

Tema 8. Introducción al análisis espectral de series temporales

Tema 8. Introducción al análisis espectral de series temporales

Tema 8. Introducción al análisis espectral de series temporales

8.1 Introducción

8.2 Densidad espectral y sus propiedades

8.3 Estimación del espectro: Periodograma

8.4 Periodograma suavizado

8.1 Densidad espectral y sus propiedades

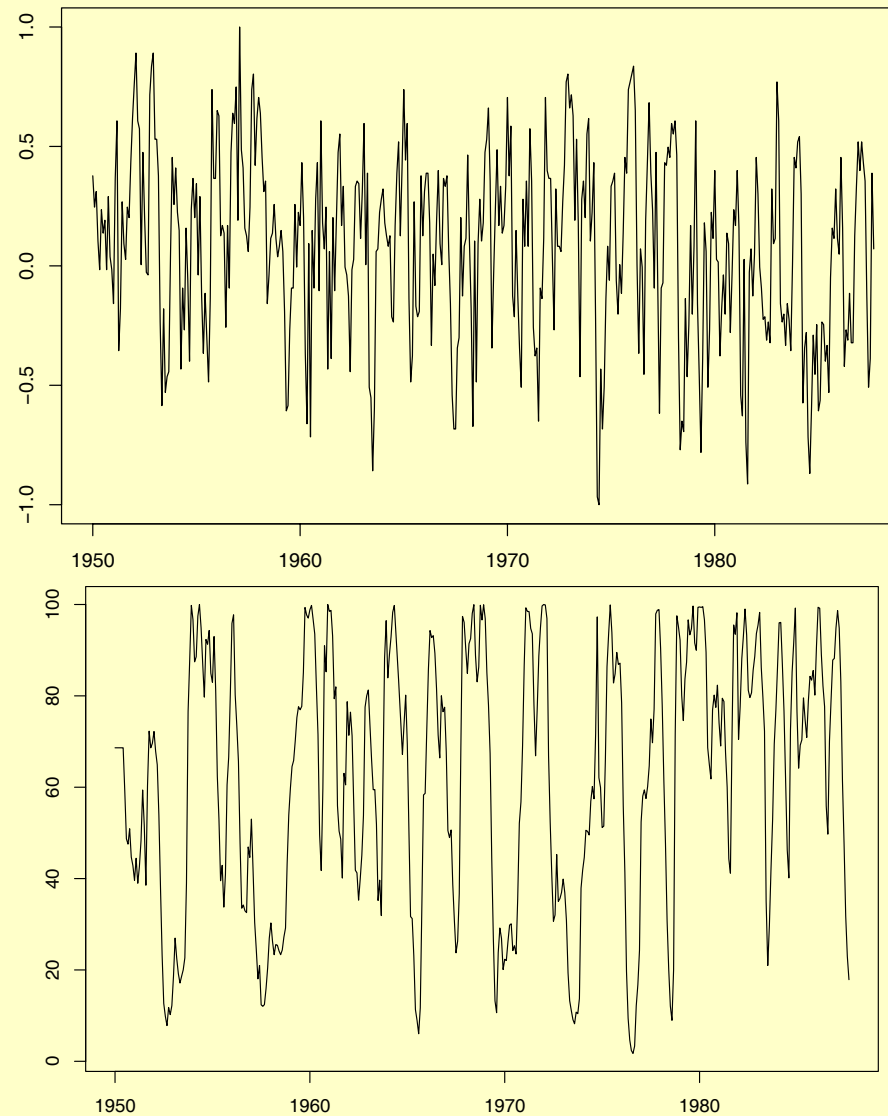


Figure 1: *Arriba.: SOI. Abajo.: Número de nuevos peces*

8.1 Densidad espectral y sus propiedades

Consideremos $\{X_t\}_{t \in T}$ un proceso estacionario con función de autocovarianzas satisfaciendo $\sum_{|h| < \infty} |\gamma_h| < \infty$.

Definición: Se define la densidad espectral de $\{X_t\}_{t \in T}$ o espectro como la función

$$f(\lambda) = \frac{1}{\pi} \sum_{|h| < \infty} \gamma_h e^{-i\lambda h}, \quad -\infty < \lambda < \infty \quad (1)$$

donde $e^{-i\lambda h} = \cos(\lambda h) + i \sin(\lambda h)$ e $i = \sqrt{-1}$.

Debemos darnos cuenta que la condición $\sum_{|h| < \infty} |\gamma_h| < \infty$ implica que la serie en (1) converge absolutamente puesto que $|e^{-i\lambda h}|^2 = \cos^2(\lambda h) + \sin^2(\lambda h) = 1$.

8.1 Densidad espectral y sus propiedades

Propiedades

1) *La densidad espectral es una función real*

8.1 Densidad espectral y sus propiedades

Propiedades

- 1) *La densidad espectral es una función real*
- 2) *La densidad espectral es una función con periodo 2π y par*

8.1 Densidad espectral y sus propiedades

Propiedades

- 1) *La densidad espectral es una función real*
- 2) *La densidad espectral es una función con periodo 2π y par*
- 3) *$f(\lambda) \geq 0$ para todo $\lambda \in [0, \pi]$.*

8.1 Densidad espectral y sus propiedades

Propiedades

- 1) *La densidad espectral es una función real*
- 2) *La densidad espectral es una función con periodo 2π y par*
- 3) *$f(\lambda) \geq 0$ para todo $\lambda \in [0, \pi]$.*
- 4) *La densidad espectral está íntimamente relacionado con la función de autocorrelación.*

8.1 Densidad espectral y sus propiedades

Ejemplos:

1) Consideremos $\{Z_t\}_{t \in T}$ un proceso de ruido blanco de varianza σ^2 , entonces la función de densidad espectral $f(\lambda) = \sigma^2/\pi$, para todo λ , i.e. todas las frecuencias son tenidas en cuenta con la misma importancia en la variabilidad de la serie.

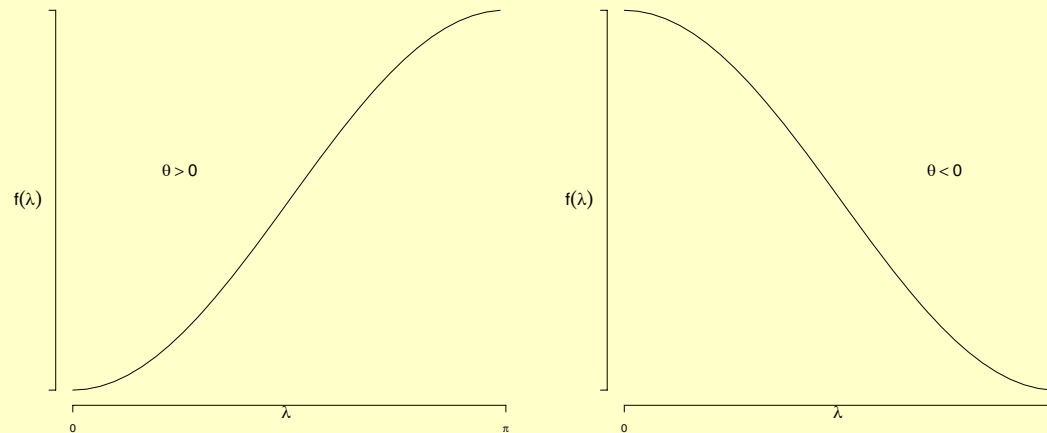


Figure 2: *Funciones de densidad espectral con MA(1) para $\theta > 0$ y $\theta < 0$*

8.1 Densidad espectral y sus propiedades

2) Consideremos $\{X_t\}_{t \in T}$ un modelo $MA(1)$, $X_t = Z_t - \theta Z_{t-1}$, con $\{Z_t\}_{t \in T}$ un proceso de ruido blanco de varianza σ^2 y $|\theta| < 1$. Ya hemos estudiado que $\gamma_0 = (1 + \theta^2)\sigma^2$, $\gamma_1 = -\theta\sigma^2$ y $\gamma_h = 0$ para $h > 1$. Así pues

$$f(\lambda) = \frac{\sigma^2(1 - 2\theta \cos(\lambda) + \theta^2)}{\pi}.$$

La forma del espectro depende del signo de θ . Cuando $\theta < 0$ la densidad espectral está concentrada en frecuencias bajas mientras que si $\theta > 0$ éste se concentra en frecuencias altas (véase figura 2).

8.1 Densidad espectral y sus propiedades

3) Consideremos $\{X_t\}_{t \in T}$ un modelo $AR(1)$, $X_t = \phi X_{t-1} + Z_t$, con $\{Z_t\}_{t \in T}$ un proceso de ruido blanco de varianza σ^2 y $|\phi| < 1$. En este caso, $\gamma_0 = \sigma^2 / (1 - \phi^2)$ y $\gamma_h = \sigma^2 \phi^h / (1 - \phi^2)$, $h \geq 1$, entonces

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{\pi(1 - \phi^2)} \left(1 + \sum_{h=1}^{\infty} \phi^h (e^{-ih\lambda} + e^{ih\lambda}) \right) \\ &= \frac{\sigma^2}{\pi(1 - \phi^2)} \left(1 + \frac{\phi e^{i\lambda}}{1 - \phi e^{i\lambda}} + \frac{\phi e^{-i\lambda}}{1 - \phi e^{-i\lambda}} \right) \\ &= \frac{\sigma^2}{\pi} (1 - 2\phi \cos(\lambda) + \phi^2)^{-1}. \end{aligned}$$

8.1 Densidad espectral y sus propiedades

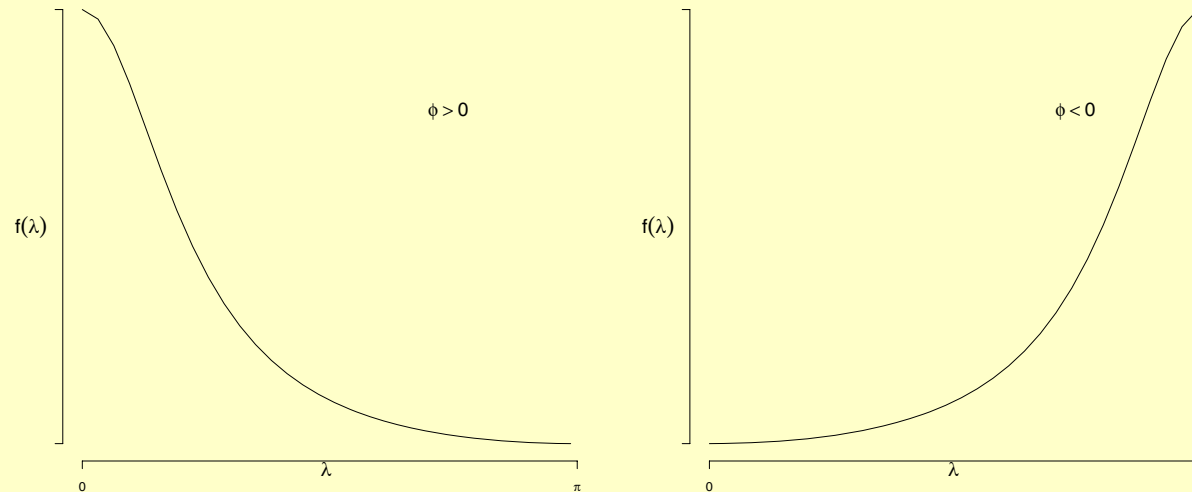


Figure 3: *Funciones de densidad espectral para un AR(1) para $\phi > 0$ y $\phi < 0$*

8.1 Densidad espectral y sus propiedades

Interpretación de la densidad espectral

Para cualquier $\lambda_1 \in (0, \pi]$

$$\int_0^{\lambda_1} f(\lambda) d\lambda \quad (2)$$

podría interpretarse como la parte de varianza de X_t que se asocia a frecuencias λ que son menores o iguales que λ_1 y consecuentemente, $f(\lambda)d\lambda$ la contribución a la varianza de frecuencias en el rango $(\lambda, \lambda + d\lambda)$.

8.1 Densidad espectral y sus propiedades

Consideremos el proceso $\{X_t\}_{t \in T}$ definido como

$$X_t = \sum_{j=1}^M (\alpha_j \cos(\omega_j t) + \delta_j \sin(\omega_j t)) \quad (3)$$

donde M es un número entero positivo, $\{\omega_j\}_{j=1}^M$ son frecuencias distintas en $[0, \pi]$, $\{\alpha_j\}_{j=1}^M$ y $\{\delta_j\}_{j=1}^M$ son variables aleatorias incorreladas de media cero y mutuamente incorreladas tales que

$$E[\alpha_j \alpha_k] = \begin{cases} 0 & \text{si } j \neq k \\ \sigma_j^2 & \text{si } j = k \end{cases}$$

$$E[\delta_j \delta_k] = \begin{cases} 0 & \text{si } j \neq k \\ \sigma_j^2 & \text{si } j = k \end{cases}$$

$$E[\alpha_j \delta_k] = 0 \text{ para todo } j, k.$$

8.1 Densidad espectral y sus propiedades

Teorema de representación espectral establece que cualquier proceso estacionario puede expresarse en términos de una generalización de (3). Concretamente, para cualquier frecuencia fija $\omega \in [0, \pi]$, se definen las variables aleatorias $\alpha(\omega)$ y $\delta(\omega)$ y se escribe, esencialmente, un proceso estacionario $\{X_t\}_{t \in T}$ con función de autocovarianzas absolutamente sumable en la forma

$$X_t = \nu + \int_0^\pi (\alpha(\omega) \cos(\omega t) + \delta(\omega) \sin(\omega t)) d\omega, \quad (4)$$

donde las variables $\alpha(\omega)$ y $\delta(\omega)$, $\omega \in [0, \pi]$ tienen media cero y además las propiedades de que para cualesquiera frecuencias $0 < \omega_1 < \omega_2 < \omega_3 < \omega_4 < \pi$, la variable $\int_{\omega_1}^{\omega_2} \alpha(\omega) d\omega$ es incorrelada con $\int_{\omega_3}^{\omega_4} \alpha(\omega) d\omega$ y la variable $\int_{\omega_1}^{\omega_2} \delta(\omega) d\omega$ es incorrelada con $\int_{\omega_3}^{\omega_4} \delta(\omega) d\omega$, y para cualesquiera $0 < \omega_1 < \omega_2 < \pi$ y $0 < \omega_3 < \omega_4 < \pi$, la variable $\int_{\omega_1}^{\omega_2} \alpha(\omega) d\omega$ es incorrelada con $\int_{\omega_3}^{\omega_4} \delta(\omega) d\omega$.

8.2 Estimación del espectro: Periodograma

Definición: Se define el periodograma basado en la muestra X_1, \dots, X_n como la función

$$I(\lambda) = \frac{1}{\pi} \sum_{|h| < n} \hat{\gamma}_h e^{-i\lambda h}, \lambda \in [0, \pi] \quad (5)$$

8.2 Estimación del espectro: Periodograma

Definición: Se define el periodograma basado en la muestra X_1, \dots, X_n como la función

$$I(\lambda) = \frac{1}{\pi} \sum_{|h| < n} \hat{\gamma}_h e^{-i\lambda h}, \lambda \in [0, \pi] \quad (5)$$

El periodograma se puede expresar como

$$I(\lambda) = \frac{1}{\pi} \left(\hat{\gamma}_0 + 2 \sum_{h=1}^{n-1} \hat{\gamma}_h \cos(\lambda h) \right), \lambda \in [0, \pi]. \quad (6)$$

8.2 Estimación del espectro: Periodograma

Definición: Se define el periodograma basado en la muestra X_1, \dots, X_n como la función

$$I(\lambda) = \frac{1}{\pi} \sum_{|h| < n} \hat{\gamma}_h e^{-i\lambda h}, \lambda \in [0, \pi] \quad (5)$$

El periodograma se puede expresar como

$$I(\lambda) = \frac{1}{\pi} \left(\hat{\gamma}_0 + 2 \sum_{h=1}^{n-1} \hat{\gamma}_h \cos(\lambda h) \right), \lambda \in [0, \pi]. \quad (6)$$

$$\int_0^{\pi} I(\lambda) d\lambda = \hat{\gamma}_0.$$

8.2 Estimación del espectro: Periodograma

Probemos que dado $\{X_1, \dots, X_n\}$, existen frecuencias $\omega_1, \dots, \omega_m$ y coeficientes $\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_m, \hat{\beta}_1, \dots, \hat{\beta}_m, m \in \mathbb{N}$ tales que

$$X_t = \hat{\alpha}_0 + \sum_{j=1}^m \hat{\alpha}_j \cos(\omega_j t) + \sum_{j=1}^m \hat{\beta}_j \sin(\omega_j t). \quad (7)$$

donde la variable $\hat{\alpha}_j \cos(\omega_j t)$ es ortogonal¹ sobre los períodos de la muestra (i.e. $t = 1, \dots, n$) a la variable $\hat{\alpha}_k \cos(\omega_k t)$, para $j \neq k$, la variable $\hat{\beta}_j \sin(\omega_j t)$ es ortogonal a la variable $\hat{\beta}_k \sin(\omega_k t)$, para $j \neq k$, y la variable $\hat{\alpha}_j \cos(\omega_j t)$ es ortogonal a $\hat{\beta}_k \sin(\omega_k t)$ para todo j y k .

¹Sean $\phi_k(t)$ y $\phi_j(t)$ funciones reales definidas sobre un dominio D , el cual es un subconjunto discreto de la recta real. Las funciones se dicen que son ortogonales sobre D si $\sum_{t \in D} \phi_k(t) \phi_j(t) = 0$ si $k \neq j$ y $\neq 0$ si $k = j$.

8.2 Estimación del espectro: Periodograma

Por simplicidad en el desarrollo matemático vamos a suponer en esta sección que n es impar. Supongamos que tenemos $n = 2m + 1$ observaciones de una serie temporal $\{X_1, \dots, X_n\}$. Definimos las frecuencias $\omega_j = 2\pi j/n$, $j = 1, \dots, m$, conocidas como frecuencias de Fourier, y consideremos el modelo de regresión

$$X_t = \alpha_0 + \sum_{j=1}^m \alpha_j \cos(\omega_j t) + \sum_{j=1}^m \beta_j \sin(\omega_j t) + \mathcal{E}_t,$$

donde $\{\alpha_j\}_{j=0}^m$ y $\{\beta_j\}_{j=0}^m$ son parámetros desconocidos. Este modelo puede escribirse como un modelo lineal general $X = Y\theta + \mathcal{E}$, donde

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad Y = \begin{pmatrix} 1 & c_{11} & s_{11} & \dots & c_{m1} & s_{m1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & c_{1n} & s_{1n} & \dots & c_{mn} & s_{mn} \end{pmatrix}, \quad \theta = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_m \\ \beta_m \end{pmatrix}$$

y $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_n)'$, con $c_{jt} = \cos(\omega_j t)$ y $s_{jt} = \sin(\omega_j t)$.

8.2 Estimación del espectro: Periodograma

$$\hat{\theta} = (Y'Y)^{-1}Y'X.$$

Para cada $j = 1, \dots, m$

$$\sum_{t=1}^n e^{i\omega_j t} = \frac{e^{i\omega_j}(1 - e^{i\omega_j n})}{(1 - e^{i\omega_j})} = 0,$$

de donde se sigue $\sum_{t=1}^n c_{jt} + i \sum_{t=1}^n s_{jt} = 0$ y por lo tanto $\sum_{t=1}^n c_{jt} = 0$ y $\sum_{t=1}^n s_{jt} = 0$. Por otro lado, usando las relaciones trigonométricas

$$\cos(\omega) \sin(\lambda) = (\cos(\omega + \lambda) + \cos(\omega - \lambda))/2,$$

$$\sin(\omega) \sin(\lambda) = (\cos(\omega - \lambda) - \cos(\omega + \lambda))/2,$$

$$\sin(\omega) \cos(\lambda) = (\sin(\omega + \lambda) + \sin(\omega - \lambda))/2,$$

8.2 Estimación del espectro: Periodograma

se sigue que, para cada $j = 1, \dots, m$

$$\sum_{t=1}^n s_{jt}c_{jt} = 2^{-1} \sum_{t=1}^n \sin(2\omega_j t) = 0$$

$$\sum_{t=1}^n c_{jt}^2 = 2^{-1} \sum_{t=1}^n (1 + \cos(2\omega_j t)) = n/2$$

$$\sum_{t=1}^n s_{jt}^2 = 2^{-1} \sum_{t=1}^n (1 - \cos(2\omega_j t)) = n/2$$

$$\sum_{t=1}^n s_{jt}c_{kt} = \sum_{t=1}^n s_{jt}s_{kt} = \sum_{t=1}^n c_{jt}c_{kt} = 0 \text{ para } j \neq k.$$

8.2 Estimación del espectro: Periodograma

Así pues,

$$\hat{\theta} = \begin{pmatrix} \bar{X} \\ (2/n) \sum_{t=1}^n c_{1t} X_t \\ \vdots \\ (2/n) \sum_{t=1}^n s_{mt} X_t \end{pmatrix}.$$

Puesto que estamos ajustando n parámetros desconocidos a n datos, el modelo se ajusta con $\mathcal{E} = 0$, de donde $\hat{X} = X$, y por lo tanto

$$\sum_{t=1}^n (X_t - \bar{X})^2 = \sum_{j=1}^m \frac{n}{2} (\hat{\alpha}_j^2 + \hat{\beta}_j^2) = \sum_{j=1}^m \frac{2}{n} \left(\left(\sum_{t=1}^n c_{jt} X_t \right)^2 + \left(\sum_{t=1}^n s_{jt} X_t \right)^2 \right)$$

8.2 Estimación del espectro: Periodograma

8.2 Estimación del espectro: Periodograma

$$\begin{aligned} & \left(\sum_{t=1}^n X_t \cos(\omega_j t) \right)^2 + \left(\sum_{t=1}^n X_t \sin(\omega_j t) \right)^2 = \\ & = \left(\sum_{t=1}^n (X_t - \bar{X}) \cos(\omega_j t) \right)^2 + \left(\sum_{t=1}^n (X_t - \bar{X}) \sin(\omega_j t) \right)^2 \\ & = \left| \sum_{t=1}^n (X_t - \bar{X}) e^{i\omega_j t} \right|^2 = \left(\sum_{t=1}^n (X_t - \bar{X}) e^{i\omega_j t} \right) \left(\sum_{s=1}^n (X_s - \bar{X}) e^{-i\omega_j s} \right) \\ & = \sum_{n=1}^n (X_t - \bar{X})^2 + 2 \sum_{k=1}^{n-1} \sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X}) \cos(\omega_j k) \\ & = \frac{1}{\pi} \left(\hat{\gamma}_0 + 2 \sum_{h=1}^{n-1} \hat{\gamma}_h \cos(\omega_j h) \right) \\ & = \pi n I(\omega_j) \end{aligned}$$

8.2 Estimación del espectro: Periodograma

Por tanto la varianza muestral $\hat{\gamma}_0 = (1/n) \sum_{t=1}^n (X_t - \bar{X})^2$ es igual a la suma de las áreas de m rectángulos, cuyas alturas son $I(\omega_1), I(\omega_2), \dots, I(\omega_m)$, cuyas anchuras son $2\pi/n$ y cuyas bases están centradas en $\omega_1, \dots, \omega_m$, i.e. $\hat{\gamma}_0 = (2\pi/n) \sum_{j=1}^m I(\omega_j)$. En conclusión, los coeficientes de la regresión tienen la propiedad de que $2^{-1}(\hat{\alpha}_j^2 + \hat{\beta}_j^2)$ representa la porción de varianza muestral que puede ser atribuida a ciclos con frecuencias ω_j . Además la cantidad $2^{-1}(\hat{\alpha}_j^2 + \hat{\beta}_j^2)$ resulta ser proporcional al periodograma evaluado en ω_j . Luego, toda serie observada $\{X_1, \dots, X_n\}$ puede ser expresada en términos de funciones periódicas tales como (7) y la porción de varianza muestral que se debe a ciclos con frecuencias ω_j puede ser encontrada a partir del periodograma.

8.2 Estimación del espectro: Periodograma

Teorema 1. Sea $\{X_t\}_{t \in T}$ un proceso lineal definido en la forma $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$ siendo $\{Z_t\}_{t \in T}$ una sucesión de variables independientes e idénticamente distribuidas con media 0 y varianza σ^2 y $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Denotemos $I_n(\omega_j)$ el periodograma de (X_1, \dots, X_n) en la frecuencia $\omega_j = 2\pi j/n$, $j = 1, 2, \dots, [n/2]$ y $f(\lambda)$ la densidad espectral de $\{X_t\}_{t \in T}$.

i) Si $f(\lambda) > 0$ para todo $\lambda \in [0, \pi]$ y si $0 < \omega_{j_1} < \omega_{j_2} < \dots < \omega_{j_m} < \pi$ ($j_i \in \{1, 2, \dots, [n/2]\}$ para $i = 1, \dots, m$) entonces el vector aleatorio $(I_n(\omega_{j_1}), \dots, I_n(\omega_{j_m}))'$ converge en distribución, cuando $n \rightarrow \infty$, a un vector de variables aleatorias independientes y exponencialmente distribuidas, la i -ésima componente del mismo con media $f(\omega_{j_i})$, $i = 1, \dots, m$.

ii) Si $\sum_{j=-\infty}^{\infty} |\psi_j| |j|^{1/2} < \infty$ y $E[Z_1^4] < \infty$, entonces

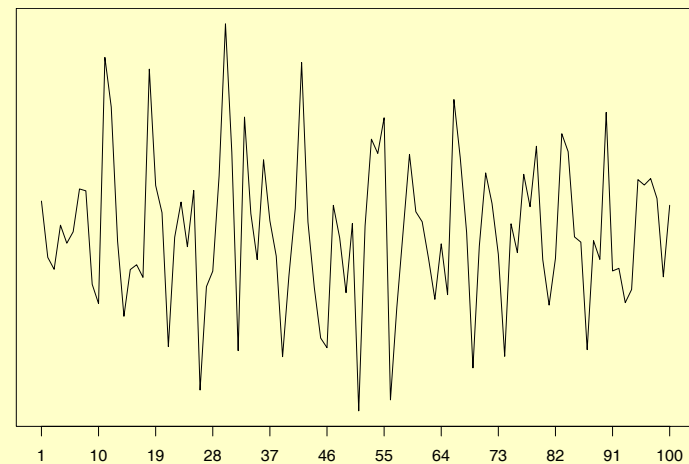
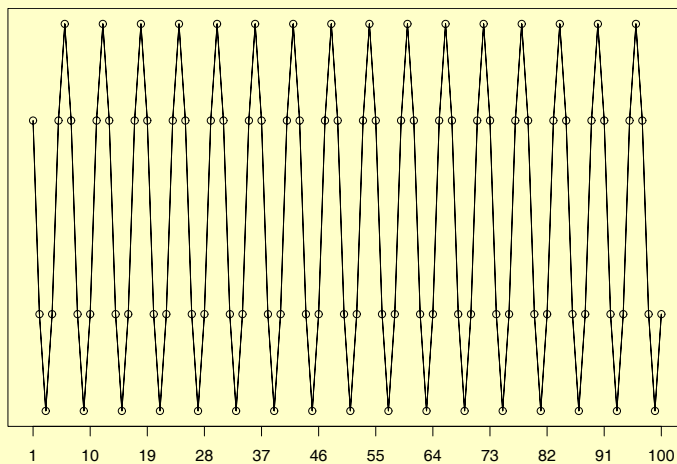
$$\text{Cov}(I_n(\omega_j), I_n(\omega_k)) = \begin{cases} 2f^2(\omega_j) + O(n^{-1/2}) & \text{si } \omega_j = \omega_k = 0 \text{ ó } \pi \\ f^2(\omega_j) + O(n^{-1/2}) & \text{si } 0 < \omega_j = \omega_k < \pi \\ O(n^{-1}) & \text{si } \omega_j \neq \omega_k \end{cases}$$

donde los términos $O(n^{-1/2})$ y $O(n^{-1})$ pueden ser acotados uniformemente en j y k por $c_1 n^{-1/2}$ y $c_2 n^{-1}$, respectivamente para algunas constantes c_1 y c_2 .

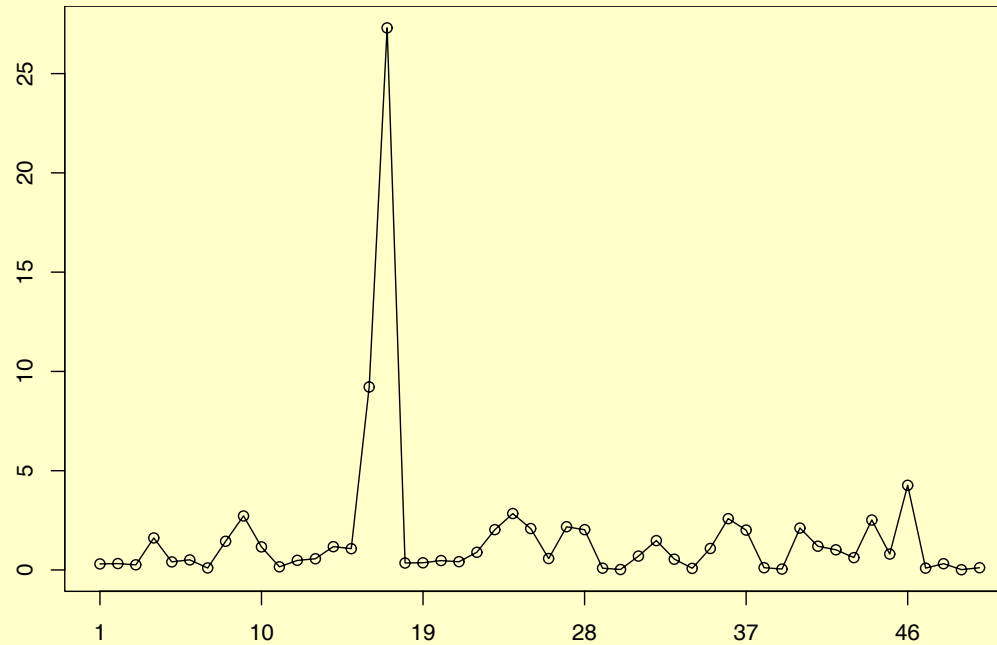
8.2 Estimación del espectro: Periodograma

Ejemplo: Periodograma de series aleatorias periódicas

Consideremos una serie temporal periódica aleatoria obtenida añadiendo una onda coseno de periodo 6 puntos a un ruido blanco i.e. $X_t = \cos(\pi t/3) + Z_t$ con $\{Z_t\}_{t \in T}$ un ruido blanco gaussiano de varianza 1.

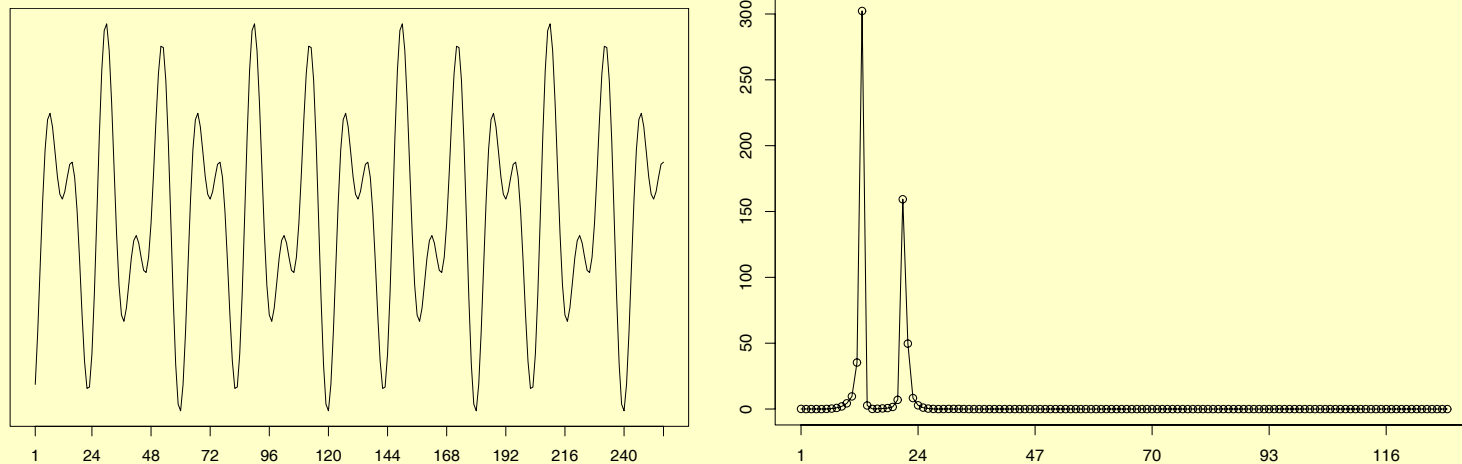


8.2 Estimación del espectro: Periodograma



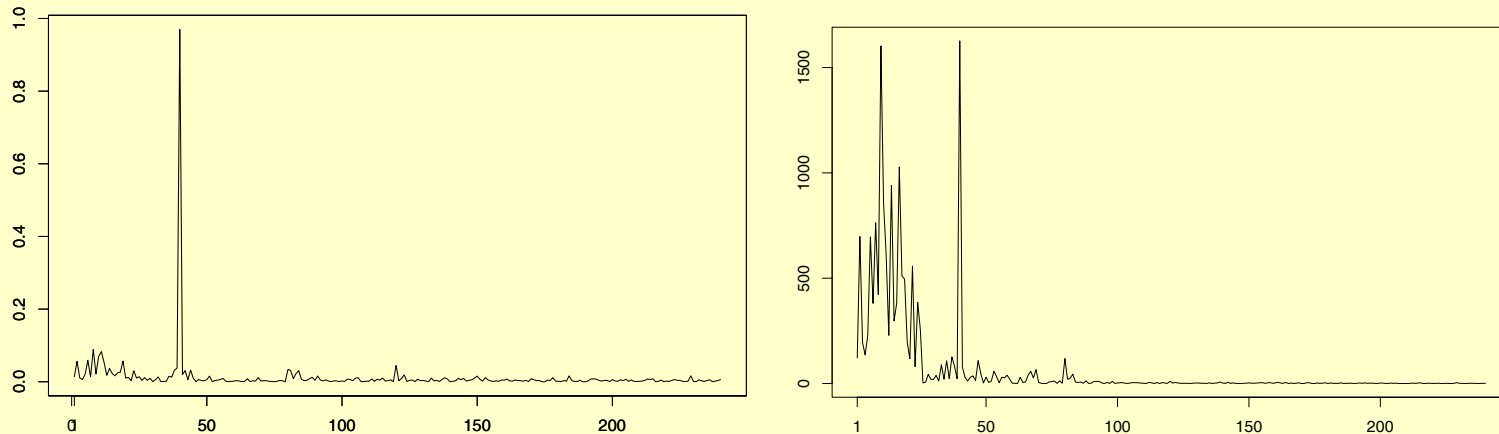
8.2 Estimación del espectro: Periodograma

Otro ejemplo de serie periódica aleatoria se puede obtener considerando el modelo (3), $X_t = \sum_{j=1}^M (\alpha_j \cos(\omega_j t) + \delta_j \sin(\omega_j t))$, con $M = 2$ y $\omega_1 = 2\pi/20$ y $\omega_2 = 2\pi/12$. La figura ?? (izquierda) muestra un trayectoria de longitud 256 de tal modelo, con valores $\alpha_1 = -2.31$, $\alpha_2 = -1.74$, $\beta_1 = -0.43$, $\beta_2 = 0.68$, obtenidos de distribuciones normales de media cero y varianzas 1 y 0.25, respectivamente.



8.2 Estimación del espectro: Periodograma

Periodogramas de las series del SOI y número de Peces



Ambas series muestran un pico común en la frecuencia $f = 0.084$ que corresponde a un periodo de 11.91 meses, obviamente la frecuencia estacional anual. Los picos pequeños en el periodograma de la serie SOI se encuentran en los periodos de 73.14 ($f = 0.014$), 51.20 ($f = 0.02$) y 42.67 ($f = 0.023$) meses, esto es entre 4 y 6 años, picos que probablemente reflejan las frecuencias del fenómeno de El Niño.

8.2 Estimación del espectro: Periodograma

Test para la presencia de periodicidades ocultas: Periodograma acumulado

Proposición 1. Si $\{X_t\}_{t \in T}$ es ruido blanco gaussiano:

i) Las variables aleatorias $(Y_1, \dots, Y_q)'$ con

$$Y_i = \frac{\sum_{k=1}^i I(\omega_k)}{\sum_{k=1}^q I(\omega_k)}, \quad i = 1, 2, \dots, q - 1$$

se distribuyen como los estadísticos de orden de una muestra de $(q - 1)$ variables aleatorias independientes e idénticamente distribuidas de acuerdo a una distribución uniforme en el intervalo $[0, 1]$.

ii) La función de distribución acumulada con saltos de tamaño $(q - 1)^{-1}$ en Y_i , $i = 1, \dots, q - 1$, es la función de distribución empírica de una muestra de tamaño $q - 1$ de una distribución uniforme en el intervalo $[0, 1]$.

8.2 Estimación del espectro: Periodograma

iii) Si definimos

$$M_q = \max_{1 \leq i \leq q} \frac{I(\omega_1)}{\sum_{i=1}^q I(\omega_i)}$$

entonces

$$P(M_q \leq a) = \sum_{j=0}^q (-1)^j \binom{q}{j} (1 - ja)_+^{q-1}, \quad (8)$$

donde $x_+ = \max\{0, x\}$.

8.2 Estimación del espectro: Periodograma

Test de Fisher para periodicidades ocultas

El apartado iii) de la proposición anterior fue usado por Fisher para contrastar la hipótesis nula de que los datos $(X_1, \dots, X_n)'$ han sido generados por un ruido blanco gaussiano contra la hipótesis alternativa de que los datos han sido generados por una sucesión de ruido blanco gaussiano contaminada por una componente periódica con frecuencia desconocida. La idea es rechazar la hipótesis nula si el periodograma contiene un valor sustancialmente más grande que el valor medio, i.e. si

$$\eta_q = \frac{\max_{1 \leq i \leq q} I(\omega_i)}{q^{-1} \sum_{i=1}^q I(\omega_i)} = qM_q$$

es suficientemente grande, siendo $q = [(n - 1)/2]$. Para aplicar el test se calcula el valor x de η_q sobre la muestra y se usa (8) para calcular

$$P(\eta_q \geq x) = 1 - \sum_{j=0}^q (-1)^j \binom{q}{j} (1 - jx/q)_+^{q-1}. \quad (9)$$

Si esta probabilidad es menor que α rechazamos la hipótesis a nivel α .

8.2 Estimación del espectro: Periodograma

Test de Kolmogorov–Smirnov

El apartado ii) de la Proposición 1 sugiere otro test para contrastar la hipótesis nula que $\{X_t\}_{t \in T}$ sea un ruido blanco gaussiano. Éste, simplemente propone, dibujar la función de distribución empírica definida en ii) y comprobar su compatibilidad con la función de distribución de la distribución uniforme en $[0, 1]$, i.e. $F(x) = x$, $0 \leq x \leq 1$, usando el test de Kolmogorov-Smirnov. Para $q > 30$ i.e. tamaños muestrales mayores que 62, una buena aproximación del test de Kolmogorov-Smirnov al nivel α es rechazar la hipótesis nula si la función de distribución empírica se sale de las cotas

$$y = x \pm k_\alpha(q - 1)^{-1/2}, \quad 0 < x < 1,$$

donde por ejemplo $k_{0.05} = 1.36$ y $k_{0.01} = 1.63$. Este procedimiento es equivalente a dibujar el periodograma acumulado estandarizado,

$$F^*(x) = \begin{cases} 0 & \text{si } x < 1 \\ Y_i & \text{si } i \leq x < i + 1, \quad i = 1, \dots, q - 1 \\ 1 & \text{si } x \geq q \end{cases}$$

y rechazar la hipótesis nula al nivel α si para cualquier $x \in [1, q]$, la función F^* se

8.2 Estimación del espectro: Periodograma

sale de las cotas

$$y = \frac{x - 1}{q - 1} \pm k_{\alpha}(q - 1)^{-1/2}.$$

8.2 Estimación del espectro: Periodograma

Retomamos el ejemplo de las series periódicas aleatorias. En este caso $q = [99/2] = 49$ y el valor del estadístico η_{49} es $x = 27.3064/1.7296 = 15.7874$. A partir de (9),

$$P(\eta_{49} \geq 15.7874) = 0.00000038$$

y en consecuencia el test propone rechazar la hipótesis nula (como tenía que ocurrir). A la luz de la figura 4, el test de Kolmogorov-Smirnov también encuentra evidencias para rechazar la hipótesis nula.

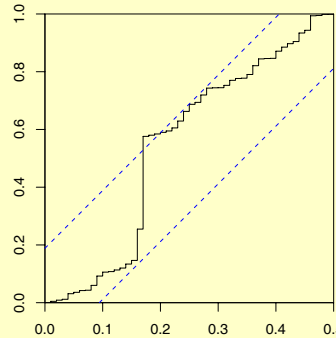


Figure 4: Representación del periodograma acumulado de la serie de la Figura ??

8.2 Estimación del espectro: Periodograma

Total de acciones del mercado bursátil español

Vamos a ilustrar algunos de los usos del análisis espectral para los datos vistos en el ejemplo que corresponden a la serie de total de acciones contratadas en el mercado bursátil español (en miles de euros), desde enero de 1991 hasta mayo de 2006 (véase figura 5 (izquierda)).

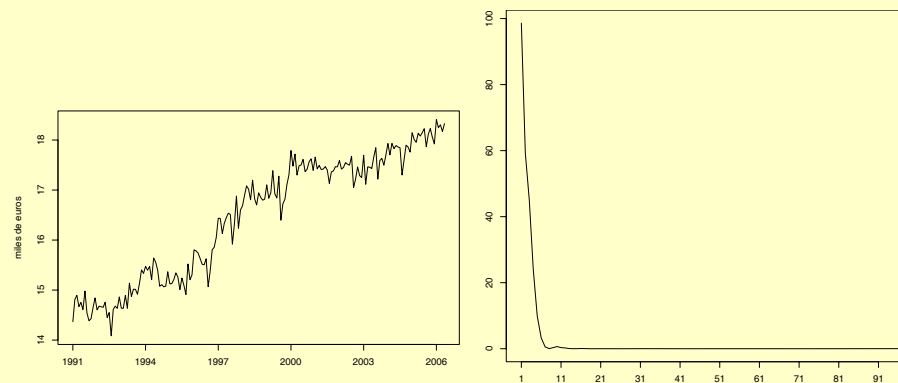
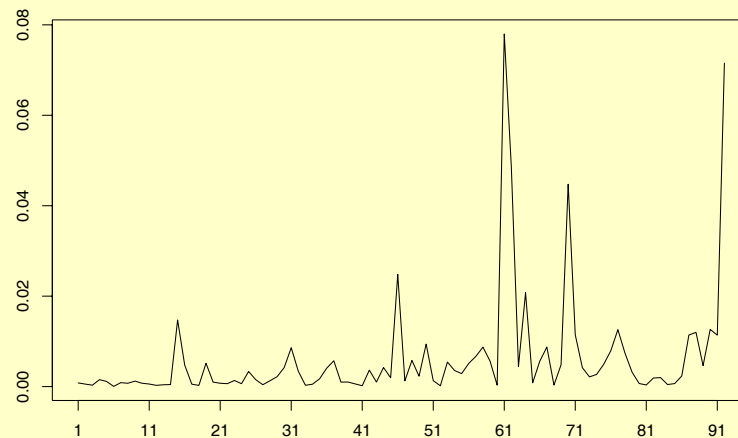


Figure 5: *Izda.: Representación de la serie del logaritmo del total de acciones en el mercado bursátil español, $\log(X_t)$. Drcha.: Periodograma de la serie de la izquierda $I(2\pi j/184)$.*

8.2 Estimación del espectro: Periodograma

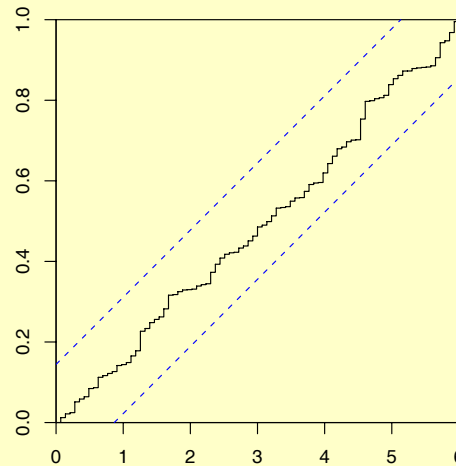
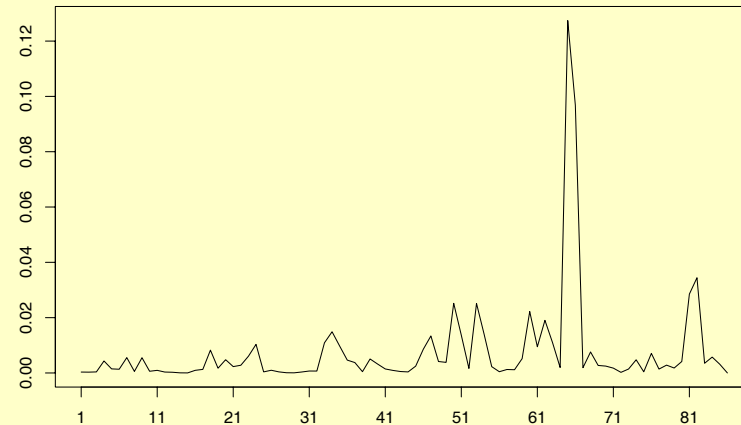
En la figura 5 (derecha) hemos representado el periodograma de los datos. En este llama la atención la magnitud del mismo en frecuencias bajas. Este hecho es típico de series que muestran un comportamiento evidente de tendencia, lo cual puede ser debido a una tendencia real o a un ciclo con un periodo mucho más grande que la longitud de la serie.



En la definición de la densidad espectral se asume que el proceso es estacionario, lo cual, obviamente, no verifican los datos que estamos considerando. Vamos, por lo tanto a analizar la serie $Y_t = \log(X_t) - \log(X_{t-1})$. El tamaño muestral de la serie es ahora 184 datos, y los primeros picos más marcados en la figura ?? corresponden aproximadamente a $j = 13, 31, 46, 61$, y por tanto a ciclos con periodos 12, 6, 4, 3, periodos correspondientes a las variaciones estacionales, recordemos que los datos

8.2 Estimación del espectro: Periodograma

son recogidos mensualmente. A continuación vamos a eliminar el efecto estacional considerando la serie transformada $W_t = Y_t - Y_{t-12}$.



En este caso, véase figura ??, el pico más señalado corresponde a $j = 65$ que

8.2 Estimación del espectro: Periodograma

corresponde a un periodo de 2.64 de nuevo justificado por el efecto estacional de la serie.

Ajustamos a la serie del logaritmo del total de acciones en el mercado bursátil español el modelo $ARIMA(2, 1, 0) \times ARIMA(0, 1, 1)_{12}$. La figura ?? representa el periodograma acumulado para la serie de residuos del modelo $ARIMA(2, 1, 0) \times ARIMA(0, 1, 1)_{12}$ mostrando las bandas de Kolmogorov-Smirnov para $\alpha = 0.05$. Éste no muestra evidencias para rechazar la hipótesis de que los residuos procedentes de este modelo no sean un ruido blanco gaussiano.

8.4 Periodograma suavizado

Estimadores en la forma

$$\hat{f}_n(\lambda) = \sum_{|j| \leq m_n} \mathcal{W}_n(j) I(g(n, \lambda) + 2\pi j/n). \quad (10)$$

donde n es el tamaño de la muestra, $g(n, \lambda)$ es el múltiplo de $2\pi/n$ más cercano a λ . La función de pesos $\mathcal{W}_n(j)$ se denomina *función de ponderación espectral* y las cantidades m_n se denominan *anchos de banda*. Supongamos las condiciones

$$m_n \rightarrow \infty \text{ y } m_n/n \rightarrow 0, \text{ cuando } n \rightarrow \infty, \quad (11)$$

$$\mathcal{W}_n(j) = \mathcal{W}_n(-j), \quad \mathcal{W}_n(j) \geq 0, \quad \text{para todo } j, \quad (12)$$

$$\sum_{|j| \leq m_n} \mathcal{W}_n(j) = 1 \quad (13)$$

y

$$\sum_{|j| \leq m_n} \mathcal{W}_n^2(j) \rightarrow 0, \text{ cuando } n \rightarrow \infty. \quad (14)$$

8.4 Periodograma suavizado

Se verifica (véase Brockwell, P. J. and Davis, R. A. (1991) p. 351) que bajo las condiciones del Teorema 1(ii),

$$\lim_{n \rightarrow \infty} E[\hat{f}_n(\lambda)] = f(\lambda)$$

y

$$\lim_{n \rightarrow \infty} \left(\sum_{|j| \leq m_n} \mathcal{W}_n^2(j) \right)^{-1} \text{Cov}(\hat{f}_n(\lambda), \hat{f}_n(\nu)) = \begin{cases} 2f^2(\lambda) & \text{si } \lambda = \nu = 0 \text{ ó } \pi, \\ f^2(\lambda) & \text{si } 0 < \lambda = \nu < \pi, \\ 0 & \text{si } \lambda \neq \nu. \end{cases}$$

8.4 Periodograma suavizado

Por otro lado a partir de (5), una forma alternativa de suavizar el periodograma es aplicar una función peso a las covarianzas muestrales, i.e.

$$\hat{f}_n(\lambda) = \frac{1}{\pi} \sum_{|h| < n} W_n(k) \hat{\gamma}_k e^{-i\lambda k}. \quad (15)$$

Dado que la función de autocovarianzas muestral es simétrica y pierde fiabilidad para k grandes, la función de pesos W_n debe ser elegida simétrica con pesos inversamente proporcionales a la magnitud de k . Así, se considera

$$\hat{f}_n^M(\lambda) = \frac{1}{\pi} \sum_{|h| \leq M} W_n(k) \hat{\gamma}_k e^{-i\lambda k}, \quad (16)$$

donde la función de pesos $W_n(k)$ se elige de manera que sea una sucesión absolutamente sumable en la forma $W_n(k) = W(k/M)$, siendo W una función acotada, par y continua satisfaciendo $|W(x)| \leq 1$, $W(0) = 1$, $W(x) = 0$ para $|x| > 1$.

8.4 Periodograma suavizado

El valor de $M \equiv M_n$ se conoce como *el punto de truncamiento* que depende del tamaño muestral n . La función peso $W_n(k)$ para las autocovarianzas es conocida como *ventana de retardo* (lag window). Al igual que la función de densidad espectral y la función de autocovarianzas están relacionadas por (??), se puede probar que si definimos

$$w(\lambda) = \frac{1}{\pi} \sum_{|h| \leq M} W_n(k) e^{-ih\lambda},$$

conocida como *ventana espectral* (spectral window), entonces se puede aproximar el estimador \hat{f}_n^M por \hat{f}_n considerando la función de ponderación espectral

$$W_n(j) = \pi w(\omega_j)/n, \quad |j| < [n/2] \quad (\omega_j = 2\pi j/n).$$

8.4 Periodograma suavizado

Se verifica que cuando $M_n/n \rightarrow 0$, cuando $n \rightarrow \infty$, bajo las condiciones del Teorema 1, el estimador $\hat{f}_n^M(\omega)$ es un estimador consistente en media cuadrática para $f(\omega)$.