6869$ e $\nu = 144.2318^2$.

Como de costume, devemos verificar se existem evidências para assumir que o modelo é adequado analisando os resíduos, que podem ser acessados no item `$residuals`.

```
res <- mod$residuals  
ts.plot(res)
```



```
acf(res)
```



```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data: res  
W = 0.99304, p-value = 0.8902
```

```
Box.test(res)
```

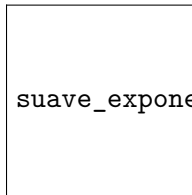
Box-Pierce test

```
data: res  
X-squared = 1.7244, df = 1, p-value = 0.1891
```

Os resíduos oscilam em torno de zero e o correlograma não mostra evidências contra a hipótese de ruído branco. O teste Shapiro-Wilks não rejeita a normalidade e o de Box-Pierce não rejeita a hipótese de ruído branco. Portanto, vamos considerar que o modelo é adequado.

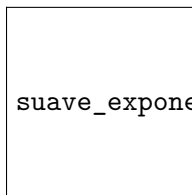
Os valores suavizados $\tilde{\mu}_0, \dots, \tilde{\mu}_t$ podem ser acessados na lista `$states`, conforme vemos abaixo.

```
ts.plot(Nile, ylab = expression(Fluxo~em~10^8~m^3) , lwd = 2)  
  
lines(mod$states, col = 'tomato', lwd = 3)
```



Por último, podemos realizar previsões com a função `forecast` conforme ilustramos abaixo para os 5 anos à frente (observe que a previsão é constante, mas, como esperado, sua variância aumenta linearmente ao longo do tempo).

```
plot( forecast(mod, 5))
```



32 Suavização Exponencial de Holt

Considere agora um modelo da forma

$$y_t = T(t) + \varepsilon_t,$$

onde $T(t)$ é uma componente de tendência. Nesse caso,

$$\mu_t = E(y_t | \mathcal{D}_t) = T(t)$$

Como vimos anteriormente, a suaziação no tempo t pode ser feita ponderando a observação y_t com sua respectiva previsão para no tempo $t - 1$:

$$\tilde{\mu}_t = \alpha y_t + (1 - \alpha) \hat{y}_{t-1}(1).$$

Como $T(t)$ é localmente linear, então é natural assumir um modelo de previsão da seguinte forma:

$$y_t(h) = \tilde{\mu}_t + h \tilde{b}_t,$$

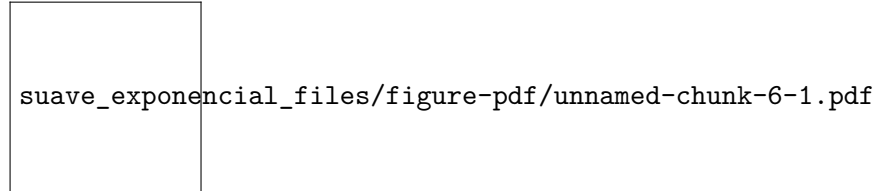
onde \tilde{b}_t é a inclinação da tendência no tempo t . Então,

$$\begin{aligned} \tilde{\mu}_t &= \alpha y_t + (1 - \alpha) \hat{y}_{t-1}(1) \\ &= \alpha y_t + (1 - \alpha) (\tilde{\mu}_{t-1} + \tilde{b}_{t-1}). \end{aligned}$$

A equação acima mostra a evolução de $\tilde{\mu}_t$, mas não há atualização para \tilde{b}_t . Com este objetivo, vamos ponderar duas fontes de informação sobre a inclinação. A primeira está no fato de que, como a tendência é localmente linear, é esperado que

$$\tilde{b}_t \approx \tilde{b}_{t-1}.$$

A segunda pode ser obtida através dos níveis $\tilde{\mu}_t$ e $\tilde{\mu}_{t-1}$. Uma vez que a relação entre esses dois é aproximadamente linear, sua diferença nos dá noção sobre a inclinação (crescimento/decrescimento) como mostra a figura abaixo.



O valor da inclinação da reta formada por a_t e $\tilde{\mu}_{t-1}$ é dado por

$$\frac{\tilde{\mu}_t - \tilde{\mu}_{t-1}}{t - (t - 1)} = \tilde{\mu}_t - \tilde{\mu}_{t-1}.$$

Combinando este com a última inclinação suavizada, teremos

$$\tilde{b}_t = \beta(\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta)\tilde{\mu}_{t-1},$$

onde $\beta \in (0, 1)$ é a constante de suavização da tendência.

Combinando as últimas equações teremos o modelo de suavização exponencial de Holt:

$$\begin{aligned}\tilde{\mu}_t &= \alpha y_t + (1 - \alpha) (\tilde{\mu}_{t-1} + \tilde{b}_{t-1}) \\ \tilde{b}_t &= \beta (\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta) \tilde{b}_{t-1}.\end{aligned}$$

O conjunto de equações acima pode ser reescrito como pode ser reescrita como

$$\begin{aligned}\tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{b}_{t-1} + \alpha (y_t - \tilde{\mu}_{t-1} - \tilde{b}_{t-1}) = \tilde{\mu}_{t-1} + \tilde{b}_{t-1} + \alpha e_t \\ \tilde{b}_t &= \tilde{b}_{t-1} + \beta (\tilde{\mu}_t - \tilde{\mu}_{t-1} - \tilde{b}_{t-1}).\end{aligned}$$

A partir da primeira equação acima, deduzimos que $\tilde{\mu}_t - \tilde{\mu}_{t-1} - \tilde{b}_{t-1} = \alpha e_t$. Substituindo essa informação na segunda equação teremos

$$\begin{aligned}\tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{b}_{t-1} + \alpha e_t \\ \tilde{b}_t &= \tilde{b}_{t-1} + \alpha \beta e_t.\end{aligned}$$

ou, em forma matricial,

$$\begin{pmatrix} \tilde{\mu}_t \\ \tilde{b}_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mu}_{t-1} \\ \tilde{b}_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha \\ \alpha\beta \end{pmatrix} e_t,$$

Note ainda que é possível escrever a equação do modelo em forma matricial:

$$y_{t+h} = \tilde{\mu}_{t+h-1} + \tilde{b}_{t+h-1} + e_{t+h-1} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mu}_{t+h-1} \\ \tilde{b}_{t+h-1} \end{pmatrix} + e_{t+h}$$

Pode-se então mostrar que este modelo possui a seguinte especificação como modelo de espaço estado:

$$\begin{aligned}F' &= \begin{pmatrix} 1 & 1 \end{pmatrix} \\ G &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ \theta'_t &= (\tilde{\mu}_t \quad \tilde{b}_t) \\ \psi' &= (\alpha \quad \alpha\beta)\end{aligned}$$

Desse modo, os parâmetros $\alpha, \beta, \tilde{\mu}_0$ e \tilde{b}_0 podem ser obtidos via método de estimação da máxima verossimilhança. Logo, a distribuição da previsão com horizonte h é dada por

$$y_{t+h} | \mathcal{D}_t \sim \text{Normal} \left(\tilde{\mu}_t + h\tilde{b}_t, \nu\sigma_h^2 \right),$$

onde

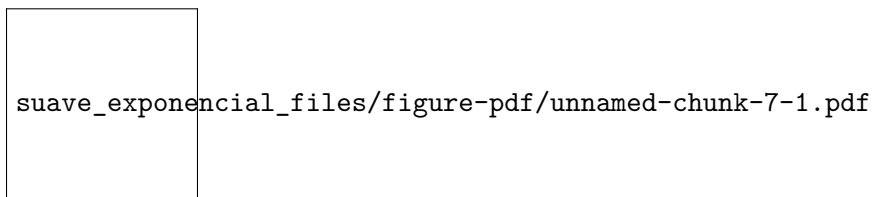
$$\sigma_h^2 = 1 + (h-1) \left(\alpha^2 + \alpha\beta h + \frac{\beta^2}{6} h(2h-1) \right).$$

32.1 Aplicação: ocorrências aeronáuticas

Voltemos ao conjunto de dados com o número mensal de ocorrências aeronáuticas, mantido pela Força Aérea Brasileira.

```
url <- 'https://www.dropbox.com/scl/fi/kq4jwbovu94u857238sus/N-mensal-de-acidentes-com-aer

x <- read.csv(url, h = T)
ocorrenciasFAB <- ts( x, start = c(2013,1), frequency=12)
ts.plot(ocorrenciasFAB, lwd = 2, xlab = 'Ano', ylab = 'No. ocorrências aeronáuticas')
```



Vamos utilizar a função `ets(y, model)`, do pacote `forecast`, onde `y` é a série temporal e `model='AAN'` representa o modelo de suavização exponencial de Holt. Vamos adicionar o argumento `damped=FALSE` - estudaremos esse argumento na seção sobre amortecimento.

```
require(forecast)
mod <- ets(ocorrenciasFAB, 'AAN', damped = FALSE)
mod
```

ETS(A,A,N)

Call:

```
ets(y = ocorrenciasFAB, model = "AAN", damped = FALSE)
```

Smoothing parameters:

alpha = 0.0651

beta = 0.0117

Initial states:

l = 61.6314

b = -0.9175

sigma: 7.7512

| | AIC | AICc | BIC |
|--|----------|----------|----------|
| | 1131.363 | 1131.863 | 1145.545 |

Acima podemos ver as estimativas dos parâmetros: $\hat{\alpha} = 0.0651$, $\hat{\beta} = 0,0117$, $\tilde{\mu}_0 = 61,6314$, $\tilde{b}_0 = -0,9175$ e $\nu = 7,7512^2$.

Como de costume, devemos verificar se existem evidências para assumir que o modelo é adequado analisando os resíduos, que podem ser acessados no item `$residuals`.

```
res <- mod$residuals
ts.plot(res)
```

suave_exponencial_files/figure-pdf/unnamed-chunk-9-1.pdf

```
acf(res)
```

suave_exponencial_files/figure-pdf/unnamed-chunk-9-2.pdf

```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data:  res
W = 0.99237, p-value = 0.7257
```

```
Box.test(res)
```

Box-Pierce test

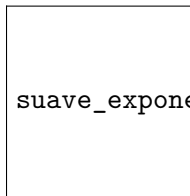
```
data: res
X-squared = 3.2409, df = 1, p-value = 0.07182
```

Os resíduos oscilam em torno de zero e o correlograma não mostra evidências contra a hipótese de ruído branco. O teste Shapiro-Wilks não rejeita a normalidade e o de Box-Pierce não rejeita a hipótese de ruído branco. Portanto, vamos considerar que o modelo é adequado.

Os valores suavizados $\tilde{\mu}_0, \dots, \tilde{\mu}_t$ podem ser acessados na lista `$states`, na primeira coluna, conforme vemos abaixo.

```
ts.plot(ocorrenciasFAB, ylab = 'No. ocorrências aeronáuticas mensal' , lwd = 2)

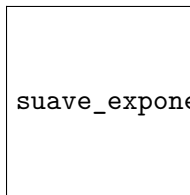
lines(mod$states[,1], col = 'tomato', lwd = 3)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-10-1.pdf

A segunda coluna de `$states` mostra os valores suavizados para a inclinação. É interessante observar o gráfico desses valores contra a linha horizontal em zero, para entender os regimes de crescimento e decrescimento da série. Abaixo mostramos o curioso padrão dessa série: a inclinação começou desacelerando desde de o começo do registro até março de 2017. Desde então, a inclinação se oscila em torno de um nível constante.

```
ts.plot(mod$states[,2], col = 'tomato', lwd = 3, ylab = 'Inclinação suavizada')
abline(h=0, lty = 2)
abline(v = 2017+3/12, lty = 2)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-11-1.pdf

Por último, podemos realizar previsões com a função `forecast` conforme ilustramos abaixo para os 6 meses à frente, completando o ano de 2023.

```
prev <- forecast(mod, 6)
plot( prev)
```

suave_exponencial_files/figure-pdf/unnamed-chunk-12-1.pdf

33 Suavização Exponencial de Holt-Winters

33.1 Definição

Considere o modelo

$$y_t = T(t) + s(t) + \varepsilon_t,$$

onde $T(t)$ e $s(t)$ são componentes de tendência e sazonalidade, com período p , respectivamente. Identificando $T(t)$ como μ_t , teremos que $E(y_t|\mathcal{D}_{t-1}) = \mu_t + s_t$, logo $y_t - \tilde{s}_t$ é uma fonte de informação para μ_t .

Considere ainda que $T(t)$ é localmente liner. Então, as previsões em curto prazo podem ser feitas através de

$$\hat{y}_t(h) = \tilde{\mu}_t + h\tilde{b}_t + \tilde{s}_{t+h}.$$

Então Outra fonte de informação sobre μ_t é a previsão $\hat{y}_{t-1}(1)$ livre de sazonalidade, dada por

$$\hat{y}_{t-1}(1) - \tilde{s}_t = \tilde{\mu}_{t-1} + \tilde{b}_{t-1}.$$

Deste modo, o sinal μ_t pode ser suavizado de modo análogo ao que foi feito no modelo de suavização exponencial simples, ponderando as duas fontes de informação:

$$\begin{aligned}\tilde{\mu}_t &= \alpha(y_t - \tilde{s}_t) + (1 - \alpha)(\hat{y}_{t-1}(1) - \tilde{s}_t) \\ &= \alpha(y_t - \tilde{s}_t) + (1 - \alpha)(\tilde{\mu}_{t-1} + \tilde{b}_{t-1}).\end{aligned}$$

Uma vez que temos o valor suavizado do nível, podemos suavizar a inclinação exatamente como foi feito no modelo de Holt:

$$\tilde{b}_t = \beta(\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta)\tilde{b}_{t-1}.$$

Vamos agora reunir duas fontes de informação sobre a componente sazonal. Primeiro, podemos descontar o sinal de tendência da série, obtendo assim informações sobre a sazonalidade:

$$y_t - \mu_t \approx s_t.$$

Vamos considerar que a componente sazonal é razoavelmente estável, ou seja $s_{t-p} \approx s_t$ (é a mesma consideração feita no modelo de suavização exponencial). Portanto, podemos suavizar s_t através da seguinte média ponderada

$$\tilde{s}_t = \gamma(y_t - \tilde{\mu}_t) + (1 - \gamma)\tilde{s}_{t-p},$$

onde $\beta \in (0, 1)$ é o suavizador sazonal.

O modelo de Holt-Winters é dado por

$$\begin{aligned}\tilde{\mu}_t &= \alpha(y_t - \tilde{s}_t) + (1 - \alpha)(\tilde{\mu}_{t-1} + \tilde{b}_{t-1}) \\ \tilde{b}_t &= \beta(\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta)\tilde{b}_{t-1}, \\ \tilde{s}_t &= \gamma(y_t - \tilde{\mu}_t) + (1 - \gamma)\tilde{s}_{t-p}.\end{aligned}$$

Agora, considere a condição inicial

$$\sum_{j=1}^p \tilde{s}_{j-p} = 0.$$

A suavização de Holt-Winter não mantém a propriedade de soma zero da componente sazonal, uma vez que

$$\sum_{j=1}^p \tilde{s}_j = \sum_{j=1}^p \gamma(y_j - \tilde{\mu}_j) + (1 - \gamma) \sum_{j=1}^p \tilde{s}_{j-p} = \sum_{j=1}^p \gamma(y_j - \tilde{\mu}_j) \neq 0$$

A solução, conhecida como normalização de Roberts-McKenzie, consiste em subtrair, da componente sazonal, o termo

$$a_t = \frac{\beta}{p}(y_t - \tilde{\mu}_t - \tilde{s}_t),$$

ou seja,

$$\tilde{s}_t = \gamma(y_t - \tilde{\mu}_t) + (1 - \gamma)\tilde{s}_{t-p} - a_t.$$

Desse modo,

$$\begin{aligned}\sum_{j=1}^p \tilde{s}_j &= \sum_{j=1}^p \gamma(y_j - \tilde{\mu}_j) + (1 - \gamma) \sum_{j=1}^p \tilde{s}_{j-p} - \sum_{j=1}^p a_j \\ &= \sum_{j=1}^p \tilde{s}_{j-p}\end{aligned}$$

Portanto, para garantir que

$$\sum_{j=1}^p s_{t+j} = 0,$$

basta que a soma dos parâmetros iniciais, $\tilde{s}_{1-p}, \dots, \tilde{s}_0$, seja nula. Portanto, o modelo de Holt-Winter com a normalização de Roberts-McKenzie é

$$\begin{aligned}\tilde{\mu}_t &= \alpha(y_t - \tilde{s}_t) + (1 - \alpha)(\tilde{\mu}_{t-1} + \tilde{b}_{t-1}) \\ \tilde{b}_t &= \beta(\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta)\tilde{b}_{t-1}, \\ \tilde{s}_t &= \gamma(y_t - \tilde{\mu}_t) + (1 - \gamma)\tilde{s}_{t-p} - \frac{\gamma}{p}(y_t - \tilde{\mu}_{t-1} - \tilde{s}_t).\end{aligned}$$

Como esperado, esse modelo pode ser reescrito na forma de espaço estado, considerando o erro de previsão $e_t = y_t - \tilde{\mu}_{t-1} - \tilde{b}_{t-1} - \tilde{s}_{t-p}$. As duas primeiras equações de estado são as mesmas do modelo de Holt:

$$\begin{pmatrix} \tilde{\mu}_t \\ \tilde{b}_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mu}_{t-1} \\ \tilde{b}_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha \\ \alpha\beta \end{pmatrix} e_t = G_{\text{Holt}} \begin{pmatrix} \tilde{\mu}_{t-1} \\ \tilde{b}_{t-1} \end{pmatrix} + \psi_{\text{Holt}} e_t$$

Já a terceira equação do modelo, com a normalização de Roberts-McKenzie, pode ser escrita como

$$\tilde{s}_t = \tilde{s}_{t-p} + \gamma \left(1 - \frac{1}{p}\right) e_t.$$

Note que essa mesma equação pode ser escrita como

$$\begin{pmatrix} \tilde{s}_t \\ \tilde{s}_{t-1} \\ \vdots \\ s_{t-p+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0_{p-1'} \\ 1 \\ \mathbf{I}_{p-1} \\ 0_{p-1} \end{pmatrix}}_P \begin{pmatrix} \tilde{s}_{t-1} \\ \tilde{s}_{t-2} \\ \vdots \\ s_{t-p} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \gamma \left(1 - \frac{1}{p}\right) e_t$$

A matriz P dada acima é uma matriz do tipo permutação. Essa em particular coloca o elemento da última coordenada na primeira e faz com que todos os outros elementos seja alocados uma coordenada à frente. Fazendo $\theta'_0 = (\tilde{\mu}_0, \tilde{b}_0, \tilde{s}_{1-p}, \dots, \tilde{s}_0)$ teremos que o modelo de Holt-Winter com correção de Roberts-McKenzie na forma de espaço estado é dado por

$$\begin{aligned}y_t &= \left(\begin{array}{cc|ccc} 1 & 1 & 1 & 0 & \cdots & 0 \end{array} \right) \theta_{t-1} + e_t \\ \theta_t &= \begin{pmatrix} G_{\text{Holt}} & 0_{2 \times p} \\ 0_{p \times 2} & P \end{pmatrix} \theta_{t-1} + \begin{pmatrix} \alpha \\ \alpha\beta \\ \gamma^* \mathbf{1}_p \end{pmatrix} e_t\end{aligned}$$

onde $\gamma^* = \gamma(1 - p^{-1})$.

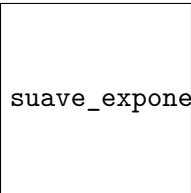
33.2 Aplicação para a série co2

Considere novamente a série co2.

```
mod <- ets(co2, model = 'AAA')
```

Abaixo, segue os gráfico da decomposição através do modelo de Holt-Winter com a normalização.


```
x <- data.frame(co2, mod$states[,1][-1], mod$states[,2][-1], mod$states[,3][-1])
names(x) <- c('Série original', 'Nível', 'Inclinação', 'Sazonalidade')
plot.ts(x)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-14-1.pdf

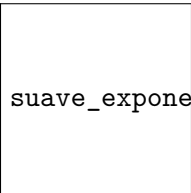
A análise dos resíduos, mostrada abaixo, mostram que o modelo está bem ajustado ao conjunto de dados.

```
res <- mod$residuals
ts.plot(res)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-15-1.pdf

```
acf(res)
```



suave_exponencial_files/figure-pdf/unnamed-chunk-15-2.pdf

```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data: res
W = 0.99742, p-value = 0.6848
```

```
Box.test(res)
```

Box-Pierce test

```
data: res
X-squared = 2.8016, df = 1, p-value = 0.09417
```

34 Modelo de Holt amortecido

Considere que a previsão um passo à frente é dada por,

$$y_t(1) = \tilde{\mu}_{t-1} + \phi \tilde{b}_{t-1} = (1 \ \phi) \begin{pmatrix} \tilde{\mu}_{t-1} \\ \tilde{b}_{t-1} \end{pmatrix}$$

onde $\phi \in (0, 1)$ tem o objetivo de amortecer a tendência dada pelo método de Holt, criando um previsão não linear. Com isso, teremos

$$\begin{aligned} \tilde{\mu}_t &= \alpha y_t + (1 - \alpha)(\tilde{\mu}_{t-1} + \phi \tilde{b}_{t-1}) \\ \tilde{b}_t &= \beta(\tilde{\mu}_t - \tilde{\mu}_{t-1}) + (1 - \beta)\phi \tilde{b}_{t-1} \end{aligned}$$

que por sua vez pode ser reescrita como

$$\begin{pmatrix} \tilde{\mu}_t \\ \tilde{b}_t \end{pmatrix} = \begin{pmatrix} 1 & \phi \\ 0 & \phi \end{pmatrix} \begin{pmatrix} \tilde{\mu}_{t-1} \\ \tilde{b}_{t-1} \end{pmatrix} + \begin{pmatrix} \alpha_t \\ \alpha\beta \end{pmatrix} e_t$$

e, identificando a matriz acima com o termo ϕ como G , teremos que

$$\hat{y}_t(h) = F' G^{h-1} \theta_t = \tilde{\mu}_t + (\phi + \dots + \phi^h) \tilde{b}_t$$

Abaixo, segue um esboço dessa função para alguns valores de ϕ .

35 Modelos com componentes multiplicativas

35.1 Tendência multiplicativa

Suponha que

$$\hat{y}_t(h) = \tilde{\mu}_t \tilde{b}_t^h$$

o que implica que o logaritmo de y_t possui tendência localmente linear. Podemos adaptar o método de Holt, suavizando, como de costume, $\tilde{\mu}_t$ por

$$\begin{aligned}\tilde{\mu}_t &= \alpha y_t + (1 - \alpha) \hat{y}_{t-1}(1) \\ &= \alpha \tilde{\mu}_t + (1 - \alpha) \tilde{\mu}_{t-1} \tilde{b}_{t-1}\end{aligned}$$

Além disso, como o logaritmo da tendência é localmente linear, teremos que

$$\log \tilde{\mu}_t - \log \tilde{\mu}_{t-1} \approx \log \tilde{b}_t$$

ou ainda

$$\tilde{b}_t \approx \frac{\tilde{\mu}_t}{\tilde{\mu}_{t-1}},$$

logo, considerando que $\tilde{b}_t \approx \tilde{b}_{t-1}$, podemos suavizar b_t do seguinte modo:

$$\tilde{b}_t = \beta \frac{\tilde{\mu}_t}{\tilde{\mu}_{t-1}} + (1 - \beta) \tilde{b}_{t-1}.$$

Deste modo, o modelo com tendência multiplicativa é dado por

$$\begin{aligned}\tilde{\mu}_t &= \alpha y_t + (1 - \alpha) \tilde{\mu}_{t-1} \tilde{b}_{t-1} \\ \tilde{b}_t &= \beta \frac{\tilde{\mu}_t}{\tilde{\mu}_{t-1}} + (1 - \beta) \tilde{b}_{t-1}\end{aligned}$$

e, na forma de modelo de espaço estado,

$$\begin{aligned}y_t &= \tilde{\mu}_{t-1} \tilde{b}_{t-1} + e_t \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} \tilde{b}_{t-1} + \alpha e_t \\ \tilde{b}_t &= \tilde{b}_{t-1} + \alpha \beta \frac{e_t}{\tilde{\mu}_{t-1}}\end{aligned}$$

As estimativas de máxima verossimilhança para $\tilde{\mu}_0, \alpha, \nu$ são obtidas via maximização da função de verossimilhança, conforme visto na introdução. Já a distribuição da função de previsão

Note que também existe uma versão deste modelo com o amortecedor. Supondo que

$$\log \hat{y}_t(h) = \log(\tilde{\mu}_{t-1}) + (\phi + \dots + \phi^h) \log \tilde{b}_{t-1}$$

teremos que a função de previsão com o amortecedor é dada por

$$\hat{y}_t(h) = \tilde{\mu}_{t-1} \tilde{b}_{t-1}^{(\phi + \dots + \phi^h)}$$

que por sua vez possui a seguinte forma em espaço estado:

$$\begin{aligned} y_t &= \tilde{\mu}_{t-1} \tilde{b}_{t-1}^\phi + e_t \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} \tilde{b}_{t-1}^\phi + \alpha e_t \\ \tilde{b}_t &= \tilde{b}_{t-1}^\phi + \alpha \beta \frac{e_t}{\tilde{\mu}_{t-1}} \end{aligned}$$

35.2 Tendência aditiva e sazonalidade multiplicativa

teste

Modelos ARMA

36 O polinômio de defasagens

Definition .6. Considere uma sequência $\{a_t\}$. O operador defasagem é definido como $Ba_t = a_{t-1}$. Em inglês este operador é conhecido como *backshift*.

Example .15. Considere a série 11, 6, 3, 7, 15, 8.

$$Ba_2 = a_1 = 11$$

$$Ba_6 = a_5 = 15$$

■

Seguem algumas propriedades fundamentais do operador defasagem

- $Bc = c$
- $B^k a_t = a_{t-k}$.
- $B(c + ba_{t-1}) = c + bBa_t$ (operador linear)
- $(B^m - B^n)a_t = a_{t-m} - a_{t-n}$.

- $B^{-1}a_t = a_{t+1}$

Note que o operador B^{-1} leva a_t para um passo a frente (como em uma previsão). É comum encontrar a definição $F = B^{-1}$, onde a letra F é escolhida por causa do termo *forecast* (previsão).

Example .16. Seja B o operado defasagem. Então, definimos um polinômio de defasagens como

$$\psi(B) = a_0 + a_1B + \cdots + a_nB^n.$$

Note que $\psi(B)$ é um modo sucinto para escrever

$$a(B)x_t = a_0x_t + a_1x_{t-1} + \cdots + a_nx_{t-n}.$$

O polinômio de defasagens pode ser operado como um polinômio regular. Por exemplo, se $a(B) = 1 - aB$ e $b(B) = 1 - bB$, fazer

$$\begin{aligned} a(B)b(B)x_t &= a(B)[(1 - bB)x_t] = a(B)(x_t - bx_{t-1}) \\ &= a(B)x_t - ba(B)x_{t-1} = (1 - aB)x_t - b(1 - aB)x_{t-1} \\ &= x_t - (a + b)x_{t-1} + abx_{t-2} \end{aligned}$$

é equivalente a encontrar

$$c(B) = (1 - aB)(1 - bB) = 1 - (a + b)B + abB^2$$

e calcular

$$c(B)x_t = x_t - (a + b)x_{t-1} + abx_{t-2}.$$

Dizemos que o polinômio de defasagens $\psi(B)$ possui inversa se existe uma função $\psi^{-1}(B)$ tal que $\psi(B)\psi^{-1}(B)x_t = x_t$.

Example .17. Para entender corretamente os modelos que serão propostos, é fundamental entender o que é a inversa do polinômio de defasagens. O caso mais importante, que será a chave para os demais, é o baseado no seguinte polinômio:

$$\psi(B) = 1 - \phi B.$$

Suponha que

$$y_t = (1 - \phi B)x_t = x_t - \phi x_{t-1}$$

Estamos procurando que qual situação existe $\psi^{-1}(B)$ tal que

$$x_t = \psi^{-1}(B)y_t.$$

Como

$$y_{t-1} = (1 - \phi B)x_{t-1} = x_{t-1} - \phi x_{t-2},$$

teremos que

$$y_t = x_t - \phi[y_{t-1} - \phi x_{t-2}] = x_t - \phi y_{t-1} - \phi^2 x_{t-2}.$$

Note que podemos continuar iterando as equações acima, obtendo

$$y_t = x_t - \sum_{j=1}^{t-1} \phi^j y_{t-j} - \phi^t x_0.$$

Assuma que $|\phi| < 1$ e que t é grande, o que implica que $\phi^t x_0$ é desprezível. Então

$$y_t = x_t - \sum_{j=1}^{t-1} \phi^j y_{t-j} = x_t - \sum_{j=1}^{t-1} \phi^j B^j y_t$$

ou ainda

$$\left(1 - \sum_{j=1}^{t-1} \phi^j B^j\right) y_t = x_t$$

Agora, multiplique os dois lados da equação acima por $1 - \phi B$. Então

$$(1 - \phi B) \left(1 - \sum_{j=1}^{t-1} \phi^j B^j\right) y_t = (1 - \phi B)x_t = y_t$$

Deste modo,

$$\psi^{-1}(B) = \lim_{t \rightarrow \infty} \left(1 - \sum_{j=1}^t \phi^j B^j\right) = 1 - \sum_{j=1}^{\infty} \phi^j B^j$$

■

No exemplo anterior, o fato de $|\phi| < 1$ garante que a inversa do polinômio de defasagens existem uma vez que $\lim_{t \rightarrow \infty} \sum_{j=1}^t \phi^j B^j$ é convergente.

O resultado geral é baseado no seguinte teorema.

Proposition .1. *Seja T uma matriz quadrada qualquer e seja*

$$S_n = \sum_{j=1}^n T^j.$$

A série S_n converge quando $n \rightarrow \infty$ se e somente se todos os autovalores de T são menores que um em módulo. Nesse caso, $T^j \rightarrow \mathbf{0}$ quando $j \rightarrow \infty$ e

$$S_n \rightarrow (1 - T)^{-1}.$$

A proposição acima é a chave para demonstrar o seguinte teorema

Theorem .2. *Seja*

$$y_t = x_t - \sum_{j=1}^p \phi_j x_{t-j} = \left(1 - \sum_{j=1}^p \phi_j B^j \right) x_t = \phi(B).$$

Então, existe $\phi^{-1}(B)$ se e somente se o módulo das raízes de $\phi(B)$ são maiores que um.

Proof. Faremos a demonstração para o caso $p = 2$, mas os mesmos passos podem ser seguidos para demonstrar o caso geral.

Seja

$$y_t = x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = (1 - \phi_1 B - \phi_2 B^2) x_t.$$

Começemos notando que

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = \begin{pmatrix} y_t \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \end{pmatrix}$$

Fazendo $z_t = (x_t \ x_{t-1})$, teremos

$$z_t = \begin{pmatrix} y_t \\ 0 \end{pmatrix} + \underbrace{\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}}_A z_{t-1}.$$

Utilizando essa relação recursiva, teremos

$$z_t = A^t z_0 + \sum_{j=0}^{t-1} A^j \begin{pmatrix} y_{t-j} \\ 0 \end{pmatrix}$$

e, notando que $x_t = (1 \ 0) z_t$, teremos

$$x_t = (1 \ 0) A^t z_0 + (1 \ 0) \sum_{j=0}^{t-1} A^j \begin{pmatrix} y_{t-j} \\ 0 \end{pmatrix}$$

Suponha que os autovalores de A são, em módulo, maiores que um. Então, pela Proposition ??, para t suficientemente grande,

$$\begin{aligned} x_t &= (1 \ 0) \sum_{j=1}^{\infty} A^j \begin{pmatrix} y_{t-j} \\ 0 \end{pmatrix} = (1 \ 0) \sum_{j=1}^{\infty} A^j B^j \begin{pmatrix} y_t \\ 0 \end{pmatrix} \\ &= \sum_{j=1}^{\infty} (1 \ 0) A^j B^j \begin{pmatrix} 1 \\ 0 \end{pmatrix} y_t = \phi^{-1}(B) y_t \end{aligned}$$

Agora, observe que os autovalores de A são obtidos através da solução de

$$\begin{aligned} 0 &= \left| \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} - \lambda \mathbf{I} \right| - \lambda(\phi_1 - \lambda) - \phi_2 \\ &= \lambda^2 - \lambda\phi_1 - \phi_2 \\ &= 1 - \frac{1}{\lambda}\phi_1 - \phi_2 \frac{1}{\lambda^2} \end{aligned}$$

Fazendo $\lambda = 1/B$, teremos que a equação acima se torna

$$0 = 1 - \frac{1}{\lambda}\phi_1 - \phi_2 \frac{1}{\lambda^2} \equiv 1 - \phi_1 B - \phi_2 B^2 = \phi(B)$$

logo, se o módulo das raízes de $\phi(B)$ são maiores que um, então o módulo dos autovalores de A são menores que um e, portanto, existe $\phi^{-1}(B)$.

□

37 O modelo autorregressivo

O modelo autorregressivo de ordem p , ou $AR(p)$, é dado por

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t \quad (12)$$

onde $\{\varepsilon_t\}$ é um ruído branco, tipicamente $\text{Normal}(0, \nu)$. Nesse modelo, a contribuição da observação y_{t-j} em y_t é dada por ϕ_j , que é invariante no tempo.

Utilizando o polinômio de defasagens, pode-se escrever o modelo $AR(p)$ como

$$\phi(B)y_t = \varepsilon_t, \quad (13)$$

onde $\phi(B) = 1 - B\phi_1 - \dots - \phi_p B^p$. Se o módulo das raízes desse polinômio são maiores que um, então existe $\phi^{-1}(B)$, ou seja

$$y_t = \phi^{-1}(B)\varepsilon_t = \left(\sum_{j=1}^{\infty} \psi_j B^j \right) \varepsilon_t$$

e o processo será estacionário, com média e variância iguais à

$$\begin{aligned} E(y_t) &= \left(\sum_{j=1}^{\infty} \psi_j B^j \right) E(\varepsilon_t) = 0 \\ Var(y_t) &= Var \left(\sum_{j=1}^{\infty} \psi_j B^j \varepsilon_t \right) = \nu \sum_{j=1}^{\infty} \psi_j^2 \end{aligned}$$

e a função de auto covariância é dada por

$$\begin{aligned}
\gamma(h) &= Cov(y_t, y_{t-h}) \\
&= Cov \left(\left(\sum_{j=1}^{\infty} \psi_j B^j \right) \varepsilon_t, \left(\sum_{k=1}^{\infty} \psi_k B^k \right) \varepsilon_{t-h} \right) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \psi_j \psi_k Cov(\varepsilon_{t-j}, \varepsilon_{t-k-h}) \\
&= \nu \sum_{j=1}^{\infty} \psi_j \psi_{j+h}
\end{aligned} \tag{14}$$

Example .18. Considere o processo $AR(1)$ abaixo:

$$x_t(1 - B\phi) = \varepsilon_{t-1}.$$

Seja \dot{B} a raiz do polinômio $1 - B\phi$. Temos que

$$\phi(\dot{B}) = 0 \Rightarrow 1 - \phi\dot{B} = 0 \Rightarrow \dot{B} = \frac{1}{\phi}$$

Logo, o processo $AR(1)$ é estacionário se

$$|\dot{B}| > 1 \Rightarrow \left| \frac{1}{\phi} \right| > 1 \Rightarrow |\phi| < 1$$

Nesse caso, já vimos que $\phi^{-1}(B) = 1 - \sum_{j=1}^{\infty} \phi^j B^j$. Então, pela equação (Equation ??), identificando $\psi_j = \phi^j$, teremos que

$$\gamma(h) = \nu \sum_{j=1}^{\infty} \phi^j \phi^{j+h} = \nu \phi^h \frac{\phi^2}{1 - \phi^2}$$

Assim, a função de autocorrelação é dada por

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h.$$

Note que:

- Se $\phi \in (0, 1)$, então $\rho(h)$ decai exponencialmente para zero
- Se $\phi \in (-1, 0)$, então $\rho(h)$ decai exponencialmente para zero, mas alternando o sinal.

processo_linear_geral_files/figure-pdf/unnamed-chunk-1-1.pdf

■

Example .19. Considere que x_t é um AR(2), ou seja,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t = \phi(B)\varepsilon_t,$$

onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$. As raízes desse polinômio são

$$\dot{B} = \frac{1}{2\phi_2} \left[\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2} \right]$$

e, considerando que $|\dot{B}| > 1$, o processo será estacionário se (ϕ_1, ϕ_2) pertence ao triângulo delimitado pelos vértices $(-2, -1)$, $(0, 1)$, $(2, -1)$. É interessante notarmos que, se $\phi_2 < 0$, é possível que as raízes de $\phi(B)$ sejam um par de números complexos conjugados.

■

37.1 Função de autocorrelação para o $AR(p)$

Consideremos o processo

$$x_t = \sum_{j=1}^p \phi_j x_{t-j} + \varepsilon_t.$$

e suponha que ele é estacionário. Multiplicando ambos os lados por x_{t-h} e aplicando a esperança, teremos

$$E(x_{t-h}x_t) = \sum_{j=1}^p \phi_j E(x_{t-h}x_{t-j})$$

e, como $\gamma(h) = E(x_t x_{t-h})$, teremos

$$\gamma(h) = \sum_{j=1}^p \phi_j \gamma(h-j) \tag{15}$$

Dividindo ambos os lados por $\gamma(0)$, teremos

$$\rho(h) = \sum_{j=1}^p \phi_j \rho(h-j)$$

Essa relação pode ser utilizada para encontrar a função de autocorrelação do processo sem a necessidade de encontrar a inversa de $\phi(B)$.

Example .20. Considere o processo $AR(2)$ abaixo:

$$x_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

Multiplicando a equação acima por x_{t-1} em ambos os lados teremos

$$x_t x_{t-1} = \phi_1 x_{t-1}^2 + \phi_2 x_{t-2} x_{t-1} + \varepsilon_t x_{t-1}.$$

Calculando o valor esperado, temos

$$\begin{aligned} \gamma(1) &= \phi_1 E(x_{t-1}^2) + \phi_2 E(x_{t-2} x_{t-1}) + E(\varepsilon_t x_{t-1}) \\ &= \phi_1 \gamma(0) + \phi_2 \gamma(1). \end{aligned}$$

Dividindo ambos os lados por $\gamma(0)$ teremos

$$\rho(1) = \phi_1 + \phi_2 \rho(1),$$

logo,

$$\rho(1) = \frac{\phi_1}{1 - \phi_2}$$

De modo análogo, teremos

$$x_{t-2} x_t = x_{t-2} (\phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t).$$

Aplicando a esperança, teremos

$$\gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0).$$

e dividindo os dois lados por $\gamma(0)$ teremos

$$\rho(2) = \phi_1 \rho(1) + \phi_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

Com os valores de $\rho(1)$ e $\rho(2)$, podemos encontrar $\rho(3)$:

$$\rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1)$$

e assim sucessivamente.

■

Em geral, a função de autocorrelação do processo $AR(p)$ pode ser escrita como

$$\rho(h) = \sum_{j=1}^p c_j \left(\frac{1}{\dot{B}_j} \right)^h$$

onde \dot{B} , novamente, é a raiz de $\phi(B) = 1 - \sum_{j=1}^p \phi_j B^j$.

O comportamento da função de autocorrelação depende das raízes de $\phi(B)$.

- As raízes reais se comportam de acordo com o que já foi visto no $AR(1)$, decaindo exponencialmente.
- As raízes complexas tem um comportamento de onda abafada exponencialmente.

Example .21. Vamos mostrar esse resultado para o caso $p = 2$ (o caso geral é análogo). Teremos que

$$\begin{pmatrix} \rho(h) \\ \rho(h-1) \end{pmatrix} = \underbrace{\begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}}_T \begin{pmatrix} \rho(h-1) \\ \rho(h-2) \end{pmatrix} = \dots T^h \begin{pmatrix} \rho(0) \\ \rho(-1) \end{pmatrix}$$

Já discutimos, na primeira seção, que os autovalores de T são equivalentes aos recíprocos das raízes de $\phi(B)$. Agora, considere a decomposição de T na forma canônica

$$T = Q\Lambda Q^{-1},$$

onde Q é a matriz cujas colunas são formadas pelos autovetores de T e Λ é uma matriz diagonal formada pelos autovalores. Note que

$$A^2 = Q\Lambda Q^{-1}Q\Lambda Q^{-1} = Q\Lambda^2 Q^{-1}$$

e, de modo análogo,

$$A^h = Q\Lambda Q^{-1}Q\Lambda Q^{-1} = Q\Lambda^h Q^{-1}$$

Como Λ^h é a matriz diagonal com os autovalores elevados à potência h , os elementos de A^h devem ser necessariamente do tipo $\sum_{j=1}^2 c_j \lambda_j^h$.

O polinômio característico da matriz T é

$$\lambda^2 - \phi_1 \lambda - \phi_2$$

e, denominando suas raízes por λ_1 e λ_2 , teremos

$$\rho(h) = c_1 \lambda_1^h + c_2 \lambda_2^h.$$

- Se λ_1 é real então λ_2 também é real e o decaimento de $\rho(h)$ será exponencial, podendo alternar o sinal se algum dos autovalores for negativo.

- Se λ_1 é complexo, então λ_2 será seu conjugado. Em coordenadas polares, teremos $\lambda_1 = r e^{i\omega}$ e $\lambda_2 = r e^{-i\omega}$. Deste modo,

$$\begin{aligned} \rho(h) &= r^h [c_1 e^{ih\omega} + c_2 e^{-ih\omega}] \\ &= r^h [(c_1 + c_2) \cos(h\omega) + i(c_1 - c_2) \sin(h\omega)] \end{aligned}$$

logo, $\rho(h)$ terá um comportamento de onda com a amplitude decrescendo exponencialmente. A próxima figura mostra esse efeito.



37.2 A função de autocorrelação parcial (PACF)

Inicialmente, considere a regressão

$$\hat{x}_t = \sum_{j=1}^p \beta_j x_{t-j}.$$

onde β é o valor de minimiza

$$E(x_t - \hat{x}_t)^2 = E \left(x_t - \sum_{j=1}^p \beta_j x_{t-j} \right)^2. \quad (16)$$

Para calcular ϕ_{hh} devemos minimizar

$$E(x_{t+h} - \hat{x}_{t+h})^2 = \gamma(0) + \sum_{j=1}^p \sum_{k=1}^p \beta_j \beta_k \gamma(j-k) - 2 \sum_{j=1}^p \beta_j \gamma(j)$$

e, encontrando o ponto crítico após derivar em β , concluímos que o valor de β que minimiza (Equation ??) satisfaz.

$$\gamma(j) = \sum_{k=1}^p \beta_k \gamma(j-k).$$

Contudo, comparando a equação acima com a identidade dada na Equation ??, temos que $\beta_k = \phi_k$, ou seja,

$$\hat{x}_t = \sum_{j=1}^p \phi_j x_{t-j}.$$

É interessante notarmos que chegaremos ao mesmo resultado de definirmos

$$\hat{x}_t = \sum_{j=1}^p \phi_j x_{t+j}.$$

Vamos alternar entre essas duas formas de acordo com a conveniência.

Definition .7. A função de autocorrelação parcial, denotada por ϕ_{hh} é

$$\phi_{11} = \text{corr}(x_{t+1}, x_t),$$

e, para $h \geq 2$,

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t).$$

Além disso, se x_t é um processo gaussiano, então

$$\phi_{hh} = \text{corr}(x_{t+h}, x_t | x_{t+1}, \dots, x_{t+h-1}).$$

■

Intuitivamente, a função de autocorrelação parcial calcula a correlação entre x_t e x_{t+h} eliminando a dependência linear entre os valores intermediários x_s , $t < s < t + h$.

Example .22. Considere o processo AR(1) estacionário. Como $\gamma(h) = \phi^2$, é imediato, pela definição, que $\phi_{11} = \gamma(1)/\gamma(0) = \phi$.

Para $h \geq 2$, vamos definir $\hat{x}_{t+h} = \phi x_{t+h-1}$ e $\hat{x}_t = \phi x_{t+1}$. Então,

$$\begin{aligned} \phi_{hh} &= \text{Cov}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) \\ &= \text{Cov}(x_{t+h} - \phi x_{t+h-1}, x_t - \phi x_{t+1}) \\ &= \gamma(h) - \phi\gamma(h-1) - \phi\gamma(h-1) + \phi^2\gamma(h-2) \end{aligned}$$

Pela Equation ??,

$$\begin{aligned} \gamma(h) &= \phi\gamma(h-1) \\ \gamma(h-1) &= \phi\gamma(h-2) \end{aligned}$$

logo, $\phi_{hh} = 0$ para todo $h \geq 2$

■

Example .23. Considere o processo AR(2) estacionário. Para $h = 2$, defina

$$\begin{aligned} \hat{x}_{t+2} &= \phi_1 x_{t+1} + \phi_2 x_t \\ \hat{x}_t &= \phi_1 x_{t+1} + \phi_2 x_{t+2} \end{aligned}$$

Então

$$\begin{aligned} \phi_{22} &= \text{Cov}(x_{t+2} - \hat{x}_{t+2}, x_t - \hat{x}_t) \\ &= \text{Cov}(x_{t+2} - \phi_1 x_{t+1} - \phi_2 x_t, x_t - \phi_1 x_{t+1} - \phi_2 x_{t+2}) \\ &= \text{Cov}(\varepsilon_{t+2}, x_t - \phi_1 x_{t+1} - \phi_2 x_{t+2}) \\ &= -\phi_2 \text{Cov}(\varepsilon_{t+2}, x_{t+2}) \end{aligned}$$

Por último, para qualquer $h > 2$,

$$\begin{aligned}
\phi_{hh} &= Cov(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) \\
&= Cov(x_{t+h} - \phi_1 x_{t+h-1} - \phi_2 x_{t+h-2}, x_t - \phi_1 x_{t+1} - \phi_2 x_{t+2}) \\
&= Cov(\varepsilon_{t+2}, x_t - \phi_1 x_{t+1} - \phi_2 x_{t+2}) \\
&= 0
\end{aligned}$$

■

Seja x_t um processo AR(p) estacionário. Faça,

$$\begin{aligned}
\hat{x}_{t+h} &= \sum_{j=1}^p \phi_j x_{t+h-j} \\
\hat{x}_t &= \sum_{j=1}^p \phi_j x_{t+j}
\end{aligned}$$

Então,

$$\begin{aligned}
\phi_{hh} &= Cov(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) \\
&= Cov\left(x_{t+h} - \sum_{j=1}^p \phi_j x_{t+h-j}, x_t - \sum_{j=1}^p \phi_j x_{t-j}\right) \\
&= Cov\left(\varepsilon_{t+h}, x_t - \sum_{j=1}^p \phi_j x_{t-j}\right) \\
&= \sum_{j=1}^p \phi_j Cov(\varepsilon_{t+h}, x_{t-j})
\end{aligned}$$

logo, ϕ_{hh} é nulo sempre que $h > j$.

Definition .8. Gráfico da autocorrelação parcial

O gráfico da função de autocorrelação parcial é construído a partir dos pares (h, ϕ_{hh}) , para $h = 1, 2, \dots$

No exemplo abaixo, uma amostra de tamanho 100 foi gerada de um processo AR(2) estacionário e o gráfico da função de autocorrelação parcial gerado. Note que há apenas duas autocorrelações significativas, levantando evidências de que a ordem do modelo autoregressivo é 2.

```

set.seed(1)
x = c(0,0)
for(i in 3:100) x[i] = x[i-1] -.8*x[i-2]+rnorm(1,0,.1)

pacf(x, xlab = 'Defasagem', ylab = 'Autocorrelação parcial', main = '', lwd = 2)

```

processo_linear_geral_files/figure-pdf/unnamed-chunk-3-1.pdf

37.3 Método de estimação de Yule-Walker

O método de estimação de Yule-Walker para estimar os parâmetros do modelo $AR(p)$ consiste na aplicação do método dos momentos. Sua principal vantagem é a ausência de suposição sobre a distribuição dos erros do modelo, exigindo apenas a condição de estacionaridade.

Para o processo $AR(p)$ estacionário, sabemos que

$$\gamma(h) = \sum_{j=1}^p \phi_j \gamma(h-j).$$

Considere as equações do modelo $AR(p)$:

$$x_t = \sum_{j=1}^p \phi_j x_{t-j} + \varepsilon_t.$$

Podemos multiplicar ambos os lados por x_t (ou seja x_{t-h} , com $h = 0$). Após aplicar a esperança teremos

$$E(x_t^2) = \sum_{j=1}^p \phi_j E(x_{t-j} x_t) + E(\varepsilon_t x_t).$$

Aqui há um detalhe importante:

$$E(\varepsilon_t x_t) = E\left(\varepsilon_t \left[\sum_{j=1}^p \phi_j x_{t-j} + \varepsilon_t\right]\right) = E(\varepsilon_t^2) = \nu$$

Portanto,

$$\gamma(h) = \begin{cases} \sum_{j=1}^p \phi_j \gamma(h-j), & \text{se } h > 0 \\ \sum_{j=1}^p \phi_j \gamma(j) + \nu, & \text{se } h = 0 \end{cases}$$

O método de estimação de Yule-Walker consiste em substituir $\gamma(1), \dots, \gamma(p)$ pelas autocovariâncias amostrais:

$$\hat{\gamma}(1) = \phi_1 \hat{\gamma}(0) + \phi_1 \hat{\gamma}(-1) + \dots + \phi_p \hat{\gamma}(1-p)$$

$$\hat{\gamma}(2) = \phi_1 \hat{\gamma}(1) + \phi_1 \hat{\gamma}(0) + \dots + \phi_p \hat{\gamma}(2-p)$$

$$\vdots = \vdots$$

$$\hat{\gamma}(p) = \phi_1 \hat{\gamma}(p-1) + \phi_1 \hat{\gamma}(p-2) + \dots + \phi_p \hat{\gamma}(0)$$

Nota: lembre-se que $\gamma(-i) = \gamma(i)$. Podemos escrever o sistema na forma matricial:

$$\underbrace{\begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(-1) & \dots & \hat{\gamma}(1-p) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(2-p) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \dots & \hat{\gamma}(0) \end{pmatrix}}_{\hat{\mathbf{\Gamma}}} \underbrace{\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix}}_{\hat{\phi}} = \underbrace{\begin{pmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix}}_{\hat{\gamma}}.$$

e os estimadores de Yule-Walker para os parâmetros autoregressivos são dados por

$$\hat{\phi} = \hat{\mathbf{\Gamma}}^{-1} \hat{\gamma}.$$

Voltando para a Equação (??), temos que

$$\gamma(0) = \sum_{j=1}^p \phi_j \gamma(j) + \nu$$

Portanto, um estimador para ν é dado por

$$\hat{\nu} = \hat{\gamma}(0) - \sum_{j=1}^p \hat{\phi}_j \hat{\gamma}(j)$$

37.4 Estimador de máxima verossimilhança

Observe que

$$x_t | \mathcal{D}_{t-1} \sim x_t | x_{t-1}, \dots, x_{t-p}$$

ou seja, $\{x_t\}$ é uma cadeia de Markov de ordem p . Deste modo, para $n > p$,

$$\begin{aligned} f(x_1, \dots, x_n) &= f(x_n | \mathcal{D}_{n-1}) f(\mathcal{D}_{n-1}) = f(x_n | x_{n-1}, \dots, x_{n-p}) f(\mathcal{D}_{n-1}) \\ &= \prod_{t=p+1}^n f(x_t | x_{t-1}, \dots, x_{t-p}) f(x_1, \dots, x_p). \end{aligned}$$

Agora, suponha que $\varepsilon_t \sim \text{Normal}(0, \nu)$. Então,

$$L(\phi_1, \dots, \phi_p, \nu) = \left(\frac{1}{2\pi\nu} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\nu} \sum_{t=p+1}^n \left(x_t - \sum_{j=1}^p \phi_j x_{t-j} \right)^2 \right\} \\ \times f(x_1, \dots, x_p | \phi_1, \dots, \phi_p, \nu).$$

O método da máxima verossimilhança condicional consiste em ignorar o termo $f(x_1, \dots, x_p | \phi_1, \dots, \phi_p, \nu)$. Com isso, o restante da função de verossimilhança pode ser escrito como um modelo linear, uma vez que

$$x_t = \sum_{j=1}^p \phi_j x_{t-j} + \varepsilon_t = \underbrace{(x_{t-1} \ \cdots \ x_{t-p})}_{F'_t} \underbrace{\begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}}_{\beta}$$

e $\hat{\beta}$ e $\hat{\nu}$ podem ser obtidos pelos estimadores já discutidos na seção sobre modelos lineares.

Já para o método da máxima verossimilhança, precisamos especificar a distribuição do termo adicional $f(x_1, \dots, x_p | \phi_1, \dots, \phi_p)$. Desde que o processo seja estacionário, é possível mostrar que

$$x_1, \dots, x_p \sim \text{Normal}(\mathbf{0}_p, \Sigma(\phi_1, \dots, \phi_p))$$

onde $\Sigma(\phi_1, \dots, \phi_p)$ é a matriz formada pelas autocovariâncias.

A maximização da função de verossimilhança é feita numericamente e pode-se utilizar as estimativas obtidas no método da máxima verossimilhança condicional como ponto de partida para o algoritmo de maximização.

37.5 Previsão

Considere a amostra \mathcal{D}_t do modelo AR(1). Então,

$$\begin{aligned} x_{t+h} &= \phi x_{t+h-1} + \varepsilon_{t+h} \\ &= \phi^2 x_{t+h-2} + \phi \varepsilon_{t+h-1} + \varepsilon_{t+h} = \cdots \\ &= \phi^h x_t + \sum_{j=0}^{h-1} \phi^j \varepsilon_{t+h-j} \end{aligned}$$

logo

$$E(x_{t+h} | \mathcal{D}_t) = \phi^h x_t$$

e

$$Var(x_{t+h}|\mathcal{D}_t) = \nu \sum_{j=0}^{h-1} \phi^{2j} = \nu \frac{1 - \phi^{2h}}{1 - \phi^2}$$

Note que, se o processo é estacionário, então $|\phi| < 1$ e a previsão converge para zero exponencialmente. Na prática, tanto a previsão quanto a variância são calculadas substituindo ϕ por sua estimativa.

Para o processo $AR(p)$, seja $z_t = (x_t, \dots, x_{t-p+1})$. Então,

$$z_t = T z_{t-1} + \varepsilon_t$$

onde

$$T = \left(\begin{array}{ccc|c} \phi_1 & \cdots & \phi_{p-1} & \phi_p \\ \hline \mathbf{I}_{p-1} & & & \mathbf{0}_{p-1} \end{array} \right)$$

e

$$\varepsilon_t = \left(\begin{array}{c} \varepsilon_t \\ \mathbf{0}_{p-1} \end{array} \right)$$

Logo, dado \mathcal{D}_t ,

$$z_{t+h} = T^h z_t + \sum_{j=0}^{h-1} T^j \varepsilon_{t+h-j}$$

Note que $x_{t+h} = (1|\mathbf{0}'_{p-1})z_{t+h}$, logo

$$E(x_{t+h}|\mathcal{D}_t) = (1|\mathbf{0}'_{p-1})T^h z_t$$

e

$$Var(x_{t+h}|\mathcal{D}_t) = \nu \sum_{j=0}^{h-1} \left(1 \mid \mathbf{0}'_{p-1} \right) T^j \mathcal{E} T'^j \left(\frac{1}{\mathbf{0}_{p-1}} \right)$$

onde \mathcal{E} é uma matriz com o valor 1 no elemento a_{11} e zero nos demais.

37.6 Processo autorregressivo com deriva

Até o momento, assumimos que o processo $AR(p)$ é estacionário com $E(x_t) = 0$. Quando $\mu = E(x_t) \neq 0$, temos um parâmetro adicional, denominado deriva e escrito como

$$x_t - \mu = \sum_{j=1}^p \phi_j (x_{t-j} - \mu) + \varepsilon_t$$

Observe que a deriva não altera a estrutura de autocorrelação do processo, uma vez que podemos definir $y_t = x_t - \mu$, onde y_t é um $AR(p)$ com $E(y_t) = 0$, o que implica em

$$\gamma(h) = Cov(x_t, x_{t-1}) = Cov(x_t - \mu, x_{t-h} - \mu) = Cov(y_t, y_{t-h}).$$

Contudo, a função de previsão do processo deve ser ajustada. Defina $z_t = (x_t - \mu, \dots, x_{t-p+1} - \mu)$. Já vimos que

$$z_{t+h} = T^h z_t + \sum_{j=0}^{h-1} T^j \varepsilon_{t+h-j}$$

onde T e ε_t foram definidos na seção anterior. Note que $x_{t+h} - \mu = (1|\mathbf{0}'_{p-1})z_{t+h}$, logo

$$E(x_{t+h}|\mathcal{D}_t) = \mu + (1|\mathbf{0}'_{p-1})T^h z_t$$

Em termos de estimação, a função de verossimilhança deve ser alterada para

$$L(\mu, \phi_1, \dots, \phi_p, \nu) = \left(\frac{1}{2\pi\nu} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\nu} \sum_{t=p+1}^n (x_t - \mu - \sum_{j=1}^p \phi_j (x_{t-j} - \mu))^2 \right\} \\ \times f(x_1, \dots, x_p | \mu, \phi_1, \dots, \phi_p, \nu).$$

onde

$$x_1, \dots, x_p \sim \text{Normal}(\mu \mathbf{1}_p, \Sigma(\phi_1, \dots, \phi_h))$$

37.7 Exemplo: número anual de terremotos

A série abaixo apresenta o número anual de terremotos de magnitude maior ou igual à 7 na escala Richter.

```
require(gsheet)
```

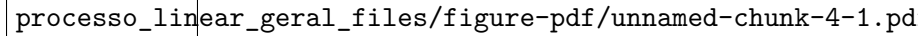
Carregando pacotes exigidos: gsheets

Warning: package 'gsheet' was built under R version 4.3.2

```
url <- 'https://docs.google.com/spreadsheets/d/1PPf1n0jwh1fnr1VtFW2DKN6PY9s9TXEQtn6ZX3igCB

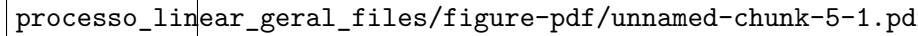
dados <- gsheets2tbl(url)
terr <- unlist(dados[,2])
terr <- ts(terr, start = 1900)

ts.plot(terr, ylab = 'No terremotos', xlab = 'Ano', lwd = 2)
```

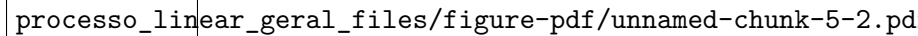


Note que o processo parece ser estacionário. Vamos explorar as funções de autocorrelação nas duas figuras abaixo. O correlograma mostra um comportamento tendendo para zero após 5 defasagens enquanto que o gráfico da função de autocorrelação parcial apresenta um único valor significativo. Temos evidências o modelo AR(1) pode ser adequado.

```
acf(terr)
```



```
pacf(terr)
```



A função `arima(x, order = c(p,0,0))` estima os parâmetros do modelo $AR(p)$ já considerando a deriva (caso contrário, utilize o argumento `include.mean=FALSE`). Abaixo apresentamos o modelo ajustado.

```
mod <- arima(terr, c(1,0,0))
mod
```

Call:
`arima(x = terr, order = c(1, 0, 0))`

Coefficients:

| | ar1 | intercept |
|------|--------|-----------|
| | 0.5433 | 19.8907 |
| s.e. | 0.0840 | 1.3180 |

sigma^2 estimated as 36.7: log likelihood = -318.98, aic = 643.97

Podemos verificar se o ajuste é adequado analisando os resíduos. Note que agora possuímos uma ferramenta nova, o gráfico da função de autocorrelação parcial. Abaixo, os gráficos das funções de autocorrelação não apresentam valores significativos. O teste de Box-Pierce não rejeita a hipótese de um processo estacionário e o teste de Shapiro-Wilks não rejeita a hipótese de ruído branco gaussiano. Portanto, temos evidências de que o modelo AR(1) com erros gaussianos é adequado para essa série.

```
acf(mod$residuals)
```

processo_linear_geral_files/figure-pdf/unnamed-chunk-7-1.pdf

```
pacf(mod$residuals)
```

processo_linear_geral_files/figure-pdf/unnamed-chunk-7-2.pdf

```
Box.test(mod$residuals)
```

Box-Pierce test

```
data: mod$residuals  
X-squared = 0.92715, df = 1, p-value = 0.3356
```

```
shapiro.test(mod$residuals)
```

Shapiro-Wilk normality test

```
data: mod$residuals  
W = 0.97813, p-value = 0.09826
```

As estimativas encontradas foram $\hat{\phi} = 0,5$, $\hat{\nu} = 36,7$ e $\hat{\mu} = 19$. Podemos utilizar a função `forecast` do pacote de mesmo nome para fazer previsões. Note que a medida que o horizonte de previsão cresce, o modelo converge para a deriva.

```
require(forecast)
```

Carregando pacotes exigidos: forecast

Warning: package 'forecast' was built under R version 4.3.1

Registered S3 method overwritten by 'quantmod':

```
method          from  
as.zoo.data.frame zoo
```

```
prev <- forecast(mod, 10)  
plot(prev)
```

processo_linear_geral_files/figure-pdf/unnamed-chunk-8-1.pdf

38 Médias móveis

O processo

$$x_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j},$$

onde $\{\varepsilon_t\}$ é um ruído branco, é denominado média móvel de ordem q (onde MA é a sigla de *moving average*). Note que esse modelo não está relacionado com o método de suavização de mesmo nome. Também podemos definir o modelo utilizando um polinômio de defasagens:

$$x_t = \left(1 + \sum_{j=1}^q \theta_j B^j \right) \varepsilon_t = \theta(B) \varepsilon_t$$

O processo $MA(q)$ é estacionário. De fato

$$E(x_t) = E \left(\varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \right) = 0,$$

$$Var(x_t) = \nu \left(1 + \sum_{j=1}^q \theta_j^2 \right).$$

Agora, defina $\psi_0 = 1$, $\psi_j = \theta_j$, para $j = 1, \dots, q$ e $\psi_j = 0$ para $j > q$. Então,

$$\begin{aligned} Cov(x_t, x_{t-h}) &= \sum_{j=0}^q \sum_{k=0}^q \psi_j \psi_k Cov(\varepsilon_{t-j}, \varepsilon_{t-k-h}) \\ &= \nu \sum_{k=0}^{q-h} \psi_{k+h} \psi_k \\ &= \nu \left[\theta_h + \sum_{j=1}^{q-h} \theta_j \theta_{j+h} \right] \end{aligned}$$

A função de autocorrelação desse processo é dada por

$$\rho(h) = \begin{cases} \frac{\theta_h + \sum_{j=1}^{q-h} \theta_j \theta_{j+h}}{1 + \sum_{j=1}^q \theta_j^2}, & 0 < h \leq q \\ 0, & \text{caso contrário} \end{cases}$$

Portanto, o correlograma pode ser utilizado para determinar o valor que q , uma vez que são esperadas q autocorrelações não nulas.

38.1 Inversibilidade do processo de média móvel

Nem sempre o processo $MA(q)$ é inversível. De fato, note que

$$x_t = \varepsilon_t + \theta \varepsilon_{t-1} = (1 + \theta B) \varepsilon_t \Rightarrow x_t (1 + \theta B)^{-1} = \varepsilon_t.$$

Sabemos que

$$\theta^{-1}(B) = \sum_{j=0}^{\infty} \theta^j B^j,$$

quando $|\theta| < 1$. Portanto, para este processo ser inversível é necessário que $|\theta| < 1$. Observe que

$$\theta^{-1}(B)x_t = \varepsilon_t \Rightarrow x_t = \sum_{j=1}^{\infty} \theta^j x_{t-j} + \varepsilon_t,$$

logo, a inversa do processo MA(1) é equivalente ao modelo AR(∞).

Proposition .2. *O processo MA(q) é inversível se e somente se o módulo das raízes do polinômio $\theta(B)$ forem maiores que um. ■*

38.2 Autocorrelação parcial do MA(q)

Considere que o processo MA(q) inversível. Então, sabemos que

$$x_t = \sum_{j=0}^{\infty} \psi_j x_{t-j} + \varepsilon_t$$

existe e é estacionário. Defina

$$\begin{aligned} \hat{x}_{t+h} &= \sum_{j=1}^p \phi_j x_{t+h-j} \\ \hat{x}_t &= \sum_{j=1}^p \phi_j x_{t+j} \end{aligned}$$

Então,

$$\begin{aligned} \phi_{hh} &= Cov(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) \\ &= Cov\left(x_{t+h} - \sum_{j=1}^{\infty} \phi_j x_{t+h-j}, x_t - \sum_{j=1}^{\infty} \phi_j x_{t+j}\right) \\ &= Cov\left(\varepsilon_{t+h}, x_t - \sum_{j=1}^{\infty} \phi_j x_{t+j}\right) \\ &= - \sum_{j=h}^{\infty} \psi_j Cov(\varepsilon_{t+h}, x_{t+j}) \\ &= - \sum_{j=h}^{\infty} \psi_j Cov(\varepsilon_{t+h}, x_{t+j}) \end{aligned}$$

e, como o processo é inversível, ϕ_{hh} deve decair exponencialmente para zero.

Considere o processo MA(1) inversível. É possível mostrar que

$$\text{Cov}(\varepsilon_t, x_{t+j}) = \nu \theta^j$$

Sabemos que

$$x_t = \sum_{j=1}^{\infty} \psi_j x_{t-j} + \varepsilon_t,$$

onde $\psi_j = \theta^j$. Então,

$$\phi_{hh} = - \sum_{j=h}^{\infty} \theta^j \text{Cov}(\varepsilon_{t+h}, x_{t+j}) = - \sum_{j=h}^{\infty} \theta^{2j} \nu$$

38.3 Estimação dos parâmetros do MA(q)

Sabemos que

$$f(x_1, \dots, x_n) = f(x_1) \prod_{t=2}^n f(x_t | x_1, \dots, x_{t-1}).$$

Considere inicialmente o modelo MA(1) e considere que ε_0 é conhecido. Então, $x_1 = \varepsilon_1 + \theta \varepsilon_0 \sim \text{Normal}(\theta \varepsilon_0, \nu)$.

Além disso, após observar x_1 , podemos escrever

$$\hat{\varepsilon}_1 = x_1 - \theta \varepsilon_0.$$

Como

$$x_2 | \mathcal{D}_1 \sim (\varepsilon_2 + \theta \varepsilon_1 | \varepsilon_0, x_1) \sim \varepsilon_1 + \theta \hat{\varepsilon}_1 \sim \text{Normal}(\theta \hat{\varepsilon}_1, \nu).$$

De um modo geral, fazendo

$$\hat{\varepsilon}_t = x_t - \theta \hat{\varepsilon}_{t-1}$$

teremos

$$(x_t | \mathcal{D}_{t-1}) \sim \text{Normal}(\theta \hat{\varepsilon}_{t-1}, \nu)$$

e podemos escrever a verossimilhança,

$$L(\theta, \nu, \varepsilon_0) = \left(\frac{1}{2\pi\nu} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\nu} \left[(x_1 - \theta \varepsilon_0)^2 + \sum_{t=2}^n (x_t - \theta \hat{\varepsilon}_{t-1})^2 \right] \right\}$$

Para um modelo $MA(q)$, supomos que $\varepsilon_{-q+1}, \dots, \varepsilon_0$ são conhecidos. Fazendo

$$\hat{\varepsilon}_t = x_t - \sum_{j=1}^q \theta_j \hat{\varepsilon}_{t-j}, \quad (17)$$

e

$$\lambda'_t = -(\hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q})$$

teremos que

$$x_t | \mathcal{D}_{t-1} \sim \text{Normal}(\lambda'_t \boldsymbol{\theta}, \nu)$$

onde $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_q)$. Logo,

$$L(\boldsymbol{\theta}, \nu, \varepsilon_{-q+1}, \dots, \varepsilon_0) \propto \left(\frac{1}{\nu}\right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2\nu} \sum_{t=1}^n (y_t - \lambda'_t \boldsymbol{\theta})^2 \right\} \quad (18)$$

Em relação aos valores $\varepsilon_{-q+1}, \dots, \varepsilon_0$, temos duas estratégias de otimização: - Fazer $\varepsilon_{-q+1} = \dots = \varepsilon_0 = 0$: isto é equivalente a dizer que todos estes valores são iguais à sua média. - Estimar $\varepsilon_{-q+1}, \dots, \varepsilon_0$: isto altera o número de parâmetros para $2q + 1$. Neste caso, a estratégia anterior pode ser utilizada como valores iniciais para o otimizador.

39 Modelo autorregressivo com médias móveis

O modelo autorregressivo com médias móveis (ARMA) é dado por

$$x_t = \sum_{j=1}^p \phi_j x_{t-1} + \varepsilon_t + \sum_{k=1}^q \theta_k \varepsilon_{t-k}, \quad (19)$$

onde (p, q) são as ordens da parte autorregressiva e das médias móveis, respectivamente. Definimos $ARMA(0, q) = MA(q)$ e, de modo análogo, definimos $ARMA(p, 0) = AR(p)$.

Utilizando o operador defasagem teremos

$$\begin{aligned} x_t &= \sum_{j=1}^p \phi_j x_{t-1} + \varepsilon_t + \sum_{k=1}^q \theta_k \varepsilon_{t-k} \\ &= \sum_{j=1}^p \phi_j B^j x_t + \varepsilon_t + \sum_{k=1}^q \theta_k B^k \varepsilon_t \end{aligned}$$

o que implica em

$$\underbrace{\left(1 - \sum_{j=1}^p \phi_j B^j\right)}_{\phi(B)} x_t = \underbrace{\left(1 - \sum_{j=1}^p \theta_j B^j\right)}_{\theta(B)} \varepsilon_t$$

O modelo ARMA será estacionário se o módulo das raízes de $\phi(B)$ forem maiores que um e será inversível se o mesmo ocorrer com o módulo das raízes de $\theta(B)$.

39.1 Função de autocorrelação

A função de auto covariância cruzada é dada por

$$\gamma_{x\varepsilon}(h) = E(\varepsilon_t x_{t-h}). \quad (20)$$

Note que x_{t-h} depende apenas dos ruídos que ocorreram até o tempo $t - h$. Portanto

$$\begin{aligned} \gamma_{x\varepsilon}(h) &= 0, \quad h > 0 \\ \gamma_{x\varepsilon}(h) &\neq 0, \quad h \leq 0 \end{aligned}$$

Considere o processo ARMA(1,1), dado por

$$x_t = \varepsilon_t + \phi x_{t-1} + \theta \varepsilon_{t-1}.$$

Teremos que

$$\begin{aligned} \gamma_{x\varepsilon}(0) &= E(x_t \varepsilon_t) = E([\varepsilon_t + \phi x_{t-1} + \theta \varepsilon_{t-1}] \varepsilon_t) \\ &= \underbrace{E(\varepsilon_t^2)}_{\nu} + \phi \underbrace{E(x_{t-1} \varepsilon_t)}_{\gamma_{x\varepsilon}(1)=0} + \theta \underbrace{E(\varepsilon_t \varepsilon_{t-1})}_{=0} \\ &= \nu \end{aligned}$$

Agora, multiplicando a equação do modelo ARMA(1,1) em ambos os lados por x_{t-h} , e aplicando a esperança teremos

$$\underbrace{E(x_{t-h} x_t)}_{\gamma(h)} = \underbrace{E(x_{t-h} \varepsilon_t)}_{\gamma_{x\varepsilon}(h)} + \phi \underbrace{E(x_{t-1} x_{t-h})}_{\gamma(h-1)} + \theta \underbrace{E(x_{t-h} \varepsilon_{t-1})}_{\gamma_{x\varepsilon}(h-1)}. \quad (21)$$

Fazendo $h = 0$ na equação acima, teremos

$$\gamma(0) = \phi \gamma(1) + \theta \gamma_{x\varepsilon}(-1)$$

e, como

$$\gamma_{x\varepsilon}(-1) = E(x_{t+1}\varepsilon_t) =$$

Fazendo $h = 1$ teremos

$$\gamma(1) = \phi\gamma(0) + \theta\nu$$

Para qualquer $h \geq 2$,

$$\gamma(h) = \phi\gamma(h-1),$$

ou ainda

$$\gamma(h) = \phi^{h-1}\gamma(1) = \phi^h\gamma(0) + \theta\nu.$$

Assim,

$$\rho(h) = \{$$

Note que a primeira auto covariância depende do parâmetro de média móvel.

Note ainda que após a primeira defasagem, a função de auto covariância se comporta como um

\end{frame}

Para o caso geral:

- As q primeiras autocorrelações dependerão dos parâmetros ϕ_1, \dots, ϕ_p e $\theta_1, \dots, \theta_q$.
- Se $q < p$, toda a função de autocorrelação consistirá de uma mistura de exponenciais ou senos amortecidos, dependendo das raízes do polinômio $\phi(B)$.
- Se $q \geq p$, então as $q - p + 1$ primeiras autocorrelações não seguirão o padrão descrito acima.

Função de autocorrelação parcial Se comporta como um processo $MA(q)$.

Verossimilhança

- Dados, y_0, \dots, y_{-p+1} e $\hat{\varepsilon}_0, \dots, \hat{\varepsilon}_{-q+1}$, temos que

$$y_t = (y_{t-1}, \dots, y_{t-p})'\phi - (\hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q})'\theta + \varepsilon_t$$

- A construção da verossimilhança pode ser feita de modo análogo ao modelo $MA(q)$.
- Considere o modelo $ARMA(1, 1)$ e que estão fixados y_0 e $\varepsilon_0 = \hat{\varepsilon}_0$.

- $t = 1$, teremos

$$\begin{aligned} y_1|y_0, \hat{\varepsilon}_0 &\sim \phi y_0 + \varepsilon_1 - \theta \varepsilon_0 | y_0, \hat{\varepsilon}_0 \\ &\sim \phi y_0 + \varepsilon_1 - \theta \hat{\varepsilon}_0 | y_0, \hat{\varepsilon}_0 \\ &\sim \text{Normal}(\phi y_0 - \theta \hat{\varepsilon}_0, \nu) \end{aligned}$$

Note ainda que, dados y_0 e $\hat{\varepsilon}_0$,

$$y_1 = \phi y_0 + \varepsilon_1 - \theta \hat{\varepsilon}_0 \Rightarrow \hat{\varepsilon}_1 = y_1 - \phi y_0 + \theta \hat{\varepsilon}_0$$

itemize

Para $t = 2$,

$$\begin{aligned} y_2|y_1, y_0, \hat{\varepsilon}_0 &\sim \phi y_1 + \varepsilon_2 - \theta \varepsilon_1 | y_1, y_0, \hat{\varepsilon}_0 \\ &\sim \phi y_1 + \varepsilon_2 - \theta \hat{\varepsilon}_1 | y_0, y_1, \hat{\varepsilon}_0 \\ &\sim \text{Normal}(\phi y_1 - \theta \hat{\varepsilon}_1, \nu) \end{aligned}$$

Novamente, dados y_1, y_2, y_0 e $\hat{\varepsilon}_0$, teremos

$$\hat{\varepsilon}_2 = y_2 - \phi y_1 + \theta \hat{\varepsilon}_1.$$

De modo geral, para qualquer $t \geq 1$,

$$\hat{\varepsilon}_t = y_t - \phi y_{t-1} + \theta \hat{\varepsilon}_{t-1}$$

e

$$(y_t | y_{0:t-1}, \hat{\varepsilon}_0) \sim \text{Normal}(\phi y_{t-1} - \theta \hat{\varepsilon}_{t-1}).$$

Verossimilhança para o modelo $ARMA(p, q)$ Fixe $y_{0:-p+1}$ e $\hat{\varepsilon}_{0:-q+1}$. Para $t \geq 1$, faça,

$$f'_t = (y_{t-1}, \dots, y_{t-p}, -\hat{\varepsilon}_{t-1}, \dots, -\hat{\varepsilon}_{t-q})' \quad (22)$$

$$\beta' = (\phi', \theta') \quad (23)$$

onde

$$\hat{\varepsilon}_t = y_t - f'_t \beta$$

teremos o modelo (não linear)

$$L(\beta, \nu) \propto \left(\frac{1}{\nu}\right)^{\frac{T}{2}} \exp \left\{ -\frac{1}{2\nu} (y - F'\beta)' (y - F'\beta) \right\}$$

40 Modelos integrados

40.1 O operador diferença

Vimos no começo do curso que o passeio aleatório, dado por

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

onde ε_t é um ruído branco, poderia gerar tendências aleatórias. Note que o passeio aleatório nada mais é que o processo AR(1) não estacionário. Podemos escrever esse modelo como

$$x_t - x_{t-1} = \varepsilon_t \Rightarrow (1 - B)x_t = \varepsilon_t \Rightarrow \Delta x_t = \varepsilon_t$$

Acima, $\Delta = 1 - B$ é denominado operador diferença. Observe que a série

$$y_t = \Delta x_t = \varepsilon_t$$

é um ruído branco. Desse modo, o operador elimina a tendência aleatória gerada pelo passeio.

Agora, considere que

$$x_t = T(t) + \varepsilon_t,$$

onde $T(t)$ é uma tendência localmente linear com inclinação constante, ou seja, para qualquer s na vizinhança de t ,

$$T(s) = \mu_t + bs.$$

Então

$$\Delta x_t = \mu_t + bt - (\mu_t + b(t-1)) + \varepsilon_t = b + \varepsilon_t$$

e $y_t = \Delta x_t = b + \varepsilon_t$ será um ruído branco com deriva.

Por último, para qualquer s na vizinhança de t , suponha que

$$T(s) = \mu_t + b_t s.$$

Nesse caso,

$$\Delta x_t = b_t + \varepsilon_t$$

logo, o processo $y_t = x_t$ não é um ruído branco. Contudo, como é usual que $b_t \approx b_{t+1}$, teremos

$$\Delta y_t = \Delta x_t - \Delta x_{t-1} = b_t - b_{t-1} + \varepsilon_t - \varepsilon_{t-1} \approx \varepsilon_t - \varepsilon_{t-1}$$

logo, aplicar o operador diferença pela segunda vez gera um ruído branco. Note que

$$\Delta y_t = \Delta(x_t - x_{t-1}) = \Delta(1 - B)x_t = \Delta^2 x_t.$$

Então, aplicar o operador diferença algumas vezes pode eliminar a tendência da série, gerando um ruído branco.

Considere o modelo

$$y_t = \Delta x_t = \varepsilon_t,$$

onde ε_t é um ruído branco. Então, para $\mathcal{D}_t = \{x_1, \dots, x_t\}$,

$$E(y_{t+h}|\mathcal{D}_t) = 0 \Rightarrow E(x_{t+h}|\mathcal{D}_t) = E(x_{t+h-1}|\mathcal{D}_t)$$

e, como $E(x_t|\mathcal{D}_t) = x_t$, teremos

$$E(x_{t+h}|\mathcal{D}_t) = x_t.$$

Portanto, esse modelo tem como previsão para qualquer horizonte o último valor observado, o que é equivalente ao modelo de suavização exponencial com $\alpha = 1$.

Considere agora o modelo

$$y_t = \Delta^2 x_t = \varepsilon_t,$$

onde ε_t é um ruído branco. Observe que

$$E(y_{t+1}|\mathcal{D}_t) = 0 \Rightarrow E(x_{t+1}|\mathcal{D}_t) = x_t + (x_t - x_{t-1}).$$

É simples mostrar por indução que

$$E(x_{t+h}|\mathcal{D}_t) = x_t + h(x_t - x_{t-1}),$$

que é equivalente ao método de Holt com nível $m_t = x_t$ e inclinação $b_t = (x_t - x_{t-1})$.

■.

40.2 Modelo autorregressivo integrado

Considere agora uma série temporal x_t na qual foram necessárias d diferenças para eliminar a tendência. Contudo, a série $y_t = \Delta^d x_t$ resultante é um ruído branco, mas sim um processo AR(p). Teremos

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t \Rightarrow \left(1 - \sum_{j=1}^p \phi_j B^j\right) y_t = \varepsilon_t$$

ou, de modo equivalente,

$$\left(1 - \sum_{j=1}^p \phi_j B^j\right) \Delta^d x_t = \varepsilon_t.$$

O modelo acima é denominado autorregressivo integrado, e sua notação é ARI(p, d), onde p é a ordem do modelo autorregressivo resultante das d diferenças da série original.

As estimativas são obtidas criando a série $y_t = \Delta^d x_t$ e estimando os parâmetros do modelo $\text{AR}(p)$ resultante.

A previsão do processo $\text{ARI}(p, d)$ pode facilmente ser realizada a partir do modelo $\text{AR}(p)$. Por exemplo, para $y_t = \Delta x_t$,

$$E(y_{t+1}|\mathcal{D}_t) = \sum_{j=1}^p \phi_j y_{t-j}$$

o que implica em

$$E(x_{t+1}|\mathcal{D}_t) = x_t + \sum_{j=1}^p \phi_j (x_{t-j} - x_{t-j-1})$$

Example .24. Previsão para o modelo $\text{ARI}(1,1)$

Seja x_t uma série temporal segundo o processo $\text{ARI}(1,1)$. Então, Δy_t é um processo $\text{AR}(1)$ e $E(y_{t+h}|\mathcal{D}_t) = \phi^h y_t$. Disto, notemos que

$$E(x_{t+1} - x_t|\mathcal{D}_t) = \phi(x_t - x_{t-1}) \Rightarrow E(x_{t+1}|\mathcal{D}_t) = x_t + \phi(x_t - x_{t-1}),$$

e

$$\begin{aligned} E(x_{t+2} - x_{t+1}|\mathcal{D}_t) &= \phi^2(x_t - x_{t-1}) \\ \Rightarrow E(x_{t+2}|\mathcal{D}_t) &= \phi^2(x_t - x_{t-1}) + E(x_{t+1}|\mathcal{D}_t) \\ \Rightarrow E(x_{t+2}|\mathcal{D}_t) &= \phi^2(x_t - x_{t-1}) + \phi(x_t - x_{t-1}) + x_t. \end{aligned}$$

Por indução, é imediato que

$$E(x_{t+h}|\mathcal{D}_t) = x_t + \sum_{j=1}^h \phi^j (x_t - x_{t-1}).$$

Identificando x_t como a estimativa para nível da série no tempo t e $(x_t - x_{t-1})$ como a inclinação da tendência, temos que a função de previsão é um modelo com tendência amortecida.

■

40.3 Modelo autorregressivo sazonal

Seja

$$x_t = g(t) + \varepsilon_t,$$

onde $g(\cdot)$ é uma função periódica com período p e ε_t um ruído branco. Note que

$$x_t - x_{t-p} = g(t) - g(t-p) + \varepsilon_t - \varepsilon_{t-p} = \varepsilon_t - \varepsilon_{t-p},$$

logo, a diferença acima gerou uma série estacionária. Observe que

$$x_t - x_{t-p} = (1 - B^p)x_t.$$

O operador $1 - B^p$ é conhecido como diferença sazonal.

Sem perda de generalidade, assumamos que $D = 3$. Então, a relação entre x_t e x_{t-3} pode ser obtida usando a seguinte transformação:

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}^2 \begin{pmatrix} x_{t-4} \\ x_{t-5} \\ x_{t-6} \end{pmatrix}$$

A matriz acima é denominada permutação, e já foi utilizada anteriormente. Para o caso geral,

$$\begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}^h \begin{pmatrix} x_{t-(3h-2)} \\ x_{t-(3h-1)} \\ x_{t-3h} \end{pmatrix}$$

ou ainda,

$$x_t = \sum_{j=1}^3 c_j \lambda_j^h x_{t-(3h-j+1)},$$

onde λ_j são os autovalores da matriz acima. Acontece que

$$\left| \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} - \lambda \mathbf{I}_3 \right| = 0 \Rightarrow \lambda^3 = 1$$

Teremos que $\lambda = 1$ é a solução real, enquanto que

$$\lambda = \exp \left\{ \frac{\pi}{3} i \right\} = \cos \left(\frac{2\pi}{3} \right) + i \sin \left(\frac{2\pi}{3} \right).$$

e

$$\lambda = \exp \left\{ \frac{2\pi}{3} i \right\} = \cos \left(\frac{4\pi}{3} \right) + i \sin \left(\frac{4\pi}{3} \right).$$

são as soluções complexas.

Na prática, para qualquer período p , o autovalores existirão $p - 1$ autovalores da matriz de permutação satisfazendo

$$\lambda = \cos(2\pi k/p) + i \sin(2\pi k/p),$$

com $k = 1, \dots, D - 1$.

Dizemos que x_t é um processo autorregressivo sazonal se $y_t = (1 - B^p)^D x_t$ é um processo autorregressivo.

Assim como no caso do modelo integrado, os parâmetros são estimados a partir do modelo $AR(P)$ considerando a amostra $y_t = (1 - B^p)^D x_t$.

A previsão é

Example .25. Seja x_t um processo autorregressivo sazonal, de período 12. Então,

$$(1 - B^{12})x_t = \Phi(1 - B^{12})x_{t-1} + \varepsilon_t,$$

ou ainda,

$$x_t = x_{t-12} + \Phi x_{t-1} - \Phi x_{t-13} + \varepsilon_t$$

41 O modelo ARMA

O modelo auto regressivo de média móvel é dado por

$$x_t = \sum_{j=1}^p \phi_j x_{t-1} + \sum_{k=1}^q \theta_k \varepsilon_{t-k} + \varepsilon_t,$$

onde (p, q) são as ordens da parte autorregressiva e da média móvel, respectivamente. Definimos o modelo de média móveis como $ARMA(0, q)$ e o autorregressivo como $AR(p)$.

Utilizando o operador defasagem teremos

$$\begin{aligned} x_t - \sum_{j=1}^p \phi_j x_{t-1} &= \varepsilon_t + \sum_{k=1}^q \theta_k \varepsilon_{t-k} \\ \Rightarrow \underbrace{\left(1 - \sum_{j=1}^p \phi_j B^j\right)}_{\phi(B)} x_t &= \underbrace{\left(1 + \sum_{k=1}^q \theta_k B^k\right)}_{\theta(B)} \varepsilon_t. \end{aligned}$$

O modelo ARMA será estacionário se o módulo das raízes de $\phi(B)$ forem maiores que um e será inversível se as raízes de $\theta(B)$ também o forem.

41.1 A função de autocorrelação

A função de autocovariância cruzada é dada por

$$\gamma_{x\varepsilon}(h) = E(\varepsilon_t x_{t-h}). \quad (24)$$

Note que x_{t-h} depende apenas dos ruídos que ocorreram até o tempo $t - h$. Portanto

$$\begin{aligned} \gamma_{x\varepsilon}(h) &= 0, \quad h > 0 \\ \gamma_{x\varepsilon}(h) &\neq 0, \quad h \leq 0 \end{aligned}$$

Example .26. Considere o processo ARMA(1,1), dado por

$$x_t = \varepsilon_t + \phi x_{t-1} + \theta \varepsilon_{t-1}.$$

Teremos que

$$\begin{aligned} \gamma_{x\varepsilon}(0) &= E(x_t \varepsilon_t) = E([\varepsilon_t + \phi x_{t-1} - \theta \varepsilon_{t-1}] \varepsilon_t) \\ &= \underbrace{E(\varepsilon_t^2)}_{\nu} + \phi \underbrace{E(x_{t-1} \varepsilon_t)}_{\gamma_{x\varepsilon}(1)=0} - \theta \underbrace{E(\varepsilon_t \varepsilon_{t-1})}_{=0} \\ &= \nu \end{aligned}$$

■

Considere o processo ARMA(1,1), dado por

$$x_t = \varepsilon_t + \phi x_{t-1} + \theta \varepsilon_{t-1}.$$

Multiplicando ambos os lados por x_{t-h} , e aplicando a esperança teremos

$$\underbrace{E(x_{t-h} x_t)}_{\gamma(h)} = \underbrace{E(x_{t-h} \varepsilon_t)}_{\gamma_{x\varepsilon}(h)} + \phi \underbrace{E(x_{t-1} x_{t-h})}_{\gamma(h-1)} + \theta \underbrace{E(x_{t-h} \varepsilon_{t-1})}_{\gamma_{x\varepsilon}(h-1)}$$

Fazendo $h = 0$ na equação acima teremos

$$\gamma(0) = \phi \gamma(1) + \theta \gamma_{x\varepsilon}(-1).$$

Fazendo $h = 1$ teremos

$$\gamma(1) = \phi \gamma(0) + \theta \nu.$$

Para qualquer $h \geq 2$,

$$\gamma(h) = \phi\gamma(h-1),$$

ou ainda

$$\gamma(h) = \phi^{h-1}\gamma(1),$$

e

$$\rho(h) = \phi^h + \nu\theta\phi^{h-1}\frac{\nu}{\gamma(0)}.$$

Note que a função de autocovariância se comporta como um modelo AR(1) tradicional, com um decaimento exponencial (podendo ser alternado ou não).

Para o caso geral: - As q primeiras autocorrelações dependerão dos parâmetros ϕ_1, \dots, ϕ_p e $\theta_1, \dots, \theta_q$. - Se $q < p$, toda a função de autocorrelação consistirá de uma mistura de exponenciais ou cossenos amortecidos, dependendo das raízes do polinômio $\phi(B)$. - Se $q \geq p$, então as $q - p + 1$ primeiras autocorrelações não seguirão o padrão descrito acima.

Para construir a função de verossimilhança, observe que

$$y_t = \phi(B)x_t = \theta(B)\varepsilon_t,$$

ou seja, y_1, \dots, y_n é um modelo de média móvel e sua função de verossimilhança pode ser construída desse a partir desse modelo.

41.2 Modelos ARMA integrados e sazonais

Dizemos que x_t é um processo autorregressivo integrado de médias móveis se

$$\phi(B)\Delta^d x_t = \theta(B)\varepsilon_t,$$

e denotamos esse modelo por ARIMA(p, d, q). A previsão desse modelo é dada por

$$\hat{y}_t(h) = E(y_{t+h}|y_t, \dots, y_{t-(p+d)}).$$

Seja x_t uma série temporal sazonal com período s . Considere que o processo x_t, x_{t-s}, x_{t-2s} é um ARIMA(P, D, Q):

$$\Phi(B)(1 - B^P)^D x_t = \Theta(B)\eta_t,$$

e que o processo $\eta_t \sim \text{ARIMA}(p, d, q)$. Então,

$$\theta(B)(1 - B)^d \eta_t = \theta(B)\varepsilon_t \Rightarrow \eta_t = [\theta(B)(1 - B)^d]^{-1} \theta(b)\varepsilon_t.$$

Portanto, teremos o modelo

$$\theta(B)\Phi(B)(1 - B^P)^D(1 - B)^d x_t = \Theta(B)\theta(B)\eta_t,$$

O modelo SARIMA é representado por ARIMA\$(p,d,q)(P,D,Q)_s\$ onde \$(p,d,q)\$ é a especificação

Intervenção e outliers

42 Análise de intervenção

Considere o problema no qual a série temporal x_t recebe apenas uma intervenção, no tempo T . O modelo geral é dado por

$$x_t = m_t + N_t,$$

onde o processo N_t representa a série temporal sem a intervenção e m_t representa o efeito da intervenção. Ante do tempo T , m_t é assumida ser nula. A série $\{x_t, t < T\}$ é referida como dados pré-intervenção e a série $\{x_t, t \geq T\}$ é referida como pós-intervenção.

A função degrau é definida como

$$S_t^{(T)} = \begin{cases} 1, & \text{se } t \geq T \\ 0, & \text{caso contrário} \end{cases}$$

ou seja, $S_t^{(T)} = 0$ durante a pré-intervenção e $S_t^{(T)} = 1$ durante a pós-intervenção. A função pulso, definida por

$$P_t^{(T)} = S_t^{(T)} - S_{t-1}^{(T)}$$

é a variável indicadora do tempo da intervenção. Como

$$P_t^{(T)} = (1 - B)S_t^{(T)}$$

podemos escrever o pulso em função de $S_t^{(T)}$.

Quando a intervenção possui um resultado imediato e permanente na função média, a mudança pode ser modelada por

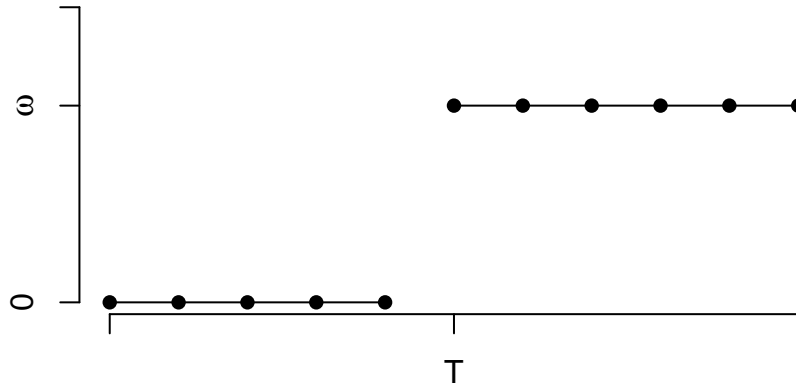
$$m_t = \omega S_t^{(T)},$$

onde ω é o parâmetro que modela da mudança permanente na média. Em termos de função pulso, teremos

$$m_t = m_{t-1} + \omega P_t^{(T)}.$$

A figura abaixo ilustra esse efeito

```
plot.new()
plot.window(xlim=c(0,11), ylim=c(0,1.5))
axis(1, at = c(0,5,10), c('','T',''))
axis(2, at = c(0,1,1.5), c(0,expression(omega),''))
lines(0:4,rep(0,5), type='o', pch=16)
lines(5:10,rep(1,6), type='o', pch=16)
```



Também é possível que a mudança na função média só ocorra após um atraso de d unidades de tempo em relação à intervenção. Se d é conhecido, teremos

$$m_t = \omega S_{t-d}^{(T)}.$$

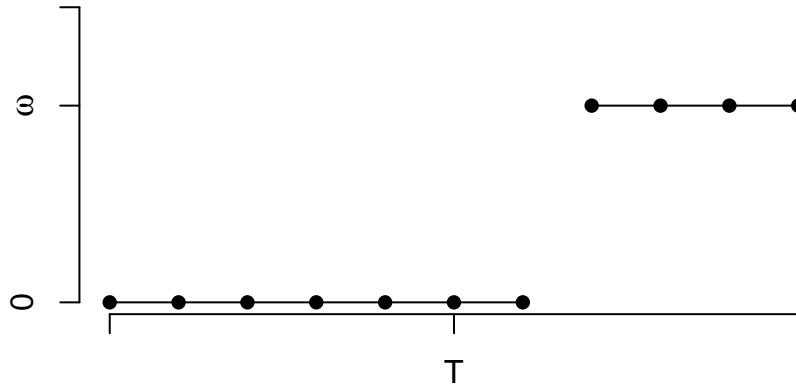
A figura abaixo apresenta essa estrutura considerando um atraso de dois dias.

```
plot.new()
plot.window(xlim=c(0,11), ylim=c(0,1.5))
axis(1, at = c(0,5,10), c('','T',''))
axis(2, at = c(0,1,1.5), c(0,expression(omega),''))
lines(0:6,rep(0,7), type= 'o', pch=16)
lines(7:10,rep(1,4), type= 'o', pch=16)
```

Em outro cenário, é possível que a intervenção possua um efeito imediato que se apaga gradualmente. Podemos reproduzir esse efeito com utilizando a estrutura

$$m_t = \delta m_{t-1} + \omega P_t^{(T)},$$

com $\delta \in (0, 1)$ e $m_0 = 0$. Uma vez que $m_t = \omega \delta^{T-t}$ para $t \geq T$, m_t decrescerá exponencialmente para zero, eliminando o efeito da intervenção. A mesma estrutura anterior pode ser reescrita como



$$m_t = (1 - \delta B)^{-1} \omega (1 - B) S_t^{(T)}$$

A figura abaixo ilustra essa estrutura.

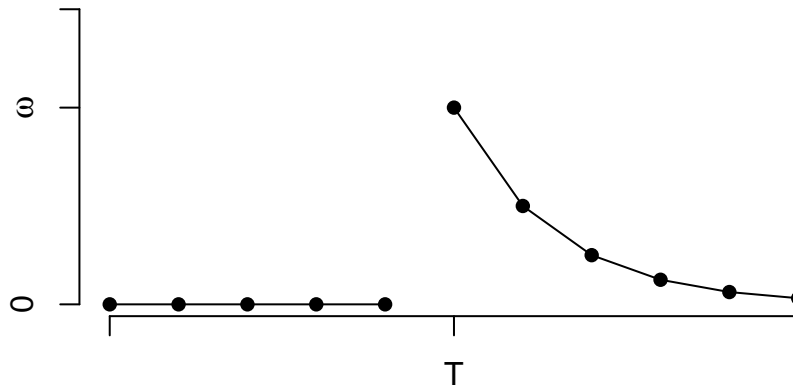
```
plot.new()
plot.window(xlim=c(0,11), ylim=c(0,1.5))
axis(1, at = c(0,5,10), c('','T',''))
axis(2, at = c(0,1,1.5), c(0,expression(omega),''))
lines(0:4,rep(0,5), type= 'o', pch=16)
m = 0
for(i in 1:6) m[i+1] = .5*m[i]+ifelse(i==1,1,0)
lines(5:10,m[-1], type= 'o', pch=16)
```

As estruturas acima podem ser combinadas para criar novas. Por exemplo, suponha que no tempo T a função média se modifica e posteriormente esse efeito diminui, se tornando uma nova constante. Podemos imaginar que a função m_t deve ser composta de duas partes

$$m_t = m_{1,t} + m_{2,t},$$

onde

$$m_{1,t} = \omega_1 S_t^{(T)}$$



representa a nova mudança a ser observada em médio prazo e

$$m_{2,t} = \delta m_{2,t-1} + \omega_2 P_t^{(T)}$$

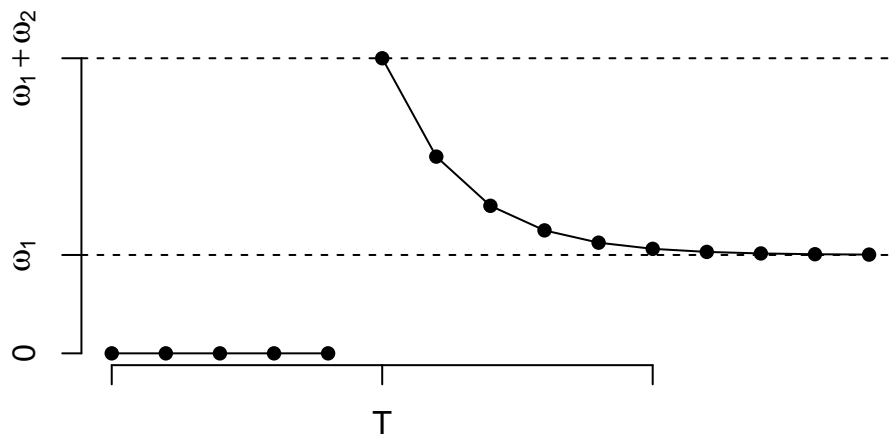
representa a mudança momentânea, que será eliminada no curto prazo. A figura abaixo ilustra essa situação.

```
plot.new()
plot.window(xlim=c(0,14), ylim=c(0,1.5))
axis(1, at = c(0,5,10), c('','T',''))
axis(2, at = c(0,1.5,.5), c(0,expression(omega[1]+omega[2]),expression(omega[1])))
lines(0:4,rep(0,5), type= 'o', pch=16)
m1 = 0
for(i in 1:10) m1[i+1] = .5*m1[i]+ifelse(i==1,1,0)
lines(5:14,.5+m1[-1], type= 'o', pch=16)
abline(h=.5, lty = 2)
abline(h=1.5, lty = 2)
```

Vamos discutir a estimação dos parâmetros da intervenção considerando um modelo ARIMA (sazonal ou não). Note que

$$x_t - m_t = N_t,$$

onde N_t é um modelo ARIMA. Sejam (θ, ϕ) os parâmetros do modelo ARIMA e ω os parâmetros da intervenção. Então,



$$L(\theta, \phi, \omega | x_1, \dots, x_n) = L_{\text{ARIMA}}(\theta, \phi | x_1 - m_1, \dots, x_n - m_n)$$

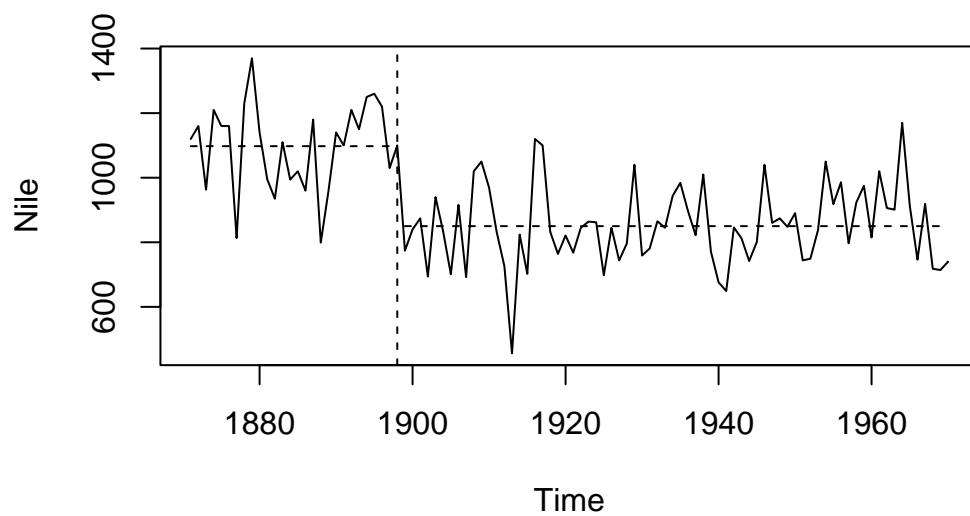
e podemos utilizar a mesma função de verossimilhança para implementar a intervenção.

42.1 Exemplo: Nível do Nilo

Considere mais uma vez a série anual do fluxo do Rio Nilo, mostrada abaixo. A linha pontilhada vertical, no ano de 1898, mostra uma mudança de nível na série. As linhas pontilhadas horizontais mostram a média amostral da série antes e depois de 1898. Note que houve uma redução no volume de aproximadamente $247 \times 10^8 m^3$.

```
ts.plot(Nile)
abline( v= 1898, lty = 2)
mediaAntes  <- mean(Nile[time(Nile)<1898])
mediaDepois <- mean(Nile[time(Nile)>1898])

segments(1871,mediaAntes,1898, mediaAntes, lty = 2)
segments(1899,mediaDepois,1970, mediaDepois, lty = 2)
```



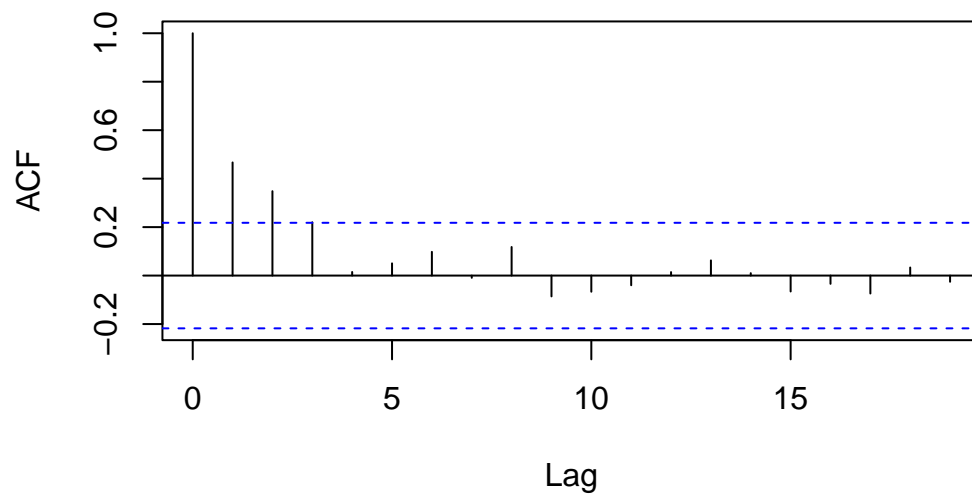
```
(mediaAntes-mediaDepois)
```

```
[1] 247.6944
```

Abaixo, mostramos que a série após 1890 é consistente com um modelo AR(1) com deriva.

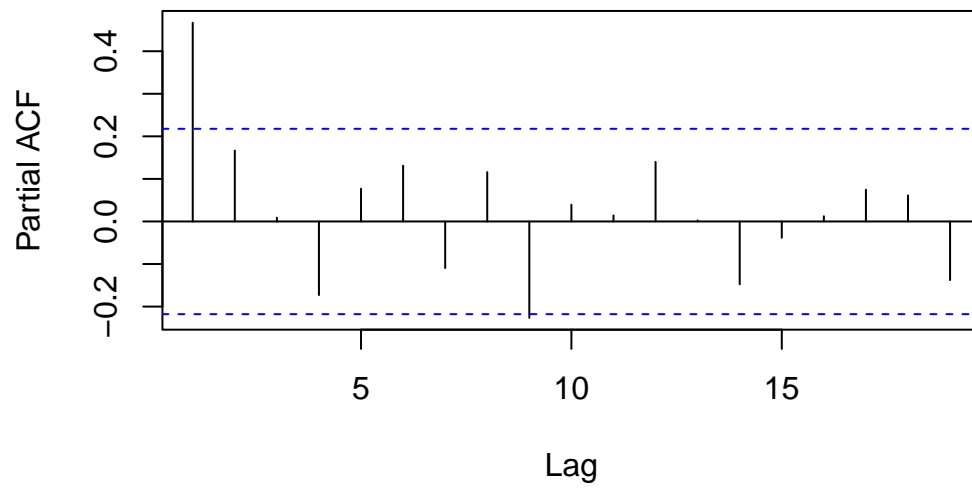
```
nile1890 <- window(Nile, start = 1890)
acf(nile1890)
```

Series Nile1890



```
pacf(Nile1890)
```

Series Nile1890



Vamos considerar o modelo de intervenção

$$m_t = \omega S_{t-1}^{(1898)}.$$

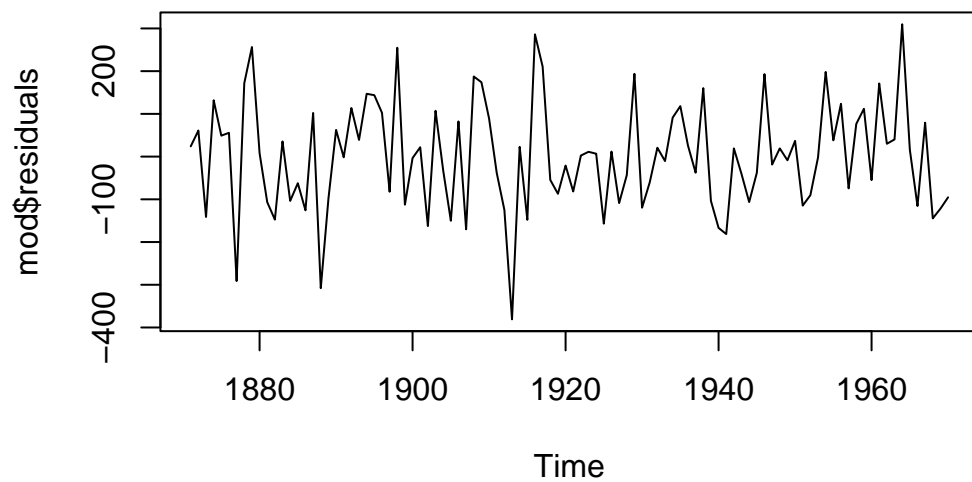
```
La <- function(omega){  
  St <- array(1,length(Nile))  
  St[time(Nile)<1898]<-0  
  mod <- arima(Nile-omega*St, c(1,0,0))  
  mod$loglik  
}  
  
optimize(La, c(-500,500), maximum = TRUE)
```

```
$maximum  
[1] -241.857
```

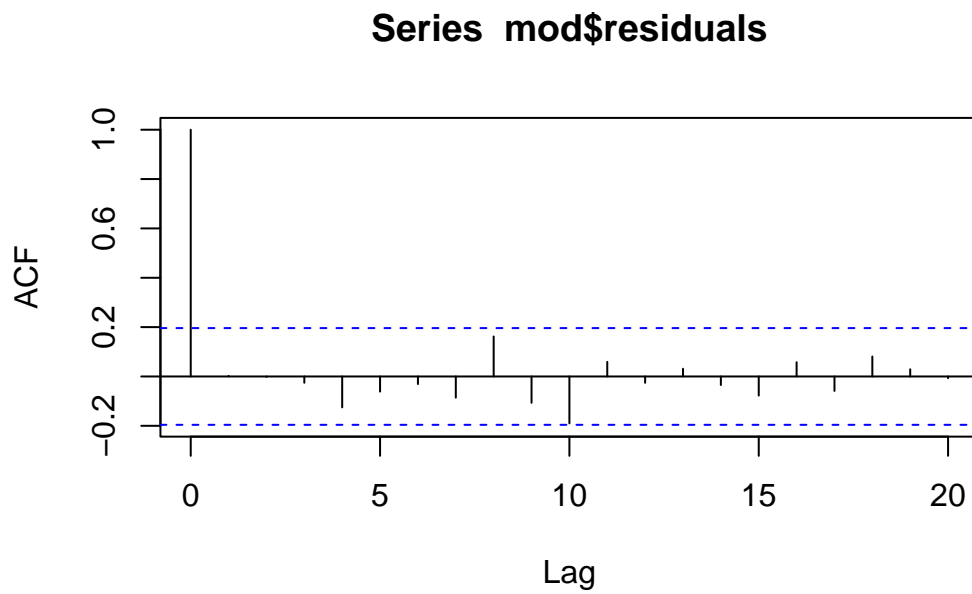
```
$objective  
[1] -626.84
```

A estimativa de máxima verossimilhança para ω é $-241,857$. Abaixo, analisamos os resíduos do modelo ajustado

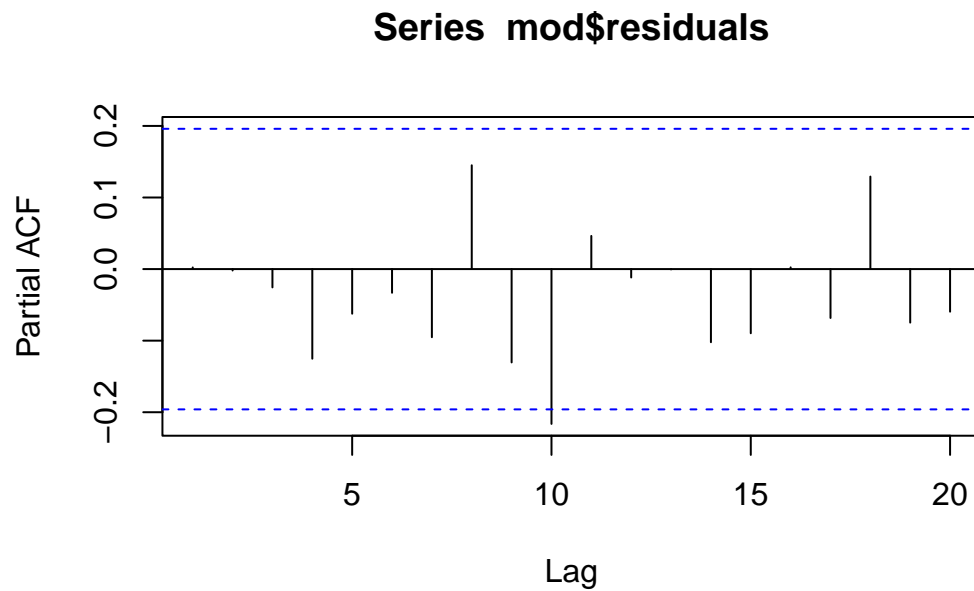
```
omega_chapeu <- -241.857  
St <- array(1,length(Nile))  
St[time(Nile)<1898]<-0  
mod <- arima(Nile-omega_chapeu*St, c(1,0,0))  
  
ts.plot(mod$residuals)
```



```
acf(mod$residuals)
```



```
pacf(mod$residuals)
```



```
Box.test(mod$residuals)
```

Box-Pierce test

```
data: mod$residuals  
X-squared = 0.00053994, df = 1, p-value = 0.9815
```

```
shapiro.test(mod$residuals)
```

Shapiro-Wilk normality test

```
data: mod$residuals  
W = 0.98747, p-value = 0.4697
```

42.2 Exemplo: óbitos maternos durante a pandemia

Durante a pandemia de COVID-19, Manaus passou por duas grandes ondas de mortes. A primeira, ocorreu no início da pandemia, em abril de 2020. A segunda, e mais impactante, ocorreu em janeiro de 2021, pela falta de tubos de oxigênio.

A série abaixo representa o número de óbitos maternos e de mulheres em idade fértil, 10 a 49 anos - MIF - na cidade de Manaus, entre 2005 e 2022.

```
require(gsheet)
```

Carregando pacotes exigidos: gsheet

Warning: package 'gsheet' was built under R version 4.3.2

```
url <- 'https://docs.google.com/spreadsheets/d/1Cd9sxyI7Hm058fCXP7Sw4WzKyuRrCbal7bWz16T8sm
x <- gsheet2tbl(url)
obt <- ts(x$N, frequency = 12, start=c(2005,1))

ts.plot(obt, ylab = 'Óbitos de MIF')
```

intervencao_files/figure-pdf/unnamed-chunk-9-1.pdf

A pandemia é o nosso evento de intervenção e o período pré-pandêmico, logo a série até março de 2020 constiuiará nossos dados pré-intervenção, cuja série é mostrada abaixo

```
obt2020_3 <- window(obt, end = c(2020,3))
ts.plot(obt2020_3, ylab= 'No. óbitos', title = 'Óbitos de MIF, pré-pandemia')
```

Warning in xy.coords(x = matrix(rep.int(tx, k), ncol = k), y = x, log = log, :
NAs introduzidos por coerção

Warning in xy.coords(x, y): NAs introduzidos por coerção

intervencao_files/figure-pdf/unnamed-chunk-10-1.pdf

Podemos notar uma leve tendência na série. O correlograma e a autocorrelação parcial amostral da primeira diferença da série mostram uma estrutura compatível com um modelo ARIMA(4,1,1).

```
tend <- loess( obt2020_3 ~ time(obt2020_3))
tend <- ts(tend$fitted, start = start(obt2020_3), frequency = frequency(obt2020_3))
ts.plot(obt2020_3)
lines(tend)
```

intervencao_files/figure-pdf/unnamed-chunk-11-1.pdf

Abaixo, estudamos o comportamento da série sem a tendência, em busca de sazonalidade. Não há evidências de sazonalidade nessa etapa da análise.

```
require(TSA)
```

Carregando pacotes exigidos: TSA

Warning: package 'TSA' was built under R version 4.3.2

Attaching package: 'TSA'

The following objects are masked from 'package:stats':

acf, arima

The following object is masked from 'package:utils':

tar

```
semTend <- obt2020_3 - tend  
ts.plot(semTend)
```

intervencao_files/figure-pdf/unnamed-chunk-12-1.pdf

```
periodogram(semTend)
```

intervencao_files/figure-pdf/unnamed-chunk-12-2.pdf

```
monthplot(semTend)
```


intervencao_files/figure-pdf/unnamed-chunk-12-3.pdf

Vamos analisar a primeira diferença da série. Há sinais de que a parte estacionária é composta por uma média móvel e ordem 1 e uma sazonal e uma autoregressiva.

```
d1 <- diff(obt2020_3,1)  
acf(d1)
```

intervencao_files/figure-pdf/unnamed-chunk-13-1.pdf

```
pacf(d1)
```



intervencao_files/figure-pdf/unnamed-chunk-13-2.pdf

Vamos tentar ajustar o modelo $\text{ARIMA}(1,1,1)(0,0,1)$.

```
require(forecast)
```

Carregando pacotes exigidos: forecast

Warning: package 'forecast' was built under R version 4.3.1


Registered S3 method overwritten by 'quantmod':

| | |
|-------------------|------|
| method | from |
| as.zoo.data.frame | zoo |

Registered S3 methods overwritten by 'forecast':


| | |
|--------------|------|
| method | from |
| fitted.Arima | TSA |
| plot.Arima | TSA |

```
mod <- Arima(obt2020_3,c(1,1,1),c(0,0,1))  
res <- mod$residuals  
ts.plot(res)
```




intervencao_files/figure-pdf/unnamed-chunk-14-1.pdf

```
acf(res)
```



intervencao_files/figure-pdf/unnamed-chunk-14-2.pdf

```
pacf(res)
```



intervencao_files/figure-pdf/unnamed-chunk-14-3.pdf

```
Box.test(res)
```

Box-Pierce test

```
data: res  
X-squared = 0.0070781, df = 1, p-value = 0.933
```

```
shapiro.test(res)
```

Shapiro-Wilk normality test

```
data: res  
W = 0.99353, p-value = 0.6022
```

Considere agora a previsão para o resto do ano de 2020, dada abaixo, com o que foi observado.

```
o20 <- window( obt, start = c(2020,4), end=c(2020,12))  
ts.plot( o20, ylab = 'No. óbitos MIF')  
prev <- forecast(mod,9)  
lines(prev$mean, lwd = 2, col = 'seagreen', type = 'o')  
legend('topright', c('original','previsão'),fill = c(1,'seagreen'), bty= 'n')
```

```
omega_hat <- o20[1]-prev$mean[1]
```

Note que o efeito da pandemia surge em abril, aumenta em março e desaparece logo em seguida. Então, até dezembro de 2020, vamos assumir que

$$x_t - x_{t-1} = \phi(x_{t-1} - x_{t-2}) + \theta\varepsilon_{t-1} + \Theta\varepsilon_{t-12} + \varepsilon_t + m_t,$$

onde

$$m_t = m_{1,t} + m_{2,t}$$

com

$$m_{1,t} = \delta_1 m_{1,t-1} + \omega P_t^{(2023/4)},$$

e

$$m_{2,t} = \delta_2 m_{2,t-1} + \omega_2 P_t^{(2023/5)}.$$

Abaixo, construímos a função de verossimilhança.

```
# selecionando a série até o fim de 2020
obt2020 <- window( obt, end = c(2020,12))

# função pulso
Pt1 <- ifelse( time(obt2020) == 2020.250, 1, 0)
Pt2 <- ifelse( time(obt2020) == time(obt2020)[185], 1, 0)

# função para encontrar o EMV
fun <- function(par){

# parâmetros da intervenção
delta1 <- exp(par[1]) / ( 1 + exp(par[1]) )
delta2 <- exp(par[2]) / ( 1 + exp(par[2]) )
omega1 <- exp(par[3])
omega2 <- exp(par[4])

# mt
mt = mt1 = mt2 = 0
for(i in 1:length(obt2020)){
```

```

    mt1[i+1] = delta1*mt1[i]+omega1*Pt1[i]
    mt2[i+1] = delta2*mt2[i]+omega2*Pt2[i]
  }

  mt = mt1 + mt2
  mt <- mt[-1] # removendo m0
  mt <- ts(mt, start=c(2005,1), frequency = 12)
  # estimação do ARIMA
  mod <- Arima(obt2020-mt, c(1,1,1),c(0,0,1))

  # retorne a log verossimilhança do modelo
  -mod$loglik
}

```

Vamos obter as estimativas de máxima verossimilhança.

```

opt <- optim( c(0,0,0,0), fun)
opt$convergence

```

```
[1] 0
```

```

delta1 <- exp(opt$par[1])/(1+exp(opt$par[1]))
delta2 <- exp(opt$par[2])/(1+exp(opt$par[2]))
omega1 <- exp(opt$par[3])
omega2 <- exp(opt$par[4])
c(delta1,delta2,omega1,omega2)

```

```
[1] 2.577569e-03 6.626144e-07 5.222526e+01 6.607272e+01
```

Logo em abril, os casos aumentaram em $\hat{\omega}_1 = 52$. Observe que $\delta_1 < 0,003$, logo o efeito para o mês seguinte dos casos de maio é praticamente nulo. Portanto, $\hat{\omega}_2 = 66$ representa, basicamente, o efeito de maio. Como δ_2 é praticamente nulo, eses efeito somem em junho.

```

# obtendo mt
# mt
mt = mt1 = mt2 = 0
for(i in 1:length(obt2020)){
  mt1[i+1] = delta1*mt1[i]+omega1*Pt1[i]
  mt2[i+1] = delta2*mt2[i]+omega2*Pt2[i]
}

```

```

mt = mt1 + mt2
mt <- mt[-1] # removendo m0
mt <- ts(mt, start=c(2005,1), frequency = 12)
# obtendo os resíduos
modFim <- Arima(obt2020-mt,c(1,1,1),c(0,0,1))
resFim <- modFim$residuals

```

Abaixo segue a análise usual dos resíduos

```
ts.plot(resFim)
```

intervencao_files/figure-pdf/unnamed-chunk-19-1.pdf

```
acf(resFim)
```

intervencao_files/figure-pdf/unnamed-chunk-19-2.pdf

```
pacf(resFim)
```

intervencao_files/figure-pdf/unnamed-chunk-19-3.pdf

```
Box.test(resFim);shapiro.test(resFim)
```

Box-Pierce test

```
data: resFim  
X-squared = 0.012452, df = 1, p-value = 0.9111
```

Shapiro-Wilk normality test

```
data: resFim  
W = 0.9931, p-value = 0.5056
```

Por último, podemos fazer um gráfico de m_t para verificar o efeito do início da pandemia na série

```
ts.plot(window(mt, start = c(2020,1)), type = 'h')
```

intervencao_files/figure-pdf/unnamed-chunk-20-1.pdf

43 Outliers

Outliers são observações atípicas. Segundo sua natureza, eles podem ser classificados como

- Outliers aditivos (AO): são resultantes de erros de medição/registro
- Outliers na inovação (IO): são mudanças abruptas de curta duração no processo subjacente.

Seja x_t uma série temporal. Então, se um outlier aditivo ocorre no tempo T , a série temporal pode ser descrita como x'_t , onde

$$x'_t = \begin{cases} x_t, & t \neq T \\ x_t + \omega_A P_t^{(T)}, & t = T. \end{cases}$$

Note que um outlier aditivo pode ser tratado como uma intervenção, onde $m_t = \omega_A P_t^{(T)}$.

Seja ε_t erro associado à série x_t . Um outlier na inovação ocorre quando o erro (também conhecido como inovação) no tempo T sofre uma perturbação.

$$\varepsilon'_t = \begin{cases} \varepsilon_t, & t \neq T \\ \varepsilon_t + \omega_I P_t^{(T)}, & t = T. \end{cases}$$

Considere o processo estacionário MA(∞)

$$x_t = \varepsilon_t + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}$$

Então, o processo com outlier na inovação no tempo T é escrito como

$$x'_t = \varepsilon'_t + \sum_{j=1}^{\infty} \theta_j \varepsilon'_{t-j} = \varepsilon_t + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j} + \theta_{t-T} \omega_I = x_t + \theta_{t-T} \omega_I,$$

onde $\theta_0 = 1$ e $\theta_j = 0$ se $j < 0$. Portanto, um outlier na inovação perturba todas as observações no tempo $t \geq T$, embora esse efeito diminua na medida que t se afasta de T .

Para detectar um IO, considere que um modelo foi ajustado, gerando a série de resíduos

$$\hat{\varepsilon}'_t = x'_t - E(x'_t | \mathcal{D}_{t-1}),$$

onde $\hat{\varepsilon}_t \approx N(0, \sigma^2)$. Então

$$\lambda_{1,t} = \frac{\hat{\varepsilon}'_t}{\sigma}$$

se comporta como uma distribuição normal padrão quando não há outliers. Na presença de outliers, teremos

$$\lambda_{1,t} = \frac{\hat{\varepsilon}'_t}{\sigma} = \begin{cases} \frac{\hat{\varepsilon}_t}{\sigma}, & t \neq T \\ \frac{\hat{\varepsilon}_t + \omega_I}{\sigma}, & t = T \end{cases}$$

Se T é conhecido, declaramos x_T como outlier se $|\lambda_{1,T}| > 1,96$ e $\hat{\omega}_I = \hat{\varepsilon}'_T$. Se T é desconhecido, calcula-se

$$\lambda_1 = \max_{1 \leq t \leq n} |\lambda_{1,t}|,$$

e define-se T como sendo o valor t tal que $\lambda_{1,t} = \lambda_1$. Como todos os pontos estão sendo avaliados, precisamos realizar algum teste múltiplo. O mais simples é a correção de Bonferroni, que nos leva a declarar que x_T é um outlier se λ_1 é maior que o quantil $1 - \alpha/2n$ da normal padrão. Na prática, σ deve ser estimado de modo robusto, uma vez que a variância tradicional é sensível à outliers. Uma alternativa é estimar σ por

$$\hat{\sigma} = \sqrt{\frac{2}{\pi}} \sum_{t=1}^n \frac{|\hat{\varepsilon}_t|}{n}.$$

Para o caso AO, considere que x'_t é um processo AR(∞). Então

$$\hat{\varepsilon}'_t = x'_t - \sum_{j=1}^{\infty} \phi_j x'_{t-j},$$

onde $\phi_j = 0$ para $j < t$. Note que, para $t < T$,

$$\hat{\varepsilon}'_t = x'_t - \sum_{j=1}^{\infty} \phi_j x'_{t-j} = x_t - \sum_{j=1}^{\infty} \phi_j x_{t-j},$$

Para $t = T$, teremos

$$\hat{\varepsilon}'_T = x_T + \omega_A - \sum_{j=1}^{\infty} \phi_j x_{T-j} = \omega_A + \hat{\varepsilon}_T.$$

Para $t = T + 1$, teremos

$$\hat{\varepsilon}'_{T+1} = x_{T+1} - \phi_1 \omega_A - \sum_{j=2}^{\infty} \phi_j x_{T+1-j} = -\phi_1 \omega_A + \hat{\varepsilon}_{T+1},$$

e, de modo geral, para $k > T$

$$\hat{\varepsilon}'_k = -\phi_{T-k} \omega_A + \hat{\varepsilon}_k.$$

Definindo a soma de quadrados

$$\sum_{j=T}^n {}' \hat{\varepsilon}_j^2 = \sum_{j=T}^n (\hat{\varepsilon}_j + \omega_A \pi_{T-j})^2$$

onde $\pi_0 = 1$ e $\pi_j = \phi_j$ se $j > 0$, podemos estimar ω por

$$\hat{\omega}_{T,A} = -\frac{\sum_{j=T}^n \pi_{T-j} \hat{\varepsilon}_j}{\sum_{j=T}^n \pi_{T-j}^2}$$

Como $\hat{\varepsilon}_t \approx N(0, \sigma^2)$, teremos que

$$\lambda_{2,T} = \frac{\hat{\omega}_{T,A}}{\sigma \sqrt{\sum_{j=T}^n \pi_{T-j}^2}}$$

deve se comportar como uma distribuição normal padrão na ausência de outliers. Novamente, se T for desconhecido, podemos escolher T como o valor t que gera o valor

$$\lambda_2 = \max_{1 \leq t \leq n} |\lambda_{2,t}|$$

e podemos testar sua significância comparando seu valor com o quantil $1-\alpha/2n$ da distribuição normal padrão.

Na prática, é provável que a natureza do outlier seja desconhecida. Nesses casos, tanto o teste para o AO quanto para o IO devem detectar os mesmos outliers. Uma regra que tende a funcionar é classificar o outlier como IO se $|\lambda_{1,T}| > |\lambda_{2,T}|$ ou como AO em caso contrário. Após identificar o outlier e estimar ω , o mesmo deve ser incorporado na função de verossimilhança e as estimativas devem ser realizadas novamente.

43.1 Mortes por doenças pulmonares no Reino Unido

Voltemos à série `ldeaths`. Sabemos que essa série possui sazonalidade de período 12 e uma leve tendência de decrescimento. Após fazer uma diferença para remover a tendência de uma diferença sazonal para a sazonalidade, encontramos evidências para ajustar o modelo $ARIMA(2,1,2)(0,1,1)$

```
ts.plot(ldeaths)
```

intervencao_files/figure-pdf/unnamed-chunk-21-1.pdf

```
d1_ldeaths <- diff(ldeaths,1)
D12d1_ldeaths <- diff(d1_ldeaths,12)

acf(D12d1_ldeaths)
```

intervencao_files/figure-pdf/unnamed-chunk-21-2.pdf

```
pacf(D12d1_ldeaths)
```

intervencao_files/figure-pdf/unnamed-chunk-21-3.pdf

Abaixo, apresentamos a série dos resíduos, que apresentam possíveis outliers.

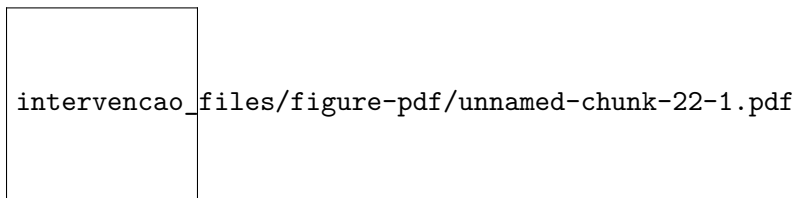
```
mod <- arima(ldeaths,c(2,1,2),c(0,1,1))
mod
```

```
Call:
arima(x = ldeaths, order = c(2, 1, 2), seasonal = c(0, 1, 1))

Coefficients:
          ar1          ar2          ma1          ma2          sma1
      -0.0145  -0.2761  -0.7617  -0.2383  -0.9999
s.e.    0.3997   0.1639   0.4327   0.4206   0.2385

sigma^2 estimated as 47226:  log likelihood = -414.79,  aic = 839.57
```

```
ts.plot(mod$residuals[-(1:12)])
```



Utilizaremos as funções `detectIO` e `detectAO`, do pacote `TSA`, para a detecção de outliers na série através dos resíduos encontrados.

```
detectAO(mod)
```

```
      [,1]
ind      72.000000
lambda2 -3.453803
```

```
detectIO(mod)
```

```
      [,1]      [,2]      [,3]
ind      26.000000 38.00000 72.000000
lambda1  6.685518 -4.41524 -3.453803
```

As observações no tempo 26 e 38 foram consideradas outliers nas inovações. Já a observação no tempo 72 recebeu as duas classificações e vamos modelar como um AO apenas por fins didáticos.

Ao considerar que o outlier no tempo 72 é aditivo, estamos modificando apenas o nível da série, assim como fizemos para a análise de intervenção. Aqui vamos utilizar a função `arimax`, do pacote `TSA` para o mesmo fim. Tal função possui o argumento `xreg` que permite o uso de variáveis regressoras (no caso, utilizaremos a função `pulso` como indicadora). Além disso, essa função possui o argumento `io` que nos permite colocar os índices os outliers nas inovações.

```
pulso <- array(0,length(ldeaths))
pulso[72] <- 1
mod2 <- arimax(ldeaths, c(2,1,2), c(0,1,1), xreg = pulso, io = c(26,38))
```

Vamos verificar se novos outliers foram revelados.

```
detectAO(mod2)
```

```
[1] "No AO detected"
```

```
detectIO(mod2)
```

```
           [,1]
ind      50.000000
lambda1  3.535189
```

Note que mais um outlier foi detectado IO, agora no tempo 50. Abaixo, construímos mais um modelo, agora com o outlier no tempo 50.

```
mod3 <- arimax(ldeaths, c(2,1,2), c(0,1,1), xreg = pulso, io = c(26,38,50))
detectAO(mod3)
```

```
[1] "No AO detected"
```

```
detectIO(mod3)
```

```
[1] "No IO detected"
```

Depois de detectar todos os outliers, vamos comparar o AIC do modelo original com o do modelo com tratamento de outliers.

```
mod$aic
```

```
[1] 839.5738
```

```
mod3$aic
```

```
[1] 773.1072
```

O modelo 3 se mostrou mais parcimonioso e vamos mantê-lo. As estimativas dos parâmetros são:

```
mod3
```

Call:

```
arimax(x = ldeaths, order = c(2, 1, 2), seasonal = c(0, 1, 1), xreg = pulso,  
       io = c(26, 38, 50))
```


Coefficients:

Warning in sqrt(diag(x\$var.coef)): NaNs produzidos

| | ar1 | ar2 | ma1 | ma2 | sma1 | xreg | I0.26 |
|------|-----------|----------|---------|---------|---------|-----------|----------|
| | 0.0010 | -0.2706 | -0.5834 | -0.3580 | -0.9590 | -596.0401 | 1383.398 |
| s.e. | 0.1331 | 0.0958 | NaN | 0.1557 | 0.0864 | 125.0698 | 148.260 |
| | I0.38 | I0.50 | | | | | |
| | -292.6500 | 499.8729 | | | | | |
| s.e. | 148.1292 | 67.7714 | | | | | |

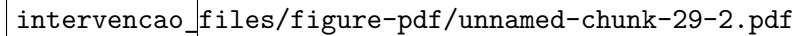
sigma^2 estimated as 14410: log likelihood = -377.55, aic = 773.11

```
resid <- mod3$residuals[-(1:12)]  
ts.plot(resid)
```



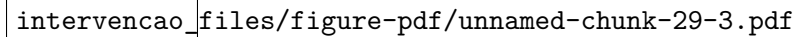
intervencao_files/figure-pdf/unnamed-chunk-29-1.pdf

```
acf(resid)
```



intervencao_files/figure-pdf/unnamed-chunk-29-2.pdf

```
pacf(resid)
```



intervencao_files/figure-pdf/unnamed-chunk-29-3.pdf

```
Box.test(resid)
```

Box-Pierce test

```
data: resid
X-squared = 0.52117, df = 1, p-value = 0.4703
```

```
shapiro.test(resid)
```

Shapiro-Wilk normality test

```
data: resid
W = 0.98455, p-value = 0.6467
```

Modelos de regressão para séries temporais

44 As funções `ts.union` e `ts.intersect`

Considere novamente a série de mortes por doenças pulmonares no Reino Unido. Além da série `ldeaths`, o pacote `datasets` possui as séries `mdeaths` e `fdeaths`, que representam as mortes

para os sexos masculino e feminino, respectivamente. Assim, como a primeira, essas duas séries são mensais, começando em janeiro de 1974 e terminando em dezembro de 1979.

Desde de dois objetos `ts` tenham o mesmo argumento `frequency`, é possível organizá-las lado a lado e um `data.frame` respeitando o tempo do registro de cada observação. Para tanto, utilizamos a função `ts.union`. Ilustramos isso com as duas novas séries apresentadas.

```
d1 <- ts.union( ldeaths, mdeaths)
head(d1)
```

| | ldeaths | mdeaths |
|------|---------|---------|
| [1,] | 3035 | 2134 |
| [2,] | 2552 | 1863 |
| [3,] | 2704 | 1877 |
| [4,] | 2554 | 1877 |
| [5,] | 2014 | 1492 |
| [6,] | 1655 | 1249 |

Observe que é possível realizar a organização mesmo que os tempos de início ou fim sejam diferentes - nesse caso, NAs serão gerados quando não houverem registros. Veja o exemplo abaixo, no qual o primeiro mês da série `mdeaths` foi removido

```
mdeaths2 <- window(mdeaths, start = c(1974,2))
d2 <- ts.union( ldeaths, mdeaths2)
head(d2)
```

| | ldeaths | mdeaths2 |
|------|---------|----------|
| [1,] | 3035 | NA |
| [2,] | 2552 | 1863 |
| [3,] | 2704 | 1877 |
| [4,] | 2554 | 1877 |
| [5,] | 2014 | 1492 |
| [6,] | 1655 | 1249 |

De modo análogo, podemos construir `data.frame` que apresenta um subconjunto das séries corresponde à interseção dos períodos registrados, utilizando a função `ts.intersect`

```
mdeaths3 <- window(mdeaths, start = c(1974,2), end = c(1974,5))
d3 <- ts.intersect( ldeaths, mdeaths3)
d3
```


| | ldeaths | mdeaths |
|----------|---------|---------|
| Feb 1974 | 2552 | 1863 |
| Mar 1974 | 2704 | 1877 |
| Apr 1974 | 2554 | 1877 |
| May 1974 | 2014 | 1492 |

45 A função de correlação cruzada e a correlação espúria

Sejam $\{x_t\}$ $\{y_t\}$ duas séries temporais. A função de covariância cruzada é dada por

$$\gamma_{r,s}(X, Y) = Cov(X_r, Y_s).$$

Se as séries forem estacionárias, a covariância cruzada é função somente da diferença $k = r - s$ e escrevemos

$$\gamma_k(X, Y) = Cov(X_{t+k}, Y_t) = Cov(X_t, Y_{t-k}).$$

Observe que a função de covariância cruzada é ímpar, uma vez que

$$\gamma_k(X, Y) \neq \gamma_{-k}(X, Y).$$

Além disso, a função de autocovariância, de X por exemplo, é dada por $\gamma_k(X, X)$.

A função de correlação cruzada é definida por

$$\rho_k(X, Y) = \frac{\gamma_k(X, Y)}{\sqrt{Var(x)Var(y)}}$$

Consideremos que

$$y_t = \beta_0 + \beta_1 x_{t-d} + \varepsilon_t,$$

onde ε_t é um ruído branco com variância ν e x_t tem distribuição normal com variância σ_X^2 e são independentes entre si e do ruído. Como,

$$Var(y_t) = \beta_1^2 Var(x_{t-d}) + Var(\varepsilon_t) = \beta_1^2 \sigma_X^2 + \nu.$$

e

$$\begin{aligned} Cov(x_t, y_{t-k}) &= Cov(x_t, \beta_0 + \beta_1 x_{t-k-d} + \varepsilon_{t-k}) \\ &= \beta_1 Cov(x_t, x_{t-k-d}) = \begin{cases} \beta_1 \sigma_X^2, & \text{se } k = -d, \\ 0, & \text{caso contrário} \end{cases} \end{aligned}$$

logo,

$$\rho_k(X, Y) = \begin{cases} \frac{\beta_1 \sigma_X}{\sqrt{\beta_1^2 \sigma_X^2 + \nu}}, & \text{se } k = -d, \\ 0, & \text{caso contrário} \end{cases}$$

■ :::

Para duas séries x_t e y_t observadas, de tamanho n , a função de correlação cruzada amostral (CCF) é definida por

$$r_k(X, Y) = \frac{\sum_{t=k+1}^n (x_t - \bar{x}_n)(y_{t-k} - \bar{y}_n)}{\sqrt{\sum_{t=1}^n (x_t - \bar{x}_n)^2} \sqrt{\sum_{t=1}^n (y_t - \bar{y}_n)^2}}$$

e a regra

pode ser utilizada para detectar se Para o modelo

$$y_t = \beta_0 + \beta_1 x_{t-d} + \varepsilon_t,$$

x é independente de y se e somente se $\beta_1 = 0$. Nesse caso, $r_k(X, Y)$ tem distribuição aproximadamente normal com média zero e variância $1/n$ e a regra

$$|r_k(X, Y)| > \frac{1,96}{\sqrt{n}}$$

pode ser utilizada para detectar se a defasagem k na regressora é significativa. Entretanto, esse modelo é muito restritivo e um modelo mais geral seria

$$y_t = \beta_0 + \beta_1 x_{t-d} + z_t,$$

onde z_t é um processo ARIMA. Nesse caso, mesmo se x_t e y_t forem processos estacionários, a variância de $r_k(X, Y)$ não é mais aproximadamente $1/n$, sendo geralmente inflada. Isso implica que, ao utilizar a regra de comparar o módulo de $r_k(X, Y)$ com $1,96/\sqrt{n}$, é possível encontrar correlações que não fazem sentido, ou seja, **espúrias**.

:::{#exm-} Como exemplo de correlação espúria, considere a produção mensal de leite e a produção de energia elétrica (em escala logaritmica) nos EUA entre janeiro de 1994 e dezembro de 2006. Note como há várias correlações significativas, quando comparadas com o limiar $1,96/\sqrt{n}$. Entretanto, não parece fazer sentido a relação entre produção de eletricidade e de leite.

```
require(TSA)
```

```
Carregando pacotes exigidos: TSA
```

```
Warning: package 'TSA' was built under R version 4.3.2
```

```
Attaching package: 'TSA'
```

The following objects are masked from 'package:stats':

acf, arima

The following object is masked from 'package:utils':

tar

```
data(milk); data(electricity)
leiteLogEle <- ts.intersect(milk, log(electricity))
plot(leiteLogEle)
```

regressao_files/figure-pdf/unnamed-chunk-4-1.pdf

```
ccf( leiteLogEle[,1] , leiteLogEle[,2] )
```

regressao_files/figure-pdf/unnamed-chunk-4-2.pdf

46 Whitening

Considere o modelo

$$y_t = \sum_{j=-\infty}^{\infty} \beta_j x_{t-j} + z_t,$$

onde z_t é um processo ARIMA. Se ambos $\{x_t\}$ e $\{y_t\}$ são estacionários, a variância de $r_k(X, Y)$ é aproximadamente

$$\frac{1}{n} \left[1 + 2 \sum_{k=1}^{\infty} \rho_k(x) \rho_k(y) \right]$$

onde $\rho_k(x)$ e $\rho_k(y)$ são as funções de autocorrelação de $\{x_t\}$ e $\{y_t\}$, respectivamente. Note que, se pelo menos um dos processos for um ruído branco, então a variância será aproximadamente $1/n$.

Suponha então $\{x_t\}$ é um processo ARIMA,

$$\phi(B)(1-B)^d x_t = \theta(B)\epsilon_t,$$

onde ϵ_t é ruído branco. Suponha ainda que o processo é inversível, ou seja, existe $\pi(B)$ tal que

$$\pi(B)x_t = \epsilon_t.$$

Então, $\pi(B)x_t$ é um ruído branco. A aplicação do filtro $\pi(B)$ em x_t para obter um ruído branco é denominado *whitening* ou *prewhitening*. Aplicando o filtro ao modelo no começo dessa seção, teremos

$$\pi(B)y_t = \sum_{j=-\infty}^{\infty} \pi(B)x_{t-j} + \pi(B)\epsilon_t.$$

Como $\pi(B)x_t$ é independente de $\pi(B)\epsilon_t$, a correlação cruzada entre $\pi(B)y_t$ e $\pi(B)x_t$ é igual à

$$\rho_k(\pi(X), \pi(Y)) = \beta_{-k} \sqrt{\frac{\text{Var}(\pi(B)x_t)}{\text{Var}(\pi(B)y_t)}},$$

ou seja, a correlação cruzada do processo *whitened* é proporcional à β_{-k} e, sob H_0 , pode-se verificar a significância da defasagem comparando o módulo da correlação cruzada amostral com $1,96/\sqrt{n}$.

Example .27. Considere novamente as séries de produção de leite e eletricidade vistas na seção anterior. Ambas possuem tendência e sazonalidade, que podem ser removidas com operadores diferença para obter um processo estacionário.

```
estacionario <- ts.intersect(diff(diff(milk,12)),
diff(diff(log(electricity),12)))
```

Abaixo, encontramos a correlação cruzada utilizando o processo *whitened*. Note que as correlações espúrias se foram e não há correlação digna de nota.

```
prewhiten( estacionario[,1], estacionario[,2])
```

regressao_files/figure-pdf/unnamed-chunk-6-1.pdf

47 Regressão para séries temporais

Sejam $\{x_t\}$ e $\{y_t\}$ séries temporais onde

$$y_t = \sum_{j=-\infty}^{\infty} \beta_j x_{t-j} + z_t,$$

onde z_t é um modelo ARIMA. Na prática, apenas um número finito dos β 's é diferente de zero. Seja $\dot{x}_t = \pi(B)x_t$, onde \dot{x}_t é um processo AR(∞). Seja $\dot{y}_t = \pi(B)y_t$. Então,

$$\dot{y}_t = \sum_{j=-\infty}^{\infty} \beta_j \dot{x}_{t-j} + \dot{z}_t,$$

onde $\dot{z}_t = \pi(B)z_t$. Como \dot{x}_t é um ruído branco e independente de \dot{z}_t , a correlação cruzada entre \dot{x}_t e \dot{y}_t na defasagem k é igual à $\beta_{-k}\sigma_{\dot{X}}/\sigma_{\dot{Y}}$.

Abaixo, segue um roteiro para uma análise preliminar: 1. Se necessário, remova o sinal de x_t de y_t , aplicando os operadores diferença necessários. 2. Encontre uma representação de um modelo AR para x_t , minimizando o AIC por exemplo, estime $\pi(B)$ 3. Obtenha $\dot{x}_t = \pi(B)x_t$ e $\dot{y}_t = \pi(B)y_t$. 4. Estime a função de correlação cruzada entre \dot{y}_t e \dot{x}_t . Determine \mathcal{K} , um conjunto de defasagens significativas 5. Encontre os estimadores de mínimos quadrados para β , minimizando

$$\sum_{t=1}^n z_t^2 = \sum_{t=1}^n \left(y_t - \sum_{j \in \mathcal{K}} \beta_j x_{t-j} \right)^2$$

6. Estime z_t pelos resíduos

$$\hat{z}_t = y_t - \sum_{j \in \mathcal{K}} \hat{\beta}_j x_{t-j}$$

e determine o modelo ARIMA estimado para z_t 7. Encontre os estimadores de máxima verossimilhança para o modelo

$$y_t = \sum_{j \in \mathcal{K}} \beta_j x_{t-j} + z_t$$

onde z_t é um modelo ARIMA com a ordem encontrada no passo anterior.

Vamos simular um processo do seguinte modelo:

$$y_t = 2x_{t-4} + z_t,$$

onde x_t é um processo AR(1) com $\phi = .8$ e z_t um ARMA(1,1) com $\phi = .2$ e $\theta = .5$.

```

set.seed(123)
z <- arima.sim( list( ar = .2, ma = .5), 105)

x <- arima.sim( list(ar = .8, ma = 0), 105)

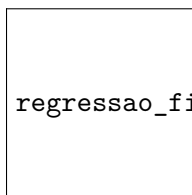
x <- ts(x, start = 1); z <- ts(z, start=1)

y <- NULL
for(t in 5:105) y[t] <- 2*x[t-4] + z[t]

x <- ts(x[5:105])
y <- ts(y[-(1:4)])

plot.ts( cbind(y,x))

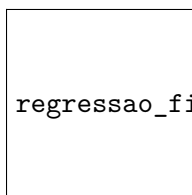
```



regressao_files/figure-pdf/unnamed-chunk-7-1.pdf

Observe abaixo que a correlação cruzada apresenta a defasagem -4 significativa e conunto com várias outras, que são espúrias.

```
ccf(x,y)
```




regressao_files/figure-pdf/unnamed-chunk-8-1.pdf

As séries x_t e y_t já parecem ser estacionárias. Abaixo, aplicamos o *whitening* e as correlações espúrias desapareceram. Temos evidências de que apenas β_4 é relevante.

```

require(TSA)
prewhiten(x,y)

```



regressao_files/figure-pdf/unnamed-chunk-9-1.pdf

Vamos ajustar o modelo de regressão

$$y_t = \beta_4 x_{t-j}$$

e obter o AIC e seus resíduos.

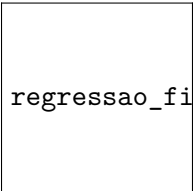
```
mod <- lm( y[5:100] ~ x[1:96] -1)
AIC(mod)
```

```
[1] 296.0433
```

```
resReg <- residuals(mod)
```

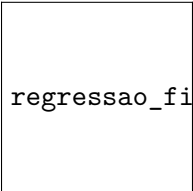
Vamos analisar os resíduos obtidos para determinar o modelo de z_t . Abaixo, vemos indícios para o modelo ARMA(1,1).

```
acf(resReg)
```



regressao_files/figure-pdf/unnamed-chunk-11-1.pdf

```
pacf(resReg)
```



regressao_files/figure-pdf/unnamed-chunk-11-2.pdf

Vamos ajustar o modelo de regressão para séries temporais:

```
modFinal <- arima(y[5:100] , order = c(1,0,1), xreg = x[1:96] , include.mean = FALSE)
modFinal$aic
```

```
[1] 256.901
```



*References