

Proceso Autorregresivo con Coeficientes Variando en el Tiempo $tvAR(p)$

Daniel Alejandro Ramírez Gómez

Trabajo de grado presentado como requisito parcial para optar al título de:
Matemático

Asesor:
PhD. Duván Humberto Cataño Salazar
Profesor del Instituto de Matemáticas

Línea de Investigación:
Estadística

Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales, Instituto de Matemáticas
El Carmen de Viboral, Colombia
2020

*Dedicado
a todas las personas que de una u otra forma me acompañaron en esta travesía,
especialmente a mis padres, hermanos y hermanas*

Agradecimientos

Primero que todo le agradezco a la Universidad de Antioquia Seccional Oriente por ser mi segundo hogar durante el transcurso de la carrera, así como también a todos los administrativos que con el tiempo se convirtieron en mi otra familia, le agradezco a Vilma Meneses por su compañía constante en la biblioteca y a Olga Lucía Merino por su enorme labor social en la seccional oriente.

Esta tesis no fue un trabajo individual y agradezco infinitamente a mi asesor Duván Cataño que a pesar de las dificultades acontecidas en el presente año tuvo la paciencia y dedicación de orientarme en esta bella área de las matemáticas. También agradezco a todos los profesores que hicieron parte de mi formación, sobre todo a los profesores Roberto Cruz, Daniel Muñoz, Natalia Cardona, Eloy Castellanos, Frank Duque, William Alvarez y Alejandro Roldan, también le agradezco a mis compañeros de clase que estuvieron a mi lado durante toda la carrera y me ayudaron a crecer no solo en el ámbito académico sino también como persona.

A todos ustedes muchas gracias.

Resumen

El estudio de series de tiempo que presentan un comportamiento estacionario o periódico han sido importantes en el análisis y extracción de información. Los procesos autorregresivos son una buena herramienta para modelar un comportamiento de este tipo; existen métodos tradicionales para la estimación de los parámetros, el problema recae cuando la serie no tiene un comportamiento estacionario, en este caso, la estimación tradicional queda corta ya que no logra capturar los cambios estructurales. En el presente trabajo se estudian las propiedades y características principales de las series estacionarias empleando modelos ARMA. Al igual que el análisis de la serie en el dominio de las frecuencias empleando la densidad espectral, con el fin de estudiar el aporte que hace el período de un fenómeno a la varianza del proceso. Por último se estudiará series no estacionaria, específicamente procesos localmente estacionarias, la cual se modelará por medio de un proceso autorregresivo con coeficientes variando en el tiempo, estimando los parámetros empleando una expansión en series de wavelets de tal manera que minimice el error cuadrático medio.

Palabras Claves: Serie de Tiempo, Proceso Autorregresivo, Localmente Estacionario, Coeficientes Variando en el Tiempo, Wavelets.

Índice general

| | |
|---|-----------|
| Agradecimientos | II |
| Resumen | III |
| 1. Introducción | 1 |
| 1.1. Preliminares | 3 |
| 1.1.1. Series estacionarias | 5 |
| 2. Proceso ARMA(p, q) | 8 |
| 2.1. Definición y propiedades. | 8 |
| 2.1.1. Función de Autocorrelación | 11 |
| 2.1.2. Función de Autocorrelación Parcial | 12 |
| 2.1.3. Estimación | 13 |
| 2.1.4. Pronóstico | 15 |
| 3. Análisis Espectral | 19 |
| 3.1. Densidad Espectral | 19 |
| 3.2. Aplicación | 28 |
| 4. Wavelets | 33 |
| 4.1. Wavelet de Haar | 34 |
| 4.2. Wavelet Daubechies | 40 |
| 5. Proceso Autorregresivo con Coeficientes Variando en el Tiempo | 43 |
| 5.1. Proceso de Estimación | 45 |
| 5.2. Estudio de Simulación | 46 |
| A. Códigos | 52 |
| Bibliografía | 56 |

Capítulo 1

Introducción

Una serie de tiempo es el resultado de observar los valores de una variable aleatoria a lo largo del tiempo en intervalos regulares y su objetivo es pronosticar información dados los datos históricos. Entre las ramas de aplicación más frecuentes se encuentra las ciencias económicas y financieras, marketing e incluso en estudios sociales considerando los valores obtenidos en el tiempo de una variable de interés para extraer información relevante.

Con el fin de estudiar el comportamiento de las series de tiempo estacionarias se investigará los modelos ARIMA estacionales los cuales se basan en escribir las observaciones del presente como combinación lineal de valores pasadas y ruidos blancos, además, se introducirá una componente estacional para el caso de una serie con comportamiento periódico [1]. No obstante, cuando la serie en consideración no es estacionaria en covarianza se estudiará una extensión de los procesos autorregresivos introduciendo una dependencia temporal en sus coeficientes, denominado proceso autorregresivo con coeficientes variando en el tiempo $ARtv(p)$ [2]. Si se asume que la serie es conducida por un $ARtv(p)$, el problema se centra en estimar los coeficientes funcionales $a_i(t) \in L_2$. Tradicionalmente se supone un comportamiento estacionario en una vecindad pequeña de un punto fijo t_0 y se estiman por el método típico de Yule-Walker. La desventaja de esta técnica es que conduce a estimaciones suaves y saltos abruptos comunes en señales son imposibles de ser detectados. En este trabajo se utilizará una técnica alternativa basada en una expansión de $a_i(t)$ en funciones ortogonales de energía finita, las cuales son generadas por la dilatación y traslación de una función de escala ϕ y una función wavelet ψ , que tienen ubicación tiempo y frecuencia [3].

Con el fin de desarrollar los objetivos planteados, el trabajo está dividido en 5 capítulos. En el primer y segundo capítulo se introducirán las nociones básicas de un espacio de probabilidad (Ω, X, \mathbb{P}) , además del modelo ARIMA con su identificación, estimación y pronóstico [1], [4]. En el capítulo 3 se analizará las propiedades de la serie en el dominio de las frecuencias y se demostrará la convergencia de la densidad espectral por medio del periodograma, que servirá para determinar las frecuencias que más aportan a la varianza del proceso y mostrando una aplicación a datos reales empleando el modelo ARIMA estacional [1],[5]. En el capítulo 4 se mostrará que a partir de la dilatación y

traslación de una función escala ϕ es posible construir una secuencia de espacios vectoriales V_j encajados cuya unión es densa en L_2 y así aproximar funciones por medio de las wavelets Haar y Daubechies [6],[7],[8], y referencias en [9],[10],[11]. Finalmente en el capítulo 5 se introducirá el proceso autorregresivo con coeficientes variando en el tiempo cuyos coeficientes funcionales serán estimados realizando una expansión en funciones de wavelets minimizando el error cuadrático medio. El desempeño de dicho modelo se estudiará vía simulación [2],[3].

1.1. Preliminares

Un proceso estocástico es una sucesión de variables de interés que al ser observada se obtiene una serie de tiempo, veremos en esta sección los conceptos necesarios para la manipulación e interpretación de estas observaciones y además se definirá medidas de dependencias que tiene una variable respecto a la otra.

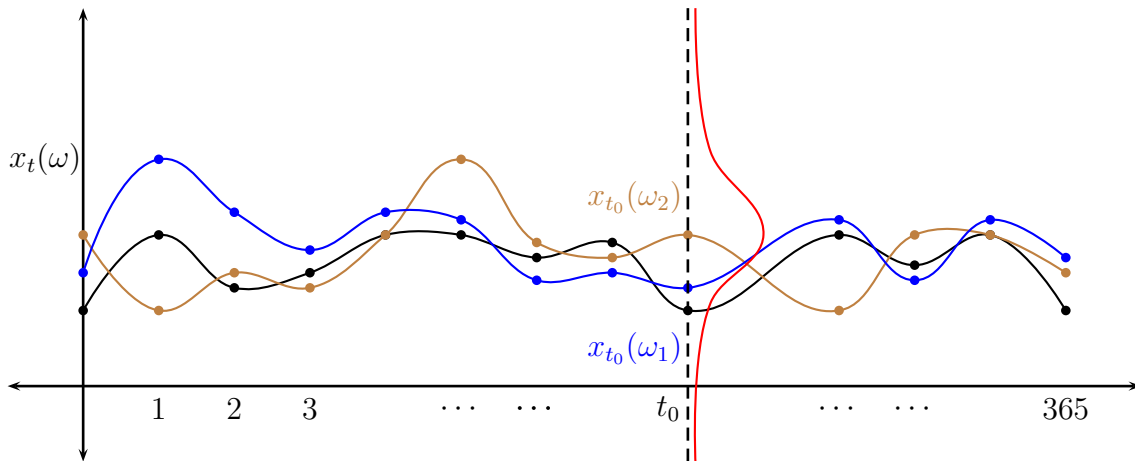
Definición 1.1.1. *Un proceso estocástico es una secuencia de variables aleatorias $(x_t)_{t \in T}$ definidas en un mismo espacio de probabilidad (Ω, X, \mathbb{P}) , donde T se denomina conjunto indicador o de tiempo y el espacio S donde las variables aleatorias toman valores es denominado espacio de estado.*

Considerando $S \subseteq \mathbb{R}$ y tomando la σ -álgebra de Borel en \mathbb{R} restringida a S , se puede considerar un proceso estocástico como una función de dos variables

$$x : T \times \Omega \longrightarrow \mathbb{R},$$

donde para cada par (t, ω) se le asocia un único valor real $x(t, \omega) = x_t(\omega)$, si se fija en el conjunto temporal un t_0 , la asignación $\omega \rightarrow x_{t_0}(\omega)$ es una variable aleatoria, mientras que si se fija un $\omega_0 \in \Omega$, la asignación $t \rightarrow x_t(\omega_0)$ se denomina realización del proceso.

Ejemplo 1.1.2. *Suponga que una entidad bancaria cuenta el número de quejas recibidas diariamente durante un año, es decir, cada año se obtiene 365 observaciones y se define x_t como el número de quejas del día t ,*



al realiza este procedimiento por varios años, es posible determinar el comportamiento probabilístico de x_{t_0} para un t_0 fijo, es otras palabras como es la distribución de x_{t_0} .

Una **serie de tiempo** $(x_t)_{t \in T}$ es una posible realización de un proceso estocástico, por ejemplo, los valores que toma la gráfica azul $(x_t(\omega_1))_{t \in T}$, constituyen una serie de tiempo que en adelante solo se denotará por x_t a no ser que se diga lo contrario.

Una descripción de una serie de tiempo observada como una colección de n variables aleatorias en t_i con $i \in \{1, \dots, n\}$ es dada por la función de distribución conjunta (FDA),

$$F_x(c_1, c_2, \dots, c_n) = \mathbb{P}(x_{t_1} \leq c_1, \dots, x_{t_n} \leq c_n), \text{ donde } (c_1, c_2, \dots, c_n) \in \mathbb{R}^n.$$

Lamentablemente obtener esta función conjunta no es tan sencillo.

Se denota la marginal de la FDA como $F_t(x) = F_{x_t}(x) = \mathbb{P}(x_t \leq x)$, y se define la función de densidad de probabilidad como

$$f_t(x) = \frac{\partial F_t(x)}{\partial x}.$$

Definición 1.1.3. *Suponiendo la existencia de la integral μ_{x_t} , la función media está definida como*

$$\mu_{x_t} = \int_{\mathbb{R}} x f_t(x) dx = \mathbb{E}(x_t),$$

si no hay ambigüedad con la serie, se denota $\mu_{x_t} = \mu_t$.

Con el fin de obtener relaciones de dependencia de la serie con ella misma se introduce la función de autocovarianza y autocorrelación, las cuales servirán de herramienta para determinar la dependencia lineal de la serie a lo largo del tiempo.

Definición 1.1.4. *Sea x_t una serie de tiempo, se define la autocovarianza como:*

$$\gamma_x(s, t) = \mathbb{E}[(x_t - \mu_t)(x_s - \mu_s)],$$

en el caso particular cuando $s = t$ se obtiene la varianza de x_t ,

$$\gamma_x(t, t) = \mathbb{E}[(x_t - \mu_t)^2] = \text{Var}(x_t).$$

Intuitivamente la función de autocovarianza mide la dependencia lineal entre dos observaciones de una misma serie en distintos momentos del tiempo, además, series suaves exhiben funciones de autocovarianza que permanece grande aun cuando $|s - t|$ es grande y series discontinuas tienden a tener funciones de covarianza casi nulas para separaciones grandes.

Si no hay confusión con el tipo de la serie se escribirá $\gamma(s, t)$ en lugar de $\gamma_x(s, t)$.

Definición 1.1.5. *Sea x_t una serie de tiempo, se define la función de autocorrelación,*

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

Al igual que en la función de autocovarianza la autocorrelación exhibe la relación lineal que existe entre dos observaciones, dándole valores en el intervalo $[-1, 1]$, más aún, si es posible predecir x_t totalmente desde x_s hasta una relación lineal $x_t = \beta_0 + \beta_1 x_s$ entonces la correlación será 1 cuando $\beta_1 > 0$ y -1 cuando $\beta_1 < 0$.

Definición 1.1.6. Una serie de tiempo w_t generada por variables aleatorias independientes e idénticamente distribuidas (i.i.d) con $\mathbb{E}(w_t) = 0$ y varianza $\sigma_w^2 < \infty$, se le llama ruido blanco y se denota como $w_t \sim Wn(0, \sigma_w^2)$.

Observación 1.1.7. De la definición anterior se tiene que

$$\mathbb{E}(w_t w_s) = \begin{cases} \sigma_w^2 & , si \ t = s \\ 0 & , si \ t \neq s \end{cases}$$

Ejemplo 1.1.8. Para el proceso conducido por una caminata aleatoria

$$x_t = \sum_{i=1}^t w_i,$$

se tiene que,

$$\begin{aligned} cov(x_t, x_s) &= cov\left(\sum_{i=1}^t w_i, \sum_{j=1}^s w_j\right) \\ &= \mathbb{E}\left(\left(\sum_{i=1}^t w_i\right)\left(\sum_{j=1}^s w_j\right)\right) \\ &= \sum_{i=1}^t \sum_{j=1}^s \mathbb{E}(w_i w_j) \\ &= \min(s, t) \sigma_w^2. \end{aligned}$$

1.1.1. Series estacionarias

Por lo general una serie de tiempo es descrita por una cantidad finita de observaciones obtenidas a lo largo de un periodo de tiempo dado, si la serie sin importar el momento en que se tomó las observaciones conserva su comportamiento o distribución en cada instante del tiempo entonces podemos decir que la serie es estacionaria en sentido estricto, ahora, si la serie conserva su valor esperado y además si la dependencia lineal entre dos momentos de la serie x_t, x_s solo depende de la distancia de s y t , la serie se considerará débilmente estacionaria, esta última suposición facilitará significativamente las propiedades estadísticas de la serie.

Definición 1.1.9. Una serie de tiempo estrictamente estacionaria es aquella para la cual el comportamiento probabilístico de cada conjunto de valores $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$ es idéntico al conjunto de tiempo desplazado $\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$ esto es,

$$\mathbb{P}(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_k} \leq c_k) = \mathbb{P}(x_{t_1+h} \leq c_1, x_{t_2+h} \leq c_2, \dots, x_{t_k+h} \leq c_k), \quad \forall k \in \mathbb{N},$$

donde los t_k es una subcolección del tiempo T y los c_k números reales.

Observación 1.1.10. *La condición de ser estrictamente estacionaria es difícil de garantizar porque necesita una estabilidad en toda la serie, y a priori no se tiene tanta información, veremos a continuación una condición menos fuerte.*

Definición 1.1.11. *Una serie de tiempo débilmente estacionaria x_t , es un proceso de varianza finita tal que:*

1. *la función de valor medio μ_t es constante o no depende del tiempo.*
2. *la función de autocovarianza $\gamma(s, t) = cov(x_s, x_t)$ depende de s y t solo a través de su distancia $|s - t|$.*

Se usará el termino estacionario para referirse a que la serie x_t es débilmente estacionaria y si es estrictamente estacionaria diremos que es estacionario en sentido estricto, además, en el sentido estricto con $k = 2$, tenemos que si la varianza existe, $\mathbb{P}(x_{t_1} \leq c_1, x_{t_2} \leq c_2) = \mathbb{P}(x_{t_1+h} \leq c_1, x_{t_2+h} \leq c_2)$ y por lo tanto se tiene que $cov(x_{t_1}, x_{t_2}) = cov(x_{t_1+h}, x_{t_2+h})$, esto es que la covarianza solo depende de la distancia de las observaciones y no del tiempo real.

Ejemplo 1.1.12. Estacionario no implica estrictamente estacionario.

Consideremos $(x_t)_{t \in \mathbb{N}}$ una sucesión de variables aleatorias i.i.d donde cada x_t tiene distribución de Cauchy, esto es que su función de densidad esta dada por.

$$f_t(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

$$\begin{aligned} \mathbb{E}(x_t) &= \int_{\mathbb{R}} x f_t(x) dx \\ &= \int_{\mathbb{R}} \frac{x dx}{\pi(1+x^2)} \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{du}{u} \\ &= +\infty - (+\infty) = ind. \end{aligned}$$

Para una serie x_t estacionaria, dado que su función media es constante se puede denotar por $\mu_t = \mu$ y la función de autocovarianza $\gamma(s, t)$, de dos observaciones x_t y x_s solo depende de la diferencia $|s - t|$, por tanto es posible simplificar la notación, si $s = t + h$, donde h representa el tiempo de retraso la autocovarianza se puede expresar como

$$\gamma(s, t) = cov(x_{t+h}, x_t) = cov(x_h, x_0) = \gamma(h, 0) = \gamma(h).$$

Definición 1.1.13. (Función de autocovarianza de una serie de tiempo estacionaria) *La autocovarianza se puede expresar como:*

$$\gamma(h) = \mathbb{E}[(x_t - \mu)(x_{t+h} - \mu)].$$

Observación 1.1.14. *El proceso definido por una caminata aleatoria $x_t = \sum_{i=1}^t w_i$ no es estacionario ya que en el ejemplo (1.1.8), $\text{cov}(x_t, x_s) = \min(s, t)\sigma_w^2$, la cual no es función de la distancia de s y t .*

Definición 1.1.15. (*Función de autocorrelación de una serie de tiempo estacionaria*) *La autocorrelación se puede expresar como:*

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}.$$

Capítulo 2

Proceso ARMA(p, q)

En el capítulo anterior se introdujo la función de autocorrelación y autocovarianza como una herramienta para la exploración temporal de la serie con un retraso h . En la siguiente sección se definirá lo que es un proceso autorregresivo y un proceso de media móvil, además se utilizará la función de autocorrelación para la identificación de estos procesos, también será necesario la introducción de una nueva función que brinde más información del orden de dependencia del proceso $AR(p)$ para su identificación, esta función se conoce como la función de autocorrelación parcial.

2.1. Definición y propiedades.

Definición 2.1.1. Sea x_t una serie de tiempo estacionaria de media cero, se dice que x_t es conducida por un proceso autorregresivo de orden p , denotado por $AR(p)$ si tiene la siguiente representación:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + w_t,$$

donde los ϕ_i son constantes, $\phi_p \neq 0$, w_t es un ruido blanco.

Como alternativa a la representación autorregresiva en la cual se considera una combinación lineal del pasado de x_t , el modelo de media móvil de orden q asume una combinación lineal del pasado de w_t para reconstruir los datos observados.

Definición 2.1.2. Sea x_t una serie de tiempo estacionaria de media cero, se dice que x_t es conducida por un proceso de media móvil de orden q , denotado por $MA(q)$ si tiene la siguiente representación:

$$x_t = \sum_{i=1}^q \theta_i w_{t-i} + w_t,$$

donde los θ_i son constantes, $\theta_q \neq 0$, w_t un ruido blanco.

Definición 2.1.3. Sea una serie de tiempo x_t , se define el operador retardo B por:

$$\begin{aligned} Bx_t &= x_{t-1}, \\ B^k x_t &= x_{t-k}, \quad \text{para } k > 1. \end{aligned}$$

Definición 2.1.4. Sean x_t, y_t series conducidas por un proceso AR(p) y un proceso MA(q), respectivamente se definen los polinomios:

- $\phi(B) = (1 - \sum_{i=1}^p \phi_i B^i)$, polinomio autoregresivo de x_t .
- $\theta(B) = (1 + \sum_{i=1}^q \theta_i B^i)$, polinomio de media móvil de y_t .
- $\Phi_P(B^s) = (1 - \sum_{i=1}^P \Phi_i B^{is})$, polinomio autorregresivo estacional de orden P .
- $\Theta_Q(B^s) = (1 + \sum_{i=1}^Q \theta_i B^{is})$, polinomio de media móvil estacional de orden Q .

El entero positivo s es denominado periodo estacional.

Observación 2.1.5. De los polinomios anteriores se puede escribir las series x_t, y_t conducidas por un procesos AR(p) y MA(q) como, $\phi(B)x_t = w_t$, $\theta(B)w_t = y_t$, respectivamente.

En la definición (2.1.1) la media de x_t es cero, en el caso en el que μ es distinto de cero, se reemplaza x_t por $x_t - \mu$

$$x_t - \mu = \sum_{i=1}^p \phi_i (x_{t-i} - \mu) + w_t$$

$$x_t = \alpha + \sum_{i=1}^p \phi_i x_{t-i} + w_t,$$

donde $\alpha = \mu(1 - \phi_1, \dots, -\phi_p)$.

Definición 2.1.6. Sea x_t estacionaria de media cero, se dice que x_t es conducida por un proceso ARMA(p, q), si tiene la siguiente representación:

$$\phi(B)x_t = \theta(B)w_t,$$

donde $\theta_q, \phi_p \neq 0$ y $\sigma_w^2 > 0$.

Observación 2.1.7. Para evitar términos redundantes ϕ y θ no tienen factores comunes.

Definición 2.1.8. Un serie x_t conducida por un proceso ARMA(p, q) $\phi(B)x_t = \theta(B)w_t$, se dice que es causal si se puede escribir como:

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t,$$

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty, \quad \psi_0 = 1.$$

Observación 2.1.9. Si x_t es una serie de tiempo causal, entonces $\text{cov}(x_t, w_t) = \sigma_w^2$ y $\text{cov}(x_t, w_{t+h}) = 0$, para $h > 0$.

El siguiente teorema brinda condiciones necesarias y suficientes para garantizar que un proceso ARMA(p, q) sea causal, para ello es necesario que las raíces del polinomio autorregresivo ϕ , estén fuera del círculo unitario.

Teorema 2.1.10. Una serie de tiempo x_t que es conducido por un proceso ARMA(p, q), es causal si $\phi(z) \neq 0$, para todo $z \leq 1$, además los coeficientes de $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$, se determinan resolviendo:

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}.$$

Demostración. Ver [1](página 531).

Definición 2.1.11. Un serie x_t conducida por un proceso ARMA(p, q), $\phi(B)x_t = \theta(B)w_t$, se dice que es invertible si se puede escribir como:

$$w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = \pi(B)x_t,$$

$$\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty, \quad \pi_0 = 1.$$

Al igual que para una serie causal el siguiente teorema brinda las condiciones necesarias y suficientes para garantizar que un proceso sea invertible, y es necesaria es que las raíces del polinomio de media móvil θ , estén fuera del círculo unitario.

Teorema 2.1.12. Una serie x_t que sigue un proceso ARMA(p, q) es invertible si $\theta(z) \neq 0, \forall(z \leq 1)$, además, los coeficientes de $w_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$ se determinan resolviendo:

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}.$$

2.1.1. Función de Autocorrelación

Si x_t es una serie de tiempo conducido por un proceso $AR(p)$ causal y teniendo en cuenta la observación (2.1.9), para $h > 0$,

$$\begin{aligned} cov(x_t, x_{t+h}) &= cov\left(x_t, \sum_{i=1}^p \phi_i x_{t+h-i} + w_{t+h}\right) \\ &= \sum_{i=1}^p \phi_i cov(x_t, x_{t+h-i}) + cov(x_t, w_{t+h}) \\ &= \sum_{i=1}^p \phi_i \gamma(h-i), \end{aligned} \quad (2.1)$$

por lo tanto al dividir por $\gamma(0)$ a ambos lados de la igualdad,

$$\rho(h) - \sum_{i=1}^p \phi_i \rho(h-i) = 0, \quad h > 0,$$

la cual es una ecuación en diferencia. Resolviendo dicha ecuación se tiene que, si $\{z_1, z_2, \dots, z_r\}$ las raíces diferentes del polinomio $\phi(z)$ (2.1.4), cada una con multiplicidad m_i respectivamente para $i = 1, 2, \dots, r$ y $\sum_{i=1}^r m_i = p$, la autocorrelación se puede expresar como:

$$\rho(h) = \sum_{i=1}^r z_i^{-h} P_i(h), \quad h \geq p,$$

donde $P_i(h)$ es un polinomio en h de grado $m_i - 1$.

Observación 2.1.13. Como el proceso es causal por el teorema (2.1.10),

$$\lim_{h \rightarrow \infty} \rho(h) = 0,$$

esta es una característica importante para identificar un $AR(p)$, ya que para un h suficientemente grande su función de autocorrelación es cero.

Ejemplo 2.1.14. La función de autocorrelación (acf) de un $MA(q)$ esta dada por,

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1 + \sum_{i=1}^q \theta_i^2}, & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h > q \end{cases}$$

Observación 2.1.15. En los procesos $MA(q)$ después del rezago q la (acf) es cero, algo que no ocurre en los $AR(p)$, por tanto se quiere buscar una función de tal manera que para un rezago mayor que p se anule.

2.1.2. Función de Autocorrelación Parcial

La idea de la autocorrelación parcial, (pacf) es obtener más información del orden de dependencia del proceso $AR(p)$. El siguiente ejemplo servirá para ilustrar el comportamiento de esta función.

Ejemplo 2.1.16. Para un $AR(1)$ causal, $x_t = \phi x_{t-1} + w_t$

$$\begin{aligned}\gamma(2) &= Cov(\phi x_{t-1} + w_t, x_{t-2}) \\ &= Cov(\phi^2 x_{t-2} + \phi w_{t-1} + w_t, x_{t-2}) \\ &= \phi^2 \gamma(0) \neq 0,\end{aligned}$$

$\gamma(2) \neq 0$ ya que x_t y x_{t-2} dependen mutuamente de x_{t-1} , quitando esta dependencia lineal se tiene que la $Cov(x_t - \phi x_{t-1}, x_{t-2} - \phi x_{t-1}) = Cov(w_t, x_{t-2} - \phi x_{t-1}) = 0$.

Sea \hat{x}_{t+h} la regresión de x_{t+h} para $h \geq 2$ en $\{x_{t+h-1}, \dots, x_{t+1}\}$ ¹, la cual se escribe como:

$$\hat{x}_{t+h} = \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2}, \dots, \beta_{h-1} x_{t+1}.$$

considerando la media cero, en caso contrario se reemplaza x_t por $x_t - \mu$, además, se denotará \hat{x}_t la regresión de x_t en $\{x_{t+1}, \dots, x_{t+h-1}\}$ entonces,

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \dots, \beta_{h-1} x_{t+h-1}.$$

Ya que la serie es estacionaria, se puede verificar que los coeficientes $\beta_1, \dots, \beta_{h-1}$ son los mismos en ambas definiciones, esto se debe a que la dependencia de x_t y x_{t+k} es igual a la dependencia de x_{t+h} y x_{t+h-k} , para $k = 1, \dots, h-1$.

Definición 2.1.17. Si el proceso es estacionario, denote ϕ_{hh} para $h \geq 1$ ²,

$$\begin{aligned}\phi_{11} &= Corr(x_{t+1}, x_t) = \rho(1), \\ \phi_{hh} &= Corr(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t).\end{aligned}$$

La autocorrelación parcial calcula la dependencia lineal de x_t y x_{t+h} quitando la dependencia que tienen compartida con $\{x_{t+1}, \dots, x_{t+h-1}\}$.

Ejemplo 2.1.18. La función de autocorrelación parcial de un proceso $AR(1)$,

$$\begin{aligned}x_t &= \phi x_{t-1} + w_t, \quad |\phi| < 1 \\ \phi_{11} &= \rho(1) = \phi.\end{aligned}$$

Para calcular ϕ_{22} es necesario calcular \hat{x}_{t+h} y \hat{x}_t ,

$$\mathbb{E}(x_{t+2} - \hat{x}_{t+2})^2 = \gamma(0) - 2\beta\gamma(1) + \beta^2\gamma(0),$$

¹Regresión en el sentido que, \hat{x}_{t+h} es la combinación lineal de $\{x_{t+h-1}, \dots, x_{t+1}\}$ tal que minimiza el error cuadrático medio, $\mathbb{E}(x_{t+h} - \sum_{j=1}^{h-1} \alpha_j x_{t+j})^2$.

² $Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}$.

donde el valor mínimo se alcanza cuando $\beta = \rho(1)$, análogamente para \hat{x}_t se va a tener que $\beta = \rho(1)$ y así,

$$\begin{aligned}\phi_{22} &= \text{Corr}(x_{t+2} - \phi x_{t-1}, x_t - \phi x_{t-1}) \\ &= \text{Corr}(w_{t+2}, x_t - \phi x_{t-1}) = 0.\end{aligned}$$

Más adelante se verá que para un AR(p), $\phi_{nn} = 0$ para $n > p$.

Afirmación 2.1.19. Para una serie x_t conducida por un proceso autorregresivo de orden p

$$\begin{aligned}x_{t+h} &= \sum_{i=1}^p \phi_i x_{t+h-i} + w_{t+h}, \\ \hat{x}_{t+h} &= \sum_{i=1}^p \phi_i x_{t+h-i}, \\ \phi_{hh} &= \text{Corr}(x_{t+h} - \hat{x}_{t+h}, x_t - \hat{x}_t) \\ \phi_{hh} &= \text{Corr}(w_{t+h}, x_t - \hat{x}_t) = 0, \text{ para } h > p.\end{aligned}$$

Por lo tanto para un modelo AR(p) la función de autocorrelación parcial es nula para un rezago mayor que p .

2.1.3. Estimación

En la sección anterior se estudió el proceso AR(p) donde los parámetros $\{\phi_i\}_{i=1}^p$ eran conocidos, en la siguiente sección se exhibirá la forma tradicional de estimación de estos parámetros, iniciando con el estimador de Yule-Walker para un proceso autorregresivo el cual es una alternativa al estimador de mínimos cuadrados, además, para una muestra suficientemente grande, la probabilidad de que sean iguales es muy alta.

Estimador de Yule-Walker

Con la formula (2.1) podemos deducir las siguientes ecuaciones.

Definición 2.1.20. Las ecuaciones de Yule-Walker están dadas por

$$\begin{aligned}\gamma(h) &= \sum_{i=1}^p \phi_i \gamma(h-i), \text{ para } h = 1, 2, 3, \dots, p, \\ \sigma_w^2 &= \gamma(0) - \phi_1 \gamma(1) - \dots, \phi_p \gamma(p),\end{aligned}$$

en notación matricial las ecuaciones de Yule-Walker son:

$$\Gamma_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p \quad \text{y} \quad \sigma_w^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p,$$

donde $\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^p$ es una matriz $p \times p$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)'$ un vector $p \times 1$, y $\boldsymbol{\gamma}_p = (\gamma(1), \dots, \gamma(p))'$, además, si se tiene una muestra $\{x_1, x_2, \dots, x_n\}$, y se aplica el

método de los momentos, reemplazando $\gamma(h)$ por la autocovarianza muestral ³ $\hat{\gamma}(h)$,

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p \quad y \quad \hat{\sigma}_w^2 = \hat{\gamma}(0) - \hat{\gamma}_p' \hat{\Gamma}_p^{-1} \hat{\gamma}_p,$$

estas ecuaciones se les conocen como el estimador de Yule-Walker.

Teorema 2.1.21. Sea x_t una serie de tiempo conducida por un proceso AR(p) causal,

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(\mathbf{0}, \sigma_w^2 \Gamma_p^{-1}), \quad \hat{\sigma}_w^2 \xrightarrow{p} \sigma_w^2.$$

Demostración. Ver[1](página 534).

Lema 2.1.22. Para un proceso autorregresivo de orden p , causal,

$$\sqrt{n} \hat{\phi}_{hh} \xrightarrow{d} N(0, 1), \quad \text{para } h > p,$$

donde $\hat{\phi}_{hh}$ es el estimador de Yule-Walker para la función de autocorrelación parcial.

Estimador de Mínimos Cuadrados

Otra técnica para estimar estos coeficientes es notar que los verdaderos valores de $\{\phi_1, \dots, \phi_p\}$, en virtud del teorema de la proyección, minimizan el error cuadrático medio,

$$(\beta_1, \beta_2, \dots, \beta_p) \mapsto \mathbb{E} \left((x_t - \mu) - \sum_{i=1}^p \beta_i (x_{t-i} - \mu) \right)^2,$$

el estimador de mínimos cuadrados está definido reemplazando la función por una combinación lineal de la observación $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, para $n > p$ y las constantes $\hat{\phi}_1, \dots, \hat{\phi}_p$ minimizan el residuo,

$$(\beta_1, \beta_2, \dots, \beta_p) \mapsto \sum_{t=p+1}^n (x_t - \beta_1 x_{t-1} - \dots - \beta_p x_{t-p})^2,$$

dado los datos $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ de un AR(p), con $n > p$ se define

$$Y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_0 & x_{-1} & \cdots & x_{1-p} \\ x_1 & x_0 & \cdots & x_{2-p} \\ \vdots & \vdots & & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-p} \end{pmatrix},$$

entonces es posible escribir el proceso como,

$$Y = \phi \mathbf{X} + W,$$

donde $W = (w_1, w_2, \dots, w_n)'$, $\phi = (\phi_1, \dots, \phi_p)'$, y la solución de mínimos cuadrados satisface que,

$$\phi^* = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' Y.$$

³La autocovarianza muestral está dada por $\hat{\gamma}(h) = n^{-1} \sum_{i=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$, $h = 0, 1, \dots, n-1$.

Observación 2.1.23. *Se puede mostrar que el estimador de Yule-Walker y el estimador de mínimos cuadrados cumple que*

$$\sqrt{n}(\hat{\phi}^* - \hat{\phi}) \xrightarrow{P} 0.$$

2.1.4. Pronóstico

La idea de esta sección es poder predecir el valor de la serie de tiempo futura x_{n+m} , para $m \geq 1$ basada en $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, asumiendo x_t estacionario y que conocemos el modelo paramétrico, para resolver lo anterior se necesitan unos conceptos del espacio de Hilbert.

Definición 2.1.24. *Para un espacio de medida (Ω, X, \mathbb{P}) se definirá $\mathcal{L}_2(\Omega, X, \mathbb{P})$ como el conjunto de todas las variables aleatorias x tales que $\|x\|^2 := \mathbb{E}(|x|^2) < \infty$. Si no hay confusión se denotará solamente por \mathcal{L}_2 .*

Sean $x, y \in \mathcal{L}_2$ y decimos que $x \sim y \iff x = y$, salvo en un conjunto de medida cero, esta relación es de equivalencia y por lo tanto induce una partición de \mathcal{L}_2 , al espacio cociente definido por esta relación se denotará como $L_2 = \mathcal{L}_2 / \sim$.

Observación 2.1.25. $\mathcal{H} = (L_2, \mathbb{E}(\cdot))$ es un espacio de Hilbert ⁴, donde el producto interior está definido como $\langle x, y \rangle = \text{Cov}(x - \mu_x, y - \mu_y)$.

Teorema 2.1.26. (Teorema de la proyección). *Sea \mathcal{M} un subconjunto cerrado de un espacio de Hilbert, sea $y \in \mathcal{H}$. Entonces y puede expresarse de manera única como:*

$$y = \hat{y} + z,$$

donde $\hat{y} \in \mathcal{M}$, tal que $\langle y - \hat{y}, w \rangle = 0$ para todo $w \in \mathcal{M}$. Adicionalmente de todos los elementos de \mathcal{M} , \hat{y} es el más cercano a y en el sentido que,

$$\|y - \hat{y}\| \leq \|y - w\|, \text{ para todo } w \in \mathcal{M},$$

y la igualdad se da si y sólo si $w = \hat{y}$.

Se denotará la proyección de y en \mathcal{M} como $\hat{y} := P_{\mathcal{M}}y$, además, el espacio cerrado de un conjunto finito de puntos $\{x_1, x_2, \dots, x_n\}$ de elementos del espacio de Hilbert es definido como el conjunto de todas las combinaciones lineales $w = \sum_{i=1}^n a_i x_i$, donde los $a_i \in \mathbb{R}$, este subespacio es denotado por $\mathcal{M} = \text{Span}\{x_1, x_2, \dots, x_n\}$.

Sea $X_0 \subseteq X$ una sub σ -álgebra de la σ -álgebra X , la colección de todas las variables $y \in L_2(\Omega, X, \mathbb{P})$ X_0 -medibles es un espacio lineal cerrado, por lo tanto cumple las hipótesis del teorema de la proyección así es posible definir la proyección de y sobre el espacio generado por X_0 .

⁴Un espacio de Hilbert es un espacio con producto interior que es completo bajo la métrica $d(x, y) = \sqrt{\langle x - y, x - y \rangle}$

Se utilizará la esperanza condicionada para pronosticar el valor futuro de una serie dado los datos conocidos y se utilizará el teorema de la proyección para determinar sistemáticamente los coeficientes que me pronostican este valor.

Definición 2.1.27. (*Esperanza condicional 1*) La proyección de $x \in L_2(\Omega, X, \mathbb{P})$ sobre el subespacio cerrado $L_2(\Omega, X_0, \mathbb{P})$ es la variable aleatoria denominada esperanza condicional de x dado X_0 y se denota como $\mathbb{E}(x|X_0)$.

La esperanza condicional de x dado X_0 se puede interpretar también como la variable aleatoria x' tal que el valor esperado sobre cualquier $A \in X_0$ es el mismo valor esperado que el de x sobre A .

Definición 2.1.28. (*Esperanza condicional 2*) La esperanza condicional de una variable aleatoria x dado X_0 no negativa o $\mathbb{E}(x) < \infty$, es una variable aleatoria X_0 -medible x' tal que $\mathbb{E}[(x - x')\mathbf{1}_A] = 0$, para todo $A \in X_0$.

Mejor predictor lineal (BLP).

Estudiaremos el caso en que $\mathcal{M} = \text{Span}\{x_1, x_2, \dots, x_n\}$, y para aligerar la notación denotaremos la esperanza condicionada de x_{n+m} dada la σ -álgebra generada por $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ tal que cada x_i es medible, como $\mathbb{E}(x_{n+m}|\mathbf{x}) = x_{n+m}^n$.

Si se tiene las observaciones $\{x_1, x_2, \dots, x_n\}$, y se quiere pronosticar x_{n+m} , para $m > 0$, en virtud del teorema (2.1.26) y la definición (2.1.27), $x_{n+m}^n \in \mathcal{M}$ y tiene la representación,

$$x_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i x_i,$$

donde los α_i son constantes tal que minimizan el error cuadrático medio, x_{n+m}^n es denominado el mejor predictor lineal (BPL).

Observación 2.1.29. Se puede mostrar que si $\{x_1, x_2, \dots, x_n\}$ son gaussianos entonces el mejor predictor lineal coincide con la esperanza condicionada, esto se debe a que para un proceso gaussiano las observaciones solo tienen dependencia lineal.

Existen varias formas de encontrar el mejor predictor lineal, los dos teoremas siguientes brindan técnicas para obtenerla, la primera se basa en la forma matricial y la segunda basándose en la definición alternativa de la esperanza condicional (2.1.28).

Teorema 2.1.30. Si $x_i \in \mathbb{R}^m$, $i = 1, 2, \dots, n$, $\mathcal{M} = \text{Span}\{x_1, x_2, \dots, x_n\}$ entonces,

$$\text{Proj}_{\mathcal{M}} \mathbf{x} = \mathbf{X}\beta,$$

donde \mathbf{X} es la $m \times n$ matriz con j^{th} columna x_j ,

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{x}.$$

Teorema 2.1.31. Ecuación de predicción. *Dados los datos x_1, x_2, \dots, x_n , el mejor predictor lineal, $x_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i x_i$ se encuentra solucionando:*

$$\mathbb{E} \left[(x_{n+m} - x_{n+m}^n) x_k \right] = 0, \quad \text{para } k \in \{0, 1, \dots, n\}, \quad x_0 = 1,$$

esta ecuación es conocida como la ecuación de predicción, la cual genera un sistema de ecuaciones, donde la solución al sistema brinda los coeficientes del mejor predictor.

Observación 2.1.32. *Si, $\mathbb{E}(x_t) = \mu$, y $k = 0$ se tiene que $\mu = \mathbb{E}(x_{n+m}) = \mathbb{E}(x_{n+m}^n)$*

$$\mathbb{E}(x_{n+m}^n) = \mathbb{E} \left(\alpha_0 + \sum_{i=1}^n \alpha_i x_i \right),$$

$$\mu = \alpha_0 + \sum_{k=1}^n \alpha_k \mathbb{E}(x_k),$$

$$\mu = \alpha_0 + \sum_{k=1}^n \alpha_k \mu,$$

despejando α_0 , se tiene que $\alpha_0 = \mu \left(1 - \sum_{k=1}^n \alpha_k \right)$ y reemplazando en el predictor

$$x_{n+m}^n = \mu + \sum_{k=1}^n \alpha_k (x_k - \mu),$$

$$x_{n+m}^n - \mu = \sum_{k=1}^n \alpha_k (x_k - \mu),$$

así se puede considerar $\mu = 0$, y en caso contrario reemplazar x_t por $x_t - \mu$.

Ejemplo 2.1.33. *Supongamos que tenemos $\{x_1, x_2, \dots, x_n\}$ y se quiere pronosticar x_{n+1} , $x_{n+1}^n = \sum_{j=1}^n \phi_{nj} x_{n+1-j}$, llamando $\alpha_j = \phi_{n, n-j+1}$ y usando la ecuación de predicción(2.1.31),*

$$\mathbb{E} \left[\left(x_{n+1} - \sum_{j=1}^n \phi_{nj} x_{n+1-j} \right) x_{n+1-k} \right] = 0, \quad \text{para } k \in \{1, 2, \dots, n\},$$

$$\mathbb{E}(x_{n+1} x_{n+1-k}) = \sum_{j=1}^n \phi_{nj} \mathbb{E}(x_{n+1} x_{n+1-k}),$$

$$\gamma(k) = \sum_{j=1}^n \phi_{nj} \gamma(k-j), \quad \text{para todo } k \in \{1, 2, \dots, n\}.$$

En la notación matricial, $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$, donde $\boldsymbol{\phi}_n = (\phi_{n1}, \dots, \phi_{nn})'$, también denotaremos el predictor de error cuadrático medio como $P_{n+1}^n = \mathbb{E}(x_{n+1} - x_{n+1}^n)^2$, donde,

$$\mathbb{E}(x_{n+1} - x_{n+1}^n)^2 = \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n.$$

Observación 2.1.34. Por el teorema (2.1.31) y el ejemplo anterior se puede justificar la afirmación (2.1.19).

Capítulo 3

Análisis Espectral

3.1. Densidad Espectral

Hasta el momento se ha visto que un proceso estacionario puede tener un comportamiento repetitivo o regular respecto al tiempo, el concepto de regularidad puede ser mejorada si se expresa en términos de las variaciones periódicas que producen la serie, uno de los objetivos principales del análisis espectral es encontrar las frecuencias dominantes de una serie. La razón de usar esta técnica alternativo es poder justificar un fenómeno empírico y una explicación que incluye unas pocas oscilaciones primarias es más sencillo de trabajar y a la hora de una explicación es físicamente más significativo que una colección de parámetros estimados.

Los comportamientos periódicos están muy poco correlacionados entre sí y por esto se facilita la escritura de probabilidades basadas en métodos estadísticos clásicos.

- Se definirá un ciclo como un periodo completo de una función seno y coseno definida en un intervalo de tiempo uniforme.
- Se medirá la frecuencia ω en ciclos por puntos de tiempo.
- El periodo descriptivo de una serie de tiempo se define como el número de puntos en un ciclo.

Considere el proceso:

$$x_t = A \cos(2\pi\omega t + \phi), \text{ para } t \in \mathbb{Z},$$

- A : Amplitud de la función.
- ϕ : la fase, es decir, el punto de inicio de la función coseno.
- Introduciendo variaciones aleatorias de tal manera que la amplitud y la fase sean variables aleatorias y usando la formula de suma de ángulos del coseno, podemos escribir el proceso x_t de la siguiente manera,

$$x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t), \tag{3.1}$$

$$\begin{aligned} U_1 &= A \cos(\phi), & U_2 &= -A \sin(\phi), \\ A &= \sqrt{U_1^2 + U_2^2}, & \phi &= \tan^{-1}\left(\frac{-U_1}{U_2}\right). \end{aligned}$$

Comúnmente se toma $U_1, U_2 \sim N(0, 1)$ independientes y por lo general una serie de tiempo puede tener más de una frecuencia con diferentes amplitudes,

$$\begin{aligned} x_t &= \sum_{k=1}^q U_{k_1} \cos(2\pi\omega_k t) + U_{k_2} \sin(2\pi\omega_k t), \\ \{U_{k_1}, U_{k_2}\}_{k=1}^q, & \text{ i.i.d con media cero y varianza } \sigma_k^2, \end{aligned}$$

si la serie x_t es estacionaria, es posible calcular la covarianza con la representación anterior la cual tendrá la forma:

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

El siguiente teorema da existencia y unicidad de una función que servirá para determinar cuáles son las frecuencias que más aportan a la varianza del proceso.

Teorema 3.1.1. *Una función $\gamma(h)$ es definida no negativa si y sólo si puede expresarse como:*

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \exp\{2\pi i \omega h\} dF(\omega),$$

donde $F(\cdot)$ es no decreciente, continua por derecha con soporte en $[-\frac{1}{2}, \frac{1}{2}]$ y única tal que $F(-\frac{1}{2}) = 0$ y $F(\frac{1}{2}) = \gamma(0)$.

Demostración. Ver [1](página 540).

Ejemplo 3.1.2. Sea ω_0 fijo, x_t estacionario con la representación de la ecuación (3.1).

$$\begin{aligned} \gamma(h) &= \sigma^2 \cos(2\pi\omega_0 t) \\ &= \sigma^2 \left(\frac{e^{-2\pi i \omega_0 h} + e^{2\pi i \omega_0 h}}{2} \right) \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} dF(\omega), \\ F(\omega) &= \begin{cases} 0 & , \text{ si } \omega < -\omega_0 \\ \frac{\sigma^2}{2} & , \text{ si } \omega \in [-\omega_0, \omega_0) \\ \sigma^2 & , \text{ si } \omega \geq \omega_0 \end{cases} \end{aligned}$$

A pesar de que F no es una función de distribución como tal, se le denomina función de distribución espectral puesto que comparte propiedades con la función de distribución, además, si la serie de $\gamma(h)$ es finita, entonces se garantiza la existencia de la función de densidad espectral que servirá para determinar las frecuencias que más aportan a la varianza del proceso.

Teorema 3.1.3. Si la función de autocovarianza de un proceso estacionario satisface,

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty,$$

entonces γ se puede representar como:

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} f(\omega) d\omega,$$

donde f es denominada la función de densidad espectral y tiene la representación:

$$f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-2\pi i \omega h}, \quad \text{para } -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

la densidad espectral de un proceso $ARMA(p, q)$, tiene una formula cerrada la cual está relacionada con la representación de un proceso causal. (2.1.10)

Teorema 3.1.4. Si $\{x_t\}$ es una serie de tiempo conducida por un proceso $ARMA(p, q)$, causal, $\phi(B)x_t = \theta(B)w_t$, entonces esta tiene función densidad espectral dada por:

$$f_x(\omega) = \sigma_w^2 \frac{|\Theta(e^{-2\pi i \omega})|^2}{|\Phi(e^{-2\pi i \omega})|^2},$$

Demostración. Si x_t es causal entonces por definición $x_t = \sum_{k \in \mathbb{N}} \psi_k \omega_{t-k}$,

$$\begin{aligned} \gamma_x(h) &= cov(x_t, x_{t+h}) \\ &= cov\left(\sum_{k \in \mathbb{N}} \psi_k \omega_{t-k}, \sum_{s \in \mathbb{N}} \psi_s \omega_{t+h-s}\right) \\ &= \sum_{k \in \mathbb{N}} \sum_{s \in \mathbb{N}} \psi_k \psi_s cov(\omega_{t-k}, \omega_{t+h-s}) \\ &= \sum_{k \in \mathbb{N}} \sum_{s \in \mathbb{N}} \psi_k \psi_s \gamma_\omega(h-s+k) \\ &= \sum_{k \in \mathbb{N}} \sum_{s \in \mathbb{N}} \psi_k \psi_s \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega (h-s+k)} f_\omega(\omega) d\omega \\ \gamma_x(h) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sum_{k \in \mathbb{N}} \psi_k e^{2\pi i \omega k}\right) \left(\sum_{s \in \mathbb{N}} \psi_s e^{-2\pi i \omega s}\right) e^{2\pi i \omega h} f_\omega(\omega) d\omega, \end{aligned}$$

igualando la representación espectral del lado izquierdo con el lado derecho, se tiene que,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} f_x(\omega) d\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| \sum_{s \in \mathbb{N}} \psi_s e^{-2\pi i \omega s} \right|^2 e^{2\pi i \omega h} f_\omega(\omega) d\omega,$$

por tanto $f_x(\omega) = \left| \sum_{s \in \mathbb{N}} \psi_s e^{-2\pi i \omega s} \right|^2 f_w(\omega)$, salvo en un conjunto de medida cero, y por teorema (2.1.10), $\sum_{s \in \mathbb{N}} \psi_s z^s = \frac{\theta(z)}{\phi(z)}$,

$$f_x(\omega) = \frac{|\theta(e^{-2\pi i \omega})|^2}{|\phi(e^{-2\pi i \omega})|^2} f_w(\omega),$$

$$f_x(\omega) = \frac{|\theta(e^{-2\pi i \omega})|^2}{|\phi(e^{-2\pi i \omega})|^2} \sigma_w^2.$$

□

En el caso particular de un $AR(p)$ la función de densidad espectral esta dada por:

$$f_x(\omega) = \frac{\sigma_w^2}{|\phi(e^{-2\pi i \omega})|^2}.$$

Ahora se quiere encontrar una función de la muestra para poder estimar a la densidad espectral $f(\omega)$.

Definición 3.1.5. *Dados los datos x_1, x_2, \dots, x_n definimos la transformada de fourier discreta (DFT) como:*

$$d(\omega_j) = n^{-\frac{1}{2}} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t}, \text{ para } j \in \{0, 1, \dots, n-1\} \text{ y } \omega_j = \frac{j}{n},$$

las frecuencias ω_j son llamados frecuencias fundamentales o de fourier, y la transformada inversa de fourier está dada por:

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

Definición 3.1.6. *Dados los datos x_1, x_2, \dots, x_n , se define el periodograma como*

$$I(\omega_j) = |d(\omega_j)|^2, \text{ para } j \in \{0, 1, \dots, n-1\}.$$

Proposición 3.1.7. *El periodograma tiene las siguientes propiedades:*

1. $I(0) = n\bar{x}^2$.
2. $d(\omega_j) = n^{-\frac{1}{2}} \sum_{t=1}^n (x_t - \bar{x}) e^{-2\pi i \omega_j t}$, para $j \neq 0$.
3. Si $\hat{\gamma}(h)$ ¹ es la función de autocovarianza muestral, entonces:

$${}^1\hat{\gamma}(h) = \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}).$$

$$I(\omega_j) = n^{-1} \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i \omega_j h}.$$

Demostración.

$$1. I(0) = |d(0)|^2 = n^{-1} \left| \sum_{t=1}^n x_t \right|^2 = n |\bar{x}|^2 = n \bar{x}^2.$$

$$2. \text{ Puesto que, } \sum_{t=1}^n e^{-2\pi i t \frac{j}{n}} = \sum_{t=1}^n \left(e^{-2\pi i \frac{j}{n}} \right)^t = e^{-2\pi i \frac{j}{n}} \frac{(1 - e^{-2\pi i j})}{1 - e^{-2\pi i \frac{j}{n}}} = 0, \text{ para } j \neq 0,$$

$$n^{-\frac{1}{2}} \sum_{t=1}^n (x_t - \bar{x}) e^{-2\pi i \omega_j t} = d(\omega_j) - n^{-\frac{1}{2}} \bar{x} \sum_{t=1}^n e^{-2\pi i \omega_j t} = d(\omega_j) + 0 = d(\omega_j).$$

3. De la definición de periodograma, se tiene que,

$$\begin{aligned} I(\omega_j) &= |d(\omega_j)|^2 = n^{-1} \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)} \\ &= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \\ &= n^{-1} \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i \omega_j h}. \end{aligned}$$

La idea es relacionar el periodograma con la función densidad espectral $f(\omega)$ de una serie de tiempo x_t estacionaria, con $\sum_{h \in \mathbb{Z}} |\gamma(h)| < \infty$, usando el mismo argumento de la demostración 3 en la proposición anterior con $w_j \neq 0$, y sacando el valor esperado a ambos lados de la igualdad se tiene que:

$$\begin{aligned} I(\omega_j) &= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \mu)(x_t - \mu) e^{-2\pi i \omega_j h}, \\ \mathbb{E}(I(\omega_j)) &= n^{-1} \sum_{h<n} (n - |h|) \gamma(h) e^{-2\pi i \omega_j h} \\ \mathbb{E}(I(\omega_j)) &= \sum_{|h|<n} \left(1 - \frac{|h|}{n} \right) \gamma(h) e^{-2\pi i \omega_j h}. \end{aligned}$$

Se quiere justificar la convergencia de $\mathbb{E}(I(\omega)) \rightarrow f(\omega)$, para cualquier frecuencia, no necesariamente fundamental, para esto consideremos las siguientes proposiciones.

Proposición 3.1.8.

$$\text{Si } \sum_{h \in \mathbb{Z}} |\gamma(k)| < \infty \quad \implies \quad \lim_{n \rightarrow \infty} \sum_{k=-n}^n \frac{|k|}{n} \gamma(k) e^{-2\pi i \omega_j k} = 0.$$

Demostración. Dado $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que para todo $n > N$.

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |\gamma(k)| - \sum_{k=-N}^N |\gamma(k)| &< \epsilon, \\ \sum_{k=-\infty}^{-(N+1)} |\gamma(k)| + \sum_{k=N+1}^{\infty} |\gamma(k)| &< \epsilon, \end{aligned}$$

sea $n > N \implies \frac{|k|}{n} < 1$ para todo $|k| < n$, por tanto,

$$\begin{aligned} 0 \leq \sum_{k=-n}^n \frac{|k|}{n} |\gamma(k)| &= \sum_{k=-N}^N \frac{|k|}{n} |\gamma(k)| + \sum_{k=N+1}^n \frac{|k|}{n} |\gamma(k)| + \sum_{k=-n}^{-(N+1)} \frac{|k|}{n} |\gamma(k)| \\ &< \sum_{k=-N}^N \frac{|k|}{n} |\gamma(k)| + \sum_{k=N+1}^n |\gamma(k)| + \sum_{k=-n}^{-(N+1)} |\gamma(k)| \\ &< \sum_{k=-N}^N \frac{|k|}{n} |\gamma(k)| + \sum_{k=N+1}^{\infty} |\gamma(k)| + \sum_{k=-\infty}^{-(N+1)} |\gamma(k)| \\ &< \sum_{k=-N}^N \frac{|k|}{n} |\gamma(k)| + \epsilon, \end{aligned}$$

como N es fijo, $\lim_{n \rightarrow \infty} \sum_{k=-N}^N \frac{|k|}{n} |\gamma(k)| = 0$ y por teorema de compresión,

$$\begin{aligned} 0 \leq \lim_{n \rightarrow \infty} \sum_{k=-n}^n \frac{|k|}{n} |\gamma(k)| \leq \epsilon &\implies \lim_{n \rightarrow \infty} \sum_{k=-n}^n \frac{|k|}{n} |\gamma(k)| = 0, \\ \left| \sum_{k=-n}^n \frac{|k|}{n} \gamma(k) e^{-2\pi i w_j k} \right| \leq \sum_{k=-n}^n \frac{|k|}{n} |\gamma(k)| &\implies \sum_{k \in \mathbb{Z}} \frac{|k|}{n} \gamma(k) e^{-2\pi i w_j k} = 0. \end{aligned}$$

□

Ya que el periodograma sólo está definido para frecuencias fundamentales, la idea es que si existe una frecuencia no necesariamente fundamental, sea posible acercarse a ella mediante una sucesión de frecuencias fundamentales, para conseguir esto, consideremos la siguiente sucesión $\omega_{j:n} = \frac{jn}{n}$ tal que,

$$\omega_{j:n} = \frac{jn}{n} \iff \left| \frac{jn}{n} - \omega \right| \leq \left| \frac{k}{n} - \omega \right|, \text{ para todo } k = 1, 2, \dots, n,$$

es decir, $\{j_n\}$ son escogidos tal que $\omega_{j:n}$ es la frecuencia fundamental más cercana a ω .

Proposición 3.1.9. *Dada una sucesión de frecuencias fundamentales $(\omega_{j:n})_{n \in \mathbb{N}}$, con $\omega_{j:n}$ como se especificó anteriormente, entonces:*

$$\omega_{j:n} \longrightarrow \omega, \text{ con, } \omega \in (0, 1).$$

Demostración.

Dado $n \in \mathbb{N}$ fijo, considere el conjunto, $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$ y para que $\omega \in (0, 1)$, se tiene que existe k , tal que $\frac{k}{n} \leq \omega \leq \frac{k+1}{n}$, ahora, dado que el tamaño del intervalo $[\frac{k}{n}, \frac{k+1}{n}]$ es de $\frac{1}{n}$ se tiene que,

$$\omega - \frac{k}{n} < \frac{1}{n}, \quad \frac{k+1}{n} - \omega < \frac{1}{n},$$

$$\text{mín} \left\{ \omega - \frac{k}{n}, \frac{k+1}{n} - \omega \right\} < \frac{1}{n},$$

de lo anterior se puede inferir que

$$\omega_{j:n} = \begin{cases} \frac{k}{n} & , \text{ si } \omega - \frac{k}{n} \leq \frac{k+1}{n} - \omega \\ \frac{k+1}{n} & , \text{ si } \omega - \frac{k}{n} > \frac{k+1}{n} - \omega \end{cases}$$

en cualquiera de los dos casos se tiene, $|\omega_{j:n} - \omega| \leq \frac{1}{2n} \implies \omega_{j:n} \rightarrow \omega$. □

Volviendo al problema de encontrar $\mathbb{E}(I(\omega))$, si se considera para una frecuencia fundamental $\omega_{j:n}$, y se aplica las dos proposiciones anteriores, se tiene que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}(I(\omega_{j:n})) &= \lim_{n \rightarrow \infty} \sum_{|h| < n} \left(1 - \frac{|h|}{n}\right) \gamma(h) e^{-2\pi i \omega_{j:n} h} \\ &= \sum_{h \in \mathbb{Z}} \gamma(h) e^{-2\pi i \omega h} \\ &= f(\omega). \end{aligned}$$

Ya se encontró un estimador para la función de densidad espectral, lo que sigue es encontrar una distribución para el periodograma y obtener un intervalo de confianza para $f(\omega)$.

Definición 3.1.10. *Dados los datos, $\{x_1, x_2, \dots, x_n\}$, se define la transformada coseno y seno respectivamente como:*

$$\begin{aligned} d_c(\omega_j) &= n^{-\frac{1}{2}} \sum_{t=1}^n x_t \cos(2\pi \omega_j t), \\ d_s(\omega_j) &= n^{-\frac{1}{2}} \sum_{t=1}^n x_t \sin(2\pi \omega_j t). \end{aligned}$$

Observación 3.1.11. *La transformada discreta de fourier y el periodograma se pueden representar por medio de las transformadas senos y cosenos de la siguiente manera:*

$$d(\omega_j) = d_c(\omega_j) - i d_s(\omega_j).$$

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

Es posible determinar una distribución para el periodograma si se considera que las observaciones sigue una distribución normal, mas específicamente, si $\mathbf{x} \sim N(\mathbf{0}, \Gamma_n)$, donde, $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ y $(\Gamma_n)_{i,j} = \gamma(i - j)$, para $i, j = 1, 2, 3, \dots, n$.

A continuación se escribirá una matriz con vectores $a, b \in \mathbb{R}^n$ arbitrarios, para posteriormente tomar los vectores apropiados que sirvan para determinar la distribución del periodograma.

Si $A = \begin{pmatrix} a' \\ b' \end{pmatrix}$, donde $a = (a_1, a_2, \dots, a_n)'$, y $b = (b_1, b_2, \dots, b_n)'$,

$$A\mathbf{x} = \begin{pmatrix} a'\mathbf{x} \\ b'\mathbf{x} \end{pmatrix}, \quad y \quad A'\Gamma_n A = \begin{pmatrix} a'\Gamma_n a & a'\Gamma_n b' \\ b'\Gamma_n a & b'\Gamma_n b \end{pmatrix},$$

y así se tiene que,

$$\begin{pmatrix} a'\mathbf{x} \\ b'\mathbf{x} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a'\Gamma_n a & a'\Gamma_n b' \\ b'\Gamma_n a & b'\Gamma_n b \end{pmatrix} \right).$$

Tomando los vectores a, b de la siguiente forma:

$$\begin{aligned} a &= n^{-\frac{1}{2}}(\cos(2\pi\omega_j), \cos(4\pi\omega_j), \dots, \cos(2n\pi\omega_j))', \\ b &= n^{-\frac{1}{2}}(\sin(2\pi\omega_j), \sin(4\pi\omega_j), \dots, \sin(2n\pi\omega_j))', \end{aligned}$$

y utilizando la linealidad del valor esperado, se tiene que,

$$\begin{aligned} a'\Gamma_n a &= n^{-1} \sum_{s=1}^n \sum_{t=1}^n \cos(2\pi\omega_j s) \cos(2\pi\omega_j t) \gamma(s-t) \\ &= n^{-1} \sum_{s=1}^n \cos(2\pi\omega_j s) \sum_{t=1}^n \cos(2\pi\omega_j t) \mathbb{E}(x_s x_t) \\ &= \mathbb{E} \left(\left(n^{-\frac{1}{2}} \sum_{t=1}^n \cos(2\pi\omega_j t) x_t \right) \left(n^{-\frac{1}{2}} \sum_{s=1}^n \cos(2\pi\omega_j s) x_s \right) \right) \\ &= \mathbb{E}(d_c(\omega_j) d_c(\omega_j)) = Cov(d_c(\omega_j), d_c(\omega_j)). \end{aligned}$$

Análogamente $b'\Gamma_n b = Cov(d_s(\omega_j), d_s(\omega_j))$ y $b'\Gamma_n a = Cov(d_s(\omega_j), d_c(\omega_j))$.

Ahora para poder definir una distribución del periodograma, es necesario que

$$\Theta := \sum_{h \in \mathbb{Z}} |h| |\gamma(h)| < \infty.$$

Definición 3.1.12. El término acotado en probabilidad, denotado por $x_n = \mathcal{O}(a_n)$, significa que para todo $\epsilon > 0$, existe $\delta(\epsilon) > 0$ tal que

$$\mathbb{P} \left\{ \frac{|x_n|}{|a_n|} > \delta(\epsilon) \right\} \leq \epsilon$$

para todo n .

Teorema 3.1.13. Sea $S_n(\omega, \omega) = \mathbb{E}(|d(\omega)|^2)$ entonces,

$$|S_n(\omega, \omega) - f(\omega)| < \frac{\Theta}{n} \iff S_n(\omega, \omega) = f(\omega) + \mathcal{O}(n^{-1}).$$

Demostración. Ver [1](página 545.)

Teorema 3.1.14. Si $\omega_k = \frac{k}{n}$ y $\omega_l = \frac{l}{n}$ con $\omega_k \neq \omega_l$, y $S_n(\omega_k, \omega_l) = \mathbb{E}(d(\omega_k)\overline{d(\omega_l)})$,

$$|S_n(\omega_k, \omega_l)| < \frac{\Theta}{n} \iff S_n(\omega_k, \omega_l) = \mathcal{O}(n^{-1}).$$

Demostración. Ver [1](página 545.)

Afirmación 3.1.15. Con los dos teoremas anteriores se puede demostrar que,

$$\text{Cov}(d_c(\omega_k), d_c(\omega_l)) = \begin{cases} \frac{1}{2}f(\omega_k) + \mathcal{O}(n^{-1}) & , \text{ si } \omega_k = \omega_l \\ \mathcal{O}(n^{-1}) & , \text{ si } \omega_k \neq \omega_l \end{cases}$$

$$\text{Cov}(d_s(\omega_k), d_s(\omega_l)) = \begin{cases} \frac{1}{2}f(\omega_k) + \mathcal{O}(n^{-1}) & , \text{ si } \omega_k = \omega_l \\ \mathcal{O}(n^{-1}) & , \text{ si } \omega_k \neq \omega_l \end{cases}$$

$$\text{Cov}(d_s(\omega_k), d_s(\omega_l)) = \mathcal{O}(n^{-1}).$$

Reemplazando lo obtenido anteriormente en la matriz $A^T \Gamma_n A$, se tiene que

$$\begin{aligned} A^T \Gamma_n A &= \begin{pmatrix} a^T \Gamma_n a^T & a^T \Gamma_n b^T \\ b^T \Gamma_n a^T & b^T \Gamma_n b \end{pmatrix} \\ &= \begin{pmatrix} \text{Cov}(d_c(\omega_j), d_c(\omega_j)) & \text{Cov}(d_s(\omega_j), d_c(\omega_j)) \\ \text{Cov}(d_c(\omega_j), d_s(\omega_j)) & \text{Cov}(d_s(\omega_j), d_s(\omega_j)) \end{pmatrix} \\ &= \begin{pmatrix} \frac{f(\omega_j)}{2} + \mathcal{O}(n^{-1}) & \mathcal{O}(n^{-1}) \\ \mathcal{O}(n^{-1}) & \frac{f(\omega_j)}{2} + \mathcal{O}(n^{-1}) \end{pmatrix}, \end{aligned}$$

regresando al comienzo, se tiene que,

$$\begin{pmatrix} d_c(\omega_j) \\ d_s(\omega_j) \end{pmatrix} \sim N(\mathbf{0}, A^T \Gamma_n A),$$

y finalmente si se toma $\omega_{j:n} \rightarrow w$, se tiene la convergencia asintótica a una variable aleatoria normal,

$$\begin{pmatrix} d_c(\omega_{j:n}) \\ d_s(\omega_{j:n}) \end{pmatrix} \xrightarrow{d} N \left(\mathbf{0}, \frac{1}{2} \begin{pmatrix} f(\omega) & 0 \\ 0 & f(\omega) \end{pmatrix} \right).$$

Dado que $I(\omega_{j:n}) = d_c(\omega_{j:n})^2 + d_s(\omega_{j:n})^2$, entonces $\frac{2I(\omega_{j:n})}{f(\omega)} \xrightarrow{d} \chi^2(2)$. El siguiente teorema generaliza lo anterior para una colección de m frecuencias diferentes y da existencia de un intervalo de confianza para la densidad espectral.

Teorema 3.1.16. Si $\Theta = \sum_{h \in \mathbb{Z}} |h| |\gamma(h)| < \infty$,

$$x_t = \sum_{j \in \mathbb{Z}} \psi_j w_{t-j}, \quad \sum_{j \in \mathbb{Z}} |\psi_j| < \infty,$$

donde $w_t \sim iid(0, \sigma_w^2)$, entonces para una colección de m distintas frecuencias $\omega_j \in (0, 1/2)$ con $\omega_{j:n} \rightarrow \omega_j$

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} \chi^2(2),$$

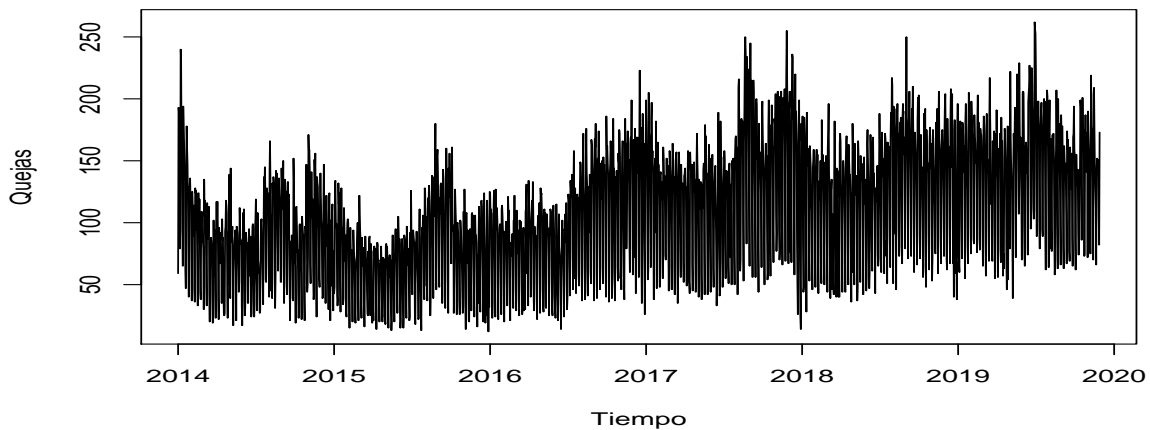
donde $f(\omega_j) > 0$ para $j = 1, 2, \dots, m$ y $\chi^2(2)$ es una distribución chi-cuadrado con 2 grados de libertad.

Demostración. Ver [1](Apéndice C.2).

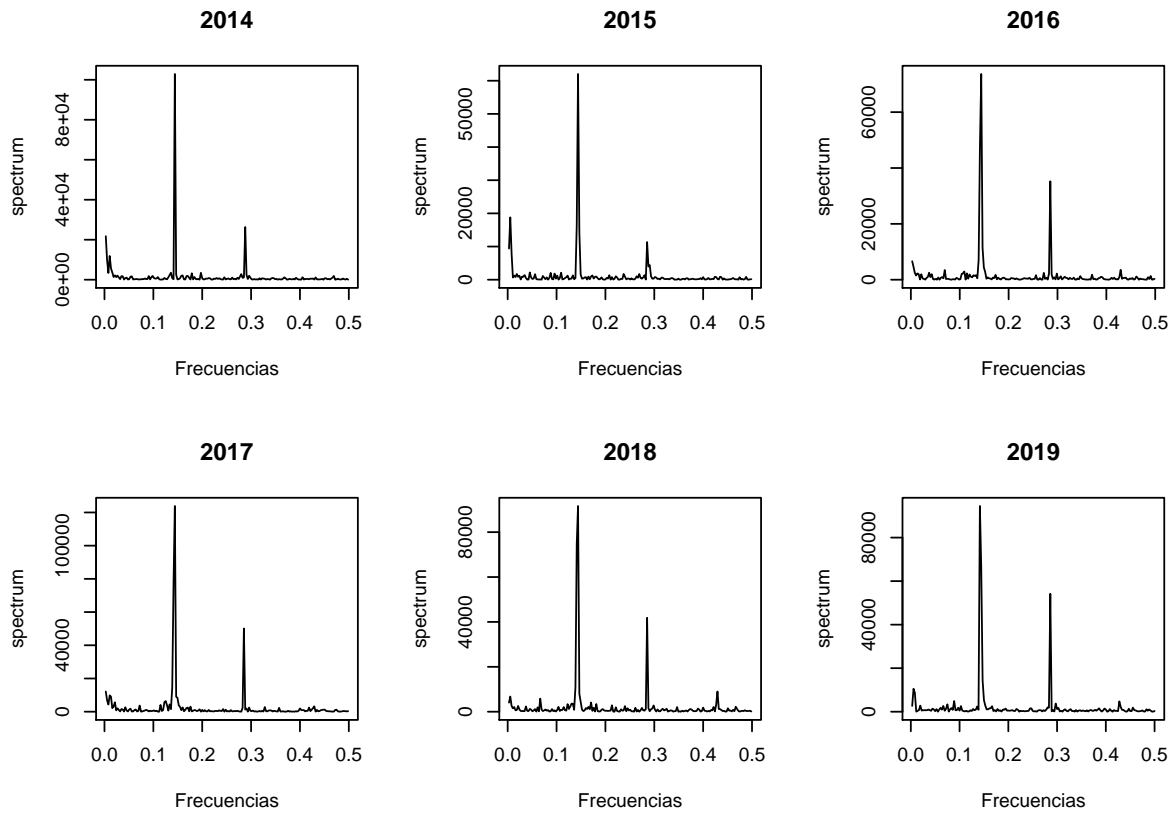
3.2. Aplicación

En la siguiente sección se presentará una serie de tiempo basada en el número de quejas diarias por parte de los usuarios de una entidad bancaria, utilizando la densidad espectral para identificar periodos ocultos en la serie y además se ajustará un modelo SARIMA el cual no solo captura los comportamientos regulares, sino también, el comportamiento estacional de la serie.

Una entidad bancaria registró diariamente desde el primero de enero del 2014 hasta el 2 de diciembre del 2019 el número de quejas recibidas por sus usuarios



En primer lugar se encontrará la frecuencia más aportante a la varianza utilizando el periodograma de la serie anual por medio del software R con la función `spec.pgram`

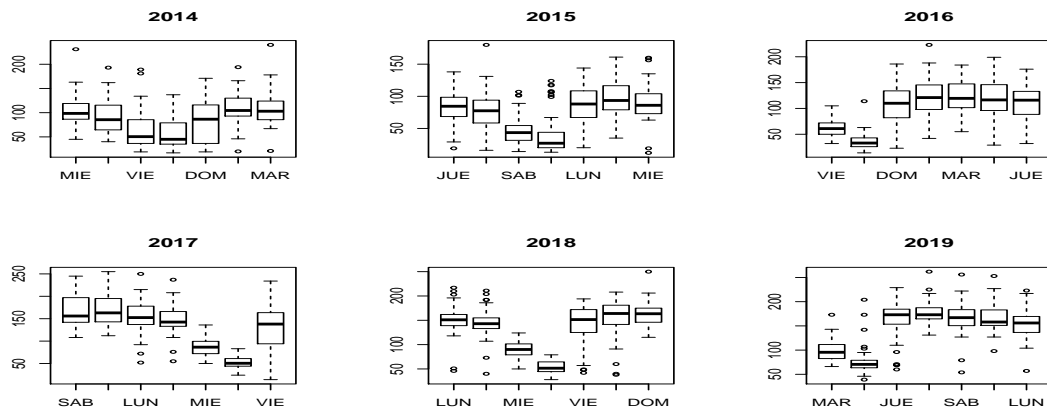


La frecuencia con la densidad espectral más alta es $\omega = 0,1430556$ y su periodo es

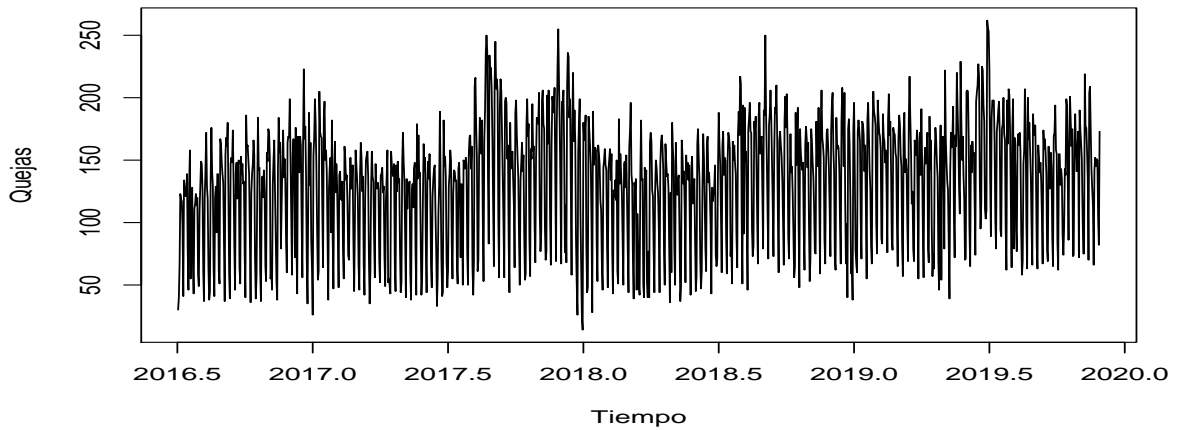
$$P = \frac{1}{\omega} = 6,990289 \approx 7,$$

lo que indica que la serie tiene un comportamiento estacional por semanas.

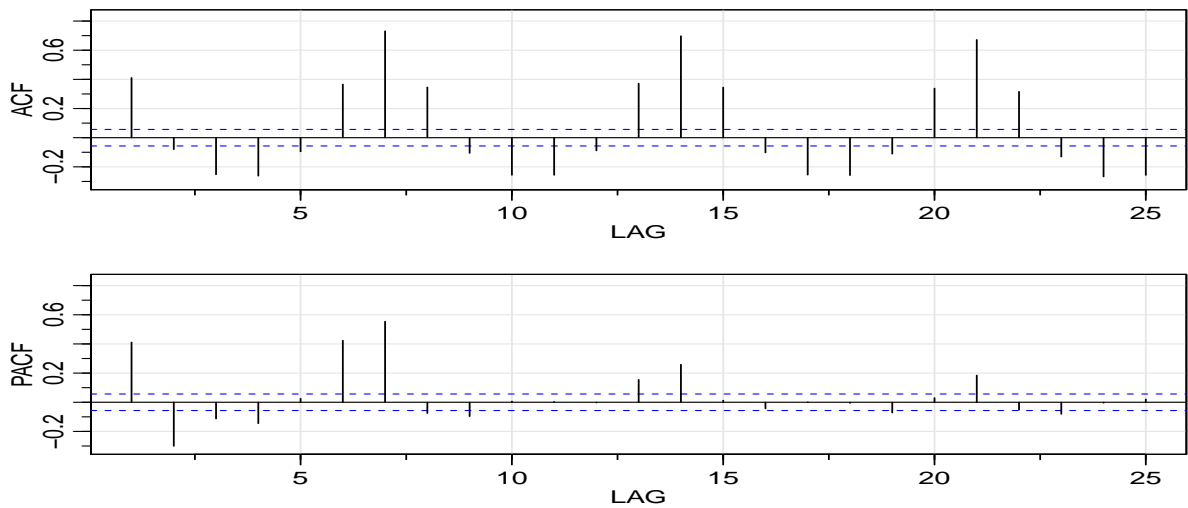
A continuación podemos ver la distribución de los días por año que presentan un comportamiento senoidal



La serie tuvo una intervención a mediados del 2016 debido a un nuevo software para el monitoreo de las transacciones, lo cual aumentó el número de usuarios y por ende un aumento de la media, por lo tanto se estudiará la serie a partir de julio del 2016.



Con ACF y PACF dadas por



se observa una fuerte correlación con rezagos múltiplos de 7 y por tanto se considerará el proceso SARIMA con los operadores (2.1.4), tomando el periodo estacional $s = 7$

$$\Phi_P(B^7)\phi(B)x_t = \Theta_Q(B^7)\theta(B)w_t,$$

este proceso es denotado por $ARIMA(p, 0, q) \times (P, 1, Q)_7$ y el objetivo es encontrar los ordenes P, p, Q, q que se ajusten mejor a la serie. Basado en el comportamiento de la

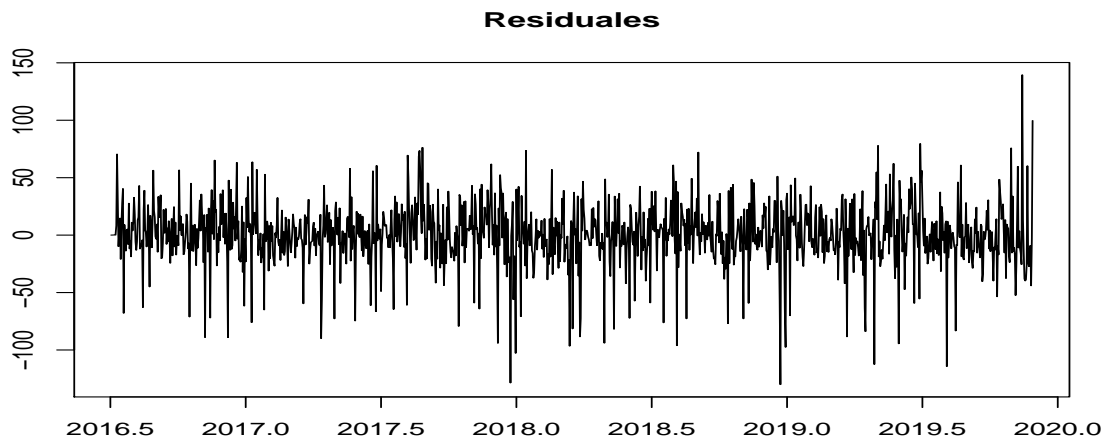
ACF y la PACF los modelos tentativos fueron

$$\begin{aligned}
 &ARIMA(1, 0, 2) \times (0, 1, 2)_7 \\
 &ARIMA(1, 0, 1) \times (1, 1, 1)_7 \\
 &ARIMA(3, 0, 0) \times (4, 1, 0)_7 \\
 &ARIMA(4, 0, 0) \times (3, 1, 0)_7 \\
 &ARIMA(3, 0, 1) \times (3, 1, 0)_7
 \end{aligned}$$

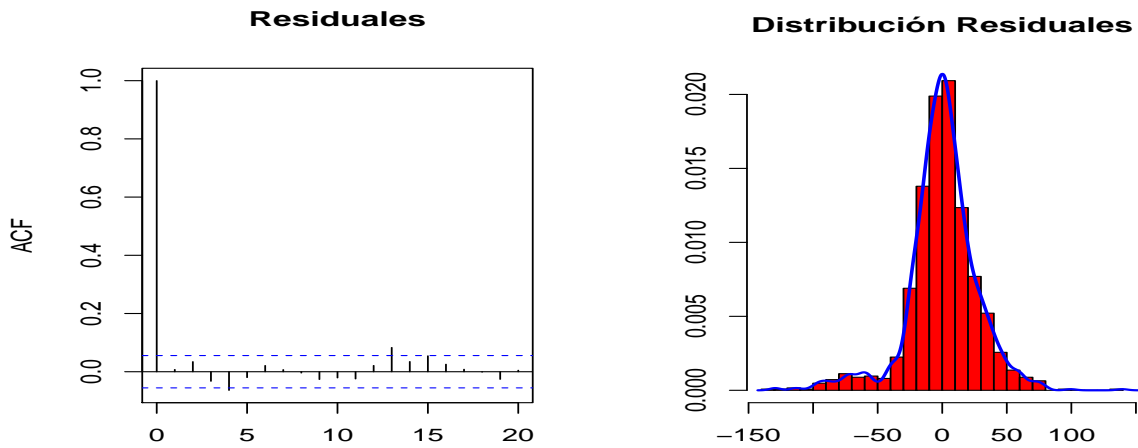
En cada uno de los modelos se realizará un diagnostico a los residuales buscando simetría respecto al cero y la no correlación, para ello se utilizará la función de autocorrelación y el histograma de los residuales, además de lo anterior se escogerá el modelo con menor criterios de información de Akaike conocido como AIC y la capacidad de predicción empleando el error porcentual medio absoluto abreviado como **mape**, realizando el análisis con el 95 % y el 80 % de la serie para dejar el 5 % y 20 % respectivamente como la validación del modelo.

| MODELO | AIC | BIC | MAPE 95 % | MAPE 80 % |
|-------------------------------------|--------|--------|--------------|--------------|
| $ARIMA(1, 0, 2) \times (0, 1, 2)_7$ | 9.4255 | 9.4542 | 0.1943 | 0.1981 |
| $ARIMA(1, 0, 1) \times (1, 1, 1)_7$ | 9.4507 | 9.4753 | 0.1965 | 0.1961 |
| $ARIMA(3, 0, 0) \times (4, 1, 0)_7$ | 9.5616 | 9.5986 | 0.1664 | 0.2225 |
| $ARIMA(4, 0, 0) \times (3, 1, 0)_7$ | 9.5713 | 9.6083 | 0.1656 | 0.2250 |
| $ARIMA(3, 0, 1) \times (3, 1, 0)_7$ | 9.5717 | 9.6087 | 0.1655 | 0.2217 |

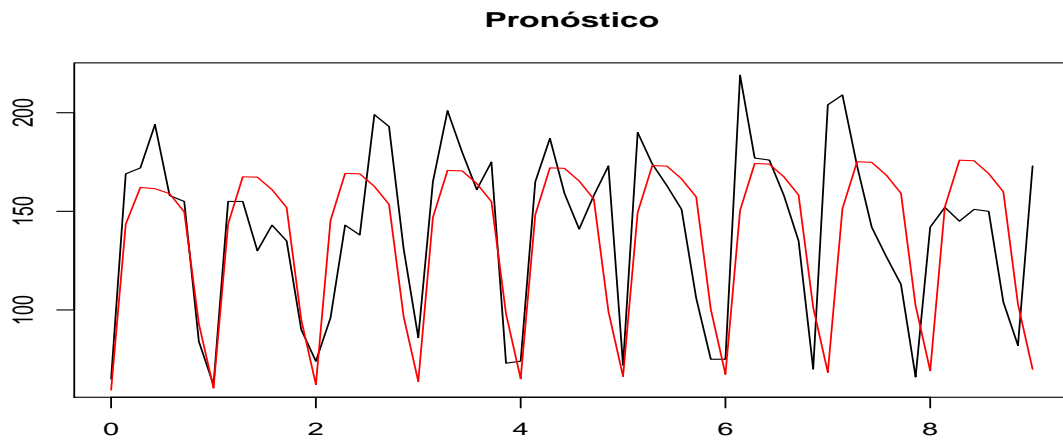
Para esta aplicación se escogió el modelo $ARIMA(1, 0, 2, 0, 1, 2)_7$, con residuales dados por



el histograma de los residuos muestra simetría al rededor del cero, además la función de autocorrelación muestra que no hay correlación aparente.



Al realizar el pronóstico con el 5% de la serie se puede observar el comportamiento de los próximos 2 meses donde el modelo captura completamente los ciclos semanales, aunque no logra capturar los datos atípicos puesto que la suposición que la varianza sea constante no se cumple, para solucionar este problema se realizó la transformación de Box-Cox para estabilizar la varianza pero al hacer esto el modelo pierde precisión y por tanto se optó por dejar la serie original.



| CUARTA SEMANA | | | | | | | |
|---------------|-------|--------|--------|--------|--------|--------|-------|
| Pronóstico | 63.74 | 146.91 | 170.72 | 170.49 | 164.08 | 154.96 | 98.02 |
| Real | 86 | 165 | 201 | 180 | 161 | 175 | 73 |

Capítulo 4

Wavelets

En 1800 Joseph Fourier descubrió que podía superponer funciones senoidales y cosenoidales para representar otras funciones y hasta el día de hoy ha sido una herramienta poderosa en el análisis de datos y filtrado de información, a pesar de tener varias limitaciones al solo considerar funciones senoidales y cosenoidales las cuales tiene energía infinita es imposible una localización temporal de la onda en un momento determinado. El análisis de wavelets u ondaletas se basa en la idea de construir un sistema ortonormal a partir de las funciones ϕ y ψ , con un soporte compacto, los cuales van a tener ubicación tiempo frecuencia, además, a partir de la función ϕ se crea una secuencia de espacios vectoriales $V_0 \subseteq V_1 \cdots \subseteq V_j, \dots$, donde V_j es generado por traslaciones y dilataciones de dicha función.

En este capítulo se exhibirán dos wavelets utilizadas frecuentemente en el análisis de datos, una de ellas es la wavelets de Haar propuesta en 1909 por Alfred Haar, la cual es la dilatación y traslación de la función indicadora del intervalo $[0, 1)$. La wavelets Daubechies que fue propuesta por Ingrid Daubechies descrita en el artículo de 1988 [9], lleva a una familia de funciones de escala con soporte compacto y suaves, pero que a diferencia de la wavelets de Haar que tiene una formula cerrada para cada una de las funciones de escala y facilita la proyección de una función de $L_2(\mathbb{R})$ en el espacio generado por estas funciones, las wavelets Daubechies no tienen una representación de esta forma y por lo cual es necesario encontrarlas mediante el algoritmo cascada que brinda una rutina para obtener una buena aproximación de la función ϕ .

Definición 4.0.1. Una colección $(V_j)_{j \in \mathbb{Z}} \subseteq L_2(\mathbb{R})$ se dice que es un análisis de multi-resolución si:

1. $V_j \subseteq V_{j+1}$.
2. $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R})$.
3. $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$.
4. $f(t) \in V_0 \iff f(2^j t) \in V_j$.

Además, existe $\phi(t) \in V_0$ tal que $\int_{\mathbb{R}} \phi(t) dt \neq 0$, llamada función de escala tal que el conjunto $\{\phi(t - k)\}_{k \in \mathbb{Z}}$ es base ortonormal de V_0 .

Un análisis de multiresolución nos da una buena manera de descomponer $L_2(\mathbb{R})$ teniendo espacios de aproximación anidados V_j , y por la propiedad (4) de la definición anterior es posible ir de un espacio a otro por medio de un reescalado de la función, además, es posible utilizar $\phi(t)$ para crear bases para V_0, V_1, \dots, V_j y por la propiedad (3) estas bases no son redundantes.

4.1. Wavelet de Haar

Para poder definir el espacio de Haar se considera la función de escala $\phi(t) = \mathbb{1}_{[0,1]}$ y en esta sección se mostrará que induce un análisis de multinivel.

$$\mathbb{1}_{[0,1]}(t) = \begin{cases} 1 & , \text{ si } t \in [0, 1) \\ 0 & , \text{ si } t \notin [0, 1) \end{cases}$$

La función indicadora $\mathbb{1}_{[0,1]}$, definirá un espacio vectorial constante en ciertos intervalos y discontinua posiblemente en los enteros, una tentativa es definir el espacio V_0 como el generado de las traslaciones de $\phi(t) = \mathbb{1}_{[0,1]}$, esto es,

$$f \in V_0 \iff f = \sum_{k \in \mathbb{Z}} a_k \phi(t - k), \quad a_k \in \mathbb{R}.$$

Es importante notar que la serie está indefinida y por tanto tiene sentido que darle una restricción para obtener una suma hasta un $N > 0$ dado, y obtener un espacio con soporte compacto.

Definición 4.1.1. Sea $\phi(t) = \mathbb{1}_{[0,1]}$, definimos el espacio V_0 como

$$V_0 = \text{span}\{\phi(t - k)\}_{k \in \mathbb{Z}} \cap L^2(\mathbb{R}).$$

Observaciones 4.1.2. Dado un entero un entero k ,

$$\phi(t - k) = \mathbb{1}_{[k, k+1)},$$

además, para cualquier par de enteros diferentes k, s , se tiene que

$$[k, k + 1) \cap [s, s + 1) = \emptyset,$$

y así $(\phi(t - k))_{k \in \mathbb{Z}}$ es un conjunto linealmente y por tanto es base de V_0 , más aun,

$$\begin{aligned} \langle \phi(t - k), \phi(t - j) \rangle &= \int_{\mathbb{R}} \phi(t - k) \phi(t - j) dt \\ &= \int_{\mathbb{R}} \mathbb{1}_{[k, k+1)} \mathbb{1}_{[j, j+1)}(t) dt \\ &= \begin{cases} 1 & , \text{ si } k = j \\ 0 & , \text{ si } k \neq j \end{cases}. \end{aligned}$$

El objetivo es construir funciones definidas a tramos en diferentes resoluciones, es decir, ya no construir funciones constantes en intervalos de la forma $[k, k+1)$, sino, por ejemplo funciones constantes en intervalos de la forma, $\left[\frac{k}{2}, \frac{k+1}{2}\right)$, para ello notemos que

$$\begin{aligned}\phi(2t) &= \begin{cases} 1 & , \text{si } 0 \leq 2t < 1 \\ 0 & , \text{e.o.c} \end{cases} \\ &= \begin{cases} 1 & , \text{si } 0 \leq t < \frac{1}{2} \\ 0 & , \text{e.o.c} \end{cases} \\ &= \mathbb{1}_{[0, \frac{1}{2})},\end{aligned}$$

$$\begin{aligned}\phi(4t) &= \begin{cases} 1 & , \text{si } 0 \leq 4t < 1 \\ 0 & , \text{e.o.c} \end{cases} \\ &= \begin{cases} 1 & , \text{si } 0 \leq t < \frac{1}{4} \\ 0 & , \text{e.o.c}, \end{cases} \\ &= \mathbb{1}_{[0, \frac{1}{4})}.\end{aligned}$$

En general para cualquier dilatación de la función indicadora se puede definir el espacio vectorial generado por sus traslaciones.

Definición 4.1.3. Sea $\phi(t) = \mathbb{1}_{[0,1)}$, se define el espacio vectorial V_j como

$$V_j = \text{span}\{\phi(2^j t - k)\} \cap L_2(\mathbb{R}). \quad (4.1)$$

Observación 4.1.4.

$$\begin{aligned}\phi(2^j t - k) &= \begin{cases} 1 & , \text{si } \frac{k}{2^j} \leq t < \frac{k+1}{2^j} \\ 0 & , \text{e.o.c} \end{cases} \\ &= \mathbb{1}_{\left[\frac{k}{2^j}, \frac{k+1}{2^j}\right)},\end{aligned}$$

$$\left\langle \phi(2^j t - k), \phi(2^j t - l) \right\rangle = \begin{cases} 2^{-j} & , \text{si } l = k \\ 0 & , \text{e.o.c} \end{cases}$$

Para cada V_j definido en la ecuación (4.1) y teniendo en cuenta la observación (4.1.4), es posible definir una base ortonormal para V_j .

Definición 4.1.5. Suponga que $\phi(t)$ es la función de escala del análisis de multiresolución $(V_j)_{j \in \mathbb{Z}}$ de $L_2(\mathbb{R})$, para $k \in \mathbb{Z}$ se define

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k).$$

Si se fija $j \in \mathbb{Z}$, el conjunto $(\phi_{j,k}(t))_{k \in \mathbb{Z}}$ es una base ortonormal para V_j con soporte compacto $\text{Supp}(\phi_{j,k}) = [\frac{k}{2^j}, \frac{k+1}{2^j}]$.

Ahora se quiere mostrar que para cada $j \in \mathbb{Z}$, $V_j \subseteq V_{j+1}$, para ello se puede notar que

$$[k, k+1) = \left[k, k + \frac{1}{2} \right) \cup \left[k + \frac{1}{2}, k + 1 \right),$$

y por tanto,

$$\begin{aligned} \mathbb{1}_{[k,k+1)} &= \mathbb{1}_{[k,k+\frac{1}{2}) \cup [k+\frac{1}{2},k+1)} \\ &= \mathbb{1}_{[k,k+\frac{1}{2})} + \mathbb{1}_{[k+\frac{1}{2},k+1)} \\ &= \mathbb{1}_{[\frac{2k}{2}, \frac{2k+1}{2})} + \mathbb{1}_{[\frac{2k+1}{2}, \frac{2k+1+1}{2})}, \end{aligned}$$

$$\phi(t - k) = \phi(2t - 2k) + \phi(2t - (2k + 1))$$

$$\phi_{0,k}(t) = \frac{1}{\sqrt{2}} (\phi_{1,2k}(t) + \phi_{1,2k+1}(t)).$$

En general para cada par de enteros j, k , la ecuación de dilatación esta dada por

$$\phi_{j-1,k}(t) = \frac{1}{\sqrt{2}} (\phi_{j,2k}(t) + \phi_{j,2k+1}(t)), \quad (4.2)$$

y dada la ecuación (4.2) se tiene que $V_{j-1} \subseteq V_j$, y servirá para ir de un espacio a otro por medio de la siguiente dilatación,

$$f(t) \in V_j \iff f(2t) \in V_{j+1}.$$

Corolario 4.1.6. Los espacios V_j generados por las dilataciones y traslaciones de la función $\phi(t) = \mathbb{1}_{[0,1)}(t)$ cumplen que

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}.$$

Demostración. Procediendo por doble inclusión,

(\supseteq) V_j Es un espacio vectorial para cada $j \in \mathbb{Z}$ entonces $0 \in V_j$ para cada $j \in \mathbb{Z}$, esto es que $0 \in \bigcap_{j \in \mathbb{Z}} V_j$.

(\subseteq) Sea $f \in \bigcap_{j \in \mathbb{Z}} V_j$, entonces $f \in V_j$ para cada $j \in \mathbb{Z}$, y por tanto, f es constante en intervalos de la forma $[\frac{k}{2^j}, \frac{k+1}{2^j})$, ya que $[\frac{k}{2^j}, \frac{k+1}{2^j}) \cap [\frac{k}{2^{j+1}}, \frac{k+1}{2^{j+1}}) \neq \emptyset$, se tiene que $f(t) = a$,

con $a \in \mathbb{R}$, además, $f \in L_2(\mathbb{R})$, esto es $\int_{\mathbb{R}} f(t)dt < \infty$, y así la única opción es $a = 0$. \square

Ya se logró definir un análisis de multiresolución con $\mathbb{1}_{[0,1)}$, ahora se mostrará la relación que existe entre V_0 y V_1 .

Proposición 4.1.7. *Supongamos que $f_1(t) \in V_1$, $f_1(t) = \sum_{k \in \mathbb{Z}} a_k \phi_{1,k}(t)$, entonces la proyección de f_1 en V_0 está dada por $f_0 = \sum_{k \in \mathbb{Z}} b_k \phi_{0,k}(t)$, donde*

$$b_k = \frac{\sqrt{2}}{2}(a_{2k} + a_{2k+1}), \text{ para todo } k \in \mathbb{Z}.$$

Demostración.

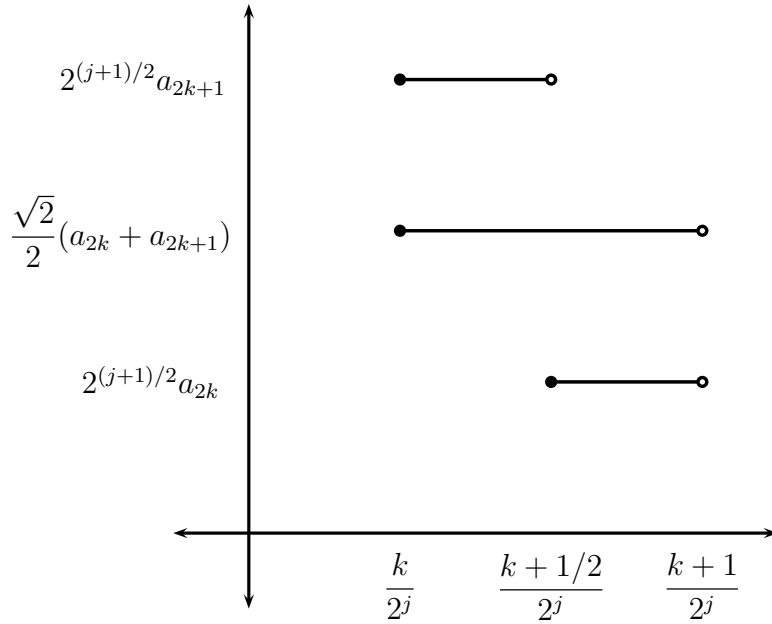
El coeficiente b_k de la proyección está dada por $b_k = \langle \phi_{0,k}, f_1 \rangle$,

$$\begin{aligned} b_k &= \langle \phi_{0,k}, f_1 \rangle = \int_{\mathbb{R}} \phi(t-k) f_1(t) dt \\ &= \int_k^{k+1} f_1(t) dt \\ &= \int_k^{k+1} \sum_{s \in \mathbb{Z}} a_s \phi_{1,s}(t) dt \\ &= \sum_{s \in \mathbb{Z}} \int_k^{k+1} a_s \phi_{1,s}(t) dt \\ &= \sum_{s \in \mathbb{Z}} \int_k^{k+1} a_s \frac{1}{\sqrt{2}} \mathbb{1}_{[\frac{s}{2}, \frac{s+1}{2})}(t) dt, \end{aligned}$$

puesto que $[k, k+1) \cap [\frac{s}{2}, \frac{s+1}{2}) \neq \emptyset$ solamente cuando $s = 2k$ y $s = 2k+1$,

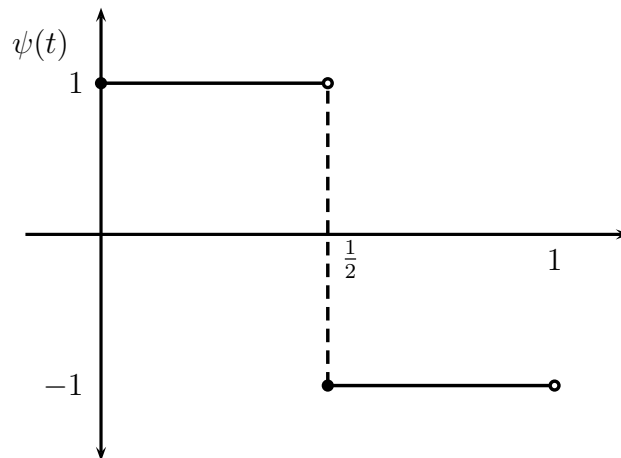
$$\begin{aligned} b_k &= \sqrt{2} a_{2k} \int_k^{k+1} \mathbb{1}_{[k, \frac{2k+1}{2})}(t) dt + \sqrt{2} a_{2k+1} \int_k^{k+1} \mathbb{1}_{[\frac{2k+1}{2}, k+1)}(t) dt \\ b_k &= \frac{\sqrt{2}}{2}(a_{2k} + a_{2k+1}). \end{aligned}$$

\square



Hasta el momento se ha construido una función $f_0(t)$ como la aproximación sobre V_0 para una función $f_1(t)$ en V_1 , supongamos que el objetivo es el contrario, es decir, se nos entregan $f_0(t)$, $f_1(t)$ y debemos encontrar una función residuo $g_0(t)$ tal que $f_0(t) + g_0(t) = f_1(t)$, para resolver este problema es necesario descomponer a $V_{j+1} = V_j \oplus W_j$, donde W_j es el complemento ortogonal de V_j sobre V_{j+1} y encontremos las funciones que generan a W_j , para esto se define la siguiente función,

$$\psi(t) = \phi(2t) - \phi(2t - 1) = \begin{cases} 1 & , \text{si } 0 \leq t < \frac{1}{2} \\ -1 & , \text{si } \frac{1}{2} \leq t < 1 \\ 0 & , \text{e.o.c} \end{cases}$$



Definición 4.1.8. Sea $\psi(t)$ la función introducida anteriormente, se define el espacio de detalle W_0 como

$$W_0 = \text{span}\{\psi(t - k)\}_{k \in \mathbb{Z}} \cap L_2(\mathbb{R}).$$

Observaciones 4.1.9.

- $(\psi(t - k))_{k \in \mathbb{Z}}$ es un conjunto linealmente independiente, más aún, es una base ortonormal para W_0 .
- W_0 es el complemento ortogonal de V_0 sobre V_1 , es decir, $V_1 = V_0 \oplus W_0$, donde esto significa que

$$V_0 \oplus W_0 = \{f_0(t) + g_0(t) : f_0 \in V_0, g_0 \in W_0\} \text{ y } V_0 \cap W_0 = \{0\}.$$

Esta idea se puede generalizar, considerando una dilatación de la función $\psi(t)$.

Definición 4.1.10. Se define para cada $j, k \in \mathbb{Z}$, las funciones

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k),$$

y el espacio vectorial W_j , como

$$W_j = \text{span}\{\psi_{j,k}\}_{k \in \mathbb{Z}} \cap L_2(\mathbb{R}).$$

Observaciones 4.1.11.

- $(\psi_{j,k})_{k \in \mathbb{Z}}$ es una base ortonormal para W_j .
- $\overline{\text{Supp}}(\psi_{j,k}) = \left[\frac{k}{2^j}, \frac{k+1}{2^j} \right]$.
- W_j es el complemento ortogonal de V_j sobre V_{j+1} .
- $V_{j+1} = V_j \oplus W_j = V_{j-1} \oplus W_{j-1} \oplus W_j = \dots = V_0 \bigoplus_{k=0}^j W_k$.

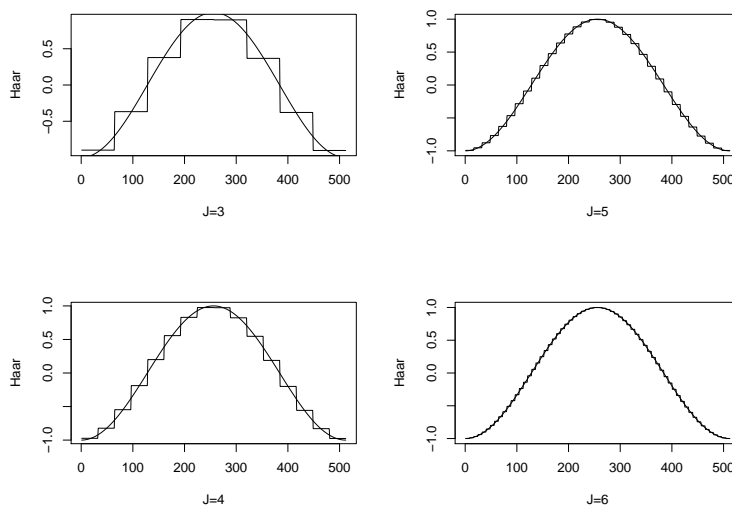
Por lo tanto cualquier función $a(u) \in L_2[0, 1]$ se puede aproximar a un nivel de resolución V_J por medio de

$$a(u) \approx \alpha_{00} \phi(u) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(u),$$

donde los coeficientes dependen del nivel de resolución, más adelante se buscará un valor apropiado para J .

Ejemplo 4.1.12. Realizar la aproximación de la siguiente función usando las wavelets Haar para los niveles de resolución $J = 3, 4, 5, 6$,

$$f(t) = \begin{cases} -\cos\left(\frac{2\pi t}{512}\right) & , \text{ si } 1 \leq t \leq 512 \\ 0 & , \text{ e.o.c} \end{cases}$$



4.2. Wavelet Daubechies

Ingrid Daubechies realizó una construcción de funciones continuas con soporte compacto cuyo método se centra en construir el símbolo de la función de escala ϕ , el cual a su vez produce la ecuación de dilatación y la función ϕ , para esto el símbolo deberá cumplir ciertas condiciones para garantizar la existencia del análisis de multiresolución.

Con las wavelets de Haar para ir de un espacio V_{j-1} al V_j se escribían los básicos de V_{j-1} en términos de los básicos de V_j , y en este caso la ecuación de dilatación es de la forma $\phi(t) = \frac{1}{\sqrt{2}}(\phi(2t) + \phi(2t - 1))$, en el caso de Daubechies se hace un análisis en el dominio de las frecuencias y por tanto, la ecuación de dilatación deberá cumplir que

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right), \quad \text{donde} \quad H(\omega) = \frac{1}{\sqrt{2}} \sum_{k=0}^L h_k e^{-ik\omega}.$$

Sí $|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1$, entonces se garantiza que ϕ genera un sistema ortogonal, además, se quiere que ϕ sea suficientemente suave, es decir, $\phi \in C^{N-1}(\mathbb{R})$, para esto Ingrid Daubechies observó que $H(\omega)$ debía tener a $\left(\frac{1+e^{-ik\omega}}{2}\right)^N$ como factor del polinomio, [10], bajo estas condiciones es posible construir $H(\omega)$, el método para construirla se puede consultar en [12].

Notación 4.2.1. La función de escala Daubechies que corresponde al símbolo $H(\omega)$ para un N dado se denotará por D_{2N} .

Ya que no existe una formula para las funciones de escala, existe una rutina llamada algoritmo de cascada introducida por Ingrid y Jeffrey Lagarias en [11]; este algoritmo inicia con una primera función ϕ_0 y se itera mediante una ecuación para obtener una

mejor aproximación $\phi_n(t)$, hasta obtener una buena aproximación de ϕ , el algoritmo utiliza la ecuación de dilatación de la siguiente manera

$$\phi_{n+1}(t) = \sqrt{2} \sum_{k=0}^M h_k \phi_n(2t - k).$$

Dado que no existe una representación explícita, éste es el único medio para tener una aproximación de $\phi(t)$ y por tanto dependiendo de la función inicial y el símbolo $H(\omega)$ se generan diferentes funciones de escala. Para más información del algoritmo y la convergencia de ϕ_n se puede consultar en, [10],[12].

Ejemplo 4.2.2. Usar el algoritmo de cascada para graficar la función de escala $\phi(t)$ de la wavelet D_4 , usando $\phi_0(t) = \mathbf{1}_{[0,1]}$.

- Se encuentra $H(\omega)$ por medio del método utilizado en [12],

$$\begin{aligned} h_0 &= \frac{1 + \sqrt{3}}{4\sqrt{2}}, & h_1 &= \frac{3 + \sqrt{3}}{4\sqrt{2}}, \\ h_2 &= \frac{3 - \sqrt{3}}{4\sqrt{2}}, & h_3 &= \frac{1 - \sqrt{3}}{4\sqrt{2}}. \end{aligned}$$

- Aplicando el algoritmo cascada,

$$\phi_{n+1}(t) = \sqrt{2} \sum_{k=0}^3 h_k \phi_n(2t - k),$$

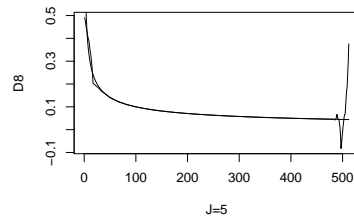
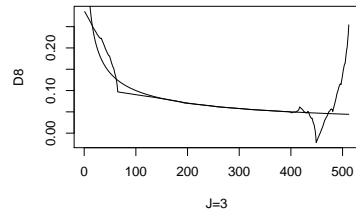
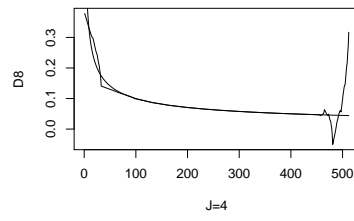
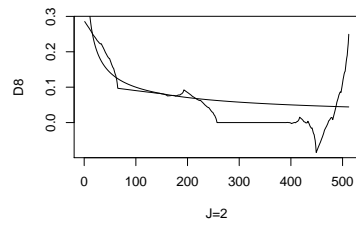
- $\phi_1(t) = \sqrt{2}h_0\phi_0(2t) + \sqrt{2}h_1\phi_0(2t - 1) + \sqrt{2}h_2\phi_0(2t - 2) + \sqrt{2}h_3\phi_0(2t - 3).$
- $\phi_1(t) = \sqrt{2}h_0\mathbf{1}_{[0, \frac{1}{2})} + \sqrt{2}h_1\mathbf{1}_{[\frac{1}{2}, 1)} + \sqrt{2}h_2\mathbf{1}_{[1, \frac{3}{2})} + \sqrt{2}h_3\mathbf{1}_{[\frac{3}{2}, 2)}.$
- $\phi_2(t) = \sqrt{2}h_0\phi_1(2t) + \sqrt{2}h_1\phi_1(2t - 1) + \sqrt{2}h_2\phi_1(2t - 2) + \sqrt{2}h_3\phi_1(2t - 3).$
- $\phi_2(t) = 2h_0^2\mathbf{1}_{[0, \frac{1}{4})} + 2h_0h_1\mathbf{1}_{[\frac{1}{2}, \frac{3}{4})} + \dots + 2h_3h_2 + 2h_3^2\mathbf{1}_{[\frac{9}{4}, 3)}.$

De esta manera recursiva es como se obtiene una aproximación $\phi_n(t)$ para la función de escala $\phi(t)$ de las wavelets D_4 y con esta función ϕ_n es posible generar un análisis de multiresolución, de la misma forma que en el caso de la wavelet de Haar.

En la práctica se utiliza la wavelet de Haar cuando la función a analizar es abrupta y tiene muchos saltos, y la wavelets Daubechies se usa cuando se quiere una aproximación mas suave.

Ejemplo 4.2.3. Realizar la aproximación de la siguiente función usando las wavelets Daubechies D_8 , a los niveles de resolución $J = 2, 3, 4, 5$,

$$f(t) = \begin{cases} t^{-\frac{1}{2}} & , si \ 1 \leq t \leq 512 \\ 0 & , si \ e.o.c \end{cases}$$



Capítulo 5

Proceso Autorregresivo con Coeficientes Variando en el Tiempo $tvAR(p)$

La estacionariedad siempre ha jugado un papel muy importante en el tratamiento de series de tiempo, por ejemplo, los procesos ARMA son utilizados cuando la serie tiene un carácter periódico o estacionario, además, el concepto de estacionariedad es la base de la teoría asintótica general; la cual garantiza que el aumento del tamaño de la muestra conduce a más información de la misma, pero lastimosamente muchas muestras no describen un comportamiento estacionario. Si se abandona este supuesto, los posibles modelos que se pueden generar son demasiados, por ejemplo, los procesos autorregresivos con coeficientes variando en el tiempo que son una clase de procesos no estacionarios donde los coeficientes autorregresivos son funciones $a_i(\cdot) \in L_2(\mathbb{R})$.

El objetivo del presente capítulo es poder estimar el comportamiento de los coeficientes autorregresivos si se tienen observaciones $\{x_1, x_2, \dots, x_T\}$ de un proceso estocástico arbitrario no estacionario, el problema es que si la serie x_t no es estacionaria y posteriormente se hace tender $T \rightarrow \infty$, es decir, extender el proceso hacia el futuro no brindará información del comportamiento del proceso al comienzo del intervalo de tiempo, es por esto, que se necesita una noción asintótica diferente.

Ejemplo 5.0.1. *Supongamos la existencia del siguiente proceso,*

$$x_t = a(t)x_{t-1} + w_t, \quad w_t \sim Wn(0, \sigma_w^2),$$

para $t = 1, 2, \dots, T$.

El objetivo es estimar la función $a(\cdot)$ definida en el intervalo $[1, T]$ de la cual solo tenemos información en los tiempos $\{1, 2, \dots, T\}$, la metodología será estudiar $a(\cdot)$ en cuadrículas más finas pero en el mismo intervalo, para ello se considera el siguiente

proceso reescalado.

$$x_{t,T} = a\left(\frac{t}{T}\right) x_{t-1,T} + w_t, \quad \text{para } t = 1, 2, \dots, T.$$

Los procesos autorregresivos con coeficientes variando en el tiempo no son estacionarios, pero entran en la clase de proceso localmente estacionarios definido en [2] donde se garantiza la existencia de dicho proceso.

Definición 5.0.2.¹ Una secuencia de procesos estocásticos $x_{t,T}$, para $t = 1, 2, \dots, T$, es llamado un proceso localmente estacionario con función de transferencia A^0 y media μ si existe una representación

$$x_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp\{i\lambda t\} A_{t,T}^0 d\xi(\lambda),$$

que cumple :

1. $\xi(\lambda)$ es un proceso estocástico definido en $[\pi, \pi]$ con $\overline{\xi(\lambda)} = \xi(-\lambda)$ y el cumulante de orden k , $\text{cum}(d\xi(\lambda_1), \dots, d\xi(\lambda_k)) = \eta\left(\sum_{j=1}^k \lambda_j\right) g_k(\lambda_1, \dots, \lambda_{k-1}) d\lambda_1 \cdots d\lambda_k$, donde $\eta(\lambda) = \sum_{k \in \mathbb{Z}} \delta(\lambda + 2\pi k)$ es la extensión 2π periódica de la función delta de Dirac.
2. Existe una constante K y una función 2π periódica $A : [0, 1] \times \mathbb{R} \rightarrow \mathbb{C}$ tal que $A(u, -\lambda) = \overline{A(u, \lambda)}$ y

$$\sup_{t, \lambda} \left| A_{t,T}^0(\lambda) - A\left(\frac{t}{T}, \lambda\right) \right| \leq KT^{-1}$$

para todo T .

$A(u, \lambda)$ y μ se asumen continua en u , además, se tiene que $|g_k(\lambda_1, \dots, \lambda_{k-1})| < C_k$ donde C_k una constante que depende de k .

El proceso autorregresivo con coeficientes variando en el tiempo fue estudiado por Subba Rao (1970), Grenier (1983), Hallin (1978) y Herteleer-de Schutter (1989).

Para simplificar las cuentas sea $\mu = 0$ y $x_{t,T}$ la solución al sistema de ecuaciones en diferencia

$$\sum_{j=0}^p a_j\left(\frac{t}{T}\right) x_{t-j,T} = \sigma\left(\frac{t}{T}\right) w_t,$$

donde $a_0(u) \equiv 1$, $w_t \sim Wn(0, 1)$,

$$a_j(u) = \begin{cases} a_j(0) & , si \ u < 0 \\ a_j(1) & , si \ u > 1 \end{cases}, \quad \sigma(u) = \begin{cases} \sigma(0) & , si \ u < 0 \\ \sigma(1) & , si \ u > 1 \end{cases}$$

¹Para mayor información [3].

Observación 5.0.3. *La densidad espectral de un proceso autorregresivo con coeficientes variando en el tiempo está dada por:*

$$f(u, \lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^p a_j(u) \exp(i\lambda j) \right|^{-2}.$$

5.1. Proceso de Estimación

En la sección anterior se introdujo el concepto de proceso autorregresivo con coeficientes variando en el tiempo y se reescala el proceso al intervalo $[0, 1]$ para poder analizar la serie en el mismo intervalo de tiempo. Por lo general las funciones $a_i(\cdot)$ son desconocidas y el objetivo de esta sección es que a partir de un proceso autorregresivo sea posible reconstruir la función $a_i(\cdot)$ por medio de una expansión de series de wavelets hasta un nivel de resolución J apropiado para el tamaño de la muestra y se analizará la función $a(\cdot)$ vía simulación para poder observar el comportamiento probabilístico de los coeficientes de wavelets, y así realizar una estimación de estos parámetros y escoger el nivel de resolución óptimo para el problema.

Supongamos que tenemos una muestra de tamaño $T > p$, $\{x_1, x_2, \dots, x_T\}$, conducida por el proceso autorregresivo con coeficientes variando en el tiempo

$$x_t = \sum_{i=1}^p a_i(t)x_{t-1} + w_t,$$

para poder realizar un análisis en el mismo intervalo de tiempo se considera el proceso reescalado

$$x_{t,T} = \sum_{i=1}^p a_i\left(\frac{t}{T}\right) x_{t-1,T} + w_t,$$

el objetivo es estimar $a_i(\cdot)$ realizando una expansión en series de wavelets hasta un nivel de resolución J , por lo tanto, el proceso se escribirá de la forma

$$x_{t,T} = \sum_{i=1}^p \left[\alpha_{0,0}^{(i)} \phi(u) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}^{(i)} \psi_{j,k}(u) \right] x_{t-1,T} + w_t, \quad (5.1)$$

$$\text{donde } a_i(u) = \alpha_{0,0}^{(i)} \phi(u) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}^{(i)} \psi_{j,k}(u), \quad u = \frac{t}{T}, \text{ para } t = 1, 2, \dots, T.$$

Debemos encontrar los coeficientes que minimizan el error cuadrático medio de (5.1), para esto se escribirá el proceso de forma apropiada para obtener una matriz Ψ con la información conocida y un vector β con los coeficientes desconocidos y poder llevar el proceso que está variando en el tiempo a la forma $\mathbf{x} = \Psi\beta + \mathbf{w}$. Para llegar a la representación anterior se define

$$\beta^{(i)} = \left(\alpha_{0,0}^i, \beta_{0,0}^i, \beta_{1,0}^i, \beta_{1,1}^i, \beta_{2,0}^i, \beta_{2,1}^i, \beta_{2,2}^i, \beta_{2,3}^i, \beta_{3,0}^i, \dots, \beta_{J-1,2^{J-1}}^i \right)',$$

para $i = 1, 2, 3, \dots, p$,

donde $\beta^{(i)}$ es el vector que contiene los coeficientes de wavelets de $a_i(u)$ y sea el vector $\boldsymbol{\beta} = (\beta^{(1)}, \beta^{(2)} \dots, \beta^{(p)})'$, con todos los coeficientes de wavelets.

considere la siguiente matriz para cada $i = 1, 2, \dots, p$

$$\boldsymbol{\Psi}^{(i)} = \begin{pmatrix} \phi\left(\frac{p+1}{T}\right)x_{p+1-i} & \psi_{0,0}\left(\frac{p+1}{T}\right)x_{p+1-i} & \cdots & \psi_{J-1,2^{J-1}}\left(\frac{p+1}{T}\right)x_{p+1-i} \\ \vdots & \vdots & \cdots & \vdots \\ \phi\left(\frac{T}{T}\right)x_{T-i} & \psi_{0,0}\left(\frac{T}{T}\right)x_{T-i} & \cdots & \psi_{J-1,2^{J-1}}\left(\frac{T}{T}\right)x_{T-i} \end{pmatrix},$$

$\boldsymbol{\Psi}^{(i)}$ es una matriz de orden, $T - p \times 2^J$, donde todas las entradas son conocidas y sea

$$\boldsymbol{\Psi} = \left(\boldsymbol{\Psi}^{(1)}, \boldsymbol{\Psi}^{(2)}, \dots, \boldsymbol{\Psi}^{(p)} \right)',$$

la concatenación de las matrices $\boldsymbol{\Psi}^{(i)}$, por tanto la observación, $\mathbf{x} = (x_{p+1}, x_{p+2}, \dots, x_T)'$ con el vector de ruido blanco $\mathbf{w} = (w_{p+1}, w_{p+2}, \dots, w_T)'$, puede ser escrita de la forma

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\beta} + \mathbf{w},$$

esta ecuación se puede resolver por medio de mínimos cuadrados generalizados y así el estimador que minimiza el error cuadrático medio está dado por

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}'\Gamma_w^{-1}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\Gamma_w^{-1}\mathbf{x},$$

donde Γ_w es la matriz de covarianzas de la serie $w_t \sim Wn(0, 1)$ y se concluye que $\Gamma_w = \mathbb{I}_{T-p}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{x}.$$

5.2. Estudio de Simulación

Supongamos que se tiene una muestra de tamaño T $(x_1, x_2, \dots, x_T)'$, conducido por el proceso

$$x_t = a\left(\frac{t}{T}\right)x_{t-1} + w_t,$$

con $w_t \sim N(0, 1)$ y se quiere estimar la función $a(\cdot)$, escribiendo su representación en series de wavelets hasta un nivel de resolución V_J , este nivel se tiene que tomar de tal

manera que asintóticamente no se pierda la eficiencia del estimador; en [3], se muestra en la sección 3 que el nivel de resolución J , tal que $2^{J-1} \leq \sqrt{T} < 2^J$ es el mejor en el sentido que se conserva la eficiencia del estimador para T grande y no hay una sobre parametrización del modelo.

Sea $u = \frac{t}{T}$ con $t \leq T$,

$$a(u) = \alpha_{00}\phi(u) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(u),$$

$$\boldsymbol{\beta} = (\alpha_{0,0}, \beta_{0,0}, \beta_{1,0}, \beta_{1,1}, \beta_{2,0}, \beta_{2,1}, \beta_{2,2}, \beta_{2,3}, \beta_{3,0}, \dots, \beta_{J-1,2^{J-1}})',$$

$$\boldsymbol{\Psi} = \begin{pmatrix} \phi(\frac{2}{T})x_1 & \psi_{0,0}(\frac{2}{T})x_1 & \cdots & \psi_{J-1,2^{J-1}}(\frac{2}{T})x_1 \\ \vdots & \vdots & \cdots & \vdots \\ \phi(\frac{T}{T})x_{T-1} & \psi_{0,0}(\frac{T}{T})x_{T-1} & \cdots & \psi_{J-1,2^{J-1}}(\frac{T}{T})x_{T-1} \end{pmatrix},$$

donde $\boldsymbol{\Psi}$ es una matriz de orden, $T - 1 \times 2^J$, $\boldsymbol{\beta}$ es un vector de orden, $2^J \times 1$, y si $\mathbf{x} = (x_2, \dots, x_T)'$, $\mathbf{w} = (w_2, \dots, w_T)'$, entonces:

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\beta} + \mathbf{w},$$

y finalmente el estimador de mínimos cuadrados está dado por:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}'\mathbf{x}.$$

Para obtener una mejor aproximación, supongamos que se tiene

$$W_t = (w_t^{(1)}, w_t^{(2)}, \dots, w_t^{(L)}) \sim N(\mathbf{0}, \mathbb{I}_L),$$

para cada $k = 1, 2, \dots, L$ conducido por el proceso

$$x_t = a\left(\frac{t}{T}\right)x_{t-1} + w_t^{(k)},$$

realizando la estimación para cada k , por el mismo método anterior se obtiene la ecuación

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\beta}^{(k)} + \mathbf{w}_t^{(k)},$$

donde $\boldsymbol{\beta}^{(k)}$ es el vector con los coeficientes de wavelets estimados con el ruido blanco $w_t^{(k)}$ y sea

$$\mathbf{w}_t^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_T^{(k)}).$$

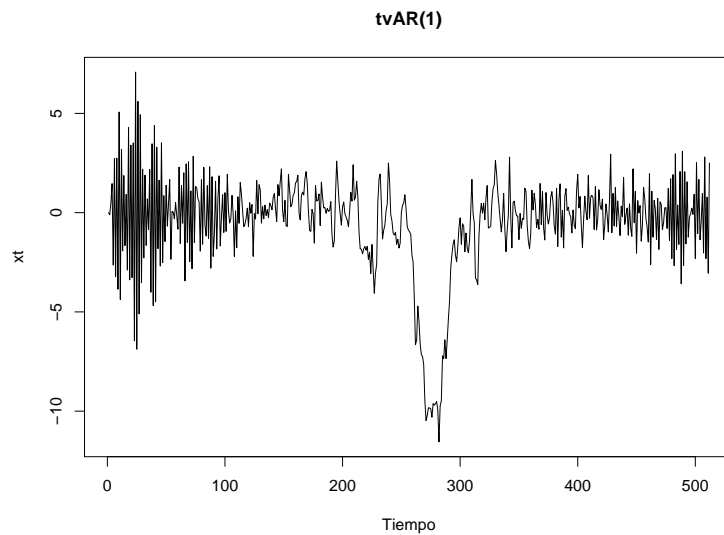
Además, el vector de coeficientes de wavelets estimado para $k = 1, 2, \dots, L$, es

$$\hat{\beta}^{(k)} = (\Psi' \Psi)^{-1} \Psi' \mathbf{x}.$$

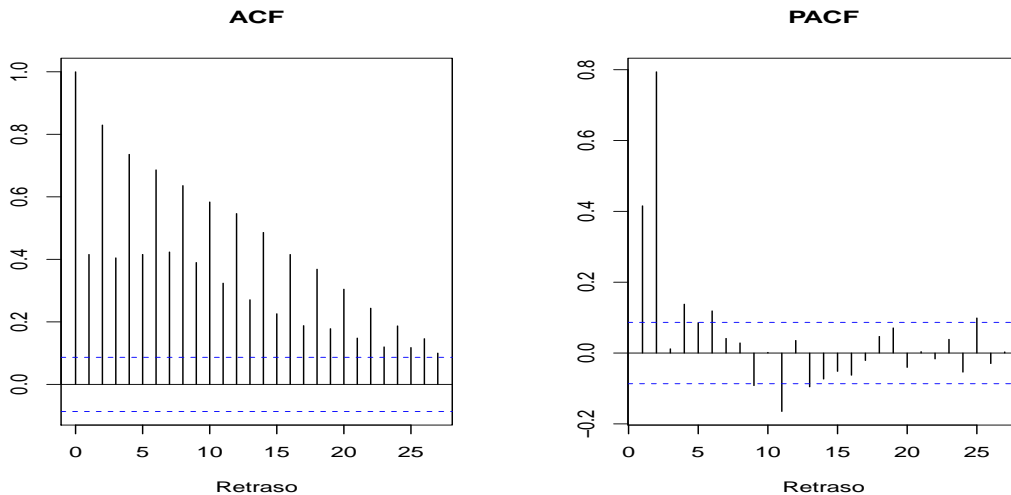
Con lo anterior es posible determinar la distribución de cada coeficiente y así escoger el más apropiado para la estimación, en el siguiente ejemplo se tomó la mediana de cada coeficiente.

Ejemplo 5.2.1. Considere el proceso conducido por $x_t = a(t)x_{t-1} + w_t$, donde

$$a(t) = \begin{cases} -\cos\left(\frac{2\pi t}{512}\right) & , \text{si } 1 \leq t \leq 512 \\ 0 & , \text{e.o.c} \end{cases}$$

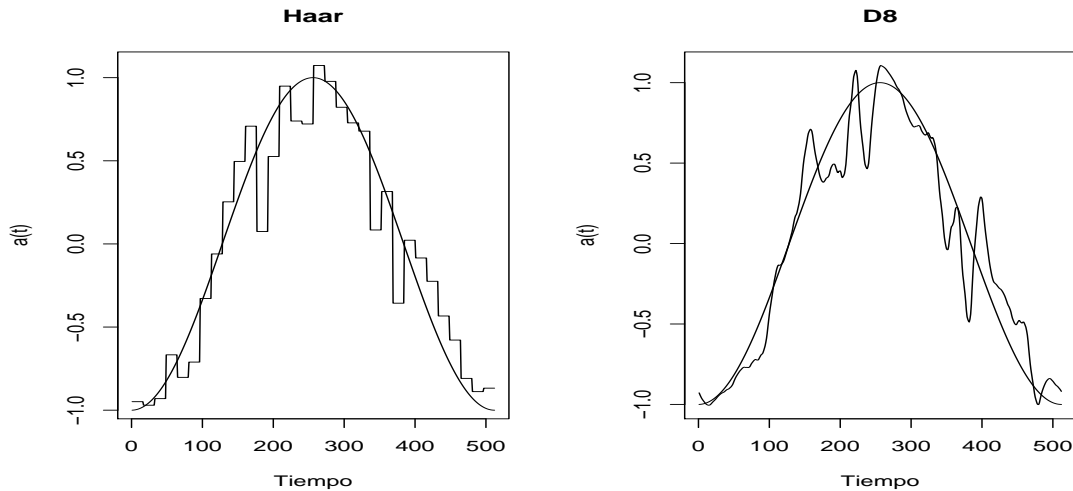


Se puede observar que el proceso no es estacionario en media y tampoco en covarianza y por tanto el análisis de series estacionarias no se puede aplicar, aunque la función de autocorrelación y autocorrelación parcial tienen un comportamiento similar a un $AR(1)$

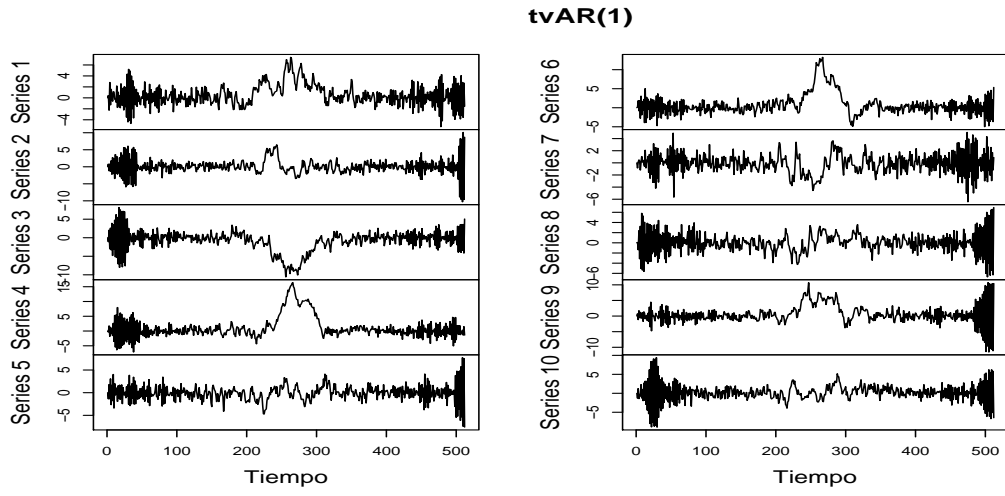


El nivel de resolución se toma $J = 5$ ya que $\sqrt{T} \approx 23$ y $2^4 \leq \sqrt{T} < 2^5$.

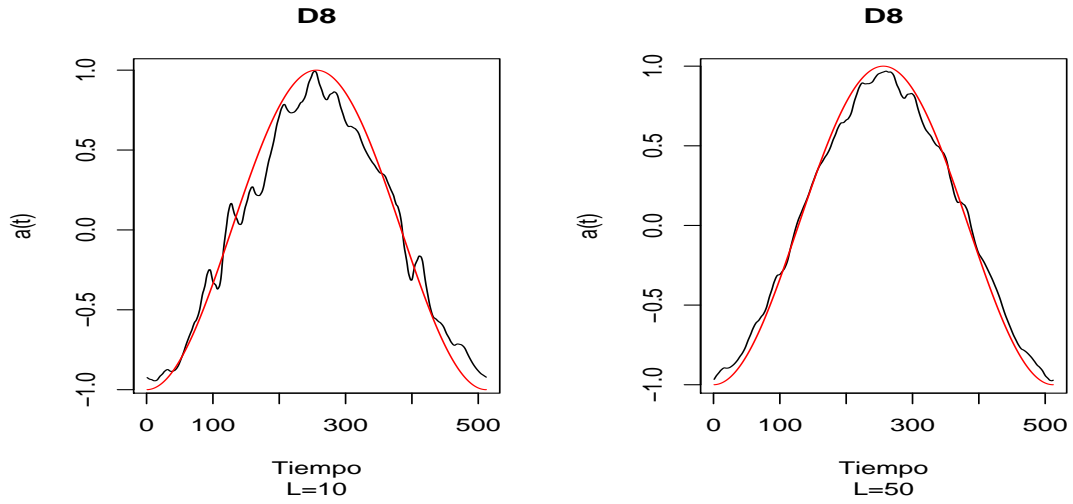
Al utilizar el software libre R, con `library(wavelets)`, se puede obtener la aproximación de las wavelets D_8 y Haar con una primera simulación



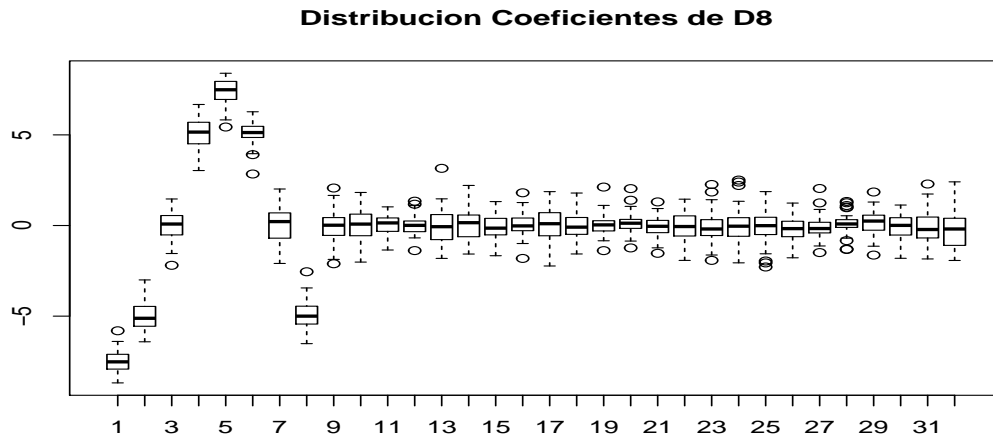
Para mejorar la estimación de D_8 , se realiza la metodología mencionada anteriormente realizando 10 y 50 simulaciones, en las cuales a cada una se realiza el proceso de estimación y para la reconstrucción se tomará el coeficiente que más se repite.



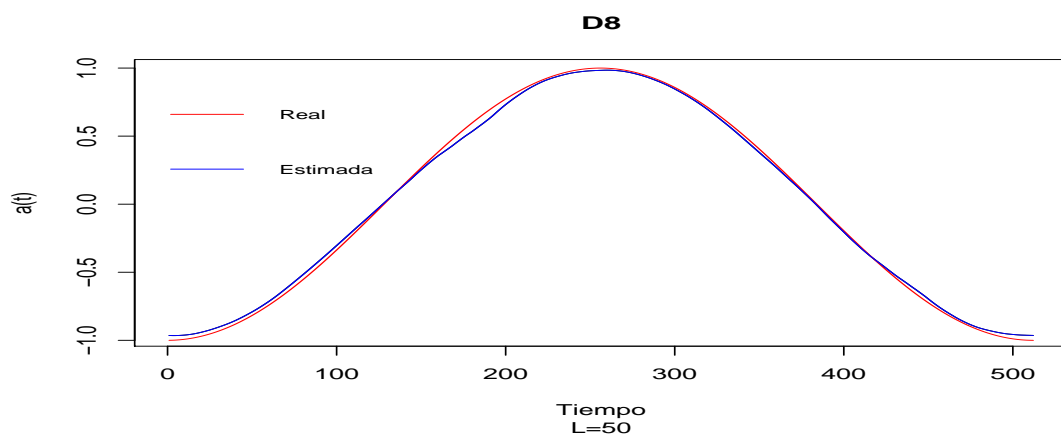
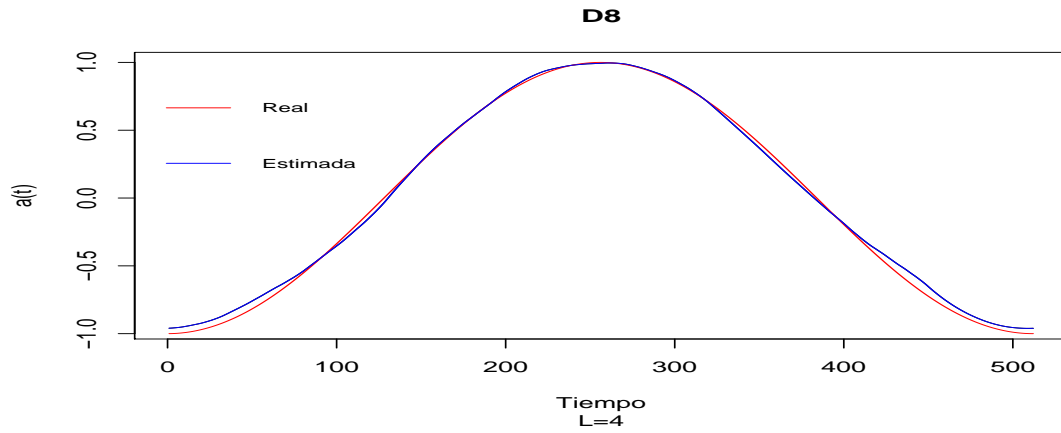
tvAR(1)



al analizar la distribución de los coeficientes de wavelets, graficando el boxplot de cada coeficiente, se tiene que



lo que sugiere es tomar un nivel de resolución más bajo, donde sólo se usen 8 coeficientes de wavelets, y así la estimación reduce considerablemente el número de iteraciones necesarias para tener una buena aproximación



Apéndice A

Códigos

entrada: TT = tamaño de la muestra , $a_t = -\cos(\frac{2\pi t}{512})$.

salida : El proceso $x_t = a(t)x_t + e_t$.

set.seed(512): fija la muestra.

rnorm(512,0,1): toma 512 aleatorios normales con $\mu = 0$ y $\sigma = 1$.

rep(0,512): una lista de ceros de tamaño 512.

for $t = 1, \dots, 511$ **do**

| $x_t = a(t)x_{t-1} + e_t$

end

return (x_t)

Algoritmo 1: Serie temporal

```
set.seed(512)
e<-rnorm(512,0,1)
TT<-512; t<-1:512
f112<-function(s){-1*cos(pi*(2*s/TT))}
at<-f112(t)
plot.ts(at)
xt<-rep(0,512)
for(i in 1:511){
  xt[i+1]<-at[i+1]*xt[i]+e[i+1]
}
```

entrada: tamanho.T = Tamaño de la muestra,
 tamanho.J = Número de parámetros,
 k = El tiempo t en el que se encuentra

salida : Vetor.Psi una matriz con los valores de las funciones wavelets en el tiempo k

rep(0,512): una lista de ceros de tamaño 512.

dwt() : Una clase de **library(waveslim)**, donde las entradas s\$5,d\$5,d\$4,d\$3,d\$2,d\$1, son los valores de los waveles del nivel de escala 1, y nivel de detalles 1, 2, 3, 4, 5 respectivamente.

wf : Wavelet que se usará: haar y d8.

```
for k = 1, ..., 32 do
|  vetor.Psi = wav.dwt[1 : 32]
end
return vetor.Psi
```

Algoritmo 2: matrix.coef

```
matrix.coef<-function(k){vetor<-rep(0,tamanho.T)
vetor[k]<-1
y=dwt(vetor,wf="d8",n.levels=5)
wav.dwt<-c(y$d5, y$d4, y$d3, y$d2, y$d1)
vetor.Psi<-wav.dwt[1:tamanho.J]
vetor.Psi}
```

entrada: tamanho.T = Tamaño de la muestra,
 tamanho.J = Número de parámetros,
 matrix.wav = Matriz con los valores de las wavelets.

salida : La solución de mínimos cuadrados de, $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\beta} + \mathbf{w}$.

A partir de la función matrix.wav que contiene los valores de las funciones de wavelets, la función, ThetaX obtiene la matriz $\mathbf{\Psi}$, multiplicando cada fila de matrix.wav por x_i , para $i = 1, 2, \dots, 511$.

Algoritmo 3:

```
ThetaX<-matrix(0,tamanho.T-1,tamanho.J)
dim(ThetaX)
for(i in 1:511){
  ThetaX[i,]<-matrix.wav[i+1,]*xt[i]
}
dim(ThetaX)

Beta<-(solve(t(ThetaX)%*%ThetaX))%*%(t(ThetaX))%*%xt[2:512]
length(Beta)
hat<-matrix.wav%*%Beta
plot.ts(hat,xlab="Tiempo",ylab="a(t)",main="D8")
lines(at)
```

Conclusiones

Se abordó el concepto de serie de tiempo estacionaria y los procesos ARMA para capturar los comportamientos regulares de la serie, se pudo analizar una serie con la información de una entidad bancaria con el número de quejas recibidas diariamente por parte de los usuarios. El análisis se realizó anualmente y se encontró que existía un comportamiento estacional en la serie cada 7 días, por lo tanto se ajustó un proceso SARIMA con un periodo de 7, el cual logró capturar completamente los ciclos semanales. Los datos atípicos o cambios estructurales entre semana no se lograron capturar ya que la varianza del proceso no era constante, a pesar de realizar transformaciones para estabilizar la varianza la predictibilidad del modelo disminuyó considerablemente y por tanto se optó por dejar la serie original.

Por último se estudió series no estacionarias, específicamente procesos localmente estacionarios, el cual se modeló por medio de un proceso autorregresivo con coeficientes variando en el tiempo y se estimó los parámetros empleando una expansión en series de wavelets de tal manera que minimice el error cuadrático medio. Los resultados mostraron que es posible reconstruir casi por completo los coeficientes por este método y dependiendo de la wavelet empleada fue necesario diferentes niveles de resolución, se observó que para estimar una función suave utilizando la wavelet de daubechies solo fue necesario un nivel de resolución de 2^3 , mientras que en el caso de la wavelet de haar fue necesario dejar la convención de tomar un nivel de resolución proporcional al tamaño de la muestra propuesto por Dalhaus. Se realizó una prueba con 10 y 50 simulaciones y a partir de la distribución de los coeficientes de wavelet se optó por utilizar la moda, la cual mostró un buen ajuste a la función real.

Un posible trabajo futuro es estudiar la identificación de un proceso localmente estacionarios, complementando con un análisis de intervención, donde se producen los cambios estructurales, además de estudiar el análisis de wavelets en dos dimensiones que son una herramienta poderosa para la reconstrucción de imágenes.

Bibliografía

- [1] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: with R examples*. Springer, 2017.
- [2] R. Dahlhaus, “On the Kullback-Leibler information divergence of locally stationary processes,” *Stochastic Processes and their Applications*, vol. 62, pp. 139–168, March 1996.
- [3] R. Dahlhaus, M. H. Neumann, R. Von Sachs, *et al.*, “Nonlinear wavelet estimation of time-varying autoregressive processes,” *Bernoulli*, vol. 5, no. 5, pp. 873–906, 1999.
- [4] L. Rincon, *Introducción a la probabilidad*. Temas selectos de matemáticas, Universidad Nacional Autónoma de México, Facultad de Ciencias, 2014.
- [5] W. W. William and W. Shyong, “Time series analysis,” 1994.
- [6] *Multiresolution Analysis*, ch. 5, pp. 179–231. John Wiley & Sons, Ltd, 2011.
- [7] *Haar Spaces*, ch. 3, pp. 85–143. John Wiley & Sons, Ltd, 2011.
- [8] Ø. Ryan, Ryan, and Peters, *Linear Algebra, Signal Processing, and Wavelets—a Unified Approach*. Springer, 2019.
- [9] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [10] I. Daubechies, *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1992.
- [11] I. Daubechies and J. C. Lagarias, “Two-scale difference equations i: Existence and global regularity of solutions,” *SIAM J. Math. Anal.*, vol. 22, p. 1388–1410, Sept. 1991.
- [12] *Daubechies Scaling Functions and Wavelets*, ch. 6, pp. 233–276. John Wiley & Sons, Ltd, 2011.