

# Análise Espectral

Airlane P. Alencar

4 de Junho de 2018



# Índice

- 1 Objetivo
- 2 Comportamento cíclico
- 3 DFT e Periodograma
- 4 Densidade espectral
- 5 Esperança do periodograma
- 6 Transformadas Cosseno e Seno
- 7 Transformadas Cosseno e Seno
- 8 Distribuição do Periodograma



## *Objetivo*

Decompor o processo estacionário em uma combinação de senóides com coeficientes aleatórios não correlacionados.

É também denominada análise no domínio da frequência em vez de no domínio do tempo.



## Comportamento cíclico

$$X_t = A\text{sen}(2\pi\nu t) + B\text{cos}(2\pi\nu t) \quad (1)$$

$$= C\text{sen}(2\pi\eta t + \phi), \quad (2)$$

onde A e B são v.a.s não correlacionadas com média 0 e variância  $\sigma^2 = 1$ .

## Expressões úteis

$$\text{sen}(a \pm b) = \text{sen}(a)\text{cos}(b) \pm \text{sen}(b)\text{cos}(a)$$

$$\text{cos}(a \pm b) = \text{cos}(a)\text{cos}(b) \mp \text{sen}(b)\text{sen}(a)$$

$$C\text{sen}(2\pi\eta t + \phi) = C\text{cos}(\phi)\text{sen}(2\pi\eta t) + C\text{sen}(\phi)\text{cos}(2\pi\eta t) \quad (3)$$

Então  $A = C\text{cos}(\phi)$ ,  $B = C\text{sen}(\phi)$ ,  $A^2 + B^2 = C^2$  e  $\text{tg}(\phi) = \frac{B}{A}$ .

Calcule a esperança, variância e covariância do processo  $X_t$ .



## Comportamento cíclico

$$X_t = A \sin(2\pi\nu t) + B \cos(2\pi\nu t)$$

$$E(X_t) = 0$$

$$\text{Var}(X_t) = \sigma^2 [\sin^2(2\pi\nu t) + \cos^2(2\pi\nu t)] = \sigma^2$$

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= \text{Cov}(A \sin(2\pi\nu t) + B \cos(2\pi\nu t), \\ &\quad A \sin(2\pi\nu(t+h)) + B \cos(2\pi\nu(t+h))) = \\ &= \sigma^2 \cos[(2\pi\nu t) - (2\pi\nu(t+h))] = \\ &= \sigma^2 \cos(2\pi\nu h) \end{aligned}$$

O processo é estacionário.



# Múltiplas frequências

## *Múltiplas frequências*

$$X_t = \sum_{k=1}^q A_k \cos(2\pi\omega_k t) + B_k \sin(2\pi\omega_k t)$$

$A_k$  e  $B_k$  variáveis não correlacionadas com média 0 e variância  $\sigma_k^2$  e  $\omega_k$  frequências distintas.

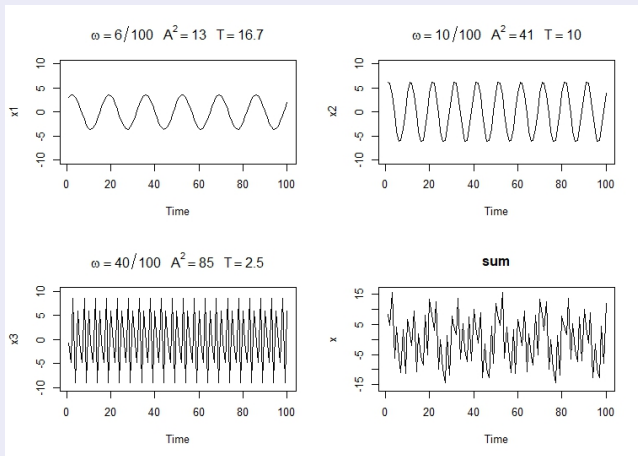
$$\gamma(h) = \sum_{k=1}^q \sigma_k^2 \cos(2\pi\omega_k h)$$

$$\gamma(0) = \sum_{k=1}^q \sigma_k^2 \tag{4}$$

# Múltiplas frequências

## Exemplo

Figura 1: Soma de 3 componentes periódicas



## Periodograma - ex. 2.8 SS

Na prática, muitas vezes é conhecido a frequência do sinal, como por exemplo um ciclo por ano ou marés,...

Quando não se sabe frequência, podemos ajustar um modelo de regressão não linear com senos e cossenos para estimar  $\omega$ .

Também podem ser testados diferentes valores de  $\omega$  e para as frequências  $\omega_j = j/n$ ,  $j = 1, \dots, n/2 - 1$ , estimando-se os coeficientes  $A_j$  e  $B_j$  obtemos ( $\hat{\beta} = (X^T X)^{-1} X^T y$ )

$$\widehat{A}(j/n) = \frac{\sum_{t=1}^n x_t \cos(2\pi t j/n)}{\sum_{t=1}^n \cos^2(2\pi t j/n)} = \frac{2}{n} \sum_{t=1}^n x_t \cos(2\pi t j/n)$$

$$\widehat{B}(j/n) = \frac{\sum_{t=1}^n x_t \sin(2\pi t j/n)}{\sum_{t=1}^n \sin^2(2\pi t j/n)} = \frac{2}{n} \sum_{t=1}^n x_t \sin(2\pi t j/n)$$

Para  $j = 0, n/2$ :  $\hat{A}(0) = 1/n \sum_{t=1}^n x_t$ ,  $\hat{A}(n/2) = 1/n \sum_{t=1}^n (-1)^t x_t$ ,  $\hat{B}(0) = \hat{B}(n/2) = 0$ .



## Periodograma - ex. 2.8 p.68 SS

Os coeficientes medem a correlação dos dados com as senóides e para a frequência  $\omega_j = j/n$

Medimos o efeito da frequência  $j/n$  usando uma medida da correlação ao quadrado, o periodograma padronizado (scaled):

$$P(j/n) = \widehat{A(j/n)}^2 + \widehat{B(j/n)}^2$$

Não é necessário ajustar uma regressão com  $n$  explicativas.

$$x_t = \sum_{j=0}^{n/2} A(j/n) \cos(2\pi t j/n) + B(j/n) \sin(2\pi t j/n)$$

Note que as variáveis explicativas formam uma base ortogonal e não há o erro.

## Periodograma - ex. 2.8 SS

Pode ser usada a transformada de Fourier discreta (DFT) com coeficientes complexos:

$$d(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \exp(-2\pi i t j/n),$$

com  $\exp(-2\pi i t j/n) = \cos(2\pi t j/n) - i \sin(2\pi t j/n)$ . Os coeficientes podem ser calculados mais rapidamente usando a FFT em vários programas.

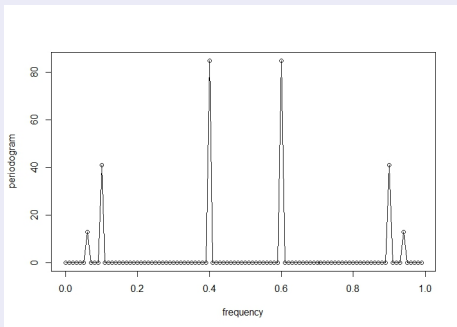
A transformada inversa de Fourier é

$$x_t = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} d(\omega_j) e^{-2\pi i \omega_j t}, \quad t = 1, \dots, n. \quad (5)$$

O periodograma é

$$I(j/n) = |d(j/n)|^2 = \frac{1}{n} \left[ \sum_{t=1}^n x_t \cos(2\pi t j/n) \right]^2 + \frac{1}{n} \left[ \sum_{t=1}^n x_t \sin(2\pi t j/n) \right]^2,$$

Figura 2: Periodograma scaled



$$P(6/100) = \widehat{A(6/100)}^2 + \widehat{B(6/100)}^2 = 2^2 + 3^2 = 13$$

A variância é decomposta nas 3 frequências em (4).

<https://robjhyndman.com/hyndsight/cyclicts/>

Dados observados a cada meia hora com sazonalidade a cada dia e semana.

Programa em R.



## Densidade espectral

Se a função de autocovariância de um processo estacionário satisfaz certa independência assintótica de modo que

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

então a densidade espectral tem representação

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, \quad -1/2 \leq \omega \leq 1/2.$$

$$e^{i\omega} = \cos(\omega) + i \sin(\omega) \text{ e } i = \sqrt{-1}.$$



$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, -1/2 \leq \omega \leq 1/2.$$

### Propriedades - Densidade espectral

$f(\omega)$  é limitada, não negativa, uniformemente contínua, é par e periódica de período 1. (Morettin e Tolo)

- $f(\omega)$  é limitada pois senos e cossenos também são e  $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ .

$$\begin{aligned} |f(\omega + \nu) - f(\omega)| &\leq \sum_h |e^{-2\pi i(\omega+\nu)h} - e^{-2\pi i\omega h}| |\gamma(h)| \\ &= \sum_h |\gamma(h)| |e^{-2\pi i\omega h}| |e^{-2\pi i\nu h} - 1| \end{aligned}$$

o último termo vai a zero quando  $\nu$  vai a zero e  $f()$  é uniformemente contínua.



$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \exp(-2\pi i \omega_j t), j = 0, 1, \dots, n-1, \quad (6)$$

onde  $\omega_j = j/n$  são as frequências fundamentais. Note que  $d(0) = n^{1/2} \bar{x}$ . Então, para  $j \neq 0$ , como  $\sum_{t=1}^n \exp(-2\pi i \omega_j t) = 0$ , podemos escrever a DFT como

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n (x_t - \bar{x}) \exp(-2\pi i \omega_j t), j = 0, 1, \dots, n-1, \quad (7)$$

E o periodograma para  $j \neq 0$ ,

$$\begin{aligned} I(j/n) &= |d(j/n)|^2 = 1/n \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) \exp^{-2\pi i \omega_j (t-s)} \\ &= 1/n \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) \exp^{-2\pi i \omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \widehat{\gamma}_h \exp^{-2\pi i \omega_j h} \end{aligned}$$



$$I(j/n) = \sum_{h=-(n-1)}^{n-1} \widehat{\gamma}_h \exp^{-2\pi i \omega_j h}$$

Dizemos que o periodograma é a densidade espectral amostral.

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, -1/2 \leq \omega \leq 1/2.$$





Esperança do periodograma, centrando as obs. com relação a  $\mu$ :

$$\begin{aligned} E[I(j/n)] &= 1/n \sum_{t=1}^n \sum_{s=1}^n E(x_t - \mu)(x_s - \mu) \exp^{-2\pi\omega_j(t-s)} \\ &= 1/n \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} E(x_{t+|h|} - \mu)(x_t - \mu) \exp^{-2\pi\omega_j h} \\ &= \sum_{h=-(n-1)}^{n-1} \left( \frac{n-|h|}{n} \right) \gamma(h) \exp^{-2\pi\omega_j h} \end{aligned}$$

Para  $\omega \neq 0$  e  $\omega_{j:n} \rightarrow \omega$  quando  $n \rightarrow \infty$  de modo que

$$E[I(\omega_{j:n})] \rightarrow f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp^{-2\pi\omega h}$$

O periodograma é assintoticamente não viesado.

Como  $I(\omega) = |d(\omega)|^2 \geq 0$ , então  $f(\omega) \geq 0$ .



Para os dados  $x_1, \dots, x_n$ , e  $\omega_j = j/n, j = 0, \dots, n - 1$ , as transformadas cosseno e seno são

$$d_c(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \cos(2\pi\omega_j t) \quad (8)$$

$$d_s(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \sin(2\pi\omega_j t) \quad (9)$$

$$(10)$$

O periodograma é  $I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j)$ .



## ANOVA Espectral

Série  $x_1, \dots, x_n$  com  $n$  ímpar e para  $m = \frac{n-1}{2}$  temos como decompor

$$x_t = a_0 + \sum_{j=1}^m a_j \cos(2\pi\omega_j t) + b_j \sin(2\pi\omega_j t)$$

e usando regressão  $a_0 = \bar{x}$ ,

$$a_j = \frac{2}{n} \sum_{t=1}^n x_t \cos(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_c(\omega_j)$$

$$b_j = \frac{2}{n} \sum_{t=1}^n x_t \sin(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_s(\omega_j)$$

Podemos escrever

$$\begin{aligned} (x_t - \bar{x}) &= \sum_{j=1}^m a_j \cos(2\pi\omega_j t) + b_j \sin(2\pi\omega_j t) \\ &= \frac{2}{\sqrt{n}} \sum_{j=1}^m d_c(\omega_j) \cos(2\pi\omega_j t) + d_s(\omega_j) \sin(2\pi\omega_j t) \end{aligned}$$



Elevando ao quadrado e depois somando em  $t$ , os termos cruzados são nulos e temos:

$$\begin{aligned}
 \sum_{t=1}^n (x_t - \bar{x})^2 &= \frac{4}{n} \sum_{t=1}^n \sum_{j=1}^m d_c^2(\omega_j) \cos^2(2\pi\omega_j t) + d_s^2(\omega_j) \sin^2(2\pi\omega_j t) \\
 &= \frac{4}{n} \left[ \sum_{j=1}^m d_c^2(\omega_j) \sum_{t=1}^n \cos^2(2\pi\omega_j t) + d_s^2(\omega_j) \sum_{t=1}^n \sin^2(2\pi\omega_j t) \right] \\
 &= 2 \left[ \sum_{j=1}^m d_c^2(\omega_j) + d_s^2(\omega_j) \right] = 2 \sum_{j=1}^m I(\omega_j),
 \end{aligned}$$

pois  $\sum_{t=1}^n \cos^2 = \sum_{t=1}^n \sin^2 = n/2$ .

ANOVA: Decomposição soma de quadrados total ( $gl=n-1$ ) em componentes de diferentes frequências aproximadamente não correlacionados e com 2 graus de liberdade para cada frequência ( $m = (n - 1)/2$ ).



$x_t \sim N \implies d_c \sim N$  e  $d_s \sim N$ .

Como achamos a esperança do periodograma, obtemos:

$$\text{cov}[d_c(\omega_j), d_c(\omega_k)] = \sum_{s=1}^n \sum_{t=1}^n \gamma(t-s) \cos(2\pi\omega_j s) \cos(2\pi\omega_k t)$$

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = \sum_{s=1}^n \sum_{t=1}^n \gamma(t-s) \cos(2\pi\omega_j s) \sin(2\pi\omega_k t)$$

$$\text{cov}[d_s(\omega_j), d_s(\omega_k)] = \sum_{s=1}^n \sum_{t=1}^n \gamma(t-s) \sin(2\pi\omega_j s) \sin(2\pi\omega_k t)$$



No apêndice C2 (Lema C3) de Shumway and Stoffer, para  $\sum_{-\infty}^{\infty} |h|\gamma(h) < \infty$ :

$$\begin{aligned} \text{cov}[d_c(\omega_j), d_c(\omega_k)] &= \begin{cases} f(\omega_j)/2 + \epsilon_n, & \omega_j = \omega_k \\ \epsilon_n, & \omega_j \neq \omega_k \end{cases} \\ \text{cov}[d_s(\omega_j), d_s(\omega_k)] &= \begin{cases} f(\omega_j)/2 + \epsilon_n, & \omega_j = \omega_k \\ \epsilon_n, & \omega_j \neq \omega_k \end{cases} \\ \text{cov}[d_c(\omega_j), d_s(\omega_k)] &= \epsilon_n \end{aligned}$$

Erro:  $|\epsilon_n| \leq \theta/n$ .

Para  $\omega_j = \omega_k = 0$  ou  $1/2$  o fator  $1/2$  é  $1$  e  $d_s(0) = d_s(1/2) = 0$ .



$x_1, \dots, x_n$  iid  $0, \sigma^2$ , pelo TCL:  $d_{c/s}(\omega_{j:n} \sim AN(0, \sigma^2))$  e são independentes.

Assim,  $\frac{2I(\omega_{j:n})}{\sigma^2} \rightarrow \chi_2^2$ .

Do caso iid para o processo linear:

P4.2: Se  $\sum_{-\infty}^{\infty} |h|\gamma(h) < \infty$ ,  $x_t = \sum_{j=-\infty}^{\infty} \psi_j a_{t-j}$ ,  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ , com  $a_t \sim iid(0, \sigma^2)$ , então para  $m$  diferentes freq  $\omega_{j:n} \rightarrow \omega_j$

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} iid \chi_2^2 \quad (11)$$

Para os quantis  $a$  e  $b$  da  $\chi_2^2$ , temos  $\alpha = P(a < \frac{2I(\omega_{j:n})}{f(\omega_j)} < b)$

IC para espectro  $f(\omega) = \left[ \frac{2I(\omega_{j:n})}{b}; \frac{2I(\omega_{j:n})}{a} \right]$

