

MAE 5871: Análise Espectral de Séries Temporais

Pedro A. Morettin

Instituto de Matemática e Estatística
Universidade de São Paulo
pam@ime.usp.br
<http://www.ime.usp.br/~pam>

Aula 8

21 de setembro de 2021

Sumário

- 1 Distribuições normais
- 2 Transformada de Fourier finita
- 3 O periodograma
- 4 Testes para periodicidade

Normal p -variada

1. As v.a. X_1, \dots, X_p , reais, têm uma distribuição conjunta normal multivariada de dimensão p (ou, simplesmente, p -variada), se sua função densidade for dada por

$$f(x_1, \dots, x_p) = (2\pi)^{-p/2} |\mathbf{A}| \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{A}(\mathbf{x} - \boldsymbol{\mu})\right\}, \quad (1)$$

onde \mathbf{A} é uma matriz $p \times p$ positiva definida, $|\mathbf{A}|$ é o determinante de \mathbf{A} , $\mathbf{x} = (x_1, \dots, x_p)'$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$.

2. Se $\mathbf{X} = (X_1, \dots, X_p)'$, segue-se que $E(\mathbf{X}) = \boldsymbol{\mu}$ e a matriz de covariâncias de \mathbf{X} é $\boldsymbol{\Sigma} = \mathbf{A}^{-1}$.
3. Se \mathbf{X} tem distribuição p -variada com média $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, escrevemos $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Pode-se verificar que a função característica de \mathbf{X} é

$$e^{i\mathbf{t}' \boldsymbol{\mu} - \mathbf{t}' \boldsymbol{\Sigma} \mathbf{t} / 2},$$

para todo vetor real \mathbf{t} , de ordem $p \times 1$.

A densidade (1) não existe se $\boldsymbol{\Sigma}$ for singular.

Normal complexa

Se $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ for um vetor aleatório $p \times 1$, então \mathbf{Z} terá uma distribuição normal complexa multivariada, se o vetor $(\mathbf{X}, \mathbf{Y})'$, de ordem $2p \times 1$, tiver uma distribuição normal multivariada, de dimensão $2p$. Há alguma ambigüidade na escolha da distribuição conjunta desse vetor e, portanto, nos fixaremos na

Definição 1. Se $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ for um vetor $p \times 1$, com componentes complexas, dizemos que \mathbf{Z} tem *distribuição normal complexa multivariada* com média $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, se e somente se o vetor $(\mathbf{X}, \mathbf{Y})'$, de ordem $2p \times 1$, com componentes reais, tiver distribuição normal $2p$ -variada, com média $(\mathcal{R}\boldsymbol{\mu}, \mathcal{I}\boldsymbol{\mu})'$ e matriz de covariâncias $\boldsymbol{\Omega} = \frac{1}{2}\mathbf{V}$, com

$$\mathbf{V} = \begin{bmatrix} \mathcal{R}\boldsymbol{\Sigma} & -\mathcal{I}\boldsymbol{\Sigma} \\ \mathcal{I}\boldsymbol{\Sigma} & \mathcal{R}\boldsymbol{\Sigma} \end{bmatrix}$$

para algum vetor $p \times 1$ $\boldsymbol{\mu}$ e alguma matriz $\boldsymbol{\Sigma}$, de ordem $p \times p$, hermitiana e não-negativa definida.

Escreveremos $\mathbf{Z} \sim N_p^c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Normal complexa

- Um caso de interesse é $p = 1$. Se $Z = X + iY$ tem distribuição normal complexa $N_1^c(\mu, \sigma)$, então X e Y serão v.a. independentes, com distribuições normais $N_1(\mu_1, \sigma/2)$ e $N_1(\mu_2, \sigma/2)$, respectivamente, onde $\mu_1 = \mathcal{R}\mu$ e $\mu_2 = \mathcal{I}\mu$.
- Um fato importante é o

Teorema 1. (Isserlis) *Se $\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4)'$ for uma v.a. com distribuição normal complexa, com média zero e matriz de covariâncias Σ , então*

$$E\{Z_1 Z_2 Z_3 Z_4\} = E\{Z_1 Z_2\}E\{Z_3 Z_4\} + E\{Z_1 Z_3\}E\{Z_2 Z_4\} + E\{Z_1 Z_4\}E\{Z_2 Z_3\}. \quad (2)$$

- No caso de v.a. normais reais com média zero, o teorema reduz-se a

$$\text{Cov}\{Z_1 Z_2, Z_3 Z_4\} = E\{Z_1 Z_3\}E\{Z_2 Z_4\} + E\{Z_1 Z_4\}E\{Z_2 Z_3\}.$$

Normal complexa

Algumas propriedades que podem ser de interesse são:

(i) Se \mathbf{Z} for uma v.a. complexa $p \times 1$, então

$$\boldsymbol{\Sigma}_{ZZ} = \boldsymbol{\Sigma}_{XX} + \boldsymbol{\Sigma}_{YY} + i(\boldsymbol{\Sigma}_{YX} - \boldsymbol{\Sigma}_{XY}).$$

(ii) Se $Z = X + iY$, então $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$.

(iii) Se $\mathbf{Z} \sim N_p^c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e se $\mathcal{I}\boldsymbol{\Sigma} = \mathbf{0}$, então as componentes de \mathbf{Z} são independentes.

Introdução

1. Nestas notas estudaremos a estimação do espectro de Fourier de um processo estacionário. A estatística básica a ser considerada é a **transformada de Fourier finita**, definida por

$$d^{(T)}(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t e^{-i\lambda t}, \quad (3)$$

$-\infty < \lambda < \infty$, dados os valores X_1, \dots, X_T do processo estacionário $\{X_t, t \in \mathbb{Z}\}$.

2. O primeiro estimador sugerido para o espectro foi o **periodograma**, que remonta a Schuster (1906), em sua análise da série de manchas solares (**sunspot numbers**). O periodograma consiste no quadrado do módulo da transformada de Fourier.
3. Como o periodograma é muito errático (tem variabilidade grande), estimadores mais estáveis são necessários. Uma possibilidade é considerar o periodograma suavizado. Outra, é considerar estimadores suavizados de covariâncias.
4. Outra possibilidade é considerar estimadores espectrais de **máxima entropia** ou **autorregressivos**. A ideia aqui é ajustar modelos autorregressivos aos dados e obter estimadores dos coeficiente e, depois, usar expressões para os espectros de modelos ARMA estacionários.

Transformada de Fourier discreta

1. Considere (3). É fácil ver que

- (i) $d^{(T)}(\lambda) = d^{(T)}(\lambda + 2\pi)$, isto é, tem período 2π ;
- (ii) $\overline{d^{(T)}(-\lambda)} = d^{(T)}(\lambda)$.

2. Portanto, basta considerar as frequências no intervalo $[-\pi, \pi]$. Embora (3) seja definida para todas as frequências nesse intervalo, ela, na prática, é calculada para frequências da forma $\lambda_\nu = \frac{2\pi\nu}{T}$, com $-\lceil(T-1)/2\rceil \leq \nu \leq \lfloor T/2 \rfloor$, chamadas **frequências de Fourier**. Obtemos a **transformada de Fourier discreta** (TFD)

$$d_\nu^{(T)} = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T X_t e^{-i2\pi\nu t/T}. \quad (4)$$

3. Vamos supor, no que segue, que X_t tem média zero, f.a.c.v. γ_τ e espectro $f(\lambda)$. Temos, então, que $E\{d_\nu^{(T)}\} = 0$. Calculemos $\text{Var}\{d_\nu^{(T)}\} = E\{|d_\nu^{(T)}|^2\}$.

1. Usando a representação espectral de X_t , obtemos

$$\begin{aligned}d_{\nu}^{(T)} &= (2\pi T)^{-1/2} \sum_{t=1}^T e^{-\lambda_{\nu} t} \int_{-\pi}^{\pi} e^{i\lambda t} dZ(\lambda) \\ &= \int_{-\pi}^{\pi} (2\pi T)^{-1/2} \sum_{t=1}^T e^{i(\lambda - \lambda_{\nu})t} dZ(\lambda).\end{aligned}$$

2. Chamando

$$\Delta^{(T)}(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T e^{i\lambda t}, \quad (5)$$

temos que

$$d_{\nu}^{(T)} = \int_{-\pi}^{\pi} \Delta^{(T)}\left(\lambda - \frac{2\pi\nu}{T}\right) dZ(\lambda). \quad (6)$$

3. É fácil ver que $|\Delta^{(T)}(\lambda)|^2$ é o núcleo de Fejér e comporta-se como uma função delta de Dirac, quando $T \rightarrow \infty$

1. Agora,

$$\begin{aligned} E\{|d_\nu^{(T)}|^2\} &= E\{d_\nu^{(T)} \overline{d_\nu^{(T)}}\} \\ &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Delta^{(T)}\left(\lambda - \frac{2\pi\nu}{T}\right) \overline{\Delta^{(T)}\left(\alpha - \frac{2\pi\nu}{T}\right)} E\{dZ(\lambda) \overline{dZ(\alpha)}\} \end{aligned}$$

2. Como $E\{dZ(\lambda) \overline{dZ(\alpha)}\} = f(\lambda) d\lambda$, se $\lambda = \alpha$, segue-se que

$$E\{|d_\nu^{(T)}|^2\} = \int_{-\pi}^{\pi} \left| \Delta^{(T)}\left(\lambda - \frac{2\pi\nu}{T}\right) \right|^2 f(\lambda) d\lambda. \quad (7)$$

3. Pela propriedade acima referida do núcleo de Fejér, segue-se que se $f(\lambda)$ for contínua, para T grande,

$$E\{|d_\nu^{(T)}|^2\} \approx f(\lambda_\nu), \quad (8)$$

e essa aproximação é tanto melhor quanto mais suave for $f(\lambda)$ na vizinhança de $\lambda_\nu = \frac{2\pi\nu}{T}$.

1. Vamos supor, agora, que X_t seja gaussiano, de média zero. Segue-se de (3) que $d_\nu^{(T)}$, sendo uma combinação linear de variáveis normais, terá uma distribuição normal complexa. Temos, então, o seguinte teorema limite central.
2. **Teorema 2.** *Se o espectro $f(\lambda)$ for contínuo, então as variáveis aleatórias $d_\nu^{(T)}$, $-[(T-1)/2] \leq \nu \leq [T/2]$, são assintoticamente independentes, quando $T \rightarrow \infty$, com distribuição assintótica normal $N_1^C(0, f(\lambda_\nu))$, se $\nu \neq 0$, $T/2$ e com distribuição assintótica $N_1(0, f(\lambda_\nu))$, se $\nu = 0$ ou $T/2$.*

1. O Teorema 2 é demonstrado sob a suposição que X_t é gaussiano. Se X_t for um processo estacionário não gaussiano, satisfazendo à condição $\sum_{\tau} |\tau| |\gamma_{\tau}| < \infty$, pode-se provar que

$$\text{Cov}\{d_{\nu}^{(T)}, d_{\kappa}^{(T)}\} = f(\lambda_{\nu}) + O(1),$$

e sob condições de regularidade adicionais sobre as frequências λ_{ν} , a variável $d_{\nu}^{(T)}$ tem distribuição normal complexa, como no Teorema 2. Para detalhes, veja Brillinger (1981, Cap. 4).

2. O Teorema 2 também é válido para uma frequência λ qualquer. Basta considerar uma sequência de inteiros $\{\nu_T\}$ e as frequências de Fourier $2\pi\nu_T/T$ convergindo para λ .
3. Outras referências para o aprofundamento desta seção são os livros de Hannan (1970), Priestley (1981), Brockwell e Davis (1991) e Fuller (1996).

O periodograma

1. A relação (8) sugere que um primeiro estimador que podemos considerar é o **periodograma**, definido por

$$I_{\nu}^{(T)} = |d_{\nu}^{(T)}|^2 = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t e^{-i\lambda_{\nu} t} \right|^2, \quad (9)$$

para frequências $\lambda_{\nu} = (2\pi\nu)/T$, $-[(T-1)/2] \leq \nu \leq [T/2]$.

2. Podemos, também, definir o periodograma para qualquer frequência $\lambda \in [-\pi, \pi]$, ou seja,

$$I^{(T)}(\lambda) = |d^{(T)}(\lambda)|^2, \quad (10)$$

mas, na prática, esse só poderá ser calculado para um número finito de frequências. Pode-se demonstrar que (10) é completamente determinado pelos seus valores nas frequências de Fourier.

3. De (8) segue-se que o periodograma é assintoticamente não viesado, isto é,

$$\lim_{T \rightarrow \infty} E\{I_{\nu}^{(T)}\} = f(\lambda_{\nu}). \quad (11)$$

Propriedades do periodograma

1. Vejamos, agora, a distribuição assintótica do periodograma. Vamos supor que o processo X_t seja gaussiano.

Teorema 3. *As ordenadas do periodograma $I_\nu^{(T)}$ são variáveis aleatórias assintoticamente independentes e com distribuição assintótica múltipla de uma variável qui-quadrado com dois graus de liberdade, se $\nu \neq 0$ e $\nu \neq T/2$, e com um grau de liberdade se $\nu = 0$ ou $\nu = T/2$.*

2. Temos, então, que **assintoticamente**,

$$\begin{aligned} E(I_\nu^{(T)}) &= f(\lambda_\nu), \\ \text{Var}(I_\nu^{(T)}) &= f^2(\lambda_\nu), \quad \nu \neq 0, \quad \nu \neq T/2, \\ \text{Var}(I_\nu^{(T)}) &= \begin{cases} 2f^2(0), & \text{se } \nu = 0 \\ 2f^2(\pi), & \text{se } \nu = T/2. \end{cases} \end{aligned} \quad (12)$$

Propriedades do periodograma

1. Vê-se, então, que embora o periodograma seja assintoticamente não viesado, ele não é consistente, o que significa que, mesmo aumentando o número de observações, a variância de $I_\nu^{(T)}$ não decresce e permanece no nível de $f^2(\lambda_\nu)$.
2. Pode-se demonstrar que as relações (13) valem para frequências quaisquer e que o Teorema 3 continua válido, removendo-se a suposição de que X_t seja gaussiano. Veja Brillinger (1981) para detalhes.
3. **Teorema 4.** Se X_t tiver média μ e $\sum_\tau |\gamma_\tau| < \infty$, então

$$E\{I^{(T)}(\lambda)\} = \int_{-\pi}^{\pi} |\Delta^{(T)}(\lambda - \alpha)|^2 f(\alpha) d\alpha + [\Delta^{(T)}(\lambda)]^2 \mu^2. \quad (13)$$

Propriedades do periodograma

Note que se $\lambda = (2\pi\nu)/T$, $\nu \neq 0$, $\nu \neq T/2$, o termo envolvendo μ no segundo membro de (13) anula-se e obtemos

$$E\{I^{(T)}(\lambda)\} = \int_{-\pi}^{\pi} |\Delta^{(T)}(\lambda - \alpha)|^2 f(\alpha) d\alpha. \quad (14)$$

Lembre também que $|\Delta^{(T)}(\cdot)|^2$ é o núcleo de Fejér, e, se $\lambda \neq 0$, $\lambda \neq \pi$, o segundo termo de (13) é pequeno e nos diz que a média do periodograma é aproximadamente igual a uma média ponderada do espectro com peso concentrado na vizinhança de λ .

Propriedades do periodograma

Teorema 4. Se λ_1 e λ_2 são frequências diferentes de zero ou π , então

$$\begin{aligned} & \text{Cov}\{I^{(T)}(\lambda_1), I^{(T)}(\lambda_2)\} \\ &= \left\{ \left[\frac{\text{sen } T[\lambda_1 + \lambda_2]/2}{T \text{sen}(\lambda_1 + \lambda_2)/2} \right]^2 + \left[\frac{\text{sen } T[\lambda_1 - \lambda_2]/2}{T \text{sen}(\lambda_1 - \lambda_2)/2} \right]^2 \right\} f^2(\lambda_1) \\ &+ O(T^{-1}). \end{aligned} \tag{15}$$

Propriedades do periodograma

Uma consequência do Teorema 4 é que, se as frequências λ_1 e λ_2 são frequências de Fourier, distintas de zero ou π , então,

$$\text{Cov}\{I^{(T)}(\lambda_1), I^{(T)}(\lambda_2)\} = O(T^{-1}). \quad (16)$$

Também, para $\lambda_1 = \lambda_2 = (2\pi\nu)/T$, obtemos de (15) que

$$\text{Var}\{I^{(T)}(\lambda_1)\} = f^2(\lambda_1) + O(T^{-1}). \quad (17)$$

A relação (16) nos diz que as ordenadas do periodograma são praticamente não correlacionadas, resultando o seu comportamento errático.

1. Para calcular a transformada de Fourier discreta são necessárias T^2 operações complexas. Por uma operação complexa entendemos uma multiplicação complexa seguida por uma adição complexa.
2. Cooley e Tukey (1965) introduziram um algoritmo, chamado **transformada rápida de Fourier**, conhecido por FFT (*fast Fourier transform*), que requer $T \log T$ operações.
3. Para séries longas, a economia de tempo é considerável. Dizemos que o algoritmo tem complexidade $O(T \log T)$.
4. Inicialmente, para o caso $T = r_1 r_2$, r_1 e r_2 inteiros e diferentes de um, é mostrado que são necessárias $T(r_1 + r_2)$ operações. Se $T = r_1 r_2 \cdots r_m$, serão necessárias $T(r_1 + r_2 + \dots + r_m)$ operações.
5. O caso $T = r^m$, com $r = 2$ ou 4 , oferece vantagens importantes sob o ponto de vista computacional. Nessa situação, o número total de operações é $Tmr = rT \log_r T$, que para $r = 2$ reduz-se a $2T \log_2 T$. Para outros detalhes, veja o artigo mencionado, Brillinger (1981) ou Bloomfield (2000).

- Veremos, mais adiante, que um procedimento para estimar $f(\lambda)$ consiste em considerar

$$\hat{f}(\lambda) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} w(\tau) \hat{\gamma}_{\tau} e^{-i\lambda\tau} \quad (18)$$

onde $\hat{\gamma}_{\tau}$ é um estimador da f.a.c.v. do processo e $w(\tau)$ é uma função peso.

- (18) é um estimador suavizado de covariâncias e corresponde a suavizar a f.a.c.v. no domínio do tempo e então transformar para o domínio de frequências.
- Outra possibilidade é obter estimadores suavizados de periodogramas, de modo que todas as computações são feitas no domínio de frequências; obtemos a TFD (4) usando uma FFT, depois obtemos o periodograma, ambas as estatísticas nas frequências de Fourier, e suavizamos ao redor de uma frequência de interesse. Aqui reside a vantagem do algoritmo, reduzindo o tempo de computação necessário.
- Se T não for uma potência de 2 ou da forma $T = r_1 \cdots r_m$, considere um inteiro adequado $T' > T$, T' uma potência de 2, e acrescente $T' - T$ zeros aos dados. Obtemos, depois, a TFD

$$d^{(T')}\left(\frac{2\pi\nu}{T'}\right) = \frac{1}{\sqrt{2\pi T'}} \sum_{t=1}^T X_t e^{-i2\pi\nu t/T'}$$

Teremos, então, um número maior de frequências do que no caso de usar T observações.

Exemplos

1. **Exemplo 1.** A Figura 1 mostra o periodograma da série de marés de Ubatuba. Notamos um pico dominante, correspondente à frequência semi-diurna (período de 12 horas) e dois picos menores, anteriores ao pico principal, correspondentes a períodos de 80 horas e 24 horas, respectivamente.
2. Vemos que esse periodograma apresenta picos, correspondendo a periodicidades presentes nos dados, mais uma componente aproximadamente constante, correspondendo a um ruído. Um modelo para essa série de marés seria constituído pela soma de um processo harmônico mais um ruído branco.

Exemplos

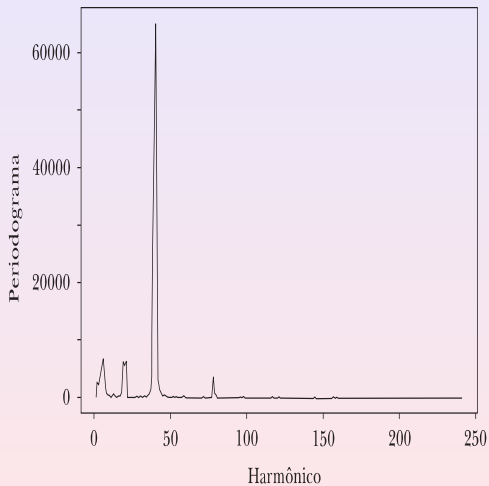


Figura 1: Periodograma da série de marés de Ubatuba.

Exemplos

1. **Exemplo 2.** Vamos retomar a série de chuvas de Fortaleza. Seu periodograma está representado na Figura 2. Vemos que há dois picos que se destacam, correspondentes aos possíveis períodos de 13 e 26 anos.
2. Utilizando testes de periodicidades apropriados, verifica-se que, de fato, esses períodos são estatisticamente significativos. Veja Morettin (2018) para uma análise completa.
3. Os gráficos envolvendo o periodograma e estimadores espectrais suavizados desse capítulo foram feitos por meio do software R e o pacote `astsa`.

Teste de Fisher

1. Vamos considerar aqui dois testes de periodicidades que utilizam as ordenadas do periodograma.
2. Suponha que um modelo adequado para a série temporal observada seja

$$Z_t = \sum_{i=1}^K R_i \cos(\omega_i t + \phi_i) + \varepsilon_t, \quad t = 1, \dots, N, \quad (19)$$

em que K , R_i e ω_i , $i = 1, \dots, K$, são constantes, $\phi_i \sim \mathcal{U}(-\pi, \pi)$ independentes, $\{\varepsilon_t\}$ sequência de ruídos brancos independentes de Z_t , com média zero e variância σ_ε^2 .

3. Utilizando uma amostra Z_1, Z_2, \dots, Z_N de (19) podemos calcular o periodograma $I_j^{(N)}$, $j = 1, \dots, [\frac{N}{2}]$. Mesmo que $R_i = 0$, todo i , é possível que ocorram picos nas ordenadas do periodograma devido a flutuações aleatórias. Em resumo, mesmo que o periodograma apresente vários picos, não poderemos concluir, *a priori*, que cada um dos picos corresponda a uma componente periódica na série Z_t .

Teste de Fisher

- Supondo que o processo Z_t seja Gaussiano, Fisher (1929) apresenta um procedimento para testar a hipótese

$$H_0 : R_i \equiv 0, \text{ todo } i \text{ (não existe periodicidade)}$$

baseado na estatística

$$g = \frac{\max I_j^{(N)}}{\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(N)}} = \frac{I^{(1)}}{\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(N)}} . \quad (20)$$

- Fisher mostrou que, para N ímpar, a distribuição exata de g , sob H_0 , é dada por

$$P(g > a) = n(1-a)^{n-1} - \binom{n}{2}(1-2a)^{n-1} + \dots + (-1)^x \binom{n}{x}(1-xa)^{n-1}, \quad (21)$$

em que

$$n = \left\lfloor \frac{N}{2} \right\rfloor \text{ e } x \text{ é o maior inteiro menor que } \frac{1}{a} \left(x = \left\lfloor \frac{1}{a} \right\rfloor \right).$$

Teste de Fisher

1. Assim, para um dado nível de significância α , podemos utilizar a equação (21) para encontrar o valor crítico $a(\alpha)$ tal que $P(g > a(\alpha)) = \alpha$. Se o valor observado da estatística (g_{obs}) for maior que $a(\alpha)$, rejeitaremos H_0 , o que significa afirmar que a série apresenta uma periodicidade igual a $1/\omega^*$, sendo ω^* a frequência em ciclos, correspondente a $I^{(1)}$.
2. Uma boa aproximação para a expressão (21) é obtida utilizando somente o primeiro termo da expansão, isto é,

$$P(g > a) \equiv n(1 - a)^{n-1}, \quad (22)$$

que pode ser utilizada para encontrar o p -valor aproximado do teste, fazendo $a = g_{\text{obs}}$.

3. Fisher (1929) obteve a expressão (21) utilizando argumentos geométricos; Grenander e Rosenblatt (1957) e Hannan (1970) apresentaram demonstrações analíticas.

Teste de Whittle

1. Whittle (1952) sugeriu uma extensão para o teste de Fisher, que fornece um teste para a segunda maior ordenada do periodograma ($I^{(2)}$), omitindo o termo $I^{(1)}$ do denominador da estatística g , isto é, considerando a estatística

$$g' = \frac{I^{(2)}}{\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(N)} - I^{(1)}} \quad (23)$$

e utilizando a distribuição g de Fisher com n substituído por $(n - 1)$.

2. Se a segunda maior ordenada for significativa, pode-se aplicar o procedimento de Whittle para testar a terceira maior ordenada ($I^{(3)}$) e assim por diante, até que se obtenha um resultado não significativo. Dessa maneira, obtemos uma estimativa do número de componentes periódicas, K , do modelo (19).
3. O teste de Whittle é bom se todas as componentes periódicas tiverem frequências iguais ou muito próximas a múltiplos de $2\pi/N$; se isto não ocorrer o poder do teste será bastante afetado. O pior caso ocorre quando uma ou mais frequências estiverem no ponto médio entre duas frequências de Fourier adjacentes, isto é, entre $2\pi j/N$ e $2\pi(j + 1)/N$.
4. Priestley (1981) apresenta outros testes de periodicidade que são adequados para processos mais gerais do que aquele representado pela expressão (19), ou seja, processos em que o termo de ruído ε_t é substituído por um modelo linear da forma $\sum_{u=0}^{\infty} g_u \varepsilon_{t-u}$.

Exemplos

Vamos aplicar o teste de Fisher à série Temperatura–Cananéia. Podemos verificar que $\max_j I_j^{(120)} = 67,05$, para $j = 10$. Utilizando (20) podemos calcular o valor da estatística de Fisher

$$g_{\text{obs}} = \frac{67,05}{\sum_{j=1}^{60} I_j^{(120)}} = \frac{67,05}{75,65} = 0,8863$$

e o p -valor, utilizando (21), é

$$\alpha^* = P(g > g_{\text{obs}}) \cong n(1 - g_{\text{obs}})^{n-1},$$

ou seja,

$$\alpha^* = P(g > 0,8863) \cong 60(1 - 0,8863)^{59} = 1,1695 \times 10^{-54},$$

que é menor do que qualquer valor usual de nível de significância. Esse p -valor nos leva a rejeitar a hipótese de não existência de periodicidade e a concluir que a série de temperatura em Cananéia possui uma periodicidade de $N/j = 120/10 = 12$ meses.

O teste de Fisher pode ser realizado utilizando a função `fisher.g.test` da biblioteca `GeneCycle` do R.

Exemplos

Vamos utilizar, agora, a série de Chuva de Fortaleza. Podemos verificar que

$$\max I_j^{(149)} = 312.911,17, \quad j = 12,$$

e

$$\sum_{j=1}^{75} I_j^{(149)} = 2.902.343,46.$$

Assim,

$$g_{\text{obs}} = \frac{312.911,17}{2.902.343,46} = 0,1078$$

e

$$\alpha^* = P(g > 0,1078) \approx 74 \times (1 - 0,1078)^{73} = 0,0179.$$

Fixando um nível de significância de 0,02, podemos concluir que existe uma periodicidade de $\frac{149}{12} = 12,41$ anos na série de chuvas em Fortaleza.

Exemplos

Para verificar a existência de uma segunda periodicidade, utilizamos a estatística g' , expressão (23), com

$$I^{(2)} = 211.756,89, \quad j = 6,$$

e

$$\sum_{j=1}^{\lfloor \frac{N}{2} \rfloor} I_j^{(149)} - I^{(1)} = 2.902.343,46 - 312.911,17 = 2.589.432,29.$$

Assim,

$$g'_{\text{obs}} = \frac{211.756,89}{2.589.432,29} = 0,08178$$

e

$$\alpha^* = P(g' > 0,0818) \approx 73(1 - 0,0818)^{72} = 0,1566,$$

indicando a não existência de uma segunda periodicidade na série.

Siegel (1980) dá uma extensão do teste de Fisher para várias periodicidades.

Veja Morettin et al. (1985) para uma aplicação à série de chuvas em Fortaleza.

Referências

- Morettin, P. A., Mesquita, A. R. and Rocha, J. G. C. (1985). Rainfall at Fortaleza in Brasil Revisited. In: Anderson, O. D. Robinson, E. A. and Ord, K. (eds.), *Time Series Analysis: Theory and Practice 6*, 67–85. Amsterdam: North-Holland.
- Morettin, P. A. (2014). *Ondas e Ondaletas*. Segunda edição. São Paulo: EDUSP.
- Morettin, P. A. and Toloi, C. M. C. (2018). *Análise de Séries Temporais*. Terceira Edição. São Paulo: Blucher.
- Percival, D. and Walden, A. (1993). *Spectral Analysis for Physical Applications*. Cambridge.
- Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*. Englewood Cliffs: Prentice Hall