



**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**DEPARTAMENTO DE ENGENHARIA ELETRÔNICA**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**GABRIEL TEIXEIRA LARA CHAVES**

**UMA ABORDAGEM DE SISTEMAS LINEARES A TEORIA ESPECTRAL DE SÉRIES**  
**TEMPORAIS ESTACIONÁRIAS E NÃO ESTACIONÁRIAS**

**BELO HORIZONTE**

**2023**

GABRIEL TEIXEIRA LARA CHAVES

UMA ABORDAGEM DE SISTEMAS LINEARES A TEORIA ESPECTRAL DE SÉRIES  
TEMPORAIS ESTACIONÁRIAS E NÃO ESTACIONÁRIAS

Monografia de graduação submetida ao curso  
de Engenharia Elétrica da Universidade Federal  
de Minas Gerais como requerimento parcial  
para obtenção do Bacharel em Engenharia  
Elétrica.

Orientador: Prof. Dr. Frederico Gualberto  
Coelho

BELO HORIZONTE

2023

GABRIEL TEIXEIRA LARA CHAVES

UMA ABORDAGEM DE SISTEMAS LINEARES A TEORIA ESPECTRAL DE SÉRIES  
TEMPORAIS ESTACIONÁRIAS E NÃO ESTACIONÁRIAS

Monografia de graduação submetida ao curso de  
Engenharia Elétrica da Universidade Federal de  
Minas Gerais como requerimento parcial para  
obtenção do Bacharel em Engenharia Elétrica.

Aprovado em: 17 de maio de 2023

BANCA EXAMINADORA

---

Prof. Dr. Frederico Gualberto Coelho (Orientador)  
Universidade Federal de Minas Gerais

---

Prof. Dr. Adriano Vilela Barbosa

## **AGRADECIMENTOS**

Obrigado !

“Alice sighed wearily. ‘I think you might do something better with the time,’ she said, ‘than waste it in asking riddles that have no answers.’ ‘If you knew time as well as I do,’ said the Hatter, ‘you wouldn’t talk about wasting it. It’s him.’ ‘I don’t know what you mean,’ said Alice. ‘Of course you don’t!’ the Hatter said, tossing his head contemptuously. ‘I dare say you never even spoke to Time!’ ‘Perhaps not,’ Alice cautiously replied, ‘but I know I have to beat time when I learn music.’ ‘Ah! That accounts for it,’ said the Hatter. ‘He won’t stand beating.’”

(L. Carrol, *Alice’s Adventures in Wonderland*)

## RESUMO

Abstract here.

**Keywords:** Time series analysis. Linear Systems. Discrete-time signal processing. Time Frequency Representations

## RESUMO

Resumo

**Palavras-chave:** Análise de Séries Temporais. Sistemas Lineares. Processamento de Sinais de Tempo Discreto. Análise em Frequência

## LISTA DE ILUSTRAÇÕES

Figura 1 – Visualização no tempo de ruído branco . . . . .	13
Figura 2 – Série temporal com tendência linear . . . . .	16
Figura 3 – Série temporal com tendência logarítmica . . . . .	16
Figura 4 – Série temporal com sazonalidade senoidal de período arbitrário . . . . .	17
Figura 5 – Série temporal com sazonalidade aditiva . . . . .	17
Figura 6 – Série temporal com sazonalidade multiplicativa . . . . .	18
Figura 7 – Visualização de correlatograma de série com tendência linear. Observe que as autocorrelações decaem lentamente ao decorrer dos atrasos, comportamento típico de tendências determinísticas. . . . .	20
Figura 8 – Visualização de correlatograma de série com sazonalidade. Observe que o padrão senoidal da série é reproduzido nas autocorrelações. . . . .	20
Figura 9 – Sinal periódico imerso em ruído e sua correspondente autocorrelação. . . . .	21
Figura 10 – Conjunto de séries demonstrando diferentes níveis de estacionariedade. . . . .	23
Figura 11 – Visualização de caminhada aleatória com $y_0 = 5$ . . . . .	25
Figura 12 – Visualização de aplicação de filtro média móvel com $m=15$ . . . . .	30
Figura 13 – Visualização de algoritmo LOESS . . . . .	31
Figura 14 – Representação de série temporal como modelo linear generalizado . . . . .	35
Figura 15 – Visualização no tempo de processos média móvel de ordens diferentes. . . . .	39
Figura 16 – Visualização do correlatograma de processos média móvel de ordens diferentes. . . . .	41
Figura 17 – Visualização de processos autoregressivos de diferentes ordens no domínio do tempo. . . . .	43
Figura 18 – Visualização do correlatograma de processos autoregressivos de ordens diferentes. . . . .	45
Figura 19 – Visualização de realização de processo ARMA(2, 1) no tempo . . . . .	50
Figura 20 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) em relação ao círculo unitário no plano L . . . . .	50
Figura 21 – Visualização de realização de processo ARMA(2, 1) com introdução de raiz unitária no tempo . . . . .	52
Figura 22 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) com introdução de raiz unitária em relação ao círculo unitário no plano L . . . . .	52
Figura 23 – Visualização de realização de processo ARMA(2, 1) derivado no tempo . . . . .	53



Figura 24 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) derivado em relação ao círculo unitário no plano L . . . . .	54
Figura 25 – Espectro de um processo MA(1) com $\beta_1 = -0.5$ . . . . .	62
Figura 26 – Espectro de um processo AR(1) com $\alpha_1 = 0.8$ . . . . .	63
Figura 27 – Espectro de um processo AR(1) com $\alpha_1 = -0.8$ . . . . .	63
Figura 28 – Espectro de um processo AR(2) com $\alpha_1 = 0.5$ and $\alpha_2 = -0.25$ . . . . .	64
Figura 29 – Espectro de um processo ARMA(4, 3) . . . . .	64
Figura 30 – Espectro de um processo MA(1) com $\beta_1 = -0.5$ diferenciado. . . . .	66

## **LISTA DE TABELAS**

# 1 INTRODUÇÃO

Séries temporais como uma categoria de dados ganhou tremenda importância no advento de *big data* por representar essencialmente qualquer tipo de informação que evolui no tempo. Grande parte da teoria envolvida no campo conhecido como análise de séries temporais é idêntica ou muito semelhante a conceitos de sistemas lineares e processamento de sinais, mas a compreensão dos temas na literatura científica e didática atual frequentemente não explora o potencial analítico dessa interseção. TODO: mention stochastic signal processing

O seguinte trabalho tem por objetivo abordar teoria clássica de séries temporais por um ponto de vista de sinais e sistemas. Em uma extensa revisão de literatura cada definição é introduzida notando paralelos em nomenclatura, conceito e papel analítico entre os dois campos. Eventualmente conceitos dificilmente capturados por quaisquer uma das áreas isoladamente são abordados de forma integrada e a simplicidade resultante é notável. Em uma reflexão teórica conclusiva o limite de comunicação dos campos é traçado. Por fim uma aplicação prática usando o corpo teórico desenvolvido é apresentada, na qual uma série de análises atípicas e informativas são realizadas por meio da abordagem desenvolvida.

O capítulo 2 inicia com a apresentação de uma base de definições e nomenclatura, sempre explorando os paralelos entre os campos abordados. Ao final do capítulo o conceito crítico de estacionariedade é apresentado e interpretado com devida atenção.

O capítulo 3 introduz teoria de decomposição, modelagem ARMA, modelagem sazonal e contém a interpretação mais direta de séries temporais como saídas de sistemas lineares. Essa interpretação é desenvolvida de forma a delimitar até onde as duas áreas podem se comunicar de forma a produzir análises úteis.

O capítulo 4 discute a representação espectral de séries temporais univariadas. São exploradas transformações lineares e não lineares. O desenvolvimento de uma representação espectral para uma realização teoricamente infinita de um processo estocástico é apresentada. Em seguida inicia-se uma discussão sobre representações não lineares por meio de uma generalização natural da função de autocorrelação e subsequentemente a classe de distribuições de Cohen.

O capítulo 5 brevemente discute testes estatísticos utilizados no capítulo 6.

O capítulo 6 apresenta um problema de extração de características para detecção de anomalias em uma série temporal multivariada com sazonalidade elaborada. O problema é abordado de forma prática - preocupações sobre sua resolução no mundo real são levantadas e ferramentas de operacionalização de modelos são utilizadas.

## 2 DEFINIÇÕES E PROPRIEDADES

### 2.1 Processo Estocástico

Dado um conjunto arbitrário  $\mathcal{T}$  um processo estocástico é uma família  $X(t, \omega)$ ,  $t \in \mathcal{T}$  e  $\omega \in \Omega$  de forma que para cada  $t \in T$ ,  $\omega \in \Omega$   $X(t, \Omega)$  é uma variável aleatória. As variáveis aleatórias podem ser reais ou complexas. Esse trabalho aborda apenas processos estocásticos reais exceto quando explicitamente mencionado.

Supõe-se que a família de variáveis aleatórias seja definida em um mesmo espaço de probabilidades  $(\Omega, \mathcal{A}, P)$  com  $\Omega$  representando um espaço amostral,  $\mathcal{A}$  uma  $\sigma$ -álgebra e  $P$  uma medida de probabilidade. Para propósitos desse trabalho podemos tomar o conjunto  $\mathcal{T}$  como  $\mathbb{R}$ , resultando em processos de tempo contínuo, e  $\mathbb{Z}$ , resultando em processos de tempo discreto.

Para cada  $t_k \in \mathcal{T}$  temos uma função de densidade de probabilidade associada à variável aleatória  $X(t_k, \omega)$  (assumindo que essa função exista). Na prática um processo aleatório no mundo real é observado ao longo de  $t$ , tal que  $\omega_k$  seja fixado ao universo em que a observação ocorre. Sob essa condição  $X(t, \omega_k)$  é considerada uma realização do processo estocástico. Realizações também são chamadas de *sample record* em alguns livros de engenharia e séries temporais na literatura estatística.

Para ilustrar os conceitos acima podemos pensar em um exemplo proposto por (BENDAT; PIERSOL, 2000) em que um gerador de ruído térmico é construído e sua tensão ao longo de um intervalo do tempo é medida. Se um outro gerador fosse construído sob condições e com propriedades idênticas sua tensão medida no mesmo intervalo de tempo não seria idêntica, assim como a tensão medida sob qualquer outro gerador idêntico. De fato cada registro de tensão é um exemplo de infinitos registros que poderiam ter ocorrido. Nessa situação os registros ou séries temporais de tensão são as realizações de um processo estocástico representativo de todas as possíveis realizações.

O adequado estudo de séries temporais é consequência de um primeiro adequado estudo sobre processos estocásticos, geradores dessas séries temporais. Essa não é a intenção desse trabalho. Como em grande parte da teoria de séries temporais estamos preocupados com o que podemos compreender ou inferir sobre o processo estocástico gerador de uma realização a partir apenas de seu único registro. Essa abordagem é essencial e de fato mais aplicável que um estudo que se preocupa excessivamente com os processos geradores devido ao fato de dados do mundo real frequentemente representarem realizações únicas. Não é possível realizar novamente

o índice Ibovespa entre 1970 e 2020 e muito menos ter acesso à realização desses índices em universos paralelos.

Ao longo desse trabalho a distinção e referência ao processo gerador de uma série temporal será feita quando necessário.

## 2.2 Série Temporal

Uma série temporal é um conjunto de observações realizadas sequencialmente no tempo, indexadas de acordo com o momento em que foram observadas. As observações representam a realização de um processo estocástico. Em alguns contextos, como análise de processos industriais, a natureza do processo subjacente é relevante para análise e modelagem de qualquer série temporal. Em outros, como análise de séries financeiras, o sistema gerador das séries é tão complexo que conhecimento sobre sua dinâmica dificilmente seria útil.

Assume-me, na linguagem de (HAMILTON, 1994), que um conjunto de amostras  $\mathbf{y}_t = (y_1, y_2, y_3 \dots y_T)$  pode ser interpretado como um segmento finito de uma sequência duplamente infinita:

$$\mathbf{y}_{t=-\infty}^{\infty} = (\dots, y_{-1}, y_0, \overbrace{y_1, y_2, y_3, \dots, y_T}^{\text{Série Observada}}, y_{T+1}, y_{T+2} \dots)$$

Apesar de parecer pouco tangível, de fato qualquer série observada é satisfatoriamente descrita dessa forma. Em um contexto industrial, por exemplo, o valor de uma variável de processo é zero até a planta ser construída e entrar em operação, assume valores representativos ou não da dinâmica de interesse (assumindo valores irrelevantes quando a planta não está em operação plena), e tendendo ao infinito retorna a zero quando a planta for desativada.

Essa interpretação de uma série temporal é importante ao implicitamente insinuar que o processo existe em um intervalo temporal mais abrangente do que o observado. É portanto necessário se questionar sobre quanto os dados representam o processo analisado e em quais intervalos de tempo.

Séries temporais são inerentemente diferentes de dados tabulares por representarem amostras de um mesmo processo estocástico ao invés de amostras aleatórias de uma população. Não faz sequer sentido discutir uma população no contexto de séries temporais uma vez que estamos restritos à realizações observadas de um processo estocástico. Essa relevante diferença

impede propriedades estatísticas agradáveis consequentes da independência amostral associada à dados tabulares apropriadamente amostrados como a lei do grandes números e o teorema do limite central.

A forma mais natural de analisar uma série temporal é visualizar seus valores no tempo, como ilustra a imagem 1.

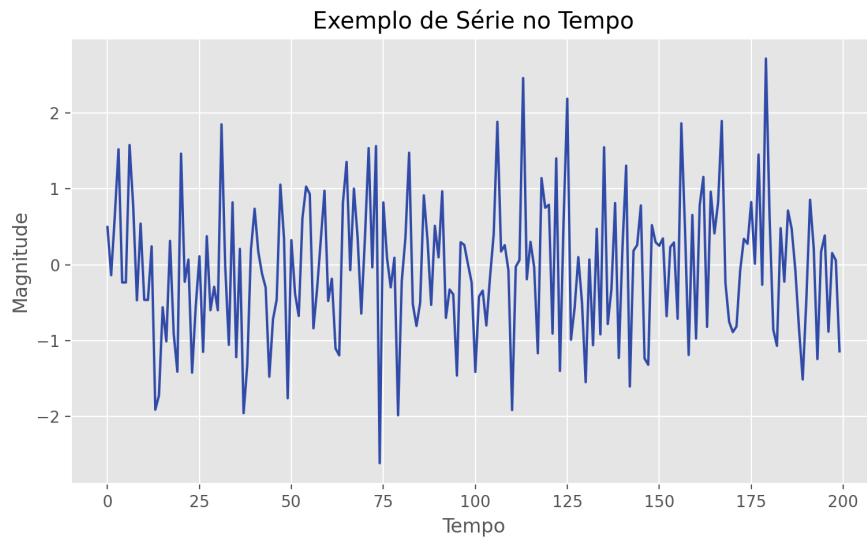


Figura 1 – Visualização no tempo de ruído branco

### 2.3 Operador de Atraso(Lag)

É importante introduzir o operador de atraso ou *lag*.

Dadas as séries  $\mathbf{y}_t = (y_1, y_2, y_3 \dots y_T)$  e  $\mathbf{x}_t = (x_0, x_1, x_2 \dots x_{T-1})$  tal que

$$\mathbf{y}_t = \mathbf{x}_{t-1}$$

isso é,

$$y_1 = x_0$$

$$y_2 = x_1$$

$$\vdots$$

$$y_T = x_{T-1}$$

Podemos definir  $\mathbf{x}_t$  em função de  $\mathbf{y}_t$  como:

$$\mathbf{x}_t = L\mathbf{y}_t$$

tal que

$$\mathbf{y}_{t-1} = L\mathbf{y}_t$$

Observamos que o operador de atraso atrasa uma série temporal em uma unidade de tempo. Uma breve divagação matemática (HAMILTON, 1994) permite definir o operador com propriedades muito semelhantes às de multiplicação dos números reais, como associatividade, comutatividade e distribuição. Para atrasar múltiplas unidades de tempo temos que:

$$L(L(\mathbf{y}_t)) = L(\mathbf{y}_{t-1}) = \mathbf{y}_{t-2} = L^2\mathbf{y}_t$$

de forma que

$$L^n\mathbf{y} = \mathbf{y}_{t-n}$$

Um uso importante do operador, decorrente de suas propriedades algébricas, é exemplificado na seguinte expansão

$$(aL^2 + bL^3)\mathbf{y}_t = a\mathbf{y}_{t-2} + b\mathbf{y}_{t-3}$$

Conhecimento do operador de atraso é importante para compreender a literatura de séries temporais e facilita comunicação objetiva de análises cotidianas. O presente trabalho usa do operador para descrever uma série de modelos.

Como nota final é importante mencionar que alguns livros (CHATFIELD; XING, 2019) (STOFFER, 2017) usam a letra  $B$  para denotar o operador de atraso e que na maior parte dos recursos *online* o operador é referido por seu nome em inglês, *lag*.

A analogia entre o operador de atraso e a variável complexa  $e^{-j\omega} = z^{-1}$  é clara, com a relevante diferença que  $z^{-1}$  atrasa um sinal em uma unidade de tempo se a operação for realizada no domínio  $z$  enquanto o operador de atraso atua diretamente no domínio do tempo. Essa característica permite que filtros com equações de recorrência complexas sejam representados de forma compacta no domínio do tempo por meio de polinômios de atraso.

Uma propriedade interessante decorrente da equivalência entre  $z^{-1}$  e  $L$  é que pode se pensar em um plano  $L$  cuja análise é igualmente informativa à do plano  $z$ , notando que o espaço é de certa forma invertido. Uma análise da posição dos polos de um sistema representado por meio de um polinômio em  $L$  conclui que o sistema é instável se tais polos estiverem fora do círculo unitário, contrário do que conhecemos do plano  $z$ .

Por fim vale mencionar que alguns autores como (AGUIRRE, 2015) usam a notação  $q^{-1}$  para esse operador.

## 2.4 Operador de Diferença

O operador de diferenças  $\nabla$  ou  $\Delta$  é o equivalente discreto da operação contínua de derivação e opera sob uma série temporal  $\mathbf{y}_t$  da seguinte forma:

$$\nabla \mathbf{y}_t = (1 - L)\mathbf{y}_t = \mathbf{y}_t - L\mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-1}$$

O operador possui propriedades de associatividade e distribuição, tal que

$$\nabla^2 \mathbf{y}_t = \nabla(\nabla(\mathbf{y}_t)) = \nabla(\mathbf{y}_t - \mathbf{y}_{t-1}) = \nabla \mathbf{y}_t - \nabla \mathbf{y}_{t-1} = \mathbf{y}_t - 2\mathbf{y}_{t-1} + \mathbf{y}_{t-2}$$

## 2.5 Tendência(*Trend*)

A variação do valor esperado de um processo estocástico é denominado tendência. A partir de uma série temporal definimos tendência como a variação de sua média amostral. A imagem 2 ilustra uma série com tendência linear. Observa-se que ao longo do tempo a média das observações cresce linearmente. Tendências de séries reais frequentemente seguem um perfil



logarítmico (CHATFIELD; XING, 2019), como na figura 3. Nesse caso uma transformação exponencial da série, isso é, a aplicação de uma função exponencial a cada observação, tornaria a tendência linear.

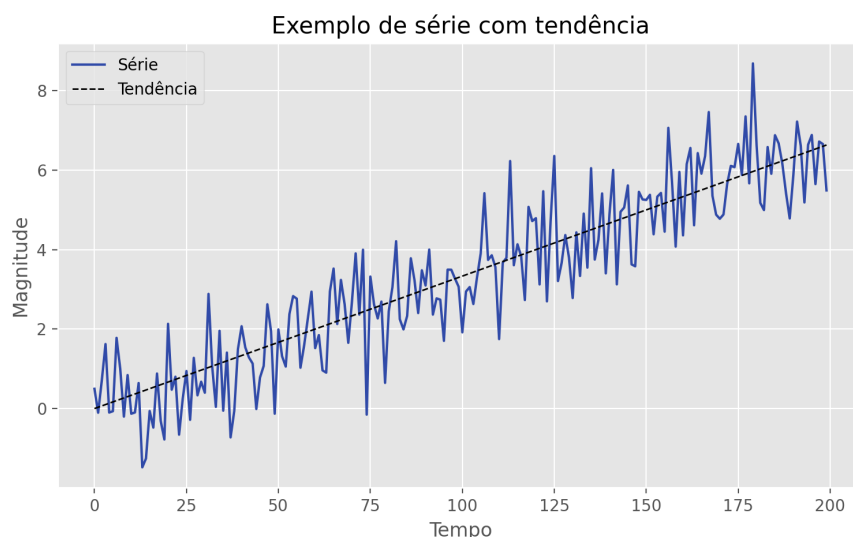


Figura 2 – Série temporal com tendência linear

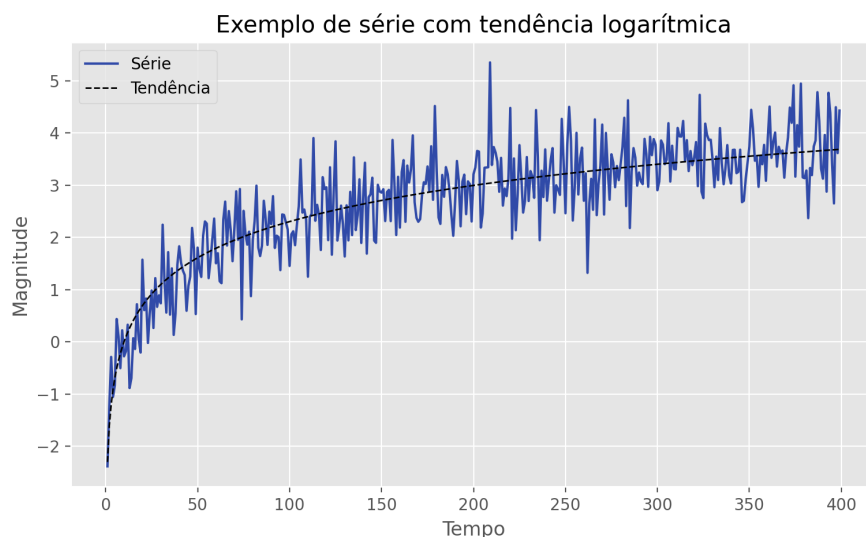


Figura 3 – Série temporal com tendência logarítmica

## 2.6 Sazonalidade

A variação periódica de média móvel das observações de uma série temporal é denominada sazonalidade. Em séries no contexto de finanças sazonalidade frequentemente segue ciclos de calendário como anual, mensal, semestral, etc. No contexto mais amplo de

séries temporais sazonalidade apresenta período arbitrário, apesar da linguagem em torno dessa propriedade estar muito associada aos períodos anteriormente mencionados.

Um exemplo de série temporal com sazonalidade é ilustrado na figura 4.

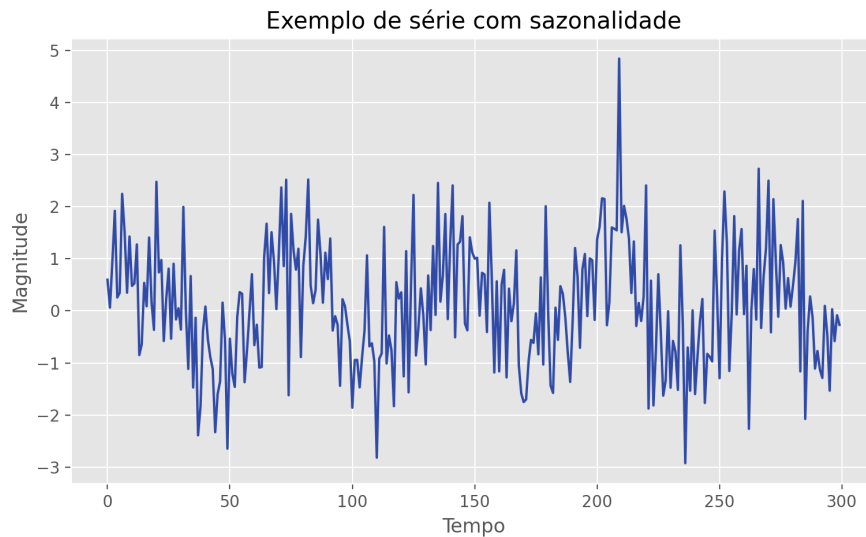


Figura 4 – Série temporal com sazonalidade senoidal de período arbitrário

Na presença de tendência sazonalidade pode ser considerada aditiva, se sua variação for constante em torno da tendência, e multiplicativa, se sua variação depender do valor da tendência. Exemplos de sazonalidade aditiva e multiplicativa são dados pelas figuras 5 e 6, respectivamente. Discernir entre os dois tipos de sazonalidade é importante para modelagem.

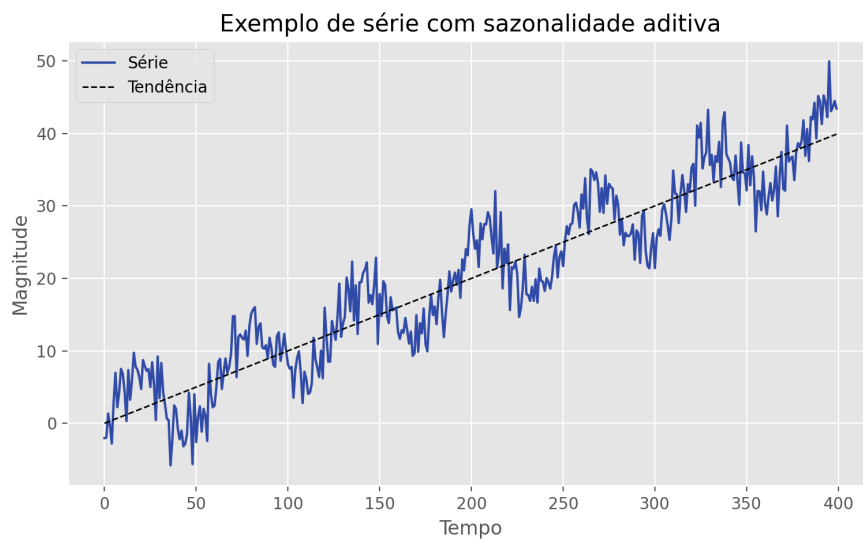


Figura 5 – Série temporal com sazonalidade aditiva

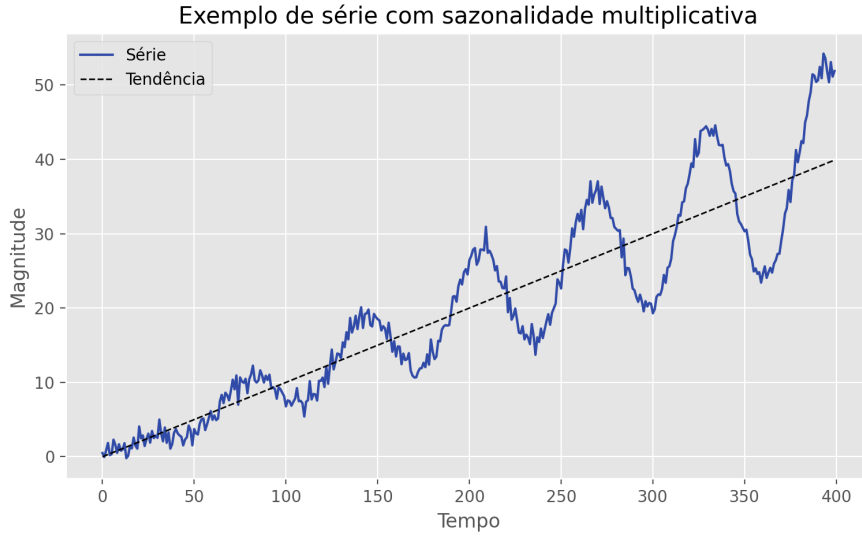


Figura 6 – Série temporal com sazonalidade multiplicativa

## 2.7 Autocorrelação

A função de autocorrelação é definida para processos estocásticos como a correlação de Pearson entre valores do processo em instantes de tempo diferentes. A função de autocovariância entre os instantes de tempo  $t_1$  e  $t_2$  é dada pela seguinte equação

$$\gamma_{xx}(t_1, t_2) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] \quad (2.1)$$

Normalizando a autocovariância obtemos a autocorrelação

$$\rho_{xx}(t_1, t_2) = \frac{\gamma_{xx}(t_1, t_2)}{\sigma_{t_1} \sigma_{t_2}}$$

Para processos estacionários (seção 2.8) a autocovariância, e consequentemente a autocorrelação, é função apenas do atraso  $\tau = |t_1 - t_2|$ . Temos então que

$$\rho_{xx}(\tau) = \frac{\gamma_{xx}(\tau)}{\sigma_{t_1} \sigma_{t_2}}$$

Para uma série temporal, isso é, uma única realização de um processo estocástico, a função de autocorrelação estacionária (tipicamente chamada apenas de função de autocorrelação) pode ser definida diretamente a partir da definição de correlação amostral sob as seguintes premissas

1. O processo estocástico gerador da série temporal é estacionário
2. O número de observações  $N$  de  $\mathbf{y}_t$  é suficientemente grande ( $N \approx 100$ )

resultando na equação 2.2, onde o subscrito duplo é omitido. Note que o atraso é discreto, indicado por  $k$ .

$$\rho_y(k) = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}, \quad k = 0, 1, 2, \dots \quad (2.2)$$

Alguns comentários sobre a nomenclatura da função são apropriados. A literatura de engenharia tende a usar os termos autocovariância e autocorrelação de forma intercambiável para designar a definição de autocovariância apresentada. A literatura estatística assume as definições abordadas acima, que serão usada no restante desse trabalho.

Além disso, a função de autocorrelação como apresentada pela equação 2.1 é definida para um processo estocástico não necessariamente estacionário, apesar do termo ser usado para descrever a equação 2.2. A aplicação da equação que assume estacionariedade em uma série não estacionária resulta em correlações informativas praticamente apenas disso. Uma modificação estratégica da função dada por 2.1 resulta na chamada função de autocorrelação instantânea (seção 4.2.1), que é usada para representar séries temporais não estacionárias.

### 2.7.1 Autocorrelação Parcial

É interessante mencionar a existência de autocorrelação parcial nessa seção juntamente de uma descrição em alto nível do que esse valor representa. Sua definição formal será apresentada na seção 3.2.4.

Autocorrelação parcial foi introduzida por Box e Jenkins em (BOX; JENKINS, 2016) como uma ferramenta auxiliar na identificação de modelos. O valor  $\phi_{kk}$  é definido como a correlação parcial entre  $\mathbf{y}_t$  e  $\mathbf{y}_{t-k}$ , isso é, a correlação restante entre  $\mathbf{y}_t$  e  $\mathbf{y}_{t-k}$  após levar em consideração a contribuição de  $\mathbf{y}_t, \mathbf{y}_{t-1} \dots \mathbf{y}_{t-k+1}$ .

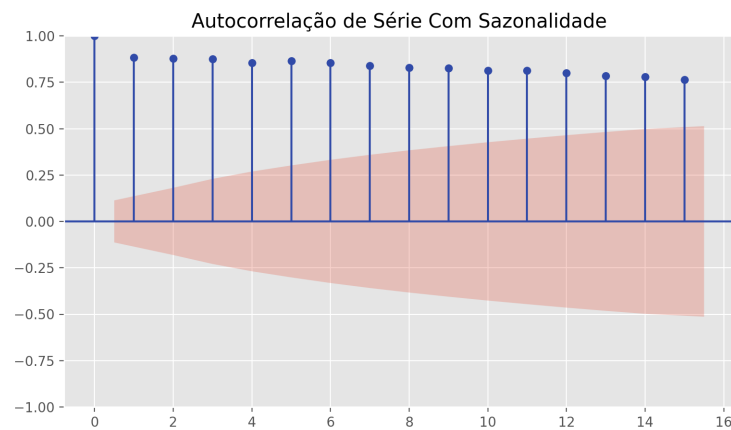


Figura 7 – Visualização de correlatograma de série com tendência linear. Observe que as autocorrelações decaem lentamente ao decorrer dos atrasos, comportamento típico de tendências determinísticas.

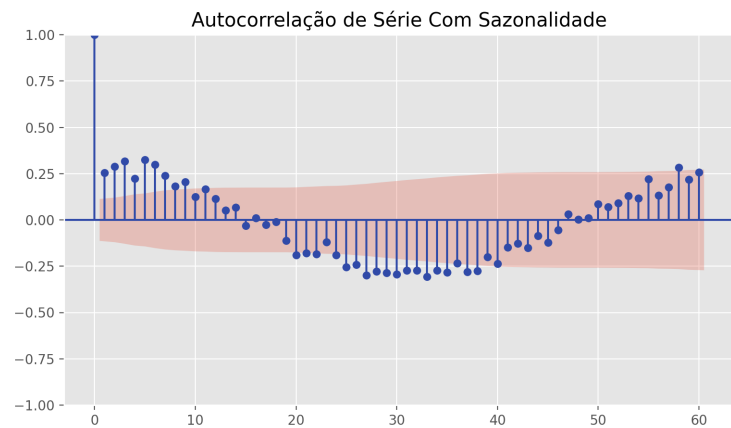


Figura 8 – Visualização de correlatograma de série com sazonalidade. Observe que o padrão senoidal da série é reproduzido nas autocorrelações.

### 2.7.2 Correlatograma

Um correlatograma é um gráfico de barras representativo da autocorrelação ou autocovariância em  $k$  amostras de uma série temporal ( $k = 0, 1, 2, \dots$ ), de forma que a primeira barra represente a autocorrelação entre  $y_t$  e si mesmo (sempre igual à 1), a segunda entre  $y_t$  e  $y_{t-1}$ , a terceira entre  $y_t$  e  $y_{t-2}$ , e assim por diante. O correlatograma da série visualizada pela figura 2 é ilustrado na figura 7.

A figura 7 informa um intervalo de relevância dado por um sombreamento vermelho. Qualquer valor de autocorrelação dentro desse intervalo é estatisticamente insignificante e pode ser considerado igual a zero.

O correlatograma é uma ferramenta indispensável em análise de séries temporais

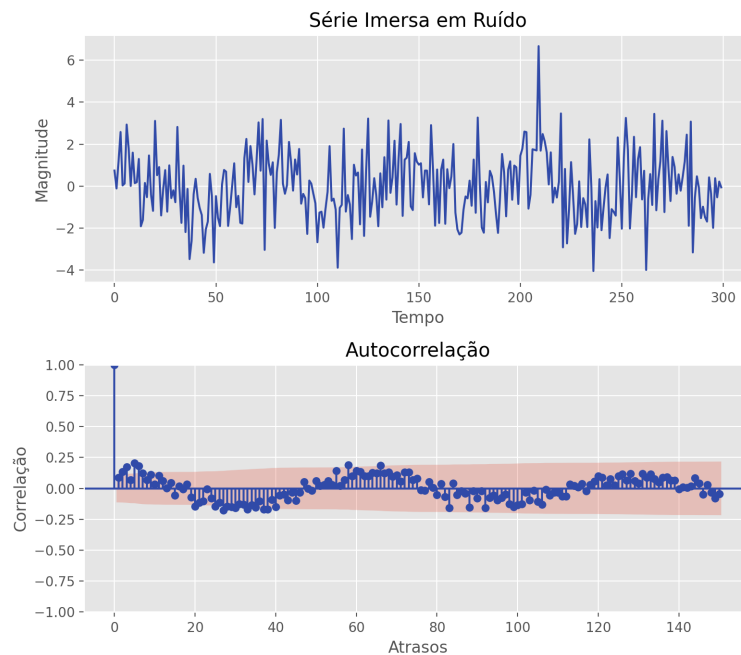


Figura 9 – Sinal periódico imerso em ruído e sua correspondente autocorrelação.

para tarefas como detecção de estacionariedade, identificação de sazonalidade, análise de resíduo, engenharia de características, escolha de modelo e identificação de ruído branco (seção 2.9), entre outros. No contexto de identificação de sistemas autocovariância e autocorrelação e portanto o correlalograma desempenham um importante papel na identificação de propriedades de sinais e sistemas imersos em ruído devido à robustez ao ruído da operação de correlação cruzada (AGUIRRE, 2015). A figura 9 apresenta um sinal imerso em ruído cuja natureza periódica subjacente se torna mais visível por meio de seu correlalograma.

É importante mencionar que o correlalograma de uma série com tendência determinística, como a da figura 7, apresenta o comportamento observado de autocorrelações altas com pouca atenuação ao longo dos atrasos. De forma análoga o correlalograma de uma série com sazonalidade apresenta periodicidade que reproduz seu padrão temporal, como ilustra a figura 8, correlalograma da série da figura 4. O primeiro correlalograma é informativo até certo ponto: informa simplesmente que a série apresenta tendência. Para analisar tais séries de forma mais produtiva, a fim de elaborar um possível modelo, por exemplo, é importante que a série seja estacionária (seção 2.8). É inclusive afirmado em alguns textos da literatura estatística (CHATFIELD; XING, 2019), que um correlalograma só faz que sentido se a série associada for estacionária, observação mais geral e rigorosa das restrições de uso da equação 2.2.

Na engenharia a análise do correlalograma de sinais não estacionários é utilizada para investigação da adequação de tempo de amostragem, onde uma autocovariância com valores

lentamente decrescentes e um mínimo local indica que o sinal pode estar superamostrado, propriedade indesejável que pode resultar em problemas computacionais além de desperdício de memória.

## 2.8 Estacionariedade

Um processo estocástico  $\mathbf{X}(t)$  é considerado estacionário no sentido amplo se atender às seguintes três condições:

1.  $E(\mathbf{X}(t)) = \mu$
2.  $Var(\mathbf{X}(t)) = \sigma^2$
3.  $Cov[\mathbf{X}(t), \mathbf{X}(t + \tau)] = \gamma(\tau)$

e estacionário no sentido restrito se sua distribuição de probabilidade  $P(x)$  for idêntica para todos os instantes de tempo, isso é,  $P(x)_{t_i} = P(x)_{t_j} \forall i, j \in \mathcal{T}$ . Essa restrição é frequentemente comunicada como a necessidade de todos os momentos da distribuição  $P(x)$  serem idênticos e invariantes ao tempo. Nesse trabalho, assim como em grande parte da literatura estatística e de engenharia, o termo estacionário se refere a estacionariedade no sentido amplo. Isso é parcialmente justificado pelo fato de verificação de estacionariedade no sentido amplo tipicamente ser condição suficiente para assumir estacionariedade no sentido restrito, segundo (BENDAT; PIERSOL, 2000).

A definição de estacionariedade apresentada descreve a família de sinais representada por um processo estocástico. Uma definição de estacionariedade para realizações únicas de um processo estocástico, isso é, séries temporais, demanda o estabelecimento de requisitos amostrais. Podemos dizer que uma série temporal é estacionária se suas propriedades amostrais de esperança, variância e autocovariância sejam apropriadamente invariantes de forma que flutuações em seu valor sob diferentes janelas de tempo sejam explicáveis por variações resultantes de amostragem. Sob essa perspectiva podemos traduzir os requisitos de estacionariedade para uma série  $\mathbf{x}_t$  como a seguir

1. A média da série  $\mathbf{x}_t$  é constante ao longo do tempo
2. A variância da série  $\mathbf{x}_t$  é constante ao longo do tempo
3. A autocorrelação de  $\mathbf{x}_t$  depende apenas do atraso

Um bom exemplo de graus de estacionariedade em séries temporais dado por Athana-

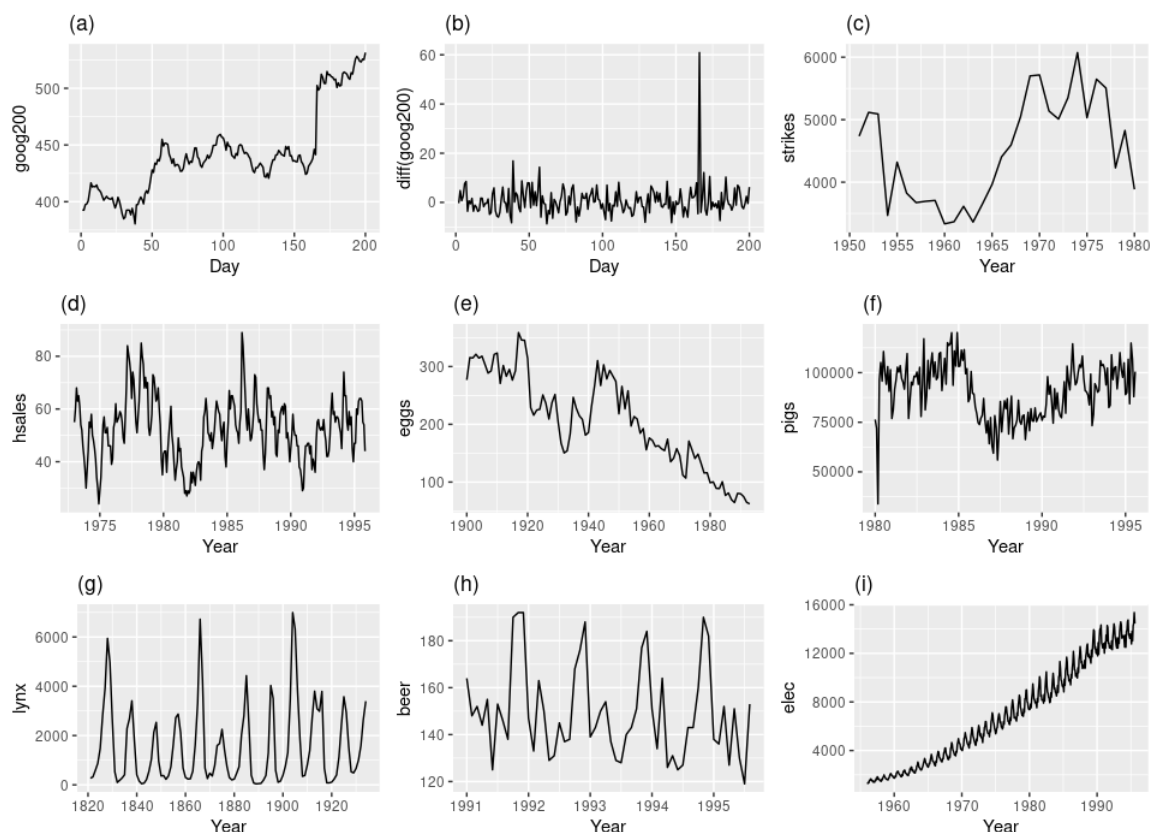


Figura 10 – Conjunto de séries demonstrando diferentes níveis de estacionariedade.

sopoulos e Hyndman (ATHANASOPOULOS; HYNDMAN, 2018) é ilustrado pela figura 10. As séries das figuras 10.a, 10.e e 10.i demonstram clara tendência, sendo portanto não estacionárias. As séries das figuras 10.d, 10.h, 10.i são igualmente não estacionárias por apresentarem clara sazonalidade, enquanto a série da figura 10.g aparenta ser sazonal mas apresenta picos aperiódicos de intensidade muito distinta, sendo considerada estacionária pelos autores. O caso da série da figura 10.g é um complicado por aparentar também violar o requisito de variância constante. Para concluir assim como os autores que a série é estacionária é necessário mais do que uma investigação visual; o texto que discute a série leva em conta seu processo gerador (ATHANASOPOULOS; HYNDMAN, 2018).

As séries das figuras 10.c e 10.f são aparentemente não estacionárias no intervalo observado por demonstrarem variação em sua média móvel mas podem ser um caso de raízes unitárias.

Por eliminação temos que apenas as séries ilustradas pelas figuras 10.b e 10.g são estacionárias, o caso de 10.b contendo uma clara anomalia.

O exemplo de análise de estacionariedade da figura 10 demonstra a imprecisão da abordagem visual para essa tarefa. Torna-se necessário o estabelecimento de procedimentos mais



objetivos para detecção de estacionariedade.

### 2.8.1 Importância de Estacionariedade

Estacionariedade é uma propriedade desejável de se observar em uma série temporal para fins analíticos e de modelagem.

Há algumas formas de compreender como estacionariedade colabora para a modelagem bem sucedida de uma série temporal.

Podemos pensar na propriedade de estacionariedade como um tipo de estrutura de dependência. Se as amostras  $X_1, X_2, \dots, X_N$  de um conjunto  $\mathbf{X}$  forem independentes entre si temos formas interessantes de modelar a função geradora de  $\mathbf{X}$  como o teorema do limite central, lei dos grandes números, etc. Há apenas uma forma de um conjunto amostral ser independente mas muitas formas de ser dependente, tornando difícil o estabelecimento de recursos eficientes para modelagem geral de processos dependentes. Séries temporais sendo naturalmente observações de processos dependentes (seção 2.2) é interessante definir estruturas de dependência que permitam o uso de tais recursos. Estacionariedade é uma estrutura de dependência que permite aplicar propriedades úteis de independência em séries temporais. Abordando a mesma ideia mais intuitivamente podemos pensar no seguinte exemplo: se um processo possui valor esperado, variância constante e autocorrelação invariante ao tempo podemos por meio da lei dos grandes números estimar seu valor esperado e variância com cada vez mais confiança a partir da média e variância amostral, respectivamente. O mesmo argumento intuitivo se estende analogamente para o caso de aprendizado de máquina, no qual por meio de teoria de aprendizado estatístico é possível argumentar que uma série estacionária é “mais fácil” de aprender.

De forma mais quantitativa o teorema de decomposição de Wold (CHATFIELD; XING, 2019) conclui que qualquer série temporal estacionária pode ser representada pela seguinte combinação linear

$$\mathbf{y}_t = \sum_{j=0}^{\infty} b_j Z_{t-j} + \eta_t$$

No qual  $\eta$  representa uma série determinística e  $Z_t$  um processo puramente aleatório (seção 2.9). O leitor reconhecerá parte da expressão acima como um processo  $MA(\infty)$  (seção 3.1.1.1). Esse resultado tem como consequência a importante conclusão que qualquer série estacionária é possivelmente aproximável por um modelo MA e portanto, via invertibilidade,

modelos AR e ARMA.

Por fim estacionariedade permite o uso de uma série de modelos que serão discutidos na seção sobre modelos estacionários. Esses métodos são bem compreendidos e implementados, facilitando sua interpretação, uso e sustentação.

### 2.8.2 *Categorias Básicas de Não Estacionariedade*

Como extensão do argumento sobre estruturas de dependência na seção anterior podemos afirmar que, sendo estacionariedade um padrão de dependência, temos infinitas formas de não estacionariedade, retornando ao caso de dependência generalizada. É interessante identificar nesse universo de dependência padrões de séries não estacionárias que são facilmente transformadas em séries estacionárias.

Uma série temporal com presença de tendência determinística, como ilustrada na figura 2, pode ser representada pela seguinte expressão:

$$y_t = e_t + f(t) + \varepsilon_t \quad \text{onde } \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

Na qual  $e_t$  representa uma série estacionária,  $f(t)$  uma função determinística do tempo e  $\varepsilon_t$  ruído branco (seção 2.9). Nota-se que  $f(t)$  é uma função monotônica arbitrária tal que  $y_t$  seja uma série não estacionária. No caso da figura 2 temos  $f(t)$  linear e na figura 3 logarítmica. Uma série temporal demonstrando esse tipo de não estacionariedade é considerada **tendência-estacionária**, uma vez que simplesmente removendo a tendência  $f(t)$  temos estacionariedade. Isso pode ser feito de várias formas, talvez com maior simplicidade diferenciando a série. Métodos mais sofisticados incluem decomposição ETS (seção 3.1) e regressão com finalidade de modelar  $f(t)$  de forma que o resíduo represente uma série estacionária  $e_t + \varepsilon_t$ .

Uma série com presença de tendência estocástica pode ser classificada de maneira semelhante. O exemplo mais simples de tal série é gerada por um passeio aleatório, definido pelo seguinte processo, visualizado pela figura 11:

$$y_t = y_{t-1} + \varepsilon_t \quad \text{onde } \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

Por meio de um desenvolvimento recursivo do processo podemos escrever:

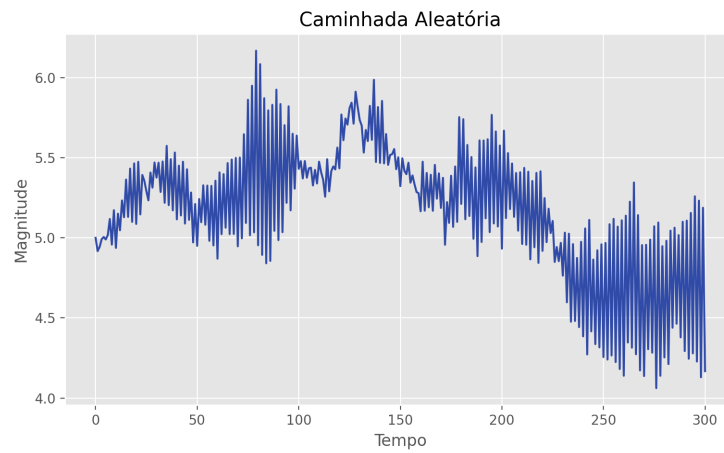


Figura 11 – Visualização de caminhada aleatória com  $y_0 = 5$

$$y_t = (y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$y_t = ((y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t$$

$$\vdots$$

$$y_t = \sum_{j=0}^{N-1} \varepsilon_{t-j} + y_0$$

Resultado a partir do qual a não estacionariedade de  $y_t$  se torna evidente, uma vez que

$$\text{var}(y_t) = \sigma^2 t$$

Além da covariância ser dependente do tempo.

Uma forma simples de tornar a série estacionária é diferenciá-la em primeira ordem, isso é, aplicar o operador de diferença primeira:

$$\nabla y_t = y_t - y_{t-1}$$

$$y_t - y_{t-1} = \varepsilon_t$$

$$\nabla y_t = \varepsilon_t$$

Sabemos pela seção 2.9 que ruído branco é um processo estacionário.

A caminhada aleatória é denominada uma série **diferença-estacionária** pelo fato da operação de diferença introduzir estacionariedade. Essa é uma forma tão comum de não estacionariedade que a ideia de “diferenciar uma série antes de fazer qualquer coisa” é proeminente entre profissionais de dados, apesar de que geralmente necessita-se apenas de estacionariedade. Essa prática é parcialmente justificada considerando que a maior parte das séries temporais “reais” são não estacionárias e frequentemente diferencialmente estacionárias.

É igualmente possível que uma série diferença-estacionária seja estacionária em sua  $n$ -ésima diferença, tal que estacionariedade seja observada por uma operação de diferenças de ordem  $n$ . A ideia de tirar sucessivas diferenças até atingir estacionariedade é fundamental no método de Box-Jenkins, por exemplo.

Séries diferença-estacionárias apresentam raízes unitárias e os dois termos são frequentemente usados nos mesmos contextos.

Podemos resumir as definições das categorias de não estacionariedade abordadas nessa seção assim como suas implicações como segue:

1. **Estacionariedade em Tendência:** Uma série é considerada tendência-estacionária se apresentar uma tendência determinística. No caso de anomalias ou eventos de perturbação séries com esse tipo de tendência retornam ao valor da tendência ao longo do tempo, “esquecendo” o evento perturbador. Esse tipo de série se torna estacionária pela remoção da tendência determinística, processo realizado por meio de regressão da tendência, por diferenciação, por decomposição, etc.
2. **Estacionariedade Diferenciável:** Uma série é considerada diferença-estacionária se apresentar uma tendência estocástica. No caso de anomalias ou eventos de perturbação séries com esse tipo de tendência são irreversivelmente afetadas, “lembrando” do evento perturbador. Esse tipo de série se torna estacionária por diferenciação em ordem  $n$ . Possui raízes unitárias e é frequentemente discutida nessa linguagem.

## 2.9 Ruído Branco

Uma série temporal  $\mathbf{x}_t$  gerada por um processo  $\mathbf{X}(t)$  é considerada ruído branco ou um processo puramente aleatório se atender às seguintes três condições:

1.  $E(\mathbf{X}(t)) = 0$
2.  $Var(\mathbf{X}(t)) = \sigma^2$
3.  $Cov[\mathbf{X}(t), \mathbf{X}(t + \tau)] = 0$

Que podem ser interpretadas da seguinte forma

1. A média da série  $\mathbf{x}_t$  é nula ao longo do tempo
2. A variância da série  $\mathbf{x}_t$  é constante ao longo do tempo
3. Não há correlação entre as amostras de  $\mathbf{x}_t$

Observa-se que ruído branco é um caso específico de estacionariedade, se diferenciando pela especificação do valor esperado e autocorrelação entre quaisquer amostras em zero. A compreensão da definição e capacidade de identificação de ruído branco é importante para análise de resíduos.

### 3 TEORIA UNIVARIADA

#### 3.1 Decomposição de Séries Temporais

Decomposição em séries temporais tipicamente descreve o processo de representação de uma série por uma combinação linear de três componentes: tendência, sazonalidade e resíduos. A decomposição de uma série  $y_t$  pode ser aditiva ou multiplicativa como expressada pelas equações 3.1 e 3.2 respectivamente.

$$y_t = S_t + T_t + R_t \quad (3.1)$$

$$y_t = S_t \cdot T_t \cdot R_t \quad (3.2)$$

A escolha de decomposição aditiva ou multiplicativa deve ser feita de acordo com o tipo de sazonalidade (seção 2.6).

Como herança de econometria grande parte dos algoritmos clássicos de decomposição (X11 (LADIRAY; QUENNEVILLE, 2001), SEATS (BIANCIONINI, 2016, capítulo 5.2) e derivados) são baseados em período de sazonalidade anuais, semestrais, trimestrais e mensais (ATHANASOPOULOS; HYNDMAN, 2018). A incapacidade desses algoritmos de processar dados com período menor os torna pouco utilizáveis no contexto de sinais elétricos, por exemplo, cujas séries tipicamente são de período inferior a um dia, apresentando padrões sazonais com período semelhante.

Antes de prosseguir é interessante comentar sobre os principais objetivos de decomposição:

1. **Análise:** O processo de pensar sobre, elaborar, ajustar e observar resultados de decomposição são altamente informativos da natureza da série sob análise. A operação é frequentemente usada para fins de análise exploratória.
2. **Indução de estacionariedade:** Como tendência sempre é e sazonalidade pode ser um padrão cuja presença qualifica não estacionariedade sua remoção pode tornar uma série “mais estacionária” ou pelo menos mais apropriada para modelagem por meio de modelos não estacionários.

3. Detecção de Anomalias: Eventos anômalos são tipicamente dissociados de e ocultos por estruturas de tendência e sazonalidade. A remoção dessas estruturas tende a expor anomalias de forma mais detectável.
4. Modelagem e Previsão: A representação de uma série por meio de três componentes distintas é interessante para problemas de previsão pela possibilidade de desenvolver modelos e previsões para cada componente de acordo com suas propriedades. Os modelos independentes tendem a ser menos complexos, mais generalizáveis e mais robustos do que um modelo adequado para a série original, se existir.

### 3.1.1 Modelagem de Tendência

Uma operação essencial em decomposição de séries temporais é a determinação de uma tendência subjacente dos dados. Os principais procedimentos para tal são abordados a seguir.

#### 3.1.1.1 Média Móvel

Um filtro de média móvel de ordem  $m$  é tipicamente expressado como:

$$MA_{t_m} = \frac{1}{m} \sum_{i=-k}^{i=k} y_{t+i} \quad (3.3)$$

A operação é visualizada pela figura 12.

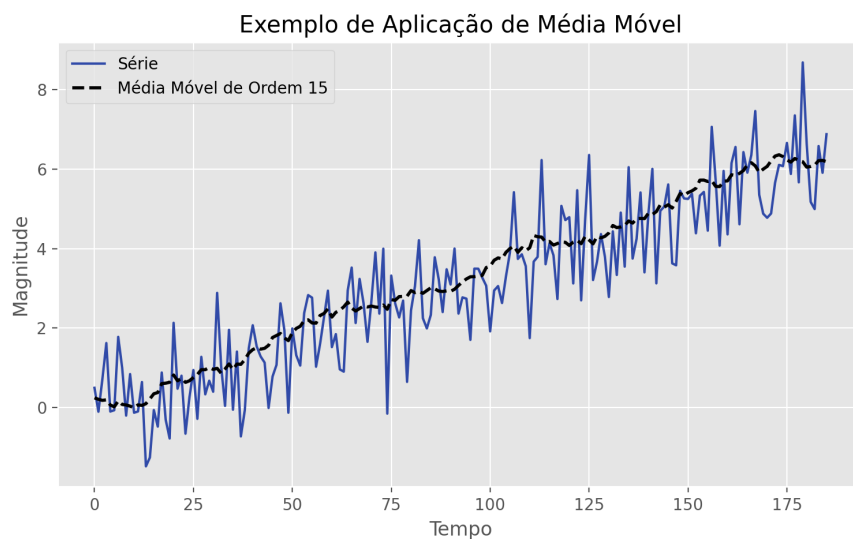


Figura 12 – Visualização de aplicação de filtro média móvel com  $m=15$

### 3.1.1.2 Regressão Localizada/Filtro de Savitzky-Golay

Uma forma mais robusta e granular de modelagem de tendência é o algoritmo LOESS, baseado em regressão localizada e conhecido em linguagem de processamento de sinais como o filtro de Savitzky-Golay. Apesar do algoritmo ser simples seu desenvolvimento será omitido por ser matematicamente verboso. É importante que o leitor conheça a existência do método LOESS, sabendo que ele oferece uma alternativa mais precisa para modelagem de tendência do que média móvel, apesar de ser potencialmente mais computacionalmente complexo. Uma breve intuição sobre seu funcionamento é apresentada a seguir.

A ideia fundamental do algoritmo é repartir a série temporal em grupos de pontos menores, calculando uma regressão (tipicamente mas não necessariamente linear) nesses grupos menores de forma a construir uma curva linear por partes que aproxima a tendência da curva original. O tamanho dos grupos é arbitrário e tipicamente informado como uma fração do tamanho da série completa e os grupos em si são determinados pelos  $n$  pontos mais próximos a um elemento chamado de “ponto focal” que é incrementado a cada iteração. O valor dos incrementos dos pontos focais também é um parâmetro.

A figura 13 ilustra o resultado do algoritmo LOESS para a mesma série analisada pela figura 12. A figura ilustra curvas de tendência modeladas via LOESS para diferentes tamanhos de grupos. Os tamanhos são informados como porcentagem do comprimento total do sinal. Observa-se que o uso de uma porcentagem menor do sinal aumenta a localidade da regressão tornando o ajuste da tendência mais sinuosa.

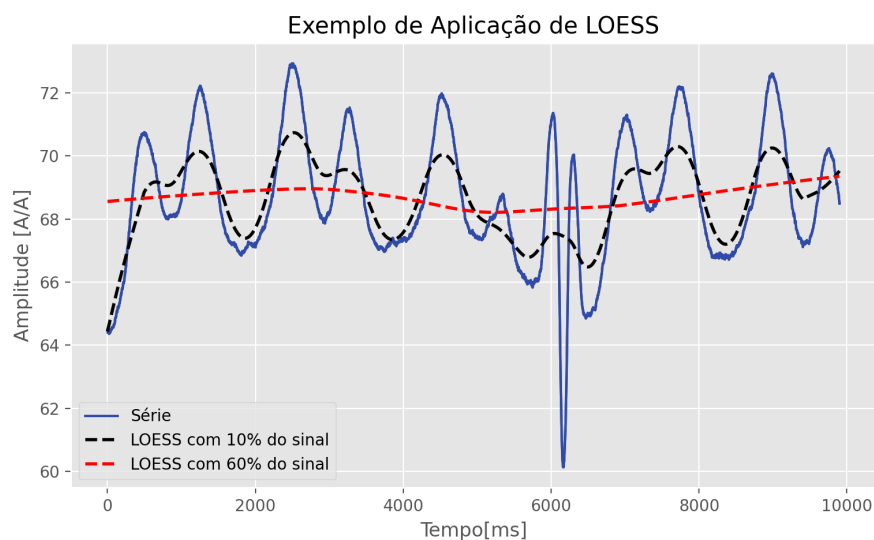


Figura 13 – Visualização de algoritmo LOESS



### 3.1.2 Modelagem de Sazonalidade(Ajuste de Sazonalidade)

A operação fundamental de decomposição é a modelagem ou ajuste de sazonalidade. Esse processo se preocupa em identificar estruturas periódicas em uma série temporal. Modelagem de sazonalidade em séries temporais é de fato uma área em si (HYLLEBERG, 1992). A seção 3.5 é dedicada à uma pequena exploração desses métodos. Abordamos a seguir sua versão mais simples, tipicamente implementada no processo de decomposição de séries temporais.

Uma possibilidade para estimativa da componente sazonal é determinar o “padrão sazonal médio” da série sem tendência. A série de tamanho  $n$  sem tendência é dividida em  $s$  segmentos de  $m$  amostras, no qual  $s = \frac{n}{m}$ . Usando livremente a vírgula como símbolo para concatenação podemos enumerar os segmentos como

$$\chi_1, \chi_2, \dots, \chi_s$$

onde um segmento arbitrário

$$\chi_i = \chi_{i_1}, \chi_{i_1} \dots \chi_{i_m}$$

O padrão sazonal modelado  $S'_t$  corresponde ao segmento médio, isso é,

$$S'_t = \left( \frac{1}{s} \sum_{k=1}^{k=s} \chi_{1_k} \right), \left( \frac{1}{s} \sum_{k=1}^{k=s} \chi_{2_k} \right), \left( \frac{1}{s} \sum_{k=1}^{k=s} \chi_{3_k} \right) \dots \left( \frac{1}{s} \sum_{k=1}^{k=s} \chi_{m_k} \right)$$

A componente sazonal  $S_t$  é então dada pela concatenação de  $m$  cópias de  $S'_t$

$$S_t = \overbrace{S'_t, S'_t, \dots, S'_t}^{s \text{ vezes}}$$

### 3.1.3 Decomposição Clássica (Aditiva)

O algoritmo de decomposição clássica aditiva é simples e capaz de identificar períodos arbitrários de sazonalidade. O procedimento é apresentado passo a passo.

#### 3.1.3.1 Passo 1

**Identificar o período de sazonalidade que se deseja modelar.**

O período  $m$  de sazonalidade corresponde à duração de um ciclo de um padrão periódico observável nos dados. A determinação da duração desse ciclo não é necessariamente

trivial, especialmente tratando de séries de alta frequência, sendo interessante inclusive usar métodos automáticos para determinação da distância entre amostras iniciais dos períodos de sazonalidade. Uma possibilidade, por exemplo, é analisar os picos da função de autocorrelação e acusar a distância entre os picos como um período amostral.

#### 3.1.3.2 *Passo 2*

##### **Modelar tendência $T_t$**

A tendência  $T_t$  é modelada por meio de algum dos métodos mencionados, por exemplo, um filtro média móvel com tamanho de janela  $w$ .

#### 3.1.3.3 *Passo 3*

##### **Remover tendência da série**

A componente sem tendência é dada por  $y_t - T_t$ . A série resultante deve possuir média aproximadamente nula.

#### 3.1.3.4 *Passo 4*

##### **Estimar componente sazonal**

A componente sazonal de  $S_t$  é estimada por meio de algum dos métodos mencionados. Para a maioria dos métodos precisamos informar o período sazonal  $m$ .

#### 3.1.3.5 *Passo 5*

##### **Calcular a componente residual**

Por fim calcula-se a componente residual  $R_t$  como

$$R_t = y_t - T_t - S_t$$

A série está enfim descomposta

$$y_t = R_t + T_t + S_t$$

### 3.1.4 Qualificação de tendência e sazonalidade

A presença de tendência ou sazonalidade é tipicamente visível imediatamente. Dúvida sobre a presença ou não dessas propriedades pode surgir se o sinal estiver imerso em ruído. Se houver presença de raízes unitárias curtos intervalos de tempo podem ser ambíguos quanto ao determinismo de sua tendência, isso é, uma tendência estocástica pode ser confundida com tendência determinística. Nesses cenários é interessante detectar a presença desses componentes por meio de testes estatísticos ou análise de correlatograma.

### 3.1.5 Quantificação de tendência e sazonalidade

#### 3.1.5.1 Força

A decomposição de uma série temporal em componentes isolados de tendência, sazonalidade e resíduos permite que a intensidade de tendência e sazonalidade sejam quantificáveis de forma elegante. Essa intensidade de tendência ou sazonalidade é tipicamente chamada de força.

Para séries com forte tendência é esperado que a componente de tendência contenha considerável variância. Uma forma interessante de quantificar a força da tendência de uma série, proposta por Hyndman et. al. (ATHANASOPOULOS; HYNDMAN, 2018), parte da observação da variância adicional introduzida pela adição do termo de tendência ao residual:

$$\frac{Var(R_t)}{Var(T_t + R_t)}$$

Espera-se que essa razão seja pequena para séries com alta tendência, isso é, haja introdução de considerável variância pela adição da componente de tendência no denominador. Podemos então definir a força  $F_T$  da tendência de uma série como

$$F_T = \max \left( 0, 1 - \frac{Var(R_t)}{Var(T_t + R_t)} \right)$$

Observe que  $F_T \in real$  limitado entre 0 e 1.

De forma identicamente análoga podemos definir a força da sazonalidade de uma série como

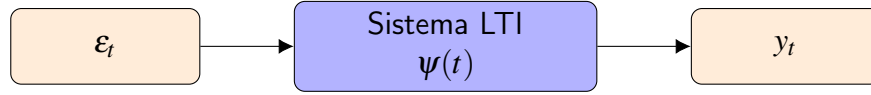


Figura 14 – Representação de série temporal como modelo linear generalizado

$$F_S = \max \left( 0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right)$$

Essas medidas são úteis ao oferecer uma interface quantitativa aos atributos tipicamente qualitativos de tendência e sazonalidade. É importante deixar claro que essas medidas necessitam que a série seja decomposta em suas componentes de tendência e sazonalidade, possivelmente dificultando sua aplicação.

## 3.2 Modelos Estacionários

### 3.2.1 O Modelo Estacionário como um Filtro Linear

Segundo Box e Jenkins (BOX; JENKINS, 2016, capítulo 1.2.1) os modelos estacionários que serão abordados nas seguintes seções foram idealizados por (YULE, 1927) como filtros lineares que modelam a série temporal de interesse por meio do processamento de ruído branco. A modelagem apropriada do sinal em questão se resume então ao apropriado ajuste da função de transferência  $\psi$  desse filtro a partir dos dados observados, isso é, o apropriado posicionamento de zeros e/ou polos.

A função de transferência  $\psi$  é definida no domínio do tempo por meio do operador de atraso  $L$  ao invés de  $z^{-1}$ . Como no contexto de filtros digitais podemos pensar em um plano  $L$ , em analogia com o plano  $Z$ , e chegar às exatas mesmas conclusões sobre o efeito do posicionamento de polos e zeros na estabilidade (inclusive marginal), invertibilidade e resposta em frequência do filtro. A diferença relevante é que o operador de atraso no domínio  $Z$ ,  $z^{-1}$ , é definido de forma inversa à referência do plano. Essa inversão não acontece no plano  $L$ , levando a uma inversão das propriedades conhecidas.

É possível também descrever certos tipos de modelos-filtros como *Finite Impulse Response*(FIR) ou *Infinite Impulse Response*(IIR), dependendo da equação de recorrência, levando às propriedades conhecidas desses tipos de filtros.

A figura 14 torna clara a interpretação traçada. Para tornar o paralelo mais claro

podemos inicialmente pensar em um sinal de entrada  $\varepsilon[n]$  ( ruído branco) transformado por uma resposta ao impulso  $h[n]$ , causal de tamanho  $m$ , no sinal  $y[n]$  pela seguinte soma de convolução

$$y[n] = \sum_{k=0}^m h[n-k]\varepsilon[k] = \sum_{k=0}^m \varepsilon[n-k]h[k]$$

Introduzindo o operador de atraso  $L$  no somatório

$$y[n] = \sum_{k=0}^m L^k h[k]\varepsilon[n]$$

observamos que temos um polinômio em  $L$  cujos coeficientes correspondem aos da resposta ao impulso do filtro. Podemos então, chaveando para notação de séries temporais, reescrever a convolução como

$$\sum_{k=0}^m \psi(k)L^k \varepsilon_t = \sum_{k=0}^m \psi(k)\varepsilon_{t-k}$$

Expandindo as somas temos

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_m \varepsilon_{t-m}$$

$$y_t = \varepsilon_t + \psi_1 L \varepsilon_t + \psi_2 L^2 \varepsilon_t + \dots + \psi_m L^m \varepsilon_t$$

$$y_t = \varepsilon_t (1 + \psi_1 L + \psi_2 L^2 + \dots + \psi_m L^m)$$

$$y_t = \psi(L)\varepsilon_t \tag{3.4}$$

onde  $\psi(L)$  é um polinômio mônico em  $L$ .

A interpretação de modelos estacionários como filtros tem suas limitações no que diz respeito às possibilidades de herança de conhecimento da área de processamento de sinais. Inicialmente podemos constatar que filtros lineares no sentido tradicional se propõem em alterar as propriedades de um sinal arbitrário que por si já contém informação. Isso leva à importância imediata da função de impulso unitário como base natural de sinais em tempo discreto e a uma enorme preocupação com a fase da resposta ao impulso do filtro devido à facilidade de distorção

de fase indesejada do sinal processado. Para o modelo-filtro a função de impulso unitário não é tão importante porque esse filtro é sempre excitado pelo mesmo tipo de sinal que não possui nenhum tipo de estrutura a se preservar: ruído branco. A estrutura de fase desse sinal é menos relevante ainda por ser aleatória. Essas diferenças tornam muita da teoria de filtros inaplicável no contexto de séries temporais. Não há sentido de projetar um modelo filtro de fase linear generalizada, por exemplo, se não há o que preservar na fase do sinal de entrada e nenhuma preocupação com atraso de grupo.

Não obstante a interpretação de modelos estacionários como filtros lineares facilita a assimilação de propriedades desses modelos por meio da linguagem de processamento de sinais e nos leva a interessantes explorações de parte da teoria de filtros aplicada à modelagem de séries temporais.

Por fim podemos notar que nesse contexto o filtro tradicional representa uma função linear do processo estocástico de ruído branco  $\varepsilon(t)$  para a saída  $Y(t)$ , qualificando a saída em si como um processo estocástico. Conforme discutido na seção 2.1 abordaremos as propriedades de realizações específicas do processo  $Y(t)$ .

### 3.2.2 *Modelo Linear Generalizado*

É interessante expandir a discussão anterior para a definição de um modelo linear generalizado (GLM).

Tomando a equação 3.4 com  $m \rightarrow \infty$  temos um modelo linear generalizado que corresponde ao processamento de ruído branco por um filtro linear com resposta ao impulso de duração infinita:

$$y_t = \psi(L)\varepsilon_t = \sum_0^{\infty} \psi_m \varepsilon_t \quad (3.5)$$

onde  $\psi(L) = 1 + L + L^2 + L^3 \dots$

A literatura estatística frequentemente se refere ao vetor de ruído branco  $\varepsilon_t$  como “choques” ou “inovações”. O teorema de Wold (WOLD, 1954) estabelece que qualquer série estacionária tem uma representação dada pela equação 3.5 tal que  $\sum_0^{\infty} \psi_k^2 < \infty$ . Esse resultado é equivalente à constatação que qualquer sinal estacionário pode ser representado pelo processamento de ruído branco por um filtro com resposta ao impulso quadrado somável. Outro ponto de vista do teorema do Wold pode ser obtido analisando a variância do GLM.

$$\sigma_{GLM}^2 = E \left[ \left( \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \right)^2 \right] = E \left[ \sum_{k=0}^{\infty} \psi_k^2 \varepsilon_{t-k}^2 \right]$$

$$\sigma_{GLM}^2 = \sum_{k=0}^{\infty} \psi_k^2 E[\varepsilon_{t-k}^2]$$

$$\sigma_{GLM}^2 = \sigma_{\varepsilon_t}^2 \sum_{k=0}^{\infty} \psi_k^2$$

Que implica a necessidade de finitude de  $\sigma_{GLM}^2$ . Construindo um vetor  $\boldsymbol{\psi}$  formado pelos coeficientes temos a variância representada pelo quadrado da norma desse vetor  $\|\boldsymbol{\psi}\|^2 = \sigma_{GLM}^2$ .

### 3.2.2.1 Autocorrelação

A fim de definir a função de  $\rho_{GLM}(\tau)$  de autocorrelação de um modelo linear generalizado definimos inicialmente sua covariância

$$\gamma_{GLM}(\tau) = E \left[ \left( \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \right) \left( \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i-\tau} \right) \right]$$

$$\gamma_{GLM}(\tau) = E \left[ \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \psi_k \varepsilon_{t-k} \psi_i \varepsilon_{t-i-\tau} \right]$$

onde temos que a esperança entre quaisquer  $\varepsilon_t - a$  e  $\varepsilon_t - b$  nula exceto para  $a = b$ , já que por definição  $\varepsilon_t$  é composto de variáveis aleatórias independentes. Estamos então interessados nos casos em que  $t - k = t - i - \tau$  tal que  $i = k - \tau$ . Substituindo as variáveis temos

$$\gamma_{GLM}(\tau) = E \left[ \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} \psi_{k-\tau} \varepsilon_{t-k} \right]$$

$$\gamma_{GLM}(\tau) = E \left[ \sum_{k=0}^{\infty} \psi_k \psi_{k-\tau} \varepsilon_{t-k}^2 \right] = \sigma_{\varepsilon_t}^2 \sum_{k=0}^{\infty} \psi_k \psi_{k-\tau}$$

A autocorrelação é então dada por

$$\rho_{GLM}(\tau) = \sum_{k=0}^{\infty} \psi_k \psi_{k-\tau}$$

### 3.2.3 Modelo Média Móvel

Um processo  $\mathbf{Y}(t, q)$  é considerado de média móvel de ordem  $q$  se uma realização  $y_t$  for definida pela equação 3.6, na qual  $\varepsilon_t$  representa a realização de um processo puramente aleatório. O processo é tipicamente chamado de  $MA(q)$ .

$$y_t = \varepsilon_t + \sum_{i=1}^{i=q} \beta_i \varepsilon_{t-i} \quad (3.6)$$

Pela definição acima observamos que o modelo média móvel corresponde a um filtro FIR excitado por ruído branco.

Um processo de média móvel, como sugerido pelo nome, é análogo a uma média móvel de de observações anteriores de uma série temporal de ruído branco. Não é de fato uma média móvel porque os coeficientes não necessariamente se somam a um.

Podemos escrever a equação 3.6 usando o operador de atraso por meio da equação 3.7. Nessa equação observamos claramente como o modelo  $MA(q)$  é um caso particular do  $GLM$ .

$$y_t = \varepsilon_t \left( 1 + \sum_{i=1}^{i=q} \beta_{t-i} L^i \right) = \varepsilon_t \phi(L) \quad (3.7)$$

A figura 15 ilustra o comportamento temporal de modelos  $MA$  para diferentes ordens.

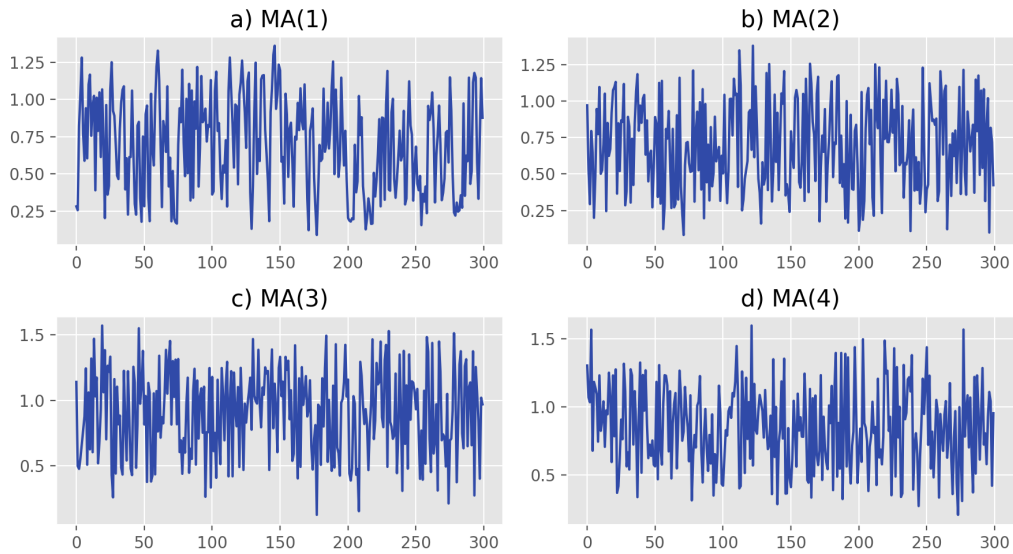


Figura 15 – Visualização no tempo de processos média móvel de ordens diferentes.

Observamos que não temos clara identidade visual dos processos de média móvel representados devido à natureza estocástica do sinal.



### 3.2.3.1 Autocorrelação

Definimos inicialmente a função de autocorrelação para o processo média móvel

$$\gamma_{MA(q)}(\tau) = E \left[ \left( \varepsilon_t + \sum_{k=1}^m \beta_k \varepsilon_{t-k} \right) \left( \varepsilon_{t-\tau} + \sum_{i=1}^m \beta_i \varepsilon_{t-i-\tau} \right) \right]$$

observando que para  $\tau > q$  temos  $\gamma_{MA(q)}(\tau) = 0$ , concluimos de forma análoga à dedução da autocorrelação do processo linear generalizado a seguinte forma

$$\gamma_{MA(q)}(\tau) = \sigma_{\varepsilon_t}^2 \sum_{k=1}^m \beta_k \beta_{k-\tau}$$

A variância é dada por  $\gamma(0)$  tal que

$$\sigma_{MA(q)}^2 = \sigma_{\varepsilon_t}^2 \sum_{k=1}^m \beta_k^2$$

Se definirmos o vetor  $\theta$  a partir dos coeficientes do polinômio  $\theta(L)$  temos  $\sigma_{MA(q)}^2 = \sigma_{\varepsilon_t}^2 \|\theta\|^2$ .

Temos autocorrelação definida então como

$$\rho_{MA(q)}(\tau) = \frac{\sum_{k=1}^m \beta_k \beta_{k-\tau}}{\|\theta\|}$$

A função de autocorrelação de um processo média móvel exibe a interessante propriedade de “cortar” após o atraso  $q$ , isso é, demonstrar autocorrelação igual a zero após um atraso de número correspondente à ordem do processo. A autocorrelação amostral de uma série temporal gerada por um processo de média móvel tende a apresentar a mesma propriedade, apesar de ser perfeitamente possível da autocorrelação amostral de um processo  $MA(q)$  cair para zero *antes* do lag  $q$  (CHATFIELD; XING, 2019).

A imagem 16 demonstra a visualização dos correlalogramas correspondentes aos modelos ilustrados no tempo pela figura 15. Observe que as autocorrelações são distintas de zero apenas para atrasos iguais ou inferiores à ordem  $q$  do processo.

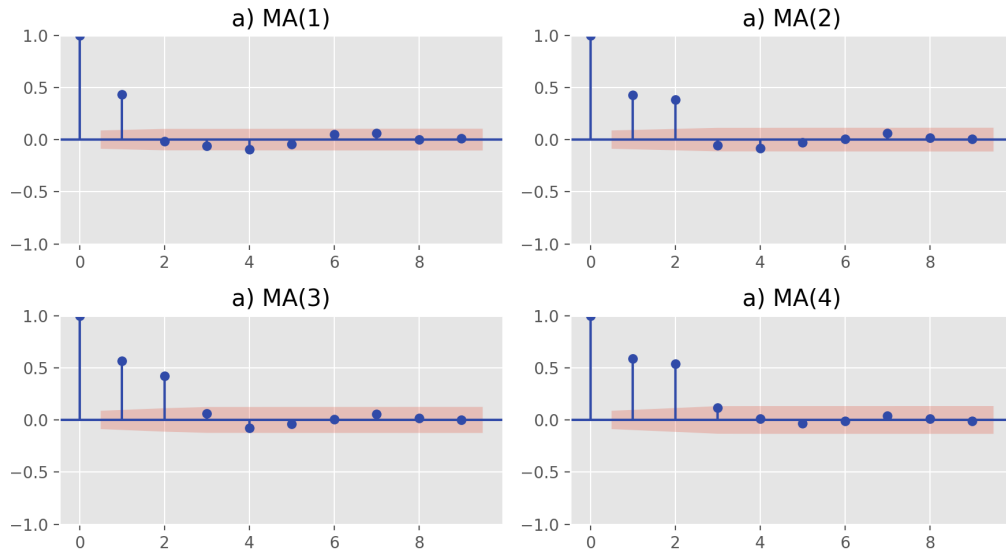


Figura 16 – Visualização do correlatograma de processos média móvel de ordens diferentes.

Essa propriedade da função de autocorrelação possui consequências diretas para modelagem de séries temporais: é possível que uma série temporal estacionária com autocorrelações iguais a zero após atraso  $k$  seja satisfatoriamente aproximada por um modelo  $MA(k)$ . A função de autocorrelação amostral é usada então como ferramenta de identificação da aplicabilidade e ordem de modelos  $MA(q)$ .

### 3.2.4 Modelo Autoregressivo

Um processo  $Y(t, p)$  é considerado autoregressivo de ordem  $p$  se uma realização  $y_t$  ele for definido pela equação 3.8, onde  $\varepsilon_t$  representa um processo puramente aleatório. O processo é frequentemente chamado de  $AR(p)$ .

$$y_t = \sum_{i=1}^{i=p} y_{t-i} \alpha_i + \varepsilon_t \quad (3.8)$$

Um processo autoregressivo de ordem  $p$  é caracterizado por uma dependência entre uma amostra de instante de tempo  $t$  e as amostras de instantes de tempo  $t - 1, t - 2, \dots, t - p$ . Como sugerido pelo nome a equação de diferenças estabelece uma relação de regressão entre uma série temporal e suas versões atrasadas no tempo.

Podemos reescrever a equação 3.8 por meio do operador de atraso resultando na equação. 3.9. O polinômio  $\alpha(L)$  é chamado de polinômio autoregressivo ou equação característica.

$$y_t = \frac{\varepsilon_t}{1 - \sum_{i=1}^{i=p} L^i \alpha_i} \quad (3.9)$$

Como exemplo temos que um processo autoregressivo de segunda ordem, isso é,  $AR(2)$ , é definido pela expressão a seguir.

$$y_t = \frac{\varepsilon_t}{(1 - L\alpha_1)} \quad (3.10)$$

O polinômio de operadores de atraso  $1 - L\alpha_1$  recebe o nome de equação característica. Ao expandir essa equação por divisão polinomial encontramos o conhecido desenvolvimento de um filtro IIR.

$$y_t = \sum_{k=0}^{\infty} \alpha^k \varepsilon_{t-k} \quad (3.11)$$

Observamos primeiramente que o parâmetro  $\alpha$  deve ser tal que a soma do lado direito da equação 3.11 seja convergente. Sabemos que essa condição corresponde à estabilidade do filtro e veremos que corresponde também à estacionariedade do sinal produzido pela filtragem. Concluimos que um sinal produzido por um processo autoregressivo estacionário pode ser modelado pelo processamento de ruído branco por um filtro IIR estável. Essa forma da equação autoregressiva é claramente um caso específico do GLM.

A divisão polinomial entre a equação 3.10 e 3.11 é generalizada como uma inversão do polinômio  $\alpha(L)$ , de forma que

$$y_t \alpha(L) = \varepsilon_t = y_t = \alpha^{-1}(L) \varepsilon_t$$

$\alpha^{-1}(L)$  é bem definido se o sistema linear definido pela função de transferência  $\alpha(L)$  for inversível, isso é, se suas raízes em  $L$  estiverem localizadas fora do círculo unitário.

Uma visualização do comportamento temporal de processos autoregressivos é dada pela figura 17. Observa-se que as séries temporais não exibem comportamento visivelmente distinguível no domínio do tempo, como no caso das séries  $MA(q)$ .

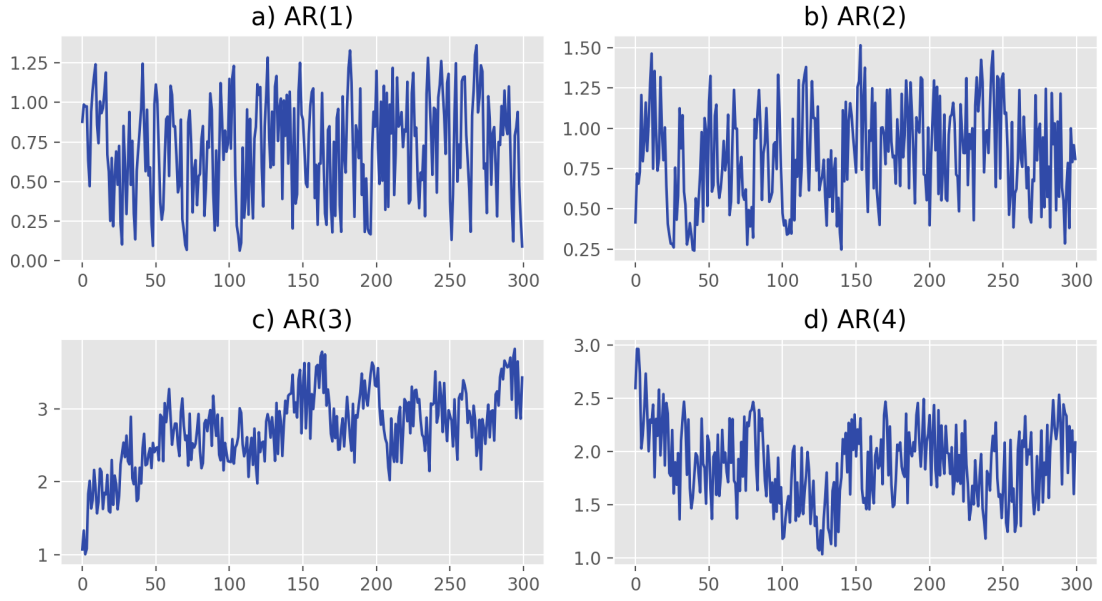


Figura 17 – Visualização de processos autoregressivos de diferentes ordens no domínio do tempo.

#### 3.2.4.1 Autocorrelação

Assumindo estacionariedade e multiplicando ambos os lados da equação 3.8 por  $y_{t-\tau}$  podemos então calcular a esperança da expressão para obter a autocovariância do processo para  $\tau \neq 0$

$$\gamma_{AR(p)}(\tau) = E \left[ \left( \sum_{k=1}^{i=p} y_{t-k} y_{t-\tau} \alpha_k \right) + (y_{t-\tau} \epsilon_t) \right]$$

$$\gamma_{AR(p)}(\tau) = E \left[ \left( \sum_{k=1}^{i=p} y_{t-k} y_{t-\tau} \alpha_k \right) \right]$$

$$\gamma_{AR(p)}(\tau) = \sum_{k=1}^{i=p} \alpha_k \gamma_{AR(p)}(\tau - k)$$

Equação que igualmente satisfaz a autocorrelação  $\rho(\tau)$  pela divisão de cada um dos termos acima por  $\gamma_{AR(p)}(0) = \sigma_{AR(p)}^2$ .

$$\rho_{AR(p)}(\tau) = \sum_{k=1}^{i=p} \alpha_k \rho_{AR(p)}(\tau - k), \tau > 0 \quad (3.12)$$

Obtemos então uma autoregressão de ordem  $p$  nas autocovariâncias também. Para calcular a variância do processo multiplicamos os termos da equação 3.8 por  $y_t$

$$\sigma_{AR(p)}^2 = \sigma_{\varepsilon_t}^2 + \sum_{k=1}^{i=p} \gamma_{AR(p)}(k)$$

$$\sigma_{AR(p)}^2 = \sigma_{\varepsilon_t}^2 + \sigma_{AR(p)}^2 \sum_{k=1}^{i=p} \rho_{AR(p)}(k)$$

$$\sigma_{AR(p)}^2 - \left( \sigma_{AR(p)}^2 \sum_{k=1}^{i=p} \rho_{AR(p)}(k) \right) = \sigma_{\varepsilon_t}^2$$

$$\sigma_{AR(p)}^2 = \frac{\sigma_{\varepsilon_t}^2}{1 - \sum_{k=1}^{i=p} \rho_{AR(p)}(k)}$$

A variância é também uma função autoregressiva da autocorrelação.

Podemos reescrever a equação 3.12 em função do polinômio de atraso (operando agora sobre  $\tau$  ao invés de  $t$ ) como

$$\alpha(L)\rho_{AR(p)}(\tau) = 0 \quad (3.13)$$

Expressando o polinômio em função de suas raízes  $G_i$  temos que

$$\alpha(L) = \prod_{k=1}^p (1 - G_k L)$$

É demonstrado em (BOX; JENKINS, 2016, capítulo 4) que a solução geral para a equação de 3.13 é dada por

$$\rho_{AR(p)}(\tau) = \sum_{k=1}^p A_k G_k^\tau \quad (3.14)$$

Essa é a expressão conclusiva da autocorrelação de um processo autoregressivo de ordem  $p$ .

Denotemos as raízes reais de  $\alpha(L)$  por  $R_h$  e as complexas conjugadas por  $C_j, C_k$ . Observamos que sob a forma 3.14 as raízes reais  $R_h$ , assumindo que  $|R_h| < 1$ , contribuem para a autocorrelação como exponenciais amortecidas  $A_h^\tau R_h$ . Raízes complexas conjugadas  $C_j, C_k$  contribuem com um senóide amortecido.

Concluimos então que a autocorrelação de um processo  $AR(p)$  é uma combinação de  $N$  exponenciais e senóides amortecidos tal que  $N \leq p$ .

Para o caso específico de um processo estacionário  $AR(1)$  temos pela equação 3.14 a seguinte relação de autocorrelação

$$\rho_{AR(1)}(\tau) = \alpha \rho_{AR(1)}(\tau - 1)$$

$$\rho_{AR(1)}(\tau) = \alpha^\tau$$

A imagem 18 demonstra a visualização dos correlalogramas correspondentes aos modelos ilustrados no tempo pela figura 17. Observe que as autocorrelações são misturas de exponenciais e senóides amortecidas.

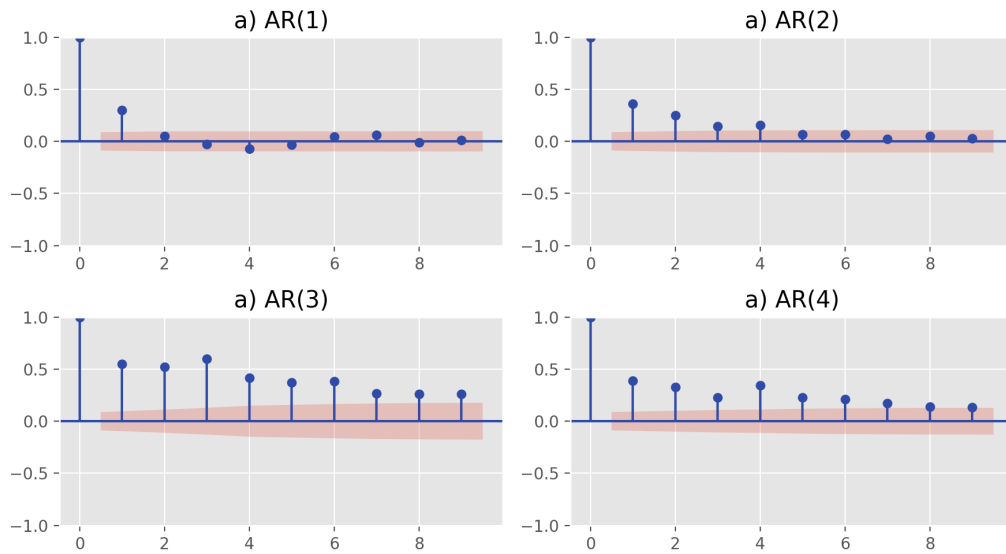


Figura 18 – Visualização do correlalograma de processos autoregressivos de ordens diferentes.

### 3.2.4.2 Equações de Yule Walker

É possível estimar os coeficientes autoregressivos  $\alpha$  desenvolvendo as equações 3.13. Com  $\rho(-k) = \rho(k)$  e  $\rho(0) = 1$  temos

$$\rho(1) = \alpha_1 + \alpha_2 \rho(1) + \alpha_3 \rho(2) \dots + \alpha_p \rho(p-1)$$

$$\rho(2) = \alpha_1 \rho(1) + \alpha_2 \rho(1) + \alpha_3 \rho(2) \dots + \alpha_p \rho(p)$$

$\vdots$

$$\rho(p) = \alpha_1 \rho(p-1) + \alpha_1 \rho(-1) + \alpha_2 \rho(-2) \dots + \alpha_p$$

Em forma matricial

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & 1 & \dots & \rho(p) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \quad (3.15)$$

As equações 3.15 são conhecidas como equações de Yule Walker, que permitem uma estimativa de coeficientes autoregressivos por meio de uma estimativa de coeficientes de autocorrelação. A forma mais intuitiva de estimar o vetor  $\alpha$  é por meio da inversão da matriz de autocorrelações defasadas, mas outros algoritmos como o de Durbin-Levinson podem ser utilizados.

### 3.2.4.3 Autocorrelação Parcial

Como mencionado na seção 2.7.1 autocorrelação parcial informa a correlação restante entre  $y_t$  e  $y_{t-\tau}$  após levar em conta a contribuição resultante dos termos intermediários  $y_{t-1}, y_{t-2} \dots y_{t-\tau+1}$ .

Autocorrelação parcial é formalmente definida para um processo autoregressivo partindo da equação 3.13, onde tomando  $p = 1$  para um processo autoregressivo de ordem 1 temos que

$$\rho(\tau) = \alpha_1 \rho(\tau-1)$$

Tomando  $\tau = p = 1$  e adicionando um subscrito adicional em  $\alpha$  para indicar a ordem do processo autoregressivo temos

$$\alpha_{1_1} = \rho(1)$$

Para  $p = 2$  temos

$$\rho(\tau) = \alpha_1 \rho(\tau - 1) + \alpha_2 \rho(\tau - 2)$$

Novamente tomando  $\tau = p = 2$  obtemos

$$\rho(2) = \alpha_{1_2} \rho(1) + \alpha_{2_2} \rho(0)$$

Estamos como no caso de  $p = 1$  interessados no valor de  $\alpha_{2_2}$ , isso é,  $\alpha_{\tau_\tau}$  com  $\tau = 2$ . Uma solução para esse valor em função das autocorrelações  $\rho$ , obtida por meio das equações de Yule Walker, é

$$\alpha_{2_2} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}$$

As autocorrelações parciais em função de  $\tau$  são então dadas por sucessivos  $\alpha_{\tau_\tau}$  para  $\tau = 1, 2, 3, \dots$ , isso é, o último coeficiente autoregressivo  $\alpha_p$  de processos autoregressivos de ordem  $p$  crescentes.

A solução desses valores em função das autocorrelações de cada processo de ordem crescente, dada por (MORETTIN, 2017), é generalizada como

$$\alpha_{\tau_\tau} = \frac{\|\mathbf{P}_\tau^*\|}{\|\mathbf{P}_\tau\|}$$

Onde a matriz  $\mathbf{P}$  é a matriz de autocorrelações de ordem  $p = \tau$  como definida na equação 3.15 e a matriz  $\mathbf{P}^*$  é obtida pela substituição da última coluna de  $\mathbf{P}$  pelo vetor de autocorrelações  $\rho$ .

Essa definição de autocorrelação parcial é bem definida para processos autoregressivos e para esse tipo de processo seu valor claramente se torna igual a zero a partir do atraso  $\tau = p$ .

A autocorrelação parcial amostral é calculada da mesma forma partindo da autocorrelação amostral como definida pela equação 2.2. A autocorrelação parcial amostral por sua vez é definida para séries temporais arbitrárias. Nesse caso sucessivos modelos autoregressivos de ordem  $p = \tau = 1, 2, 3, \dots$  são ajustados à série em questão e o último coeficiente de cada regressão é armazenado como a autocorrelação parcial amostral para o atraso  $\tau$ .



Como estamos interessados em trabalhar com séries temporais a definição acima de autocorrelação parcial amostral é o suficiente para generalizar o conceito definido sobre processos autoregressivos para realizações únicas (séries temporais) de processos estocásticos arbitrários.

### 3.2.5 ARMA

Processos ARMA, como sugerido pelo nome, são gerados pela sobreposição de processos AR e MA. Um processo  $\mathbf{Y}(t)$  é considerado  $ARMA(p, q)$  se for dado pela equação 3.16

$$y_t = \varepsilon_t + \sum_{i=1}^{i=q} \beta_i \varepsilon_{t-i} + \sum_{i=1}^{i=p} y_{t-i} \alpha_i \quad (3.16)$$

Um processo  $ARMA(2, 1)$  é portanto dado pela seguinte equação de recorrência.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

Em analogia aos casos anteriores o processo pode ser escrito por meio do operador de atraso.

$$y_t(1 - \alpha_1 L - \alpha_2 L^2) = \varepsilon_t(\beta_0 + \beta_1 L)$$

Podemos introduzir  $\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2$  e  $\theta(L) = \beta_0 + \beta_1 L$  tal que:

$$y_t = \varepsilon_t \frac{\phi(L)}{\theta(L)} \quad (3.17)$$

A equação 3.17 é prontamente generalizada para corresponder à equação 3.16 estabelecendo:

$$\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p$$

$$\theta(L) = 1 - \beta_1 L^1 \dots - \beta_q L^q$$

Observamos que a forma da equação 3.17 é idêntica à de uma função de transferência, sendo definida no domínio do tempo sobre polinômios de  $L$  ao invés de no domínio  $Z$  sobre polinômios em  $z$ . As implicações do posicionamento das raízes dos polinômios numerador e denominador decorrem de maneira análoga.

Uma propriedade interessante de modelos ARMA é que processos estacionários frequentemente podem ser modelados por modelos ARMA com menos parâmetros que modelos AR ou MA.

### 3.3 Raízes Unitárias

Como um processo ARMA é essencialmente resultante do processamento de ruído branco por um filtro linear com função de transferência  $\frac{\theta(L)}{\phi(L)}$  sabemos que a posição das raízes dos polinômios  $\theta(L)$  e  $\phi(L)$  determina suas propriedades.

O posicionamento de raízes no círculo unitário é um caso interessante de analisar. Se algum dos polos função de transferência de um processo ARMA estiverem posicionados no círculo unitário diz-se que esse processo possui raízes unitárias. Esse termo tipicamente descreve o posicionamento dos polos e não dos zeros pelo maior efeito dessas raízes na dinâmica do sistema, mas essa seção inclui uma breve discussão sobre o efeito de zeros unitários também.

Antes de prosseguir um processo  $y_t$  ARMA(2, 1) será estabelecido para exemplificar as seguintes discussões.

Diante da recorrência

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

temos

$$y_t = \frac{1 + \beta_1 L}{1 - \alpha_1 L - \alpha_2 L^2} \varepsilon_t$$

Uma realização desse processo para os parâmetros

$$\alpha_1 = -0.5, \alpha_2 = 0.25, \beta_1 = 0.8$$

é ilustrada pela figura 19 e seu diagrama de polos e zeros no plano  $L$  pela figura 20.

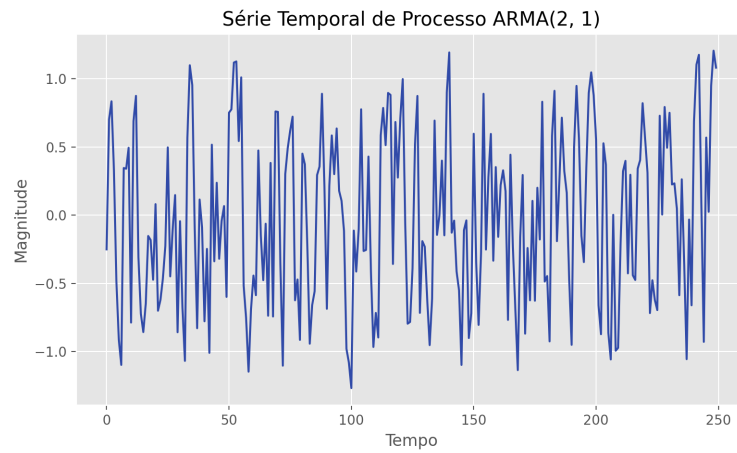


Figura 19 – Visualização de realização de processo ARMA(2, 1) no tempo

Diagrama de Polos e Zeros em L de Processo ARMA(2, 1)

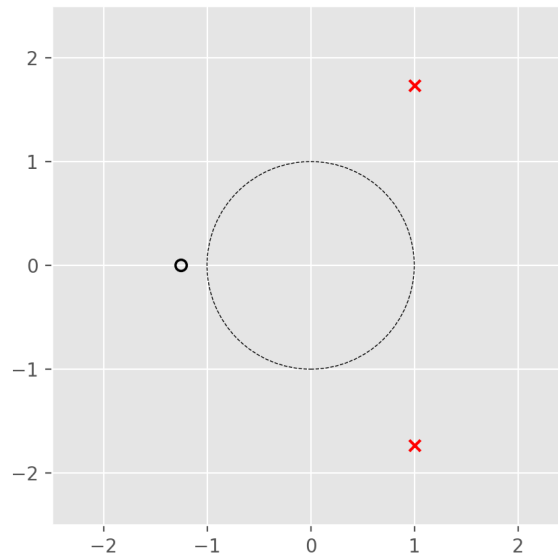


Figura 20 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) em relação ao círculo unitário no plano L

### 3.3.1 Raízes Unitárias no Polinômio Autoregressivo

Inspecionemos qualitativamente a inserção presença de uma raiz unitária no polinômio autoregressivo do processo ARMA(2, 1) apresentado. Tomando  $z_t$  como o processo

$$z_t = \frac{y_t}{(1-L)} = \frac{1+0.8L}{(1+0.5L-0.25L^2)(1-L)} \varepsilon_t$$

Temos a seguinte relação entre  $z_t$  e  $y_t$

$$(1 - L)z_t = y_t$$

$$z_t - z_{t-1} = y_t$$

$$\nabla z_t = y_t$$

Tal que  $z_t$  represente uma integração de  $y_t$ .

$$z_t = \nabla^{-1} y_t$$

Dizemos que nesse caso o processo  $z_t$  é integrado de ordem um, ou  $I(1)$ , já que o diferenciando uma vez temos em um processo estacionário. Observamos imediatamente que o processo  $z_t$  não é estacionário.

Podemos definir um processo integrado de ordem  $d$ ,  $I(d)$ , como um processo cuja diferenciação em  $d$  vezes resulta em estacionariedade. Tal processo seria gerado a partir de  $y_t$  por meio de  $d$  integrações. Essa definição implica corretamente que um processo  $I(d)$ ,  $d > 0$ , não é estacionário, já que a condição de estacionariedade para processos ARMA é pontualmente violada.

Analisando uma realização do processo  $z_t$ , por meio da figura 21, observamos um típico processo de raiz unitária, cuja não estacionariedade não é tão óbvia quanto nos casos de um filtro ARMA instável ou presença de tendência determinística. Processos com raiz unitária são diferença estacionários e possuem tendência determinística, como discutido na seção 2.8.2. Como esperado o diagrama de polos e zeros desse modelo inclui um polo adicional correspondente à raiz unitária.

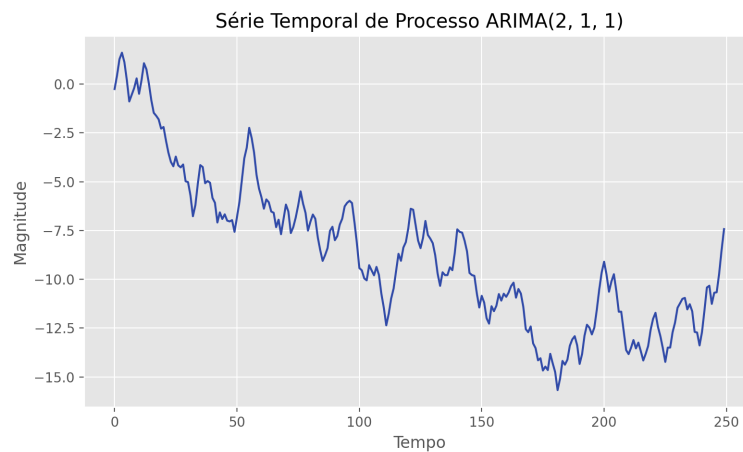


Figura 21 – Visualização de realização de processo ARMA(2, 1) com introdução de raiz unitária no tempo

Diagrama de Polos e Zeros em L de Processo ARIMA(2, 1, 1)

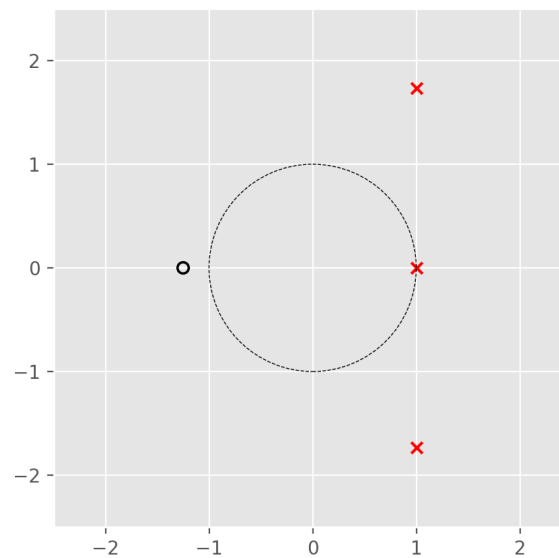


Figura 22 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) com introdução de raiz unitária em relação ao círculo unitário no plano L

Processos de raízes unitárias são especialmente interessantes por sua capacidade de representar fenômenos do mundo real de forma eficiente. Logo serão reconhecimentos como processos da classe ARIMA.

No processamento de sinais determinísticos sistemas lineares com raiz unitária levam ao caso de estabilidade marginal, onde a resposta ao impulso do sistema é um sinal de potência ao invés de um sinal de energia. A ideia de uma resposta ao impulso de energia infinita sob presença de polos integradores manifesta-se no contexto de sinais aleatórios quando é dito que

processos de tendência estocástica (de estacionariedade diferenciável, como discutido na seção 2.8.2) são afetados irreversivelmente por eventos perturbadores. A “lembrança” desses eventos é justamente a resposta de duração infinita do evento perturbador, apropriadamente modelado por funções impulso ou degrau. Essa interpretação é importante no campo de análise de intervenção, que procura incorporar efeitos de eventos perturbadores reais em modelos de séries temporais.

### 3.3.2 Raízes Unitárias no Polinômio Média Móvel

A presença de uma raiz unitária no polinômio de média móvel de um filtro ARMA gera efeitos menos dramáticos em sua dinâmica. Antes de tudo nota-se que a introdução de raízes unitárias de médias móveis tornam um sistema não inversível, já que sua inversão tornaria essa raiz um polo.

As figuras 23 e 24 ilustram uma realização do processo  $y_t$  com adição de raiz unitária de média móvel e o diagrama de polos e zeros resultante, respectivamente.

Observamos que a série com raiz de média móvel adicional aparenta ter uma distribuição de potência mais enviesada para altas frequências, de forma recíproca ao de raiz unitária autoregressiva, em que há introdução de componentes de baixa frequência. Isso será discutido apropriadamente no capítulo 4.

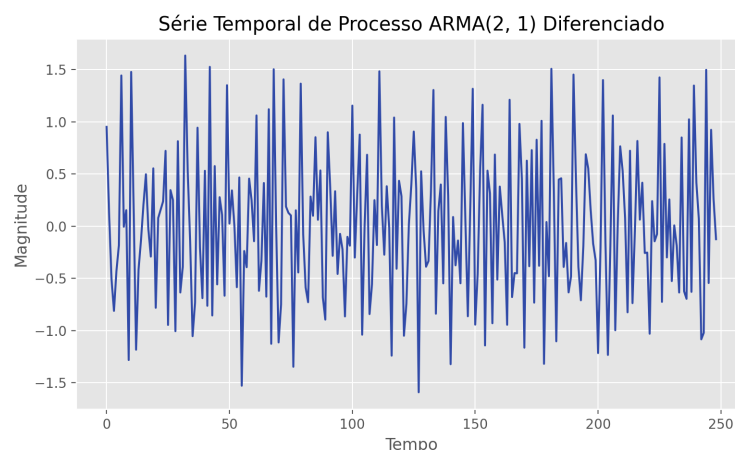


Figura 23 – Visualização de realização de processo ARMA(2, 1) derivado no tempo

Diagrama de Polos e Zeros em  $L$  de Processo ARMA(2, 1)

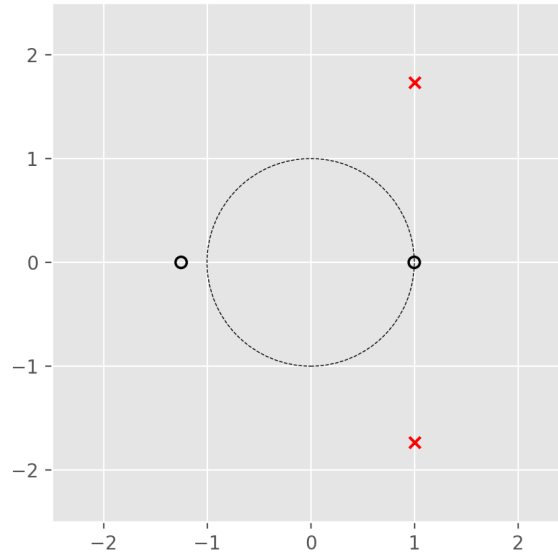


Figura 24 – Visualização de posicionamento de polos e zeros de processo ARMA(2, 1) derivado em relação ao círculo unitário no plano  $L$

Tomando o processo  $w_t$  como

$$w_t = (1 - L)y_t = \frac{1 - 0.2L + L^2}{1 + 0.5L - 0.25L^2}\varepsilon_t$$

Temos  $w_t = y_t - y_{t-1} = \nabla y_t$  tal que  $w_t$  seja a derivada de  $y_t$ . Vemos que ato de diferenciar uma série, em uma tentativa de introduzir estacionariedade por exemplo, corresponde à introdução de uma raiz unitária no polinômio de média móvel. Conseguimos dessa forma compreender a indução de estacionariedade em uma série não estacionária com raiz unitária autoregressiva por meio de sua derivação como a sobreposição de  $d$  zeros aos  $d$  polos integradores do processo original. Isso é ilustrado por uma comparação das figuras 22 e 24, onde vemos que a diferenciação da série  $z_t$  introduziria um zero unitário de forma a “cancelar” seu polo unitário.

A presença de raízes unitárias de média móvel podem indicar que a série foi diferenciada mais vezes que necessário ou que diferenciação é a operação inadequada para indução de estacionariedade, isso é, a série não é diferença estacionária.

Um exemplo é a série tendência estacionária  $m_t$  a seguir:

$$m_t = \mu + \eta t + \varepsilon_t$$

$$\nabla m_t = \varepsilon_t - \varepsilon_{t-1} + \eta$$

$$\nabla m_t = (1 - L)\varepsilon_t + \eta$$

Em  $\nabla m_t$  temos uma raiz unitária no polinômio autoregressivo, resultando em uma série não inversível. A tendência determinística poderia ter sido removida por regressão resultando em um processo estacionário e inversível.

### 3.4 Modelo ARIMA

Também chamados de modelos ARMA Integrados modelos ARIMA são essencialmente modelos ARMA com tentativas de remoção de não estacionariedade. Assumindo que a série em questão seja diferença-estacionária o modelo ARIMA induz estacionariedade ajustando um modelo ARMA com raízes unitárias adicionais. A expectativa é que o processo original contenha  $d$  polos unitários que serão então sobrepostos por  $d$  zeros unitários, resultando em processo estacionário.

A distinção entre o modelo ARMA e ARIMA é a substituição de  $y_t$  por  $\nabla^d y_t$ .

Definindo  $w_t = \nabla^d y_t$  para  $d = 0, 1, 2, \dots$  temos a definição de um processo  $ARIMA(p, d, q)$  dada pela equação 3.18. Observe que a equação é idêntica à equação 3.16 com  $w_t$  ao invés de  $y_t$  e que a saída é  $w_t$ , demandando uma integração para previsão de  $y_t$ . O nome do modelo referencia essa operação de reconstrução de  $y_t$ .

$$w_t = \varepsilon_t + \sum_{i=1}^{i=q} \beta_i \varepsilon_{t-i} + \sum_{i=1}^{i=p} w_{t-i} \alpha_i \quad (3.18)$$

Podemos escrever um modelo  $ARIMA(p, d, q)$  por meio do operador de atrasos por meio da equação 3.19, onde:

$$\phi(L) = 1 - \alpha_1 L - \alpha_2 L^2 \dots - \alpha_p L^p$$

$$\theta(L) = 1 - \beta_0 - \beta_1 L^1 \dots - \beta_q L^q$$



Nessa forma o paralelo entre modelos ARMA e ARIMA se torna mais claro.

$$y_t = \varepsilon_t \frac{\theta(L)}{\phi(L)} \frac{1}{(1-L)^d} \quad (3.19)$$

O modelo ARIMA segue a metodologia Box-Jenkins de modelagem, assumindo diferença-estacionariedade e tentando induzir estacionariedade por meio de sucessivas operações de diferenças. Pela discussão acima é claro que um modelo  $ARIMA(p, 0, q)$  corresponde a um modelo  $ARMA(p, q)$ .

O leitor agora reconhece o sistema da figura 22 como  $ARIMA(2, 1, 1)$ .

### 3.5 Modelos de Sazonalidade

#### 3.5.1 Diferenciação Sazonal

O operador de diferença sazonal  $\nabla_s$  estende o operador de diferenças para subtração entre elementos não adjacentes e age sob uma série temporal  $\mathbf{y}_t$  da seguinte forma:

$$\nabla \mathbf{y}_t = (1 - L^s) \mathbf{y}_t = \mathbf{y}_t - L^s \mathbf{y}_t = \mathbf{y}_t - \mathbf{y}_{t-s}$$

O operador mantém as propriedades da diferença simples e pode ser empregado em ordens elevadas da mesma forma:

$$\nabla_4^2 y_t = \nabla_4 \nabla_4 y_t = \nabla_4 (y_t - y_{t-4}) = y_t - 2y_{t-4} + y_{t-8}$$

#### 3.5.2 SARIMA

O modelo ARIMA pode ser estendido de forma natural para acomodar padrões de sazonalidade. O modelo SARIMA (Seasonal ARIMA) parte da observação que saídas de sinais com alta sazonalidade podem possuir alta correlação com saídas anteriores em uma distância corresponde ao padrão de sazonalidade, como constatado no capítulo 1.

O modelo SARIMA então estabelece, em adição à operação de diferenciação do modelo ARIMA, uma operação de diferenciação sazonal ao definir a componente  $w_t$  da seguinte forma:

$$w_t = \nabla_s^D \nabla^d y_t$$

Como na definição de  $w_t$  para o modelo ARIMA temos uma diferenciação tradicional de ordem  $d$  seguida de uma diferenciação sazonal de período  $s$  e ordem  $D$ .

Como exemplo um  $w_t$  definido em função de  $d = 1$ ,  $D = 1$  e  $s = 24$  assume a seguinte forma.

$$w_t = \nabla_{12}^1 \nabla^1 y_t = \nabla_2^1 (y_t - y_{t-1})$$

$$w_t = (y_t - y_{t-24}) - (y_{t-1} - y_{t-25})$$

Em seguida componentes autoregressivos e de média móvel são introduzidos com atrasos em múltiplos de  $s$ . O modelo final é descrito como  $SARIMA(p, d, q)(P, D, Q)_s$ , com  $P$  e  $D$  referenciando os componentes *AR* e *MA* sazonais. Para ilustrar melhor essa notação podemos analisar a expressão de um modelo específico para depois generalizar.

Um modelo  $SARIMA(1, 0, 1)(2, 1, 1)_{12}$  é dado pela seguinte expressão.

$$w_t = \varepsilon_t + \overbrace{\alpha_1 w_{t-1}}^{p=1} + \overbrace{\alpha_2 w_{t-12}}^{P=1} + \overbrace{\alpha_3 w_{t-24}}^{P=2} + \overbrace{\beta_1 \varepsilon_{t-1}}^{q=1} + \overbrace{\beta_2 \varepsilon_{t-12}}^{Q=1}$$

A escrita desse modelo por meio do operador de atraso é mais clara.

$$w_t = \varepsilon_t \frac{\beta_1 L^1 + \beta_2 L^{12}}{(1 - \alpha_1 L^1 + \alpha_2 L^{12} + \alpha_3 L^{24})}$$

Podemos então generalizar o modelo na forma do operador de atraso por meio da equação 3.20.

$$w_t = \varepsilon_t \frac{\theta_q(L) \theta_Q(L)}{\phi_p(L) \phi_P(L^s)} \quad (3.20)$$

Com os polinômios de atraso definidos como nos casos anteriores.

A diferenciação sazonal inicial do modelo SARIMA é um dos exemplos mais simples de um passo essencial no processamento de séries com padrões sazonais: a modelagem

do padrão de sazonalidade em si, que de fato é um campo próprio com livros como (LADIRAY; QUENNEVILLE, 2001). Diferenciação sazonal é limitada em sua capacidade de expressar múltiplos padrões de sazonalidade. Outra forma mais flexível de modelagem de sazonalidade será abordada na subseção seguinte.

### 3.5.3 Variáveis de Fourier

Versões sazonais de modelos ARIMA, como SARIMA, são pouco eficientes para modelar padrões sazonais com as seguintes características:

- Múltiplas periodicidades (ATHANASOPOULOS; HYNDMAN, 2018)
- Períodos múltiplos fracionários do tempo de amostragem (HYNDMAN, 2014a).
- Curto tempo de amostragem (de um dia ou inferior) (ATHANASOPOULOS; HYNDMAN, 2018)
- Períodos maiores que algumas centenas do tempo de amostragem (HYNDMAN, 2014b).

Sinais elétricos quase sempre possuem as duas últimas propriedades, tornando a exploração de métodos alternativos indispensável para este trabalho.

A inclusão de variáveis de Fourier como regressores é capaz de modelar padrões sazonais como esses de forma mais flexível. Essas variáveis exógenas aos ao modelo são somas de senos e cossenos, ambos reais, que oscilam em múltiplos de uma frequência fundamental definida como a frequência do padrão sazonal que se deseja modelar.

Para modelar um padrão de período  $m$  tomamos as variáveis dadas pela série  $F_t$

$$F_t = \sum_{k=1}^K \left( \alpha_k \frac{\sin(2\pi kt)}{m} + \beta_k \frac{\cos(2\pi kt)}{m} \right)$$

onde harmônicos do período sazonal  $m$  são incluídos pelo incremento de  $k$ . Para incluir múltiplos períodos sazonais as variáveis são generalizadas para diferentes valores de  $m$  como a seguir:

$$F_t = \sum_{i=1}^M \sum_{k=1}^K \left( \alpha_k \frac{\sin(2\pi kt)}{m_i} + \beta_k \frac{\cos(2\pi kt)}{m_i} \right)$$

Temos alguns parâmetros não regressores:  $K$ , a ordem dos harmônicos,  $m_i$ , os períodos dos padrões que se deseja modelar.  $M$  é simplesmente o número de padrões. A determinação de  $m_i$  pode ser realizada por conhecimento a priori do fenômeno em questão ou

métodos quantitativos como análise espectral dos dados amostrais. Já  $K$  é melhor determinado por métodos de seleção de modelo como AIC e BIC (MCQUARRIE; TSAI, 1998).

A série exógena pode então ser incluída como informação auxiliar em um modelo ARIMA resultando em um modelo sazonal  $y_t$  em função de  $(p, d, q, m_i, k)$ :

$$y_t = \varepsilon_t \frac{\theta(L)}{\phi(L)} \frac{1}{(1-L)^d} + F_t$$

$$y_t = \varepsilon_t \frac{\theta(L)}{\phi(L)} \frac{1}{(1-L)^d} + \sum_{i=1}^M \sum_{k=1}^K \left( \alpha_k \frac{\sin(2\pi kt)}{m_i} + \beta_k \frac{\cos(2\pi kt)}{m_i} \right)$$

## 4 TEORIA ESPECTRAL UNIVARIADA

### 4.1 Análise Estacionária

Sinais estacionários no sentido amplo são, por definição (seção 2.8), sinais de potência. Como a transformada de Fourier é bem definida apenas para sinais de energia finita sinais estocásticos estacionários não possuem uma transformada de Fourier no sentido tradicional. Para desenvolver uma representação espectral desse tipo de sinal é necessário definir o conceito de densidade de potência espectral e concluir que essa função é proporcional ao quadrado da magnitude de uma transformada de Fourier hipotética.

#### 4.1.0.1 Densidade de Potência Espectral

Constatamos inicialmente o teorema de Parseval, em que  $F\{\}$  representa a transformada de Fourier

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F\{x(t)\}(\omega)|^2 d\omega$$

Estendendo essa definição para potência de sinal temos

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \frac{1}{2\pi} \int_{-T}^T |F\{x(t)\}(\omega)|^2 d\omega$$

Note que apesar de  $F\{x(t)\}$  não ser bem definida a relação acima ainda é válida se  $|F\{x(t)\}|^2$  for descrita de uma forma diferente, o que será feito em breve.

A potência de um sinal pode ser reescrita representando a transformada de Fourier de  $x(t)$  por  $X(\omega)$  como

$$P = \lim_{T \rightarrow \infty} \frac{1}{2T} \frac{1}{2\pi} \int_{-T}^T |X(\omega)|^2 d\omega$$

Onde  $\lim_{T \rightarrow \infty} \frac{1}{2\pi} \frac{1}{2T} |X(\omega)|$  é reconhecida como uma função de densidade. A função densidade de potência espectral é finalmente definida como

$$S_x(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \frac{1}{2T} |X(\omega)|$$

O nome dessa função é bem informativo para sua interpretação:  $S_x(\omega)$  representa a contribuição das componentes de frequência de  $x(t)$  localizadas em  $\omega + d\omega$  para a potência do sinal como um todo. A definição de  $|X(\omega)|^2$  necessária para que essa função faça sentido é fornecida pelo teorema de Wiener-Khinchin.

#### 4.1.1 Teorema de Wiener-Khinchin

O teorema de Wiener Khinchin pode ser desenvolvido da seguinte maneira

$$|X(\omega)|^2 = X(\omega)X^*(\omega) = F(F^{-1}(X(\omega))) * (F^{-1}(X^*(\omega))) = F(x(t) * x^*(-t)) = F(x(t) * x(-t))$$

Examinando a parte mais à direita dessa igualdade observamos que a função que está sendo transformada corresponde à convolução de  $x(t)$  com uma versão espelhada de si mesmo. Isso é precisamente a definição de autocorrelação. Assumindo ergodicidade podemos agora expressar a magnitude ao quadrado da transformada de Fourier de  $x(t)$  como a transformada de Fourier de sua função de autocorrelação.

$$|X(\omega)_T|^2 = \frac{1}{2\pi} \int_{-T}^T \rho(t) e^{-j\omega t} dt$$

Esse resultado é conhecido como o teorema de Wiener-Khinchin e permite uma representação espectral bem definida para sinais estocásticos estacionários.

Note que a transformada de Fourier da autocorrelação de um sinal real é em si puramente real, propriedade consistente com nossa noção de magnitude ao quadrado.

#### 4.1.2 Espectro de um Processo ARMA

Tomando a magnitude ao quadrado da transformada Z da forma de recorrência geral de um processo ARMA ( 3.17) obtemos a seguinte função de transferência

$$H(z) = \frac{1 + \sum_i^q b_k z^{-k}}{1 + \sum_i^q a_k z^{-k}}$$

que é excitada por ruído branco de forma a gerar uma realização de um processo ARMA. Podemos agora expressar a densidade de potência espectral de um processo ARMA como

$$S_{ARMA}(\omega) = |H(z)|^2 S_{\varepsilon}$$

$$S_{ARMA}(\omega) = \frac{\sigma^2 |1 + \sum_{k=1}^q b_k e^{-j\omega k}|^2}{2\pi |1 + \sum_{k=1}^p a_k e^{-j\omega k}|^2} \quad (4.1)$$

Essa definição é usada como uma forma de estimação paramétrica de espectro: os parâmetros são inferidos no domínio do tempo e usados pela relação acima para sugerir um espectro.

Visualizaremos agora o espectro de alguns processos ARMA.

#### 4.1.2.1 MA(1)

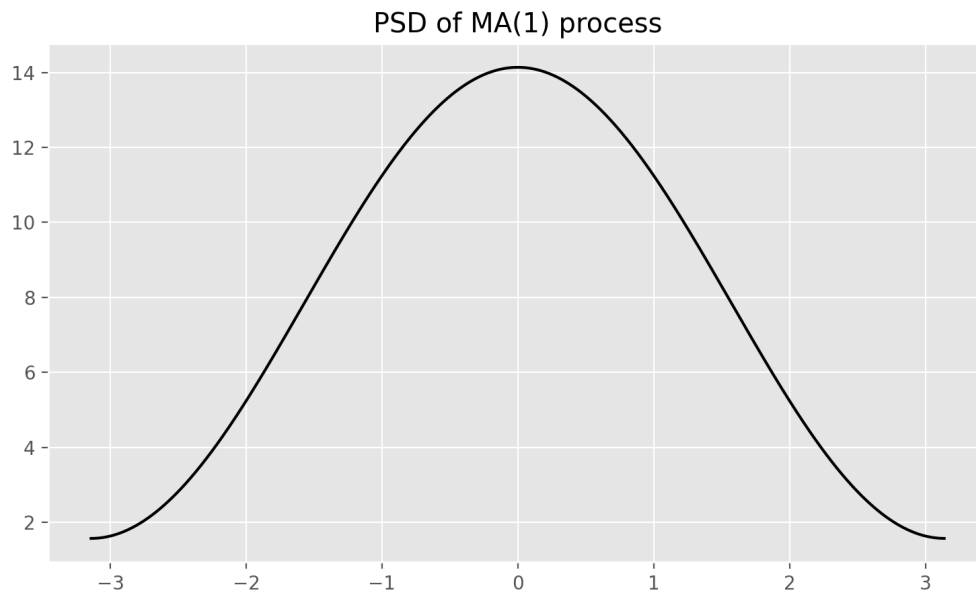


Figura 25 – Espectro de um processo MA(1) com  $\beta_1 = -0.5$

#### 4.1.2.2 $AR(1)$

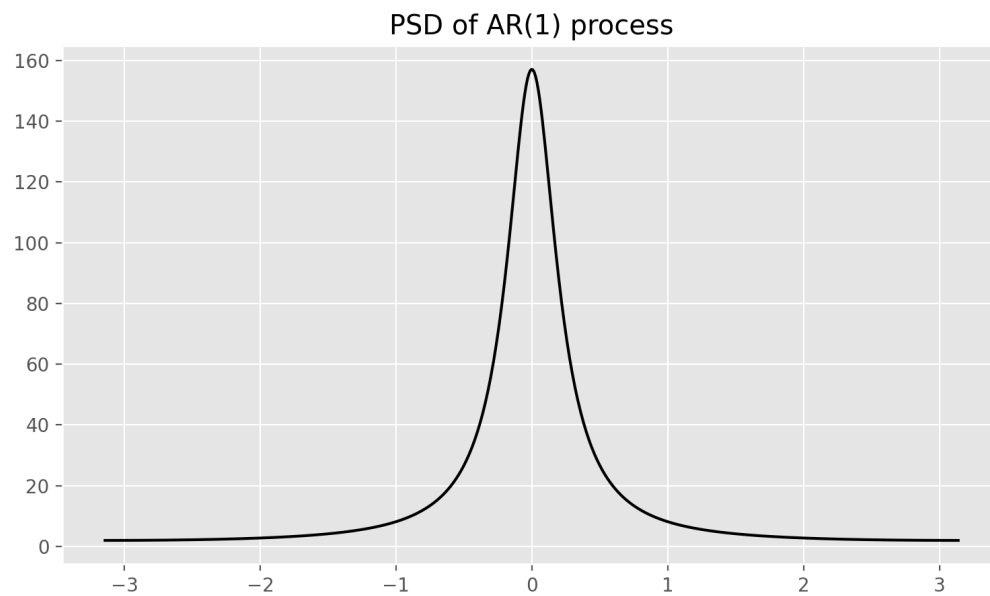


Figura 26 – Espectro de um processo AR(1) com  $\alpha_1 = 0.8$

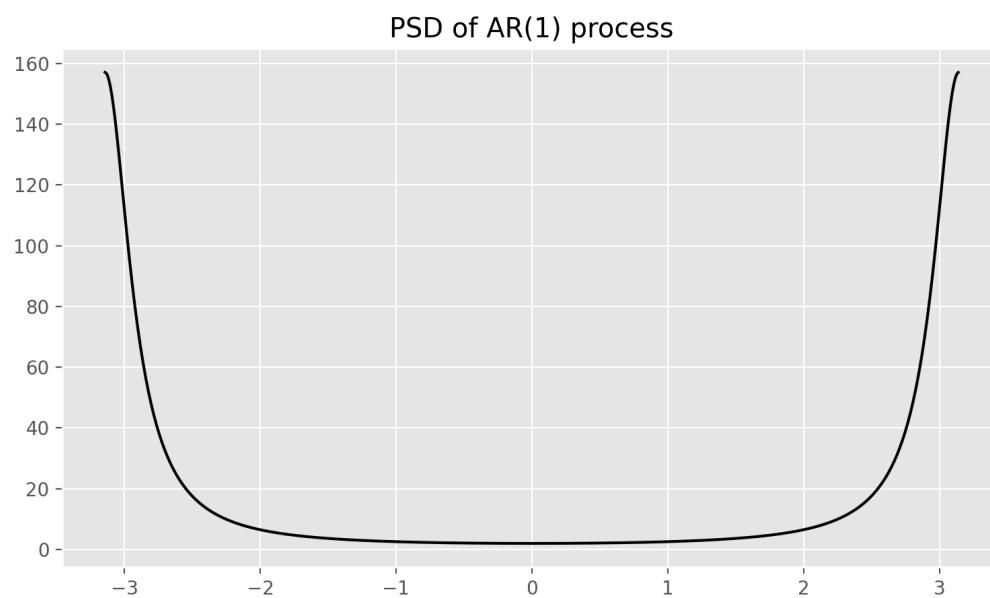


Figura 27 – Espectro de um processo AR(1) com  $\alpha_1 = -0.8$



#### 4.1.2.3 $AR(2)$

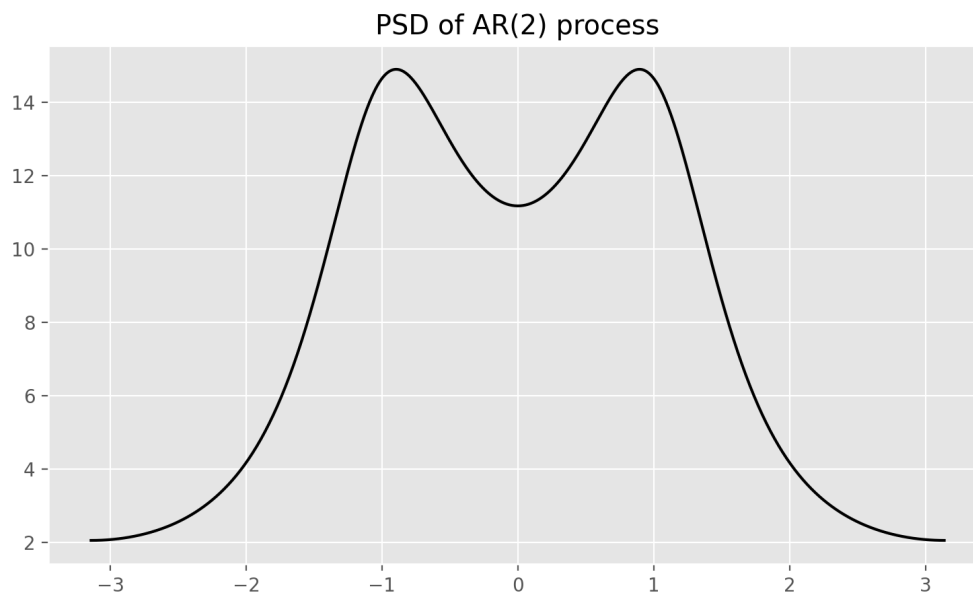


Figura 28 – Espectro de um processo  $AR(2)$  com  $\alpha_1 = 0.5$  and  $\alpha_2 = -0.25$

#### 4.1.2.4 $ARMA(4, 3)$

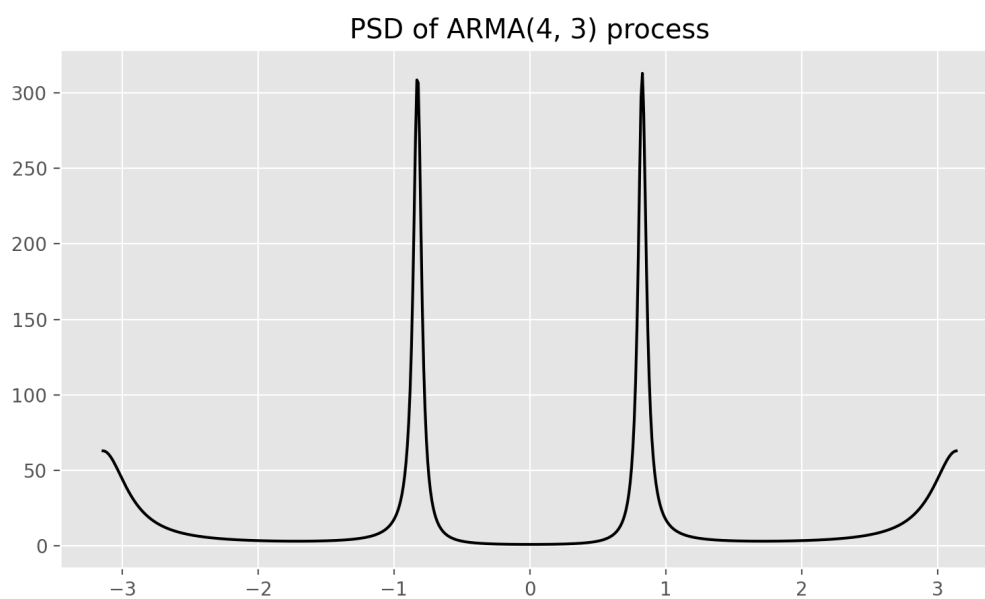


Figura 29 – Espectro de um processo  $ARMA(4, 3)$

### 4.1.3 Efeitos de raízes unitárias em espectros ARMA

A introdução de raízes unitárias no polinômio de média móvel ou autoregressivo de um processo ARMA tem claros efeitos em seu conteúdo espectral, como mencionado na subseção 3.3.2. O impacto dessas operações no espectro de um processo pode ser compreendido pelas alterações resultantes ao gráfico de polos e zeros ou pela modificação da série temporal.

Uma análise do gráfico de polos e zeros de um sistema ARMA torna evidente o efeito de integração na função de transferência do sistema: um polo é introduzido em  $e^{j\omega} = 0$  de forma a fornecer energia às componentes de baixa frequência do sinal. Reciprocamente podemos compreender diferenciação, que é a introdução de um zero em  $e^{j\omega} = 0$ , como a supressão de componentes de baixa frequência, resultado em uma operação análoga à filtragem passa-altas.

No domínio do tempo o efeito de diferenciação pode ser observado comparando as figuras 19 e 23. A série diferenciada claramente tem mais energia distribuída em torno de componentes de alta frequência. Isso é de fato verdadeiro para qualquer sinal: diferenciação no domínio do tempo age de forma a enviesar o conteúdo espectral para altas frequências. O recíproco também é verdadeiro: integral um sinal tende seu conteúdo espectral para baixas frequências por meio da introdução de tendências estocásticas.

Apresenta-se agora uma versão diferenciada do espectro do processo MA(1) da figura 25. Note que o sinal passa por uma filtragem passa altas, como mencionado.

#### 4.1.3.1 *MA(1) Diferenciado*

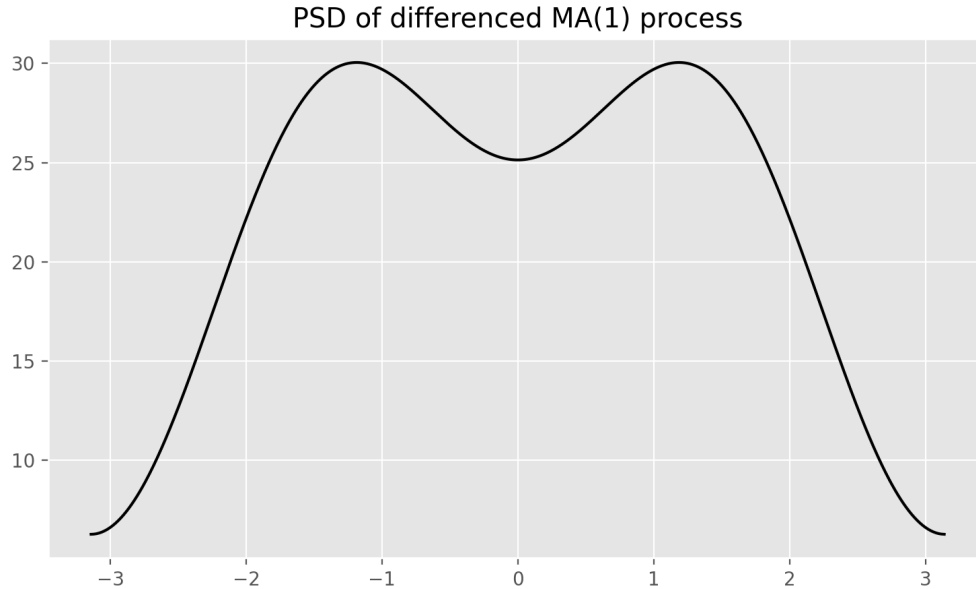


Figura 30 – Espectro de um processo MA(1) com  $\beta_1 = -0.5$  diferenciado.

## 4.2 Representações Não Estacionárias

O fato da grande maioria dos sinais produzidos por fontes reais serem não estacionários leva à necessidade de desenvolver algum tipo de representação espectral não estacionária. A densidade de potência espectral, definida pela transformada de Wiener-Khinchin, é incapaz de representar variações espectrais ao longo do tempo. Uma ideia natural para gerar uma “densidade de potência espectral variante no tempo” é tomar a transformada de Fourier de uma função de autocorrelação variante no tempo. Tal função é chamada de função de autocorrelação instantânea. A representação resultante é não linear, gerando termos cruzados prejudiciais para o objetivo da transformada.

### 4.2.1 *Função de Autocorrelação Instantânea*

A função de autocorrelação instantânea é, na realidade, só a função de autocorrelação de um sinal não estacionário escrita em uma forma específica. Interessantemente o termo função de autocorrelação passou a denominar uma função dependente apenas de um atraso  $\tau$ , como é o caso da autocorrelação para sinais estacionários, ao invés de uma função dependente de dois instantes no tempo  $t_1$  e  $t_2$ . Inicialmente reescrevemos a função de autocorrelação geral  $R_{xx}$  de

um sinal  $x(t)$  como

$$R_{xx}(t_1, t_2) = E[x(t_1)x(t_2)]$$

Que também pode ser escrita em função de um instante de tempo  $t$  e um atraso  $\tau$  como

$$R_x(t, \tau) = E[x(t)x(t - \tau)]$$

Que é ligeiramente mais conveniente para computações. Note que se  $x(t)$  é estacionário a dependência no tempo é removida porque  $R_{xx}$  possui o mesmo valor para todo  $t$ .

Um pequeno ajuste de notação nos leva a

$$\mathcal{R}_x(t, \tau) = E \left[ x \left( t - \frac{\tau}{2} \right) x \left( t + \frac{\tau}{2} \right) \right]$$

Usando um novo símbolo para indicar que enfim chegamos à função de autocorrelação instantânea  $\mathcal{R}_x$ .

Uma representação em tempo frequência agora é natural. Já que o teorema de Wiener-Khinchin afirma que o espectro de um sinal estacionário é a transformada de Fourier de sua função de autocorrelação podemos de forma análoga assumir que a representação espectral de um sinal não estacionário será dado pela transformada de Fourier ao longo do eixo  $\tau$  da função de autocorrelação instantânea. Isso nos leva à distribuição de Wigner Ville.

#### **4.2.2 Distribuição de Wigner-Ville**

Definimos a distribuição de Wigner-Ville (WVD) como

$$\mathcal{W}_x(t, f) = \int_{-\infty}^{\infty} \mathcal{R}_x(t, \tau) e^{-j\omega\tau} d\tau$$

$$\mathcal{W}_x(t, f) = \int_{-\infty}^{\infty} x \left( t - \frac{\tau}{2} \right) x \left( t + \frac{\tau}{2} \right) e^{-j\omega\tau} d\tau$$

Essa representação natural pode ser considerada uma densidade de potência espectral instantânea. Sabe-se que a WVD otimiza a resolução tempo frequência (SCHOOL, 2021), uma propriedade desejável de representações tempo frequência (TFRs). Veremos agora que isso não ocorre sem problemas.

#### 4.2.2.1 Termos cruzados

Por construção a WVD é uma representação quadrática. Pelo princípio de superposição quadrática (HLAWATSCH; BOUDREAUX-BARTELS, 1992) sabemos que se  $x(t) = \mu x_1(t) + \lambda x_2(t)$  a representação WVD de  $x(t)$  é dada por

$$\mathcal{W}_x = \mu^2 \mathcal{W}_{x_1} + \lambda^2 \mathcal{W}_{x_2} + 2(\lambda\mu)^2 (\mathcal{W}_{x_1, x_2})$$

Onde  $\mathcal{W}_{z,y}$  representa a WVD cruzada entre  $x$  e  $y$ . Já que todo sinal de complexidade relevante é uma combinação linear na base de  $\cos(t)$  e  $\sin(t)$  esperamos que uma quantidade considerável de componentes cruzados seja introduzida. Esse é conhecido como o problema de termos cruzados das representações de Wigner Ville e é um dos motivos da transformada não ser a melhor escolha para grande parte de sinais reais apesar de sua ótima resolução tempo-frequência.

É conhecido que os termos cruzados possuem padrões de alta frequência (MARTINEZ-HERRERA *et al.*, 2023), levando à possibilidade de tornar a WVD mais representativa de seus auto termos por meio de filtragem. As diferentes formas possíveis e úteis de filtrar a WVD leva a um subconjunto da classe de distribuição de Cohen.

#### 4.2.3 Smoothed Pseudo Wigner Ville Distributions

A maioria dos membros da classe de distribuições de Cohen são essencialmente versões filtradas da WVD (SCHOOL, 2021). Um caso particularmente útil é conhecido como a *Smoothed Pseudo Wigner Ville Distribution* (Smoothed Pseudo Wigner Ville Distribution).

$$SPWVD(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_t(t - \tau) h_f(f - \phi) d\tau d\phi$$

em que  $h_t$  representa o filtro aplicado ao longo do tempo e  $h_f$  ao longo da frequência.

Se o projeto dos filtros for bem sucedido a supressão de termos cruzados é suficientemente bem sucedida a ponto de justificar o uso da SPWVD ao invés de métodos mais simples

de representação em tempo frequência.

## 5 TESTES ESTATÍSTICOS

### 5.1 Teste de *Dickey-Fuller*

Os testes de *Dickey-Fuller* testam a hipótese nula de presença de raízes unitárias no processo gerador de uma série temporal com a hipótese alternativa de estacionariedade.

O teste mais simples de *Dickey-Fuller* assume que o processo gerador da série temporal em questão é dado pela equação 5.1, na qual  $\varepsilon$  é ruído branco.

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (5.1)$$

A hipótese nula do teste é a presença de raízes unitárias em  $y_t$ , isso é,  $\phi = 1$ , e a hipótese alternativa é  $\phi < 1$ , correspondente à estacionariedade.

$$\begin{cases} H_0 : \phi = 1, \text{ não estacionariedade (raízes unitárias, possível diferença-estacionariedade)} \\ H_1 : \phi < 1, \text{ estacionariedade} \end{cases}$$

Em seguida  $Ly_t$  é subtraído de ambos os lados da equação 5.1, resultando no desenvolvimento a seguir:

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + \varepsilon$$

$$\nabla y_t = (\phi - 1)y_{t-1} + \varepsilon$$

$$\nabla y_t = \delta y_t + \varepsilon$$

As hipóteses do teste são agora reformuladas para as seguintes

$$\begin{cases} H_0 : \delta = 0, \text{ não estacionariedade (raízes unitárias, diferença-estacionariedade)} \\ H_1 : \delta < 0, \text{ estacionariedade} \end{cases}$$

TODO: definir estatística de teste

A estatística de teste é computada e comparada com um valor crítico proveniente da distribuição de *Dickey-Fuller* (geralmente sob  $p = 0.5$ ) para rejeição ou não da hipótese nula.

O modelo da série temporal, dado explicitamente pela equação 5.1, pode ser alterado para testar raízes unitárias com constante e com constante e tendência determinística no tempo por meio das equações 5.2 e 5.3, respectivamente. Note que ambas as equações são apresentadas em função de  $\delta$ . É mais comum testar por apenas raízes unitárias, uma vez que uma análise subjetiva visual, por correlatograma ou decomposição juntamente de remoção de tendência ou sazonalidade é tipicamente realizada antes de um teste estatístico de estacionariedade.

$$\nabla y_t = \delta y_{t-1} + u_t + a_0 \quad (5.2)$$

$$\nabla y_t = \delta y_{t-1} + u_t + a_0 + a_1 t \quad (5.3)$$

O teste aumentado de *Dickey-Fuller*(ADF) modela o processo de forma mais geral, incluindo na equação 5.1 termos representativos de processos estacionários arbitrários. O teste aumentado é projetado para remover autocorrelação do processo de validação de hipótese. De forma identicamente análoga ao teste de DF temos expansões do ADF para incluir constantes e tendências como nas equações 5.2 e 5.3 apesar de que, como no teste de DF, essas variações são pouco usadas. A estatística de teste do ADF é negativa, isso é, quanto menor seu valor maior a rejeição da hipótese nula de não estacionariedade (maior certeza de estacionariedade).

Como em qualquer teste de hipótese um valor  $p$  maior que 0.05 indica falha em rejeitar a hipótese nula, nesse caso correspondendo à impossibilidade de constatar estacionariedade. Um valor  $p$  menor ou igual a 0.05 indica rejeição da hipótese nula, correspondendo à conclusão que a série sob análise é estacionária.

Na prática o teste mais usado é o ADF que é uma simples extensão mais robusta do teste de *Dickey-Fuller*. Implementações eficiente e populares existem para R e Python.

### 5.1.1 Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

TODO: check this

O teste KPSS desempenha uma função semelhante ao ADF com a relevante diferença de inerentemente modelar uma tendência linear no tempo por meio da equação 5.4. Seu desenvolvimento matemático é análogo porém mais trabalhoso que o caso do ADF e será portanto omitido.



$$y_t = \phi y_{t-1} + \varepsilon_t + \beta t \quad (5.4)$$

Com  $\varepsilon_t$  representando ruído branco. O teste em seguida define as seguintes hipóteses:

$$\begin{cases} H_0 : \text{a série apresenta tendência-estacionariedade} \\ H_1 : \text{a série apresenta raízes unitárias} \end{cases}$$

Como em qualquer teste de hipótese um valor  $p$  maior que 0.05 indica falha em rejeitar a hipótese nula, nesse caso correspondendo à impossibilidade de constatar que a série não apresenta tendência estacionariedade, boa evidência de que a série é tendência-estacionária. Um valor  $p$  menor ou igual a 0.05 indica rejeição da hipótese nula, correspondendo à conclusão que a série sob análise possui raízes unitárias e é portanto não estacionária.

Observe que há uma diferença crítica: a alternativa nula não postula não estacionariedade, como no caso do ADF, mas sim tendência-estacionariedade (seção 2.8.2) decorrente diretamente da inclusão de tendência linear no modelo da equação 5.4. A diferença principal da alteração da hipótese nula é que o KPSS é usado para investigar presença de estacionariedade sob uma tendência determinística (tendência-estacionariedade) e o ADF (tipicamente) de estacionariedade propriamente dita. O KPSS é bem implementado em R e Python.

## 5.2 Box-Pierce

## 5.3 Ljung-Box-Pierce

## **6 APLICAÇÕES**

## **7 CONCLUSÃO**

## REFERÊNCIAS

- AGUIRRE, L. A. **Introdução à Identificação de Sistemas. Técnicas Lineares e não Lineares Aplicadas a Sistemas. Teoria e Aplicação.** [S.l.]: UFMG, 2015.
- ATHANASOPOULOS, G.; HYNDMAN, R. **Forecasting: Principles and Practice.** [S.l.]: O'Reily, 2018.
- BENDAT, J.; PIERSOL, G. **Random Data Analysis and Measurement Procedures.** [S.l.]: John Wiley & Sons, 2000.
- BIANCIONINI, S. **Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation.** [S.l.]: Springer, 2016.
- BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and control.** [S.l.]: Wiley, 2016.
- CHATFIELD, C.; XING, H. **The Analysis of Time Series: An introduction with r.** [S.l.]: CRC Press, 2019.
- HAMILTON, J. **Time Series Analysis.** [S.l.]: Princeton University Press, 1994.
- HLAWATSCH, F.; BOUDREAUX-BARTELS, G. F. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 1992.
- HYLLEBERG, S. **Modelling seasonality.** [S.l.]: Oxford University Press, 1992.
- HYNDMAN, R. **Forecasting Weekly Data.** 2014. Available at: <<https://robjhyndman.com/hyndsight/longseasonality/>>.
- HYNDMAN, R. **Forecasting with long seasonal periods.** 2014. Available at: <<https://robjhyndman.com/hyndsight/longseasonality/>>.
- LADIRAY, D.; QUENNEVILLE, B. **Seasonal Adjustment with the X-11 method.** [S.l.]: Springer, 2001.
- MARTINEZ-HERRERA, A. L.; LEDESMA-CARRILLO, L. M.; FERRUCHO-ALVAREZ, E. R.; LOPEZ-RAMIREZ, M.; RODRIGUEZ-DONATE, C.; CABAL-YEPEZ, E. Tunable hyperbolic cohen-class kernel for cross-term diminishing in time–frequency distributions. **Mechanical Systems and Signal Processing**, Elsevier, v. 186, p. 109850, 2023.
- MCQUARRIE, A. D.; TSAI, C.-L. **Regression and time series model selection.** [S.l.]: World Scientific, 1998.
- MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras.** [S.l.]: Editora Blucher, 2017.
- SCHOOL, S. Fourier, gabor, morlet or wigner: Comparison of time-frequency transforms. *arXiv*, 2021.
- STOFFER, R. S. D. **Time Series Analysis and Its Applications.** [S.l.]: Springer, 2017.
- WOLD, H. **A Study In the Analysis of Stationary Time Series.** [S.l.]: Alqwist & Wiskell, 1954.