



Universidad de Oviedo

Predicción de series temporales utilizando la
integral de Choquet discreta

David Álvarez González

Dirigido por: Juan Baz González y Susana Montes Rodríguez

UNIVERSIDAD DE OVIEDO

Facultad de Ciencias

Grado en Matemáticas

Junio de 2023

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.3. Estructura del trabajo	3
2. Conceptos básicos	4
2.1. Estadística y probabilidad básica	4
2.1.1. Probabilidad condicionada	7
2.2. Variables aleatorias	8
2.3. Vectores aleatorios	13
2.4. Distribución normal	15
2.5. Estimadores	16
2.6. Errores de predicción	18
2.7. Conjuntos parcialmente ordenados y el diagrama de Hasse . .	18
2.7.1. Diagramas de Hasse	20
3. Series temporales	23
3.1. Modelos autorregresivos	29
3.1.1. Modelo autorregresivo de primer orden $AR(1)$	30
3.1.2. Modelo autorregresivo de segundo orden $AR(2)$	31
3.2. Modelos de media móvil	36
3.2.1. Modelo de media móvil de primer orden $MA(1)$	36

3.2.2. Modelo de media móvil de segundo orden MA(2) . . .	37
3.3. Modelo autorregresivo de media móvil ARMA(p,q)	39
3.4. Modelo autorregresivo integrado de media móvil ARIMA(p,q)	39
3.5. Componente estacional, el modelo SARIMA	40
4. Integral de Choquet Discreta	44
4.1. Propiedades de la integral de Choquet	48
4.2. El diagrama de Hasse de una capacidad	50
5. El modelo de proceso estocástico basado en la integral de Choquet discreta	53
6. Estudio del modelo con series temporales reales	59
6.1. El número de realizaciones previas k	59
6.2. La capacidad m	61
6.3. La constante Φ	61
6.4. Primera comparación	62
6.5. Análisis comparativo para varias series	63
7. Conclusiones	65

Capítulo 1

Introducción

1.1. Motivación

Una serie temporal es una secuencia de observaciones obtenidas a lo largo del tiempo[16]. En su estudio, se busca analizar la dependencia de las observaciones con el tiempo, así como predecir su comportamiento en tiempos futuros (pronóstico).

El análisis de series temporales es ampliamente utilizado en una gran cantidad de campos para intentar predecir ciertas variables claves y anticiparse a su cambio. En economía, por ejemplo, se usa para hacer un estudio de impuestos sobre el capital [1] o de precios de productos [6].

Por otra parte, la integral de Choquet discreta, que es un concepto no tan habitual en el área de estadística, es una función que toma n valores y devuelve otro valor basándose en una capacidad, es decir, una medida no aditiva, que permite modelizar la interacción entre los valores de entrada [4] [5].

La motivación de este trabajo de fin de grado es aunar ambos conceptos de una forma inédita en la literatura, de forma que se generalice el modelo autorregresivo, en tanto que la integral de Choquet puede verse como una media generalizada. Así pues, la motivación principal es proporcionar un

nuevo método de predicción de series temporales para el que se verificará su aplicabilidad y utilidad en el análisis de este tipo de datos.

1.2. Objetivos

El principal objetivo de este trabajo de fin de grado es investigar sobre series temporales y los modelos conocidos, así como la integral de Choquet discreta vista como función de agregación y medible. Posteriormente se intentará unir ambos conceptos proponiendo un modelo de predicción de series temporales que use la integral de Choquet discreta para estudiarlo y aplicarlo a distintos ejemplos, comparando su precisión con los métodos conocidos mencionados anteriormente. Esquemáticamente, los objetivos son:

- Estudiar el campo de las series temporales y los métodos existentes.
- Presentar la integral de Choquet y la integral de Choquet discreta como función medible.
- Proponer un modelo de predicción de series temporales utilizando la integral de Choquet discreta como función de agregación.
- Analizar el modelo propuesto mediante simulaciones, centrándose en el efecto de los parámetros.
- Comparar la precisión de los datos pronosticados con otros métodos existentes, como el SARIMA.

Para alcanzar dichos objetivos, se utilizarán las competencias generales o transversales adquiridas en el Grado en Matemáticas, como son la elaboración de una memoria escrita, la búsqueda y organización de documentación o la gestión óptima del tiempo de trabajo, así como algunas competencias específicas de dicho grado. En concreto, se han utilizado las competencias CE1 a CE10 de la guía docente de esta asignatura.

1.3. Estructura del trabajo

En base a los objetivos marcados la estructura del trabajo será la siguiente:

En el capítulo 2 se introducirá los conceptos básicos necesarios para entender el desarrollo del trabajo. En concreto se presentarán las herramientas estadísticas básicas utilizadas en el mismo.

En el capítulo 3 se definirá el concepto formal de serie temporal, así como ciertas magnitudes propias de estas. Se presentarán varios métodos existentes para la simulación y predicción de series temporales como el modelo autorregresivo, el de media móvil y, por último, el modelo SARIMA.

En el capítulo 4 se presentará el concepto de capacidad y la integral de Choquet, para conocer la integral de Choquet discreta. También se establecerá un modelo de representación para una capacidad basado en el diagrama de Hasse.

En el capítulo 5 se propondrá un modelo original uniendo los conceptos previos y se hará un estudio de las propiedades del mismo, así como un análisis de los parámetros que participan en el modelo, mediante la utilización de datos simulados.

En el capítulo 6 se aplicará el modelo a bases de datos reales, comparándolo con otros modelos ya existentes y por último, en el capítulo 7 se discutirán las conclusiones principales del trabajo desarrollado.

Capítulo 2

Conceptos básicos

Previo al estudio de los dos conceptos centrales de este trabajo, las series temporales y la integral de Choquet discreta, es necesaria la introducción de ciertos conceptos básicos. En este capítulo se presentarán algunos conceptos elementales de estadística y probabilidad. El mismo también servirá para fijar la notación utilizada en el resto del trabajo.

2.1. Estadística y probabilidad básica

Previa a la introducción de una serie temporal es necesario conocer la base estadística sobre la que se sustenta, por lo que se introducirán algunos conceptos de probabilidad y estadística. El objetivo de la estadística es estudiar una **población**, compuesta de **individuos** cuyas características son de interés. El subconjunto de individuos de la población sobre el que se realiza el estudio se llama **muestra** [7]. Desde el punto de vista descriptivo, se definen los conceptos fundamentales de media y varianza como sigue.

Definición 2.1. [7] *Dada una muestra x_1, \dots, x_n se define la **media muestral**, \bar{x} , como:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.1)$$

Definición 2.2. [7] Dada una muestra x_1, \dots, x_n se define la **varianza muestral**, s^2 , como:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.2)$$

En ocasiones a esta medida se la llama cuasivarianza muestral, reservando el nombre de varianza para el caso en el que el denominador del cociente sea n en lugar de $n-1$. No obstante, para los objetivos marcados, es más conveniente utilizar la denominación dada en la definición anterior.

En los dos conceptos introducidos hasta ahora, simplemente se hace un resumen descriptivo de los datos. No obstante, en este estudio es necesario introducir algunos conceptos matemáticos, que permitirán realizar un estudio formal de los problemas analizados.

Un experimento es cualquier proceso cuyo resultado esta asociado a cierta incertidumbre. El **espacio muestral** es el conjunto de todos los posibles resultados de un experimento y se denotará por Ω . Un subconjunto del espacio muestral se llama **evento** [7].

Dado un espacio muestral Ω , a un evento $A \subseteq \Omega$ se le puede asociar la probabilidad de que ocurra dicho evento A . Formalmente, para definir el concepto de probabilidad, es necesario previamente introducir el concepto de σ -álgebra.

Definición 2.3. [14] Dado un conjunto Ω , un conjunto $\emptyset \neq \mathcal{A} \subseteq \mathcal{P}(\Omega)$ se dice **anillo** si cumple:

$$A, B \in \mathcal{A} \implies A \cup B, A \setminus B \in \mathcal{A}$$

Un anillo \mathcal{A} para el que $\Omega \in \mathcal{A}$, cerrado por el paso al complementario (si $A \in \mathcal{A} \implies A^c \in \mathcal{A}$) y para uniones finitas (si $A_1, \dots, A_n \in \mathcal{A} \implies \bigcup_{i=1}^n A_i \in \mathcal{A}$) se llama **álgebra**. Un álgebra cerrada para uniones numerables se llama **σ -álgebra**, es decir:

1. $\Omega \in \mathcal{A}$

2. Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$

3. Si $A_1, \dots, A_n, \dots \in \mathbf{A}$, entonces $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$

El par (Ω, \mathcal{A}) donde Ω es un conjunto y \mathcal{A} es una σ -álgebra de Ω es un espacio medible.

A partir de ahora se llamará evento a cualquier elemento que pertenezca a una σ -álgebra asociada al espacio muestral.

Ejemplo 2.1. Dado $\Omega = \{1, 2, 3, 4, 5, 6\}$, el siguiente conjunto es una σ -álgebra de Ω ; $\mathcal{A} = \{\emptyset, \{2\}, \{4\}, \{2, 4\}, \{1, 3, 5, 6\}, \{1, 3, 4, 5, 6\}, \{1, 2, 3, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}$, y por tanto (Ω, \mathcal{A}) es un espacio medible.

Definición 2.4. [14] Dado un conjunto Ω y $\emptyset \neq \mathcal{A} \subseteq \mathcal{P}(\Omega)$, la σ -álgebra más pequeña que contiene a \mathcal{A} se dice la σ -álgebra **generada por \mathcal{A}** .

Ejemplo 2.2. Sea \mathbb{R} , y la σ -álgebra generado por $\mathcal{B} = \{(-\infty, b) | b \in \mathbb{R}\}$. Este es un ejemplo útil en el contexto de la probabilidad, si x es el resultado de algún experimento donde el espacio muestral es \mathbb{R} , ¿es cierto que $x \leq b$? Esta σ -álgebra se denomina la σ -álgebra de Borel y se denota por $\mathcal{B}_{\mathbb{R}}$.

A los elementos de \mathcal{A} se les llama conjuntos \mathcal{A} -medibles (o simplemente conjuntos medibles). En un contexto probabilístico, se les suele llamar sucesos o eventos.

Con esta estructura definida, ya es posible introducir el concepto de probabilidad:

Definición 2.5. Sea Ω un espacio muestral y \mathcal{A} una σ -álgebra de subconjuntos de Ω . Una función $P : \mathcal{A} \rightarrow \mathbb{R}$ es una **función de probabilidad sobre (Ω, \mathcal{A})** si se cumplen los siguientes axiomas:

Ax1. $0 \leq P(A), \forall A \in \mathcal{A}$

Ax2. $P(\Omega) = 1$

Ax3. Si $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ son sucesos mutuamente excluyentes, es decir, si se verifica que $A_i \cap A_j = \emptyset \quad \forall i \neq j$, entonces: $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

A la terna (Ω, \mathcal{A}, P) se la denomina espacio de probabilidad, esto es, un espacio de sucesos (espacio muestral) en el que se han definido los posibles sucesos a considerar (la σ -álgebra de sucesos) y la probabilidad de cada suceso (la función de probabilidad).

De estos axiomas se pueden deducir algunas propiedades interesantes:

Proposición 2.1. [14] Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , se verifica que:

1. $P(\emptyset) = 0$
2. Dada cualquier familia de eventos A_1, \dots, A_n se tiene que:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

3. $P(\Omega \setminus A) = 1 - P(A)$
4. Si $A \subseteq B$, entonces $P(A) \leq P(B)$
5. Para cualesquiera $A, B \in \mathcal{A}$:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2.1.1. Probabilidad condicionada

La probabilidad condicional es un concepto elemental pero muy importante, que se utiliza con mucha frecuencia en el cálculo de probabilidades.

Definición 2.6. [14] Sean A y B dos eventos teniendo B una probabilidad estrictamente positiva. La **probabilidad condicional del evento A dado el evento B** se denota $P(A|B)$ y se define como el cociente

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.3)$$

La condición de $P(B) > 0$ es necesaria para que el cociente esté bien definido.

La probabilidad condicionada, vista como función de A cumple los tres axiomas de la probabilidad [14].

- $0 \leq P(A|B), \forall A$
- $P(\Omega|B) = 1$
- $P(\cup_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B)$ si $A_i \cap A_j = \emptyset$ para todo $i \neq j$

Por tanto, dado un $B \in \mathcal{A}$ con $P(B) > 0$, la función $P(\cdot|B) : \mathcal{A} \mapsto \mathbb{R}$ tal que a cada A de \mathcal{A} le asigna el valor $P(A|B)$ es una medida de probabilidad [14].

2.2. Variables aleatorias

Una vez definido el espacio de probabilidad se puede definir una variable aleatoria como una función sobre este. Esto va a permitir considerar el resultado del experimento aleatorio como un número real tomado por la variable aleatoria [14].

Definición 2.7. [14] *Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , una **variable aleatoria** es una transformación X del espacio muestral Ω al conjunto de los números reales \mathbb{R} , esto es*

$$X : \Omega \longrightarrow \mathbb{R}$$

tal que para cualquier $x \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A} \tag{2.4}$$

La definición anterior es equivalente a considerar que una variable aleatoria es una función medible de un espacio de probabilidad al conjunto \mathbb{R}

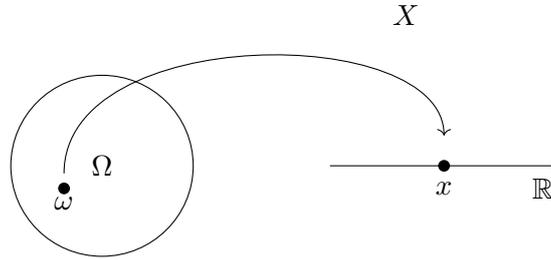


Figura 2.1: Esquema de una variable aleatoria.

con la σ -álgebra de Borel del ejemplo 2.1. X es una variable aleatoria si para cualquier conjunto de Borel B se tiene:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad (2.5)$$

Con esta transformación, se tiene una σ -álgebra para el conjunto \mathbb{R} asociada con el espacio de probabilidad inicial (Ω, \mathcal{A}, P) pero aún falta definir la probabilidad para tener un espacio de probabilidad en \mathbb{R} .

Medida de probabilidad inducida. Para cualquier intervalo en \mathbb{R} de la forma $(-\infty, x]$ se puede tomar su antimagen por X , es decir,

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\}$$

Como este conjunto pertenece a \mathcal{A} por la condición 2.4 se puede aplicar la probabilidad P , pues esta tiene como dominio \mathcal{A} . Así, mediante la función X puede trasladarse la medida de probabilidad P a intervalos de la forma $(-\infty, x]$ y puede demostrarse que ello es suficiente para extenderla a la totalidad de la σ -álgebra de Borel de \mathbb{R} [14].

A esta nueva probabilidad se le denota P_X , aunque por simplicidad, si no da lugar a confusión, se omitirá el subíndice. De esta manera se tiene un nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P)$ [14], inducido por el espacio de probabilidad inicial (Ω, \mathcal{A}, P) , tal como aparece representado en la figura 2.2.

Considerando el conjunto de valores que puede tomar una variable aleatoria existen dos tipos importantes de variables: discretas y continuas. In-

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$$

Figura 2.2: Espacio de probabilidad inducido.

tuitivamente, una variable aleatoria es discreta si el conjunto de valores que toma es finito o numerable. De forma coloquial, se dice continua si puede tomar una cantidad infinita no numerable de valores distintos. La formalización de estos conceptos se dará una vez introducido el concepto de función de distribución, que es una función real que caracteriza a la variable aleatoria.

Definición 2.8. [14] Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria, la **función de distribución acumulada** de la variable aleatoria X es una función $F : \mathbb{R} \rightarrow [0, 1]$ definida como

$$F(x) = P(X \leq x) \quad (2.6)$$

A partir de esta función, se puede definir el concepto de variable aleatoria discreta y continua, como sigue:

Definición 2.9. [14] Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria con función de distribución F . Se dice que una variable aleatoria X es:

- **Discreta** si existe un conjunto numerable $E \subset \mathbb{R}$ tal que $P(X \in E) = 1$.
- **Continua** si su función de distribución F es una función absolutamente continua.

Definición 2.10. [14] Sea X una variable aleatoria discreta con valores x_0, x_1, \dots . La **función de probabilidad de X** , denotada por $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ se define como:

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_0, x_1, \dots \\ 0 & \text{en otro caso} \end{cases} \quad (2.7)$$

Con esta definición, la probabilidad de un evento se reduce a una suma. Si A es un conjunto de \mathbb{R} , la probabilidad de ese evento es [14]:

$$P(X \in A) = \sum_{x \in A} f(x) \quad (2.8)$$

Definición 2.11. [14] Sea X una variable aleatoria continua. Se dice que la función integrable y no negativa $f : \mathbb{R} \rightarrow \mathbb{R}$ es la **función de densidad de X** si para cualquier intervalo $[a, b] \subseteq \mathbb{R}$ se cumple:

$$P(X \in [a, b]) = \int_a^b f(x)dx \quad (2.9)$$

Con lo cual, si la variable aleatoria es discreta la función de distribución se calcula como:

$$F(x) = \sum_{u \leq x} f(u) \quad (2.10)$$

mientras que si es una variable continua se calcula como:

$$F(x) = \int_{-\infty}^x f(u)du \quad (2.11)$$

Ahora se introducirán algunas características numéricas, llamadas parámetros, asociadas a la distribución de una variable aleatoria.

Definición 2.12. [14] Sea X una variable aleatoria discreta con función de probabilidad $f(x)$. La **esperanza de X** se define como:

$$E(X) = \sum_x xf(x) \quad (2.12)$$

suponiendo que esta suma es absolutamente convergente. Por otro lado, si X es una variable aleatoria continua con función de densidad $f(x)$, la esperanza es:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (2.13)$$

suponiendo que la integral es convergente.

Proposición 2.2. [14] Sea X una variable aleatoria discreta y sea ϕ una función tal que $\phi(X)$ es una variable aleatoria con esperanza finita. Entonces:

$$E[\phi(X)] = \sum_x \phi(x)P(X = x) \quad (2.14)$$

Proposición 2.3. [14] Sea X una variable aleatoria continua y sea ϕ una función tal que $\phi(X)$ es una variable aleatoria con esperanza finita. Entonces:

$$E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)f(x)dx \quad (2.15)$$

Algunas propiedades de la esperanza son las siguientes.

Proposición 2.4. [14] Sean X e Y dos variables aleatorias con esperanza finita y c una constante. Entonces

1. $E(c) = c$
2. $E(cX) = cE(X)$
3. Si $X \geq 0$, entonces $E(X) \geq 0$
4. $E(X + Y) = E(X) + E(Y)$

Definición 2.13. [14] Sea X una variable aleatoria discreta con función de probabilidad $f(x)$. **La varianza de X** se define como el número

$$Var(X) = \sum_x (x - E(X))^2 f(x) \quad (2.16)$$

cuando esta suma es absolutamente convergente. Para una variable aleatoria continua X con función de densidad $f(x)$ se define

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) \quad (2.17)$$

cuando la integral es convergente

Como con la esperanza, se darán algunas propiedades de la varianza.

Proposición 2.5. [14] Sea X una variable aleatoria con varianza finita y c una constante. Entonces

1. $Var(X) \geq 0$
2. $Var(c) = 0$
3. $Var(cX) = c^2 Var(X)$
4. $Var(X + c) = Var(X)$
5. $Var(X) = E(X^2) - E(X)^2$

2.3. Vectores aleatorios

En esta sección se desarrollara brevemente el tema de variables aleatorias multidimensionales, también conocidas como vectores aleatorios [14].

Definición 2.14. Un *vector aleatorio de dimensión n* es un vector de la forma (X_1, \dots, X_n) tal que X_i es una variable aleatoria para todo $i \in \{1, \dots, n\}$.

Para estos vectores se puede definir la función de probabilidad en caso de que sean discretos y la función de densidad si son continuos de manera análoga a como se ha hecho previamente para variables unidimensionales. Sin embargo, se puede definir la función de distribución con independencia de la naturaleza de la variable como sigue.

Definición 2.15. [14] La *función de distribución conjunta del vector* (X_1, \dots, X_n) , denotada por $F : \mathbb{R}^n \rightarrow [0, 1]$ se define de la siguiente manera

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (2.18)$$

A partir de la función de distribución conjunta de un vector aleatorio es posible obtener la función de distribución de una variable individual. A esta última se le llama función de distribución marginal.

Definición 2.16. [14] Sea (X_1, \dots, X_n) un vector aleatorio y sea F su función de distribución conjunta. La **función de distribución marginal de la variable** X_i como la función de una variable

$$F_{X_i}(x_i) = \overbrace{\lim_{x_1 \rightarrow \infty} \dots \lim_{x_n \rightarrow \infty}}^{\text{salvo } x_i} F(x_1, \dots, x_n) \quad (2.19)$$

De manera análoga se puede obtener la función de distribución marginal de cualquier subvector de (X_1, \dots, X_n) .

A continuación se define el concepto de variables independientes, que son aquellas para las que existe una relación determinada entre la función de distribución conjunta y sus marginales.

Definición 2.17. Se dice que las variables aleatorias X_1, \dots, X_n son **independientes** si para cualesquiera números reales x_1, \dots, x_n se cumple la igualdad

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \quad (2.20)$$

Este tipo de variables tiene propiedades específicas muy interesantes. En concreto recordamos dos relativas a la esperanza y la varianza.

Proposición 2.6. Si X e Y son dos variables aleatorias independientes entonces:

- $E(XY) = E(X)E(Y)$
- $Var(X + Y) = Var(X) + Var(Y)$

También puede extenderse el concepto de probabilidad condicionada visto en la ecuación (2.3) al caso de funciones de probabilidad y densidad y al de funciones de distribución.

Definición 2.18. [14] Sea (X_1, \dots, X_n) un vector aleatorio discreto o continuo con función de probabilidad o densidad $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$. Sea x_i un valor de la variable X_i tal que $f_{X_i}(x_i) \neq 0$, se denomina **función de**

probabilidad o densidad condicional dado $X_i = x_i$, a la función de $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ definida como:

$$f_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n | X_i}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n | x_i) = \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{f_{X_i}(x_i)} \quad (2.21)$$

La relación entre dos variables se puede medir con la covarianza.

Definición 2.19. [14] Dado un vector aleatorio de dos variables (X, Y) con esperanza finita, la **covarianza** de X e Y se define como:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (2.22)$$

Desarrollando el producto y aplicando la linealidad de la esperanza se puede obtener la siguiente igualdad, siendo esta otra forma de calcular la covarianza [14]:

$$Cov(X, Y) = E(XY) - E(X)E(Y) \quad (2.23)$$

A continuación se presenta una transformación de la misma independiente de la escala de medida de las variables.

Definición 2.20. [14] El **coeficiente de correlación/covarianza** entre las variables aleatorias X e Y con varianzas distintas de 0 y finitas se define como:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \quad (2.24)$$

2.4. Distribución normal

La distribución normal es quizá la distribución de probabilidad más importante. En este trabajo se utilizará para la parte de simulación, por lo que se hará una breve descripción de la misma. Se dice que la variable aleatoria X continua tiene una distribución normal si su función de densidad está dada por la siguiente expresión:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad \forall x \in \mathbb{R} \quad (2.25)$$

donde $\sigma \in \mathbb{R}^+$ y $\mu \in \mathbb{R}$ son dos parámetros y se denotará como $X \sim N(\mu, \sigma^2)$. La gráfica de esta función tiene forma de campana como se ve en la figura 2.3. También se puede ver en la imagen el significado geométrico de los parámetros: μ es el centro de la campana y σ , la raíz positiva de σ^2 , es la distancia entre μ y cualquiera de los dos puntos de inflexión de la curva [14].

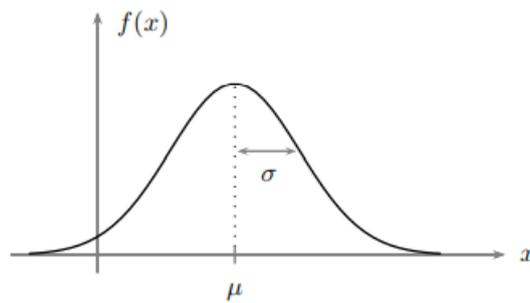


Figura 2.3

Para una variable aleatoria X con distribución $N(\mu, \sigma^2)$ se puede demostrar que [14]:

- $E(X) = \mu$
- $Var(X) = \sigma^2$

Una variable aleatoria X sigue una distribución normal estándar si sigue una distribución normal con $\mu = 0$ y $\sigma^2 = 1$.

Proposición 2.7. [14] Sea X una variable aleatoria con distribución $N(\mu, \sigma^2)$.

Entonces

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \quad (2.26)$$

2.5. Estimadores

Los parámetros son magnitudes que caracterizan a una población o definen una función. Posibles parámetros de una población son [15]:

- El promedio, μ , y todas las medidas de posición
- Su varianza, σ^2 , y su desviación estándar, σ .

En la práctica con mucha frecuencia estos parámetros no son conocidos, por lo que a partir de una muestra aleatoria se pretende estimar dichos valores suponiendo conocida la familia a la que pertenece la distribución de la población. Con este propósito se define un estadístico para referirse a una función de las observaciones de la muestra. Si este estadístico es utilizado para estimar una característica de la población, éste recibe el nombre de estimador [15].

El valor del estimador generalmente difiere del valor del parámetro que se pretende estimar. La diferencia entre ambas magnitudes es el error. La raíz de este error puede ser de muestreo (porque la muestra es aleatoria) y por otros errores ajenos al muestreo (sesgos). En lo que atañe al error de muestreo, el objetivo es conseguir un estimador con las mejores propiedades posibles. Las siguientes definiciones ayudan a escoger el mejor estimador [15].

Definición 2.21. [15] Un estimador se dice **insesgado** si su valor esperado es igual al parámetro que estima, es decir, si $\hat{\theta}$ es un estimador para θ , se cumple que $E(\hat{\theta}) = \theta$.

Definición 2.22. [15] Un estimador $\hat{\theta}$ de θ se dice **consistente** si

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| > \epsilon\} = 0, \forall \epsilon > 0$$

donde n representa el tamaño muestral.

Definición 2.23. [15] Si $\hat{\theta}_1, \dots, \hat{\theta}_n$ son estimadores del mismo parámetro, el estimador más **eficiente** es aquel con menor varianza.

Definición 2.24. Un estimador $\hat{\theta}$ de θ es **suficiente** si la distribución de la muestra aleatoria dado cualquier valor del estimador no depende de θ , es decir, si $f(x_1, \dots, x_n | \hat{\theta} = t)$ es independiente de θ .

2.6. Errores de predicción

A la hora de analizar un modelo, es necesario saber el error que comete prediciendo. Esto se hace comparando las predicciones con los datos reales que se intentan pronosticar. A continuación, se muestran una serie de errores comunes utilizados en la literatura (veáse, por ejemplo, [10]).

Sean n observaciones, y_t para las que se han hecho n predicciones \hat{y}_t , se define:

- el error medio (ME) como:

$$ME = \frac{1}{n} \sum_{t=1}^n y_t - \hat{y}_t \quad (2.27)$$

- el error absoluto medio, o la desviación absoluta media (MAD) como:

$$MAD = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2.28)$$

- el error cuadrático medio (MSE) como:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (2.29)$$

- el error medio porcentual (MPE) como:

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \quad (2.30)$$

- el error medio porcentual absoluto ($MAPE$) como:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.31)$$

2.7. Conjuntos parcialmente ordenados y el diagrama de Hasse

Una vez revisados los conceptos elementales de teoría de la probabilidad necesarios en este trabajo, se hará una breve introducción a las estructuras ordenadas y su representación.

La manera más simple de definir un conjunto parcialmente ordenado es decir que es un sistema de elementos con una relación binaria $a \leq b$ que es transitiva, reflexiva y antisimétrica. Esto quiere decir:

1. Si $a \leq b$ y $b \leq c$, entonces $a \leq c$.
2. Se tiene que $a \leq a$.
3. Nunca se dan ambos casos $a \leq b$ y $b \leq a$ a menos que $a = b$.

La definición formal de este concepto puede verse a continuación:

Definición 2.25. [12] Sea X un conjunto cualquiera. Un subconjunto S del producto cartesiano $X \times X$ se dice **orden parcial sobre X** si se satisface:

1. *Reflexividad:* $(x, x) \in S, \forall x \in X$.
2. *Transitividad:* Si $(x, y) \in S$ e $(y, z) \in S$, entonces $(x, z) \in S$.
3. *Antisimetría:* Si $(x, y) \in S$ e $(y, x) \in S$, entonces $x = y$.

Si la relación binaria es transitiva e irreflexiva, esto es transitiva y ningún elemento está relacionado consigo mismo, se dice que es un orden estricto. De estas dos relaciones se sigue la propiedad de asimetría: nunca se da $a < b$ y $b < a$ a la vez. Formalmente,

Definición 2.26. [12] Sea X un conjunto cualquiera. Un subconjunto S' del producto cartesiano $X \times X$ se dice **orden estricto sobre X** si se satisface:

1. *Irreflexividad:* $(x, x) \notin S', \forall x \in X$.
2. *Transitividad:* Si $(x, y) \in S'$ y $(y, z) \in S'$ entonces $(x, z) \in S'$.

Por conveniencia, $x \leq y$ es usado para denotar que $(x, y) \in S$. Además, $x < y$ es usado para decir que $x \leq y$ y $x \neq y$, es decir, $(x, y) \in S'$. El par (X, \leq) es un **conjunto parcialmente ordenado** si \leq es un orden

parcial sobre X . Asimismo, el par $(X, <)$ es un **conjunto estrictamente ordenado** si $<$ es un orden estricto sobre X .

Si $<$ es una relación de orden estricto en un conjunto X , entonces la relación \leq definida por:

$$a \leq b \Leftrightarrow a < b \text{ o } a = b \quad (2.32)$$

es una relación de orden parcial.

Dos elementos distintos x, y en el par (X, \leq) se dicen **comparables** si $x \leq y$ o $y \leq x$ y se dicen **incomparables** si no se cumple ninguna de esas dos condiciones. Además, si $x < y$ y no existe ningún $z \in X$ tal que $x < z < y$ se dice que **y cubre a x** y se denota como $x \leftarrow y$. Se verán ahora algunos ejemplos obtenidos de [12].

Ejemplo 2.3. 1. Sea $\mathcal{P}(X)$ el conjunto de las partes de un conjunto X .

Sea la relación $A \leq B$ dada cuando A es un subconjunto de B , donde ambos $A, B \in \mathcal{P}(X)$. Entonces $(\mathcal{P}(X), \leq)$ es un conjunto parcialmente ordenado.

2. El conjunto de los enteros con la relación mayor o igual \geq es un conjunto parcialmente ordenado.

3. Sea $X := \{1, 2, 3, 4, 6, 12\}$ y la relación $x \leq y$ dada cuando x es un divisor de y ($x, y \in X$). Entonces (X, \leq) es un conjunto parcialmente ordenado.

2.7.1. Diagramas de Hasse

Un diagrama de Hasse es una forma de representar un conjunto parcialmente ordenado y finito, para ayudar a la comprensión del conjunto de una manera visual. Sea (X, \leq) un conjunto parcialmente ordenado. Se dibuja el diagrama de Hasse del conjunto representando cada elemento de X en un punto distinto, de manera que si $x < y$ el punto y estará situado por encima del punto x . Más aún, si $x \leftarrow y$, los puntos se conectarán con una línea.

Ejemplo 2.4. [12] Dado el conjunto mencionado en el tercer punto del ejemplo 2.3 $X := \{1, 2, 3, 4, 6, 12\}$ con la misma relación. Como $2 \leq 12$ el punto 12 está por encima del punto 2, pero como 12 no cubre a 2 ($2 \leq 4$ y $4 \leq 12$) no hay ninguna línea que los una. Sin embargo, si hay línea entre el 2 y el 4 y entre el 4 y el 12. Los números 3 y 4 son incomparables, por lo que se coloca el 3 aparte del 4. La representación gráfica de este conjunto parcialmente ordenado se puede ser en la figura 2.4.

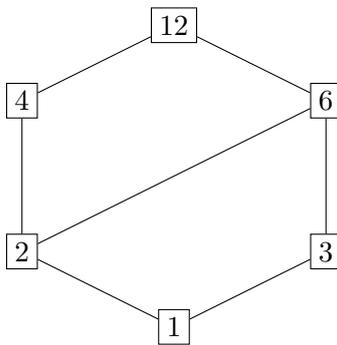


Figura 2.4: Diagrama de Hasse del conjunto $X = \{1, 2, 3, 4, 6, 12\}$.

En un diagrama de Hasse, también puede haber etiquetas en la unión de vértices, pudiendo indicar el coste que tiene pasar de una a otra, la resta de valores, etc. Siguiendo con el ejemplo 2.3, se puede asignar el valor por el que hay que multiplicar a ese número para llegar a la representación de la figura 2.5.

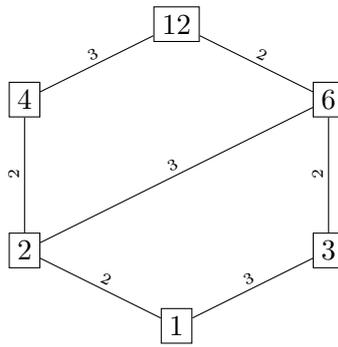


Figura 2.5: Diagrama de Hasse del conjunto $X = \{1, 2, 3, 4, 6, 12\}$ con valores asignados a las aristas.

Capítulo 3

Series temporales

Una serie temporal es una secuencia de observaciones a lo largo del tiempo. Las series temporales aparecen naturalmente en distintas disciplinas fuera de las matemáticas. En medicina, por ejemplo, pueden servir para el control de enfermedades [9], en psicología pueden usarse para modelar conductas [5] o incluso en demografía para estudiar crecimientos de poblaciones [13]. Matemáticamente, las series temporales se modelizan como observaciones de procesos estocásticos indexados en el tiempo. En lo que sigue se va a hacer un breve repaso a los conceptos fundamentales asociados a las series temporales, tomando como referencia el libro de Wei [16].

Definición 3.1. *Llamamos **proceso estocástico** a una familia de variables aleatorias indexadas en el tiempo $Z(X, t)$, donde X pertenece a un espacio de muestras y t pertenece a un conjunto de índices. A un t fijo, $Z(X, t)$ es una variable aleatoria. Para X fijo, $Z(X, t)$, vista como función de t , se le llama función de muestra o realización.*

*Así, una **serie temporal** es una función de muestra de un proceso estocástico en la cual el conjunto de índices está asociado a una evolución temporal, bien discreta (subconjunto de \mathbb{N}) o bien continua (subconjunto de \mathbb{R}).*

En este trabajo se asumirá que el conjunto de índices es el conjunto de todos los números enteros, a no ser que se especifique lo contrario. Dado un subconjunto finito de variables aleatorias $\{Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}\}$ de un proceso estocástico $\{Z(X, t), t = 0, \pm 1, \pm 2, \dots\}$. La función de distribución n -dimensional se define como sigue:

$$F_{Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}}(x_1, \dots, x_n) = P\{X : Z_{t_1} \leq x_1, \dots, Z_{t_n} \leq x_n\} \quad (3.1)$$

con $\mathbf{x} \in \mathbb{R}^n$.

Definición 3.2. *Un proceso se dice **estrictamente estacionario** si para cualquier $n=1, 2, \dots$ se cumple que:*

$$F_{Z_{t_1}, Z_{t_2}, \dots, Z_{t_n}}(x_1, \dots, x_n) = F_{Z_{t_1+k}, Z_{t_2+k}, \dots, Z_{t_n+k}}(x_1, \dots, x_n) \quad (3.2)$$

para cualesquiera t_1, \dots, t_n, k enteros.

Gráficamente, que una serie temporal sea estacionaria significa que la serie oscilan en torno a un valor en concreto, sin seguir alguna tendencia. Esta propiedad se puede ver como que la distribución se mantiene invariante respecto a traslaciones temporales.

Cabe remarcar que si se cumple 3.2 para $n = m$, también es cierto para cualquier $n \leq m$ porque la distribución de orden m determina todas las distribuciones de menor orden.

Un proceso estocástico se dice **de valor real** si solamente asume valores reales. A no ser que se mencione lo contrario, se asumirá que todos los procesos serán de valor real. Dado un proceso de valor real $\{Z_t : t = 0, \pm 1, \pm 2, \dots\}$ se pueden definir las siguientes magnitudes:

$$\mu_t = E[Z_t] \quad (3.3)$$

llamada la **función de medias**,

$$\sigma_t^2 = E[(Z_t - \mu_t)^2] \quad (3.4)$$

llamada la **función de varianzas**,

$$\gamma(t_1, t_2) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})] \quad (3.5)$$

llamada la **función de covarianzas entre Z_{t_1} y Z_{t_2}** y

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} \quad (3.6)$$

llamada la **función de correlación entre Z_{t_1} y Z_{t_2}** .

Para un proceso estrictamente estacionario, dado que la función de distribución es igual para todos los t , la media es constante, $\mu_t = \mu$, siempre que $E[|Z_t|] < \infty$. De la misma manera, si $E[Z_t^2] < \infty$ entonces $\sigma_t^2 = \sigma^2$ para todo t , y es también constante. Más aún, dado que $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ para cualesquiera t_1, t_2, k enteros, se tiene $\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k)$ y $\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k)$. Tomando $t_1 = t - k$ y $t_2 = t$, se tiene:

$$\begin{aligned} \gamma(t_1, t_2) &= \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k \\ \rho(t_1, t_2) &= \rho(t - k, t) = \rho(t, t + k) = \rho_k = \frac{\gamma_k}{\sqrt{\text{Var}(Z_t)} \sqrt{\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0} \end{aligned}$$

A estas funciones (de k) se les denomina **función de autocovarianza** y **función de autocorrelación (ACF)** respectivamente.

Por último, existe una función cuyo objetivo es estudiar la correlación entre Z_t y Z_{t+k} cuando los valores $Z_{t+1}, Z_{t+2}, \dots, Z_{t+k-1}$ son conocidos. La correlación condicional

$$\text{Corr}(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1}) \quad (3.7)$$

como función de k se denomina la **función de correlación parcial (PACF)**.

Se verá como obtener esta función de correlación parcial. Sea un proceso estrictamente estacionario $\{Z_t\}$ con $E(Z_t) = 0$. Sea definida la dependencia lineal de Z_{t+k} con $Z_{t+1}, \dots, Z_{t+k-1}$ como el mejor estimador lineal en el sentido de la media cuadrática de Z_{t+k} como función lineal de $Z_{t+1}, \dots, Z_{t+k-1}$.

Esto es, si \hat{Z}_{t+k} es el mejor estimador entonces:

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+1} + \alpha_2 Z_{t+2} + \dots + \alpha_{k-1} Z_{t+k-1} \quad (3.8)$$

donde α_i ($1 \leq i \leq k-1$) son los coeficientes de regresión lineal de mínimos cuadrados obtenidos minimizando:

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 = E(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})^2 \quad (3.9)$$

El método de minimización por diferencias da lugar al siguiente sistema de ecuaciones:

$$\gamma_i = \alpha_1 \gamma_{i-1} + \alpha_2 \gamma_{i-2} + \dots + \alpha_{k-1} \gamma_{i-k+1} \quad (1 \leq i \leq k-1) \quad (3.10)$$

Por tanto,

$$\rho_i = \alpha_1 \rho_{i-1} + \alpha_2 \rho_{i-2} + \dots + \alpha_{k-1} \rho_{i-k+1} \quad (1 \leq i \leq k-1) \quad (3.11)$$

En notación matricial,

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{pmatrix} \quad (3.12)$$

Similarmente

$$\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \dots + \beta_{k-1} Z_{t+k-1} \quad (3.13)$$

donde β_i ($1 \leq i \leq k-1$) son los coeficientes de regresión lineal de mínimos cuadrados obtenidos minimizando:

$$E(Z_t - \hat{Z}_t)^2 = E(Z_t - \beta_1 Z_{t+1} - \beta_2 Z_{t+2} - \dots - \beta_{k-1} Z_{t+k-1})^2 \quad (3.14)$$

Por tanto,

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-1} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{pmatrix} \quad (3.15)$$

lo que implica que $\alpha_i = \beta_i (1 \leq i \leq k-1)$.

Se sigue que la autocorrelación parcial entre Z_t y Z_{t+k} será igual a la autocorrelación ordinaria entre $(Z_t - \hat{Z}_t)$ y $(Z_{t+k} - \hat{Z}_{t+k})$. Por lo tanto, si P_k es la autocorrelación parcial entre Z_t y Z_{t+k} , se tiene:

$$P_k = \frac{Cov[(Z_t - \hat{Z}_t)(Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}} \quad (3.16)$$

Ahora bien,

$$\begin{aligned} Var(Z_{t+k} - \hat{Z}_{t+k}) &= E[(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})^2] = \\ &= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})] - \\ &\quad - \alpha_1 E[Z_{t+k-1}(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})] - \dots \\ &\quad \dots - \alpha_{k-1} E[Z_{t+1}(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})] = \\ &= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+1} - \alpha_2 Z_{t+2} - \dots - \alpha_{k-1} Z_{t+k-1})] \end{aligned} \quad (3.17)$$

porque todos los términos restantes son cero por la ecuación (3.10). Por tanto:

$$Var(Z_{t+k} - \hat{Z}_{t+k}) = Var(Z_t - \hat{Z}_t) = \gamma_0 + \alpha_1 \gamma_1 + \dots + \alpha_{k-1} \gamma_{k-1} \quad (3.18)$$

Después, usando que $\alpha_i = \beta_i (1 \leq i \leq k-1)$, se tiene:

$$\begin{aligned} &Cov[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})] = \\ &= E[(Z_t - \alpha_1 Z_{t+1} - \dots - \alpha_{k-1} Z_{t+k-1})(Z_{t+k} - \alpha_1 Z_{t+k-1} - \dots - \alpha_{k-1} Z_{t-1})] = \\ &= E[(Z_t - \alpha_1 Z_{t+1} - \dots - \alpha_{k-1} Z_{t+k-1})Z_{t+k}] = \\ &= \gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1 \end{aligned} \quad (3.19)$$

Por tanto:

$$P_k = \frac{\gamma_k - \alpha_1 \gamma_{k-1} - \dots - \alpha_{k-1} \gamma_1}{\gamma_0 + \alpha_1 \gamma_1 + \dots + \alpha_{k-1} \gamma_{k-1}} = \frac{\rho_k - \alpha_1 \rho_{k-1} - \dots - \alpha_{k-1} \rho_1}{1 - \alpha_1 \rho_{k-1} - \dots - \alpha_{k-1} \rho_1} \quad (3.20)$$

Solucionando el sistema de la ecuación (3.12) por Cramer se obtiene:

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_{i-1} & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_{i-2} & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}} \quad (3.21)$$

Las dos matrices son casi iguales, salvo por la columna i -ésima. Sustituyendo este α_i en (3.20) y multiplicando tanto el numerador como el denominador por el determinante

$$\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{vmatrix} \quad (3.22)$$

se puede obtener el valor de P_k como el ratio entre los dos determinantes

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}} \quad (3.23)$$

Ejemplo 3.1. Sea Z_t el doble del valor del t -ésimo lanzamiento de un dado. Si el dado se lanza 3 veces de manera independiente, tenemos un proceso estocástico $Z_t = Z(X, t)$, donde $t \in \{1, 2, 3\}$ y X pertenece al espacio muestral

$\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. Para un X particular, por ejemplo $(1, 5, 3)$, la función de muestra será $(2, 10, 6)$.

Si un dado es lanzado independientemente repetidas veces, el proceso estocástico será $Z_t = Z(X, t)$ con el conjunto de índices son los positivos enteros y el espacio muestral es $\{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \dots$. Un X particular, $(1, 3, 4, 5, \dots)$ tendrá función de muestra $(2, 6, 8, 10, \dots)$. Este proceso es claramente estrictamente estacionario.

Definición 3.3. Un proceso $\{a_t\}$ es llamado **proceso de ruido blanco** si es una secuencia de variables aleatorias incorreladas de una distribución fija con media constante $E(a_t) = \mu_a$, normalmente se asume que dicha media es 0, varianza constante $Var(a_t) = \sigma_a^2$ y γ_k dada por:

$$\gamma_k = \begin{cases} \sigma_a^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases} \quad (3.24)$$

3.1. Modelos autorregresivos

Para facilitar el análisis de series temporales se pueden introducir dos representaciones si se trabaja con procesos estacionarios. La primera es intentar escribir el proceso estacionario en el tiempo t , Z_t , en función de sus valores en tiempos anteriores, llamada **representación autorregresiva** (AR):

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \dots + a_t \quad (3.25)$$

donde a_t es un ruido blanco y $\dot{Z}_t = Z_t - \mu$.

Definiendo el operador *backshift* como $B^j x_t = x_{t-j}$ se puede reescribir 3.25 como:

$$\pi(B) \dot{Z}_t = a_t \quad (3.26)$$

donde $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$. Todo proceso autorregresivo es invertible. Para que un proceso $Z_t = \pi(B)Z_t$ sea estacionario las raíces de $\pi(B) = 0$ como

función de B deben residir fuera del disco unidad, es decir, si β es raíz de $\pi(B)$, entonces $|\beta| > 1$, siendo $|\cdot|$ la métrica euclídea estándar.

En la práctica, este modelo no es el utilizado al comenzar a modelizar las series temporales, dado que se necesita una cantidad infinita de parámetros imposibles de calcular con un número finito de observaciones. En lugar de este, se utilizan modelos con un número finito de parámetros, es decir, se usa un número de pesos finitos, $\pi_1 = \phi_1, \pi_2 = \phi_2, \dots, \pi_p = \phi_p$ y $\pi_k = 0$ para $k > p$. A este proceso resultante se le denomina **proceso autorregresivo de orden p** , y la representación sería :

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \dots + \phi_p \dot{Z}_{t-p} + a_t \quad (3.27)$$

3.1.1. Modelo autorregresivo de primer orden AR(1)

El modelo autorregresivo de primer orden se escribe:

$$(1 - \phi_1 B) \dot{Z}_t = a_t \quad (3.28)$$

o también como

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t \quad (3.29)$$

El proceso es invertible por ser autorregresivo. Para que sea estacionario, la raíz de $(1 - \phi_1 B) = 0$ tiene que caer fuera del círculo unidad, lo que se traduce en que $|\phi_1| < 1$.

Ejemplo 3.2. *Para ilustrar un ejemplo, se han simulado 250 valores de un proceso autorregresivo de orden 1, $(1 - 0.9B) \dot{Z}_t = a_t$, con el ruido blanco siendo una distribución normal $N(0, 1)$. En la figura 3.1 se muestra la simulación correspondiente.*

Las autocovarianzas se obtienen como:

$$E(\dot{Z}_{t-k} \dot{Z}_t) = E(\phi_1 \dot{Z}_{t-k} \dot{Z}_{t-1}) - E(\dot{Z}_{t-k} a_t) \Rightarrow \gamma_k = \phi_1 \gamma_{k-1} \quad (3.30)$$

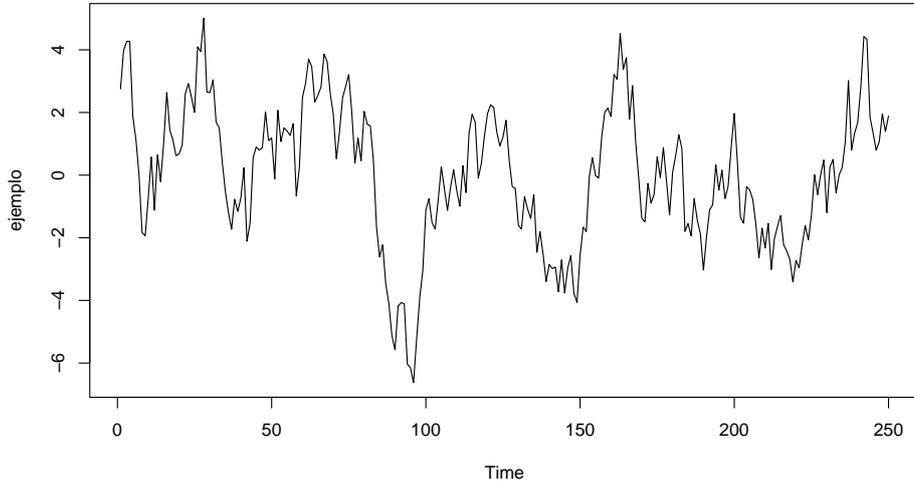


Figura 3.1: Modelo AR(1) simulado, $(1 - \phi_1 B)\dot{Z}_t = a_t$.

Y la función de autocorrelación es:

$$\rho_k = \phi_1 \rho_{k-1} = \phi_1^k \quad (3.31)$$

donde $\rho_0 = 1$. Por lo tanto, cuando el proceso es estacionario, $|\phi_1| < 1$, la función de autocorrelación decae a 0 de la forma ϕ_1^k . Si $0 < \phi_1 < 1$ la función decae al 0 tomando siempre valores positivos y si $-1 < \phi_1 < 0$ la función decae al 0 tomando alternativamente valores positivos y negativos, dependiendo de la paridad de k. Si k es impar, $\rho_k < 0$ y si k es par, $\rho_k > 0$.

3.1.2. Modelo autorregresivo de segundo orden AR(2)

En el modelo autorregresivo de segundo orden se tiene:

$$\dot{Z}_t(1 - \phi_1 B - \phi_2 B^2) = a_t \quad (3.32)$$

o dicho de otra forma que:

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + a_t \quad (3.33)$$

Este modelo, AR(2), como modelo autorregresivo es siempre invertible. Para que sea estacionario las raíces de $(1 - \phi_1 B - \phi_2 B^2) = 0$ deben estar fuera del círculo unidad. Sean B_1 y B_2 las raíces del polinomio $\phi_2 B^2 + \phi_1 B - 1 = 0$. Se tiene:

$$B_1 = \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad (3.34)$$

y

$$B_2 = \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad (3.35)$$

Por lo tanto:

$$\begin{aligned} \frac{1}{B_1} &= \frac{2\phi_2}{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}} = \frac{2\phi_2(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{(-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})} = \\ &= \frac{2\phi_2(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2})}{\phi_1^2 + 4\phi_2 - \phi_1^2} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \end{aligned} \quad (3.36)$$

y análogamente

$$\frac{1}{B_2} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \quad (3.37)$$

Despejando ϕ_1, ϕ_2 en función de B_1, B_2 se obtiene

$$\frac{2}{B_1} - \phi_1 = \sqrt{\phi_1^2 + 4\phi_2} \Rightarrow \left(\frac{2}{B_1} - \phi_1\right)^2 = \phi_1^2 + 4\phi_2 \quad (3.38)$$

$$\frac{2}{B_2} - \phi_1 = -\sqrt{\phi_1^2 + 4\phi_2} \Rightarrow \left(\phi_1 - \frac{2}{B_2}\right)^2 = \phi_1^2 + 4\phi_2 \quad (3.39)$$

Por una parte, igualando ambas expresiones

$$\left(\frac{2}{B_1} - \phi_1\right)^2 = \left(\phi_1 - \frac{2}{B_2}\right)^2 \Rightarrow \frac{2}{B_1} - \phi_1 = \phi_1 - \frac{2}{B_2} \Rightarrow \phi_1 = \frac{1}{B_1} + \frac{1}{B_2} \quad (3.40)$$

donde se ha tomado la raíz positiva para obtener la dependencia de ϕ_1 respecto a B_1, B_2 . Por otro lado

$$\frac{4}{B_1^2} + \phi_1^2 - \frac{4\phi_1}{B_1} = \phi_1^2 + 4\phi_2 \Rightarrow -4\phi_1 = 4\phi_2 B_1 - \frac{4}{B_1} \quad (3.41)$$

$$\frac{4}{B_2^2} + \phi_1^2 - \frac{4\phi_1}{B_2} = \phi_1^2 + 4\phi_2 \Rightarrow -4\phi_1 = 4\phi_2 B_2 - \frac{4}{B_2} \quad (3.42)$$

Igualando ahora ambas expresiones

$$\begin{aligned} 4\phi_2 B_1 - \frac{4}{B_1} &= 4\phi_2 B_2 - \frac{4}{B_2} \Rightarrow 4\phi_2(B_1 - B_2) = 4\frac{B_1 - B_2}{B_1 B_2} \Rightarrow \\ &\Rightarrow \phi_2 = \frac{1}{B_1 B_2} \end{aligned} \quad (3.43)$$

Las condiciones de que $|B_i| > 1$ implican que $|1/B_i| < 1$ para $i = 1, 2$ y por tanto:

$$\begin{aligned} |\phi_2| &= \left| \frac{1}{B_1 B_2} \right| < 1 \\ |\phi_1| &= \left| \frac{1}{B_1} + \frac{1}{B_2} \right| < 2 \end{aligned} \quad (3.44)$$

Por lo tanto tenemos condiciones independientes de B_1, B_2 dependiendo de si las raíces son o no complejas:

$$\begin{cases} -1 < \phi_2 < 1 \\ -2 < \phi_1 < 2 \end{cases} \quad (3.45)$$

Para las raíces reales, $\phi_1^2 + 4\phi_2 \geq 0$, por lo que

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \\ -2 < \phi_1 < 2 \end{cases} \quad (3.46)$$

Para las raíces complejas, $\phi_1^2 + 4\phi_2 < 0$, por lo que

$$\begin{cases} \phi_1^2 + 4\phi_2 < 0 \\ -1 < \phi_2 < 1 \\ -2 < \phi_1 < 2 \end{cases} \quad (3.47)$$

Essto viene representado en la figura 3.2.

Ejemplo 3.3. *Se han simulado 250 valores de un proceso autorregresivo de orden 2, $(1 - 1.5B - 0.65B^2)Z_t = a_t$, siendo a_t un ruido blanco siguiendo una distribución normal $N(0, 1)$. Se muestra en la figura 3.3.*

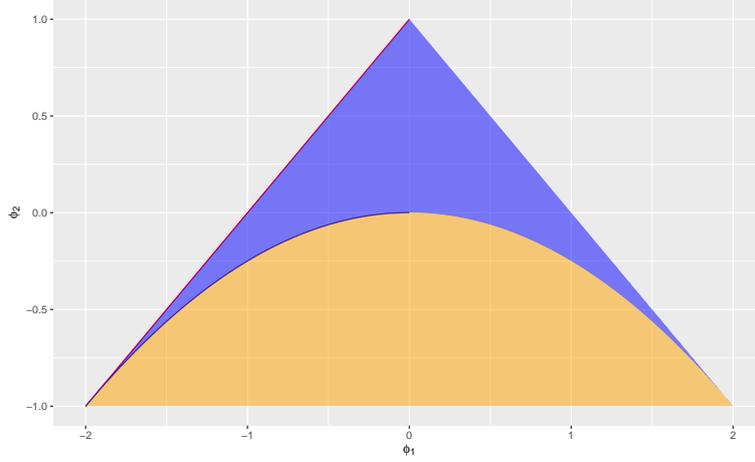


Figura 3.2: Representación gráfica las regiones estacionarias para ϕ_1 y ϕ_2 , en azul las condiciones si las raíces B_1 y B_2 son reales, en naranja si son complejas.

Las covarianzas se obtienen:

$$\begin{aligned} E(\dot{Z}_{t-k}\dot{Z}_t) &= \phi_1 E(\dot{Z}_{t-k}\dot{Z}_{t-1}) + \phi_2 E(\dot{Z}_{t-k}\dot{Z}_{t-2}) + E(\dot{Z}_{t-k}a_t) \Rightarrow \\ &\Rightarrow \gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \end{aligned} \quad (3.48)$$

Y la función de autocorrelación:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad (3.49)$$

Específicamente, cuando $k = 1, 2$:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned} \quad (3.50)$$

lo que implica:

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1-\phi_2} \\ \rho_2 &= \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1-\phi_2} \end{aligned} \quad (3.51)$$

Para resolver recursivamente el resto de ACF, es útil el siguiente teorema.

Teorema 3.1. [16] Sea $C(B)Z_t = 0$ una ecuación en diferencias lineal donde $C(B) = 1 + C_1B + C_2B^2 + \dots + C_nB^n$. Si $C(B) = \prod_{i=1}^N (1 - R_iB)^{m_i}$,

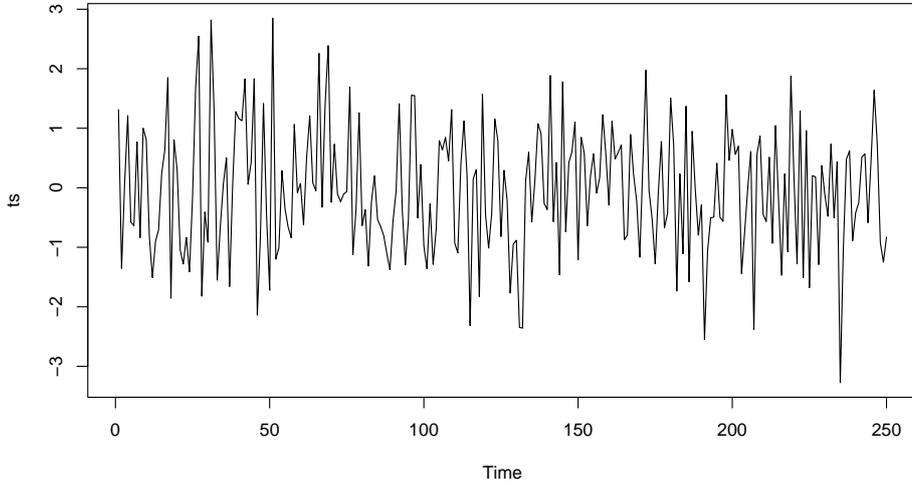


Figura 3.3: Modelo AR(2) simulado, $(1 - 1.5B - 0.65B^2)Z_t = a_t$

donde $\sum_{i=1}^N m_i = n$ y $B_i = R_i^{-1}$, $i = 1, \dots, N$ son las raíces de $C(B) = 0$ de multiplicidad m_i , entonces $Z_t = \sum_{i=1}^N \sum_{j=1}^{m_i-1} b_{ij} t^j R_i^t$. En particular, si $m_i = 1 \forall i$ y los R_i^{-1} son distintos $\forall i$, se tiene $Z_t = \sum_{i=1}^n b_i R_i^t$.

Haciendo uso de el teorema, se obtiene:

$$\rho_k = \begin{cases} b_1 \left(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \right)^k + b_2 \left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \right)^k, & \text{si } \phi_1^2 + 4\phi_2 \neq 0 \\ (b_1 + b_2 k) \left[\frac{\phi_1}{2} \right]^k & \text{si } \phi_1^2 + 4\phi_2 = 0 \end{cases} \quad (3.52)$$

donde las constantes b_1, b_2 se pueden obtener con las condiciones iniciales dadas en la ecuación (3.50). Por lo tanto, la ACF decaerá exponencialmente si las raíces de $(1 - \phi_1 B - \phi_2 B^2) = 0$ son reales y seguirá un comportamiento de un seno amortiguado si las raíces son complejas.

Generalizando este proceso que se acaba de desarrollar, se puede encontrar la función de autocorrelación de un proceso autorregresivo de orden p ,

obteniendo:

$$\rho_k = \sum_{i=1}^N \sum_{j=1}^{m_i-1} b_{ij} k^j R_i^k \quad (3.53)$$

siendo R_i^{-1} las raíces de $\phi_p(B) = 0$ con multiplicidad m_i

3.2. Modelos de media móvil

Otra manera útil es representar un proceso Z_t como combinación lineal de una secuencia de variables aleatorias incorreladas, es decir:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (3.54)$$

donde $\psi_0 = 1$, $\{a_t\}$ es un proceso de ruido blanco con media 0 y $\sum_{j=0}^{\infty} \psi_j < \infty$.

Recordando el operador *backshift* y que $\dot{Z}_t = Z_t - \mu$ se puede escribir 3.54 como:

$$\dot{Z}_t = \psi(B)a_t \quad (3.55)$$

donde $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$. A esta representación se la llama media móvil. Para que un modelo $Z_t = \psi(B)a_t$ sea invertible, las raíces de $\psi(B) = 0$ vista como una función de B tienen que residir fuera del círculo unidad. Esto es, si β es raíz de $\psi(B)$, luego $|\beta| > 1$, donde $|\cdot|$ es la métrica Euclídea estándar.

Como se ha hecho en la ecuación (3.27) se elige un número finito de pesos, $\psi_1 = -\theta_1, \dots, \psi_q = -\theta_q$ y $\psi_k = 0$ para $k > q$. Al proceso resultante se le denomina media móvil de orden q y la representación es:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.56)$$

3.2.1. Modelo de media móvil de primer orden MA(1)

Se representa de la siguiente manera:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 B)a_t \quad (3.57)$$

donde $\{a_t\}$ es un ruido blanco de media 0 y varianza constante σ_a^2 . La autocovarianza viene dada por [16]:

$$\gamma_k = \begin{cases} (1 + \theta_1^2)\sigma_a^2 & \text{si } k = 0 \\ -\theta_1\sigma_a^2 & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases} \quad (3.58)$$

y la función de autocorrelación por:

$$\rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2} & \text{si } k = 0 \\ 0 & \text{si } k > 1 \end{cases} \quad (3.59)$$

El proceso MA(1) siempre es estacionario, dado que el ruido blanco lo es y por tanto, una transformación constante en el tiempo del mismo lo será. Para que sea invertible, la raíz de $(1 + \theta_1 B) = 0$ tiene que estar fuera del círculo unidad, por lo que $|\theta_1| < 1$.

Ejemplo 3.4. *Se han simulado 250 valores de un proceso de media móvil de orden uno para ejemplificar, $Z_t = (1 - 0.5B)a_t$, siendo a_t un ruido blanco siguiendo una distribución normal $N(0, 1)$. El resultado se muestra en la figura 3.4.*

3.2.2. Modelo de media móvil de segundo orden MA(2)

El proceso de media móvil de segundo orden viene dado por:

$$\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2)a_t \quad (3.60)$$

siendo $\{a_t\}$ un ruido blanco. Para que este proceso sea invertible, las raíces de $(1 - \theta_1 B - \theta_2 B^2) = 0$ tienen que residir fuera del círculo unidad, obteniendo:

$$\begin{cases} \theta_1 + \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1 \\ -2 < \theta_1 < 2 \end{cases} \quad (3.61)$$

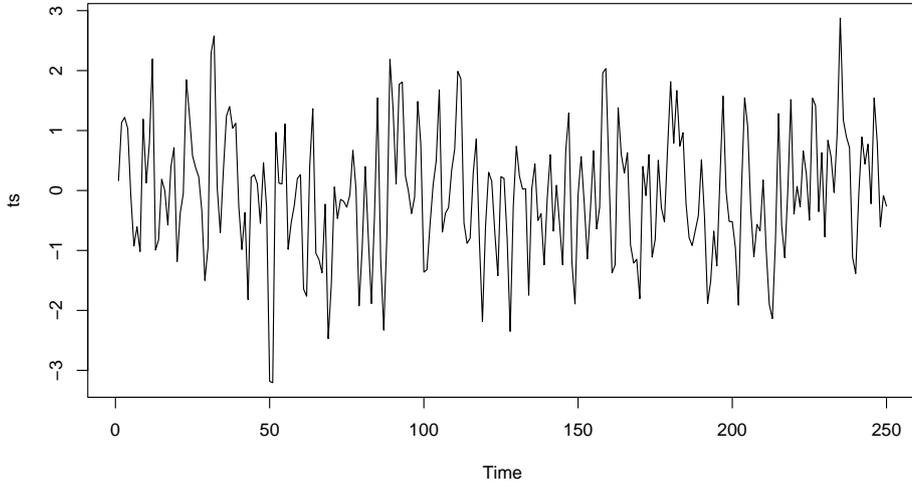


Figura 3.4: Modelo AM(1) simulado, $Z_t = (1 - 0.5B)a_t$.

La autocovarianza viene dada por:

$$\begin{aligned}
 \gamma_0 &= (1 + \theta_1^2 + \theta_2^2)\sigma_a^2 \\
 \gamma_1 &= -\theta_1(1 - \theta_2)\sigma_a^2 \\
 \gamma_2 &= -\theta_2\sigma_a^2 \\
 \gamma_k &= 0 \quad k > 2
 \end{aligned} \tag{3.62}$$

y la función de autocorrelación por:

$$\begin{cases} \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} & k = 1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} & k = 2 \\ 0 & k > 2 \end{cases} \tag{3.63}$$

Generalizando para un proceso de media móvil de orden q se obtiene:

$$\gamma_k = \begin{cases} \sigma_a^2(-\theta_k - \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q) & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \tag{3.64}$$

$$\rho_k = \begin{cases} \frac{-\theta_k - \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, 2, \dots, q \\ 0 & k > q \end{cases} \tag{3.65}$$

3.3. Modelo autorregresivo de media móvil AR-MA(p,q)

El problema de los modelos autorregresivos y de media móvil puros es que, aunque sean finitos, pueden contener una gran cantidad de parámetros para poder conseguir una buena aproximación. Una posible solución es una extensión natural de ambos modelos que combine ambas representaciones para intentar reducir ese número de parámetros, lo que deriva en:

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t \quad (3.66)$$

donde

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned}$$

Para que el modelo sea invertible se requiere que las raíces de $\theta_q(B) = 0$ estén fuera del círculo unidad, y para que sea estacionario se requiere que las raíces de $\phi_p(B) = 0$ estén fuera del círculo unidad. También se asume que las raíces de ambos son distintas. Las covarianzas y la función de autocorrelación son:

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad k \geq q + 1 \\ \rho_k &= \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad k \geq q + 1 \end{aligned} \quad (3.67)$$

3.4. Modelo autorregresivo integrado de media móvil ARIMA(p,q)

Hasta ahora se ha trabajado con modelos y series estacionarias, y se han buscado los pesos para los que esto se cumplía. Sin embargo, una serie temporal no siempre cumple esta condición, por lo que hay que modificar el método o buscar otro para convertir esas series en estacionarias. Un comportamiento frecuente en series no estacionarias es que la diferencia entre niveles de media locales si se comporta como una serie estacionaria. Este tipo

de series se denominan **series no estacionarias homogéneas**. Por lo tanto, estas se pueden convertir en series estacionarias si se toma la diferencia adecuada.

El comportamiento local de este tipo de series homogéneas es independiente del nivel, por lo tanto, si $\Phi(B)$ es el operador de autorregresión que describe el comportamiento, se tiene que:

$$\Psi(B)(Z_t + C) = \Psi(B)Z_t \quad (3.68)$$

siendo C una constante. Esta ecuación implica que $\Phi(B)$ tiene que ser de la forma:

$$\Psi(B) = \phi(B)(1 - B)^d \quad (3.69)$$

para algún $d > 0$, donde $\phi(B)$ es el operador de autorregresión estacionario. Por lo tanto, la serie $\{Z_t\}$ es no estacionaria pero la serie d -enésimamente diferenciada, $\{(1 - B)^d Z_t\}$, para un entero $d \geq 1$ si lo es. Generalmente, esta serie diferenciada sigue un proceso ARMA(p,q), por lo que se tiene:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (3.70)$$

donde a_t es un ruido blanco, $\phi_p(B)$ es el operador de autorregresión estacionario, $\theta_q(B)$ es el operador de media móvil estacionario y no comparten factores comunes.

3.5. Componente estacional, el modelo SARIMA

Es común ver ciertas series que tienen un comportamiento estacional, es decir, la serie se repite al cabo de cierto periodo de tiempo regular. El periodo más corto para este comportamiento repetitivo se dice **periodo estacional**. Por ejemplo, la serie trimestral de venta de helados es muy alta en verano, por lo que tiene un periodo estacional de 4. También la venta de juguetes mensual crece en Diciembre, su periodo estacional es de 12. Como

ilustración, se muestra en la figura 3.5 el número de pasajeros de avión, en miles, registrados entre los años 1949 y 1960.

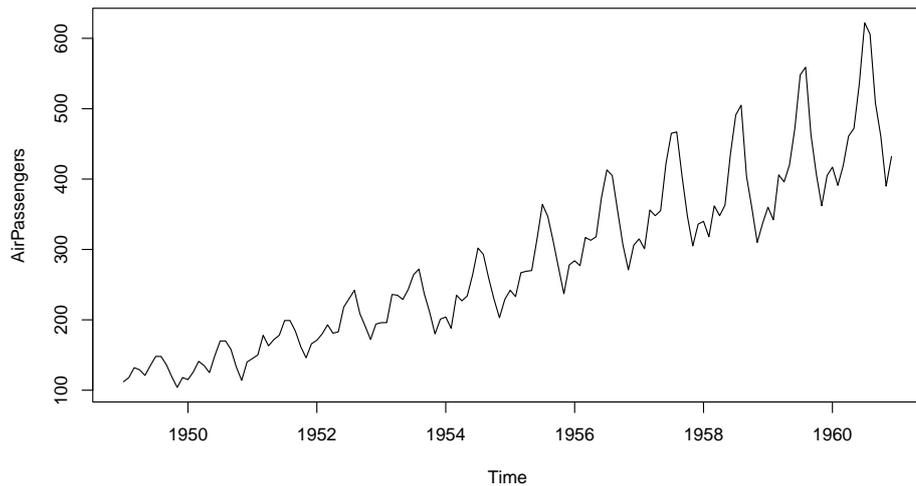


Figura 3.5: Pasajeros mensuales de avión, en miles, desde el año 1949 hasta 1960.

Esta serie tiene claramente una tendencia ascendente y además tiene un periodo estacional. El número de pasajeros aumenta en el mes de marzo para volver a disminuir y aumentar en los meses veraniegos de julio y agosto. Posteriormente vuelven a bajar los números. Este ciclo se repite cada 12 meses, por lo que su periodo estacional es de 12. Para analizar los datos, es útil agregar los datos en una tabla bidimensional, como se muestra en la tabla 3.1, para una serie estacional con periodo s . Se denomina tabla de Buys-Ballot.

La tabla indica que los valores de la serie no solo están relacionados con sus valores adyacentes en t , sino que también están relacionados con los valores que distan de la posición t un numero proporcional a s . Para ilustrar, se toma como referencia el número de pasajeros mensuales de avión, en miles, entre los años 1949 y 1960. Para pronosticar el número de pasajeros

	1	2	3	...	s	Total	Media
	Z_1	Z_2	Z_3	...	Z_s	T_1	\bar{Z}_1
	Z_{s+1}	Z_{s+2}	Z_{s+3}	...	Z_{2s}	T_2	\bar{Z}_2
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	$Z_{(n-1)s+1}$	$Z_{(n-1)s+2}$	$Z_{(n-1)s+3}$...	Z_{ns}	T_n	\bar{Z}_n
Total	T_1	T_2	T_3	...	T_n	T	T/s
Media	\bar{Z}_1	\bar{Z}_2	\bar{Z}_3	...	\bar{Z}_n	T/n	T/ns

Tabla 3.1: Tabla de Buys-Ballot para una serie estacional de periodo s .

en el mes de julio, no solo se tendrá que tener en cuenta los valores de los meses vecinos, como junio y agosto, sino que también se tendrá que tener en cuenta los valores de los meses de julio en años anteriores. En general, la tabla de Buys-Ballot implica que $\{Z_t\}$ contiene información dentro del periodo y entre periodos. Dentro del periodo representa la correlación entre $\dots, Z_{t-2}, Z_{t-1}, Z_t, Z_{t+1}, Z_{t+2}, \dots$ y entre periodos representa la correlación entre $\dots, Z_{t-2s}, Z_{t-s}, Z_t, Z_{t+s}, Z_{t+2s}, \dots$. Sea $\{Z_t\}$ una serie que no se sabe que es estacional y se ajusta a un modelo ARIMA no estacional, es decir,

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)b_t \quad (3.71)$$

Entonces $\{b_t\}$ no será un ruido blanco, puesto que contendrá la información entre periodos. Sea

$$\rho_{j_s} = \frac{E[(b_{t-j_s} - \mu_b)(b_t - \mu_b)]}{\sigma_b^2} \quad (3.72)$$

la función de autocorrelación de $\{b_t\}$ representando la relación entre periodos aún sin explicar. Se puede ver que esta relación entre periodos puede verse como otro modelo ARIMA:

$$\Phi_P(B^s)(1-B^s)^D b_t = \Theta_Q(B^s)a_t \quad (3.73)$$

donde $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^P s$ y $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^Q s$ son polinomios en B^s que no comporten raíces comunes.

Las raíces de estos polinomios caen fuera del círculo unidad y $\{a_t\}$ es un ruido blanco. Combinando las ecuaciones (3.71) y (3.73) se obtiene:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D\dot{Z}_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (3.74)$$

donde

$$\dot{Z}_t = \begin{cases} Z_t - \mu & \text{si } d = D = 0 \\ Z_t & \text{en otro caso} \end{cases} \quad (3.75)$$

A los polinomios $\phi_p(B)$ y $\theta_q(B)$ se les llama los polinomios regulares de autorregresión y media móvil y a $\Phi_P(B^s)$ y $\Theta_Q(B^s)$ los polinomios de autorregresión estacional y de media móvil estacional, respectivamente. Al modelo enseñado en la ecuación (3.74) se le denomina $ARIMA(p, d, q) \times (P, D, Q)_s$, siendo s el periodo estacional.

Capítulo 4

Integral de Choquet Discreta

En este capítulo se definirá el concepto de integral de Choquet e integral de Choquet discreta. La integral de Choquet está siempre asociada a una capacidad, por lo que es necesario introducir antes este concepto.

Definición 4.1. [11] Dado un conjunto X , una función $m : \mathcal{P}(X) \rightarrow \mathbb{R}$ tal que $m(\emptyset) = 0$ y $m(A) \leq m(B)$ siempre que $A \subseteq B$ (monotonía) se llama **capacidad**. Una capacidad que cumple que $m(X) = 1$ se dice que está normalizada.

Ejemplo 4.1. [11] Sea X el conjunto de todos los trabajadores en una empresa, produciendo todos el mismo bien. Para cada $A \subseteq X$ se considera la situación en que los miembros de A trabajan en la empresa. Cada subgrupo A tendrá una manera diferente de trabajar, pero todos lo hacen de la manera más eficiente. Sea $m(A)$ el número de bienes que produce cada grupo A en una hora. La función de conjuntos $m : \mathcal{P}(X) \rightarrow \mathbb{R}^+$ es monótona (dado que se asume que a mayor cantidad de trabajadores mayor producción de bienes) y se anula en el conjunto vacío y por tanto es una capacidad.

Una capacidad no tiene por qué ser aditiva. Volviendo al ejemplo de antes, dados dos subconjuntos A y B de trabajadores, si ambos trabajan por separado se tendría que $m(A \cup B) = m(A) + m(B)$. Pero esto no siempre

es así en tanto que los trabajadores interactúan entre ellos, por lo que la ecuación no siempre se cumple. Aún así, de la monotonía podemos deducir que $m(A \cup B) \geq m(A)$ y $m(A \cup B) \geq m(B)$, en tanto que $A, B \subseteq A \cup B$.

Una vez definida la capacidad, se precisan algunos conceptos más, antes de introducir la integral de Choquet.

Definición 4.2. Dado X un conjunto no vacío, $f : X \rightarrow \mathbb{R}$ se dice simple si su rango es finito.

Definición 4.3. Dado un conjunto no vacío X , \mathcal{A} un álgebra en X , $f : X \rightarrow \mathbb{R}$ se dice **\mathcal{A} -medible** si $\{x : f(x) > t\}$ y $\{x : f(x) \geq t\}$ pertenecen a \mathcal{A} para todo $t \in \mathbb{R}$.

Se llama $B(\mathcal{A})$ al conjunto de funciones acotadas y \mathcal{A} -medibles; $B^+(\mathcal{A})$ al conjunto de funciones acotadas, no negativas y \mathcal{A} -medibles.

Definición 4.4. Sea $f \in B^+(\mathcal{P}(X))$, m una capacidad en $(X, \mathcal{P}(X))$ la **función acumulativa** de f con respecto a m es:

$$G_{m,f}(t) = m(\{x \in X : f(x) \geq t\}) \quad (t \in \mathbb{R}) \quad (4.1)$$

Con todos estos conceptos, ya se puede introducir la definición de integral de Choquet.

Definición 4.5. Sea $f \in B^+(\mathcal{P}(X))$ y m una capacidad en $(X, \mathcal{P}(X))$. Se define la **integral de Choquet** de f con respecto a m como:

$$\int f dm = \int_0^\infty G_{m,f}(t) dt \quad (4.2)$$

donde la parte derecha de la ecuación representa la integral de Riemann.

Sea f una función simple, medible y no-negativa. Al ser simple, $\text{rang } f = \{a_1, \dots, a_n\}$. Sin pérdida de generalidad $0 \leq a_1 \leq a_2 \leq \dots \leq a_n$. Se definen los conjuntos $A_i = \{x \in X : f(x) \geq a_i\}$ para $i = 1, \dots, n$. Por la definición de función acumulativa, se deduce:

$$\int f dm = \sum_{i=1}^n (a_i - a_{i-1}) m(A_i) =: C_m(f) \quad (4.3)$$

La expresión (4.3) es conocida como **integral de Choquet discreta**. Ahora bien, si el conjunto sobre el que se define la función f es un conjunto finito $[n] := \{1, 2, \dots, n\}$ la integral de Choquet discreta es una función de agregación [4]. En este caso los conjuntos A_i serán subconjuntos de $[n]$, $A_i \subseteq [n]$, y la capacidad estará definida sobre el conjunto $[n]$. Otra forma de expresar la integral de Choquet es la siguiente [4]:

$$C_m(f) = \sum_{i=1}^n a_i(m(A_i) - m(A_{i+1})) \quad (4.4)$$

con $A_{n+1} := \emptyset$.

Teniendo en cuenta la definición de una integral como el cálculo del área bajo una función, la integral de Riemann toma rectángulos verticales, donde su alto es el valor de la función, y su ancho es el recorrido en el eje de las abscisas, que se va haciendo cada vez más pequeño. De manera equivalente, la integral de Lebesgue toma rectángulos horizontales y lo que se hace cada vez más pequeño es la altura de esos rectángulos. En este caso, se multiplica la altura de los rectángulos por el ancho combinado del conjunto de valores para los que la función está por encima de un determinado valor. La integral de Choquet, al igual que la integral de Lebesgue, toma rectángulos horizontales. Sin embargo, se deja de asumir que los anchos sean proporcionales a los elementos del conjunto, viniendo ahora determinados por los valores de la capacidad para los correspondientes subconjuntos.

En la figura 4.1 se muestra la función en la que mostraremos los ejemplos para las integrales.

Como se ha mencionado anteriormente, la integral de Riemann toma triángulos verticales, de ancho el recorrido en x y de alto el valor de la función, como se muestra en la figura 4.2. En este caso, el valor que se obtendría para la integral sería:

$$I_R = \sum_{i=1}^n \frac{1}{n} a_i \quad (4.5)$$

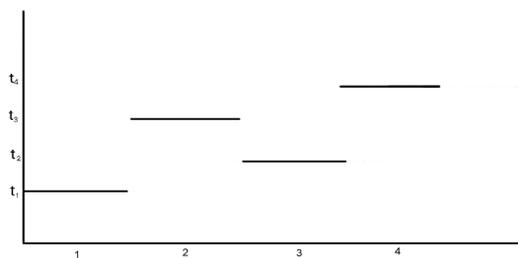


Figura 4.1: Función de ejemplo.

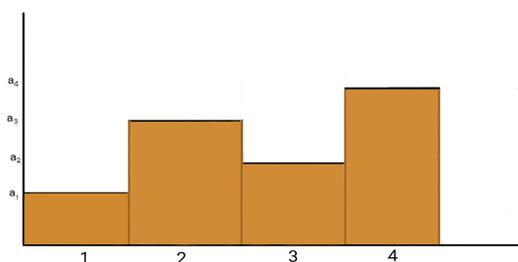


Figura 4.2: Integral de Riemann.

Tomando la integral de Lebesgue, que hace uso de rectángulos horizontales, como se muestra en la figura 4.3, la integral tendría la siguiente expresión (equivalente a la integral de Riemann):

$$I_L = \sum_{i=1}^n \frac{n-i+1}{n} (a_i - a_{i-1}) \quad (4.6)$$

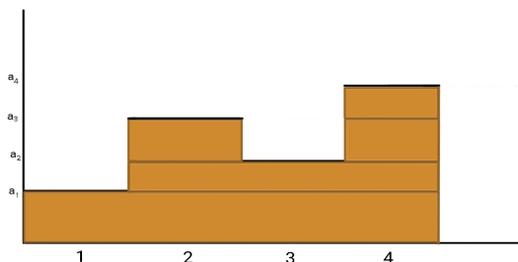


Figura 4.3: Integral de Lebesgue.

Ahora bien, la integral de Choquet, al igual que la integral de Lebesgue, trabaja con rectángulos horizontales, pero los anchos de cada rectángulo vienen dados por los valores de cada rectángulo de la capacidad. Así, la expresión para la integral sería la siguiente:

$$I_C = \sum_{i=1}^n (a_i - a_{i-1}) m(\{j : a_j \geq a_i, j = 1, \dots, n\}) \quad (4.7)$$

y viene representada en la figura 4.4.

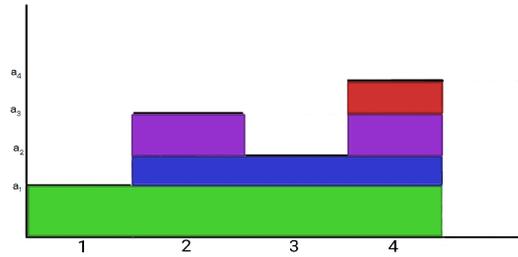


Figura 4.4: Integral de Choquet.

4.1. Propiedades de la integral de Choquet

En esta sección se darán algunas propiedades de la integral de Choquet discreta como función de agregación.

Definición 4.6. [4] Una **función de agregación** es una función $\mathbf{A}^{(n)} : \mathbb{I}^n \rightarrow \mathbb{I}$ que cumple las siguientes condiciones:

1. es no decreciente en cada variable.
2. cumple las condiciones contorno, es decir:

$$\inf_{\mathbf{x} \in \mathbb{I}^n} \mathbf{A}^{(n)}(\mathbf{x}) = \inf \mathbb{I} \quad \sup_{\mathbf{x} \in \mathbb{I}^n} \mathbf{A}^{(n)}(\mathbf{x}) = \sup \mathbb{I} \quad (4.8)$$

donde \mathbb{I} es un intervalo real no vacío.

Una función de agregación es simplemente una función que convierte n valores en uno solo, cumpliendo las anteriores características. Un ejemplo de esta es la media aritmética:

$$\mathbf{AM}^{(n)}(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^n x_i \quad (4.9)$$

Definición 4.7. [4] Sean $\mathbf{x}, \mathbf{x}' \in \mathbb{R}$. Se dice que \mathbf{x}, \mathbf{x}' son **comonótonos** si no existen i, j tales que $x_i > x_j$ y $x'_i < x'_j$.

Definición 4.8. [4] Una función $F : \mathbb{I}^n \rightarrow \mathbb{R}$ es **invariante en escala de intervalo** si para cualesquiera $r, s \in \mathbb{R}, r > 0$ se tiene:

$$F(r\mathbf{x} + s\mathbf{1}) = rF(\mathbf{x}) + s \quad (4.10)$$

para todo $\mathbf{x} \in \mathbb{I}^n$ tal que $r\mathbf{x} + s\mathbf{1} \in \mathbb{I}^n$.

Definición 4.9. Una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ se dice **idempotente** si $\forall x \in \mathbb{R}$ se cumple que $f(x\vec{1}) = x$.

Proposición 4.1. [4] La integral de Choquet satisface las siguientes propiedades:

1. Es lineal con respecto a la capacidad esto es, para cualesquiera capacidades m_1, m_2 definidas sobre $[n]$, cualesquiera $\lambda_1, \lambda_2 \in [0, \infty)$ se tiene:

$$C_{\lambda_1 m_1 + \lambda_2 m_2} = \lambda_1 C_{m_1} + \lambda_2 C_{m_2}$$

2. Es comonóticamente aditiva, es decir, para cualquier par de vectores \mathbf{x}, \mathbf{x}' comonótonos, para cualquier capacidad m sobre $[n]$ se cumple que:

$$C_m(\mathbf{x} + \mathbf{x}') = C_m(\mathbf{x}) + C_m(\mathbf{x}')$$

3. Sean m, m' dos capacidades sobre $[n]$. Entonces $m \leq m'$ si y solo si $C_m \leq C_{m'}$.

4. *Es no decreciente.*
5. *Es estrictamente creciente si y solo si m es una capacidad estrictamente monótona, es decir, $A \subsetneq B$ implica $m(A) < m(B)$.*
6. *Es idempotente si $m([n]) = 1$.*
7. *Es aditiva si y solo si m es aditiva.*
8. *Es una función interna, es decir, $\min(x) \leq C_m(x) \leq \max(x)$, $\forall x$, si m es una capacidad normalizada.*
9. *Si m es una capacidad normalizada, es invariante en escala de intervalo.*

4.2. El diagrama de Hasse de una capacidad

Se ha visto en este capítulo que la integral de Choquet es una función de agregación solo cuando el conjunto sobre el que se define es finito. Ahora bien, si $[n]$ es un conjunto finito y se define $\mathcal{P}([n])$, también será un conjunto finito. Dotado este de la relación de orden de inclusión, mostrada en el ejemplo 2.3 (punto 1) se puede hacer el diagrama de Hasse de un conjunto finito parcialmente ordenado. Si además se define una capacidad m sobre el conjunto $[n]$, es posible asignar valores a las aristas en relación a la capacidad. En este caso, a cada arista se le asignará el valor obtenido de restar el valor de la capacidad del conjunto más pequeño al valor de la capacidad del conjunto más grande. Es decir, si se tienen dos conjuntos $A, B \in \mathcal{P}([n])$ tales que B cubre a A , el valor de la arista que los une será $m(B) - m(A)$, que será siempre positivo ya que $A \subseteq B$.

Ejemplo 4.2. *Sea $X := \{A, B, C\}$ el conjunto de tres trabajadores, Alicia, Bob y Charlie respectivamente. Se quiere hacer un análisis de la productividad al fabricar cajas de madera, que será el valor de la capacidad. La*

productividad total es de una caja de madera, $m(X) = 1$. Alicia sola es capaz de producir 0.2 cajas de madera, $m(A) = 0.2$, Bob 0.5, $m(B) = 0.5$ y Charlie 0.7, $m(C) = 0.7$. Ahora bien, Alicia y Bob trabajan bien juntos, aumentando su productividad conjunta a 0.8, $m(A, B) = 0.8$, pero Charlie es una persona que no trabaja en equipo, bajando su productividad conjunta cuando se junta con el resto de gente, $m(A, C) = 0.8$ y $m(B, C) = 0.9$. Con esto se ha definido una capacidad. Se puede dibujar el diagrama de Hasse y asignar el valor de la resta de las capacidades, tal como puede verse en la figura 4.5.

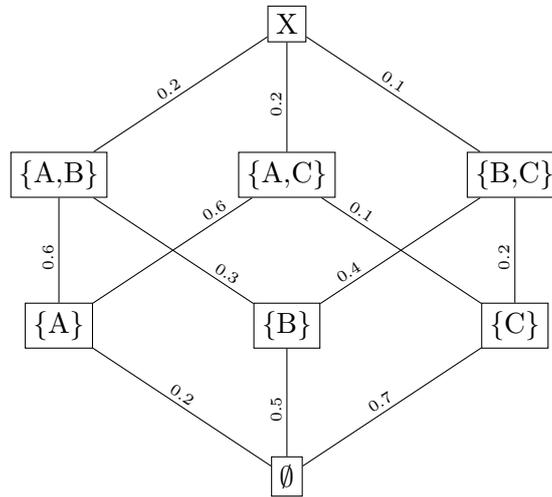


Figura 4.5: Diagrama de Hasse para una capacidad.

Definición 4.10. Sea $\mathcal{P}([n])$ y su diagrama de Hasse asociado, se define **camino ascendente** a un familia de subconjuntos de X , $\{X_1, \dots, X_n\}$ tal que $X_1 \leftarrow X_2 \leftarrow \dots \leftarrow X_n$. Se denota $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$.

Observación 1. Un diagrama de Hasse se puede ver como un grafo no orientado, siendo los vértices los conjuntos del diagrama, las aristas las relaciones de orden y los pesos los valores de las diferencias de las capacidades. Dado que un grafo no orientado se puede ver como un grafo orientado en ambos sentidos, se puede hablar de caminos. También puede verse como un

grafo orientado, poniendo la orientación en el sentido de contenido, aunque no sería una representación estándar. Con esta representación, un camino ascendente sería simplemente un camino.

Proposición 4.2. *Sea $\mathcal{P}([n])$ y su diagrama de Hasse asociado. Según el esquema mostrado (el valor de la arista se corresponde al valor de la capacidad del elemento mayor menos el valor de la capacidad del elemento menor), la suma de las aristas de cualquier camino ascendente desde el conjunto vacío, \emptyset , hasta el conjunto total, X , será igual a $m(X)$. Más aún, si la capacidad está normalizada tendrá un valor de 1.*

Demostración. Dado un camino ascendente $\emptyset \rightarrow X_1 \rightarrow \dots \rightarrow X_n \rightarrow X$, la suma de las aristas viene dada por $(m(X) - m(X_n)) + (m(X_n) - m(X_{n-1})) + \dots + (m(X_2) - m(X_1)) + (m(X_1) - m(\emptyset)) = m(X) - m(\emptyset)$. Por ser m una capacidad, $m(\emptyset) = 0$ y por lo tanto $(m(X) - m(X_n)) + (m(X_n) - m(X_{n-1})) + \dots + (m(X_2) - m(X_1)) + (m(X_1) - m(\emptyset)) = m(X)$. Si además la capacidad está normalizada $m(X) = 1$. \square

Capítulo 5

El modelo de proceso estocástico basado en la integral de Choquet discreta

Este capítulo está dedicado a la definición y estudio de un proceso estocástico temporal basado en la integral de Choquet discreta. Siguiendo la misma idea que un modelo autorregresivo, el proceso en un tiempo t , Z_t , es igual a la integral de Choquet de los k anteriores tiempos, Z_{t-1}, \dots, Z_{t-k} más un ruido blanco. La expresión viene dada por:

$$Z_t = C_m(Z_{t-1}, \dots, Z_{t-k}) + a_t$$

donde m es una capacidad normalizada, C_m su integral de Choquet discreta y a_t un ruido aleatorio.

Una vez propuesto el método, es necesario que sea estrictamente estacionario. Como bien se ha dicho anteriormente, que una serie sea estacionaria implica que oscila en torno a un valor, y no se dispara. Por lo tanto, si el método es estacionario significa que se pueden usar los valores pasados para predecir un valor futuro.

Primero es necesario probar que la integral de Choquet, vista como una función $C_m : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función medible. El espacio medible será $(\mathbb{R}^n, \mathcal{A})$ donde \mathcal{A} es el σ -álgebra de Borel en \mathbb{R}^n . Con este fin se enunciarán algunos resultados que serán útiles.

Lema 5.1. [3] Sea (X, \mathcal{A}) un espacio medible, sea A un subconjunto de X perteneciente a \mathcal{A} , sean $f, g : A \rightarrow \mathbb{R}$ y sea $\alpha \in \mathbb{R}$. Entonces $\alpha f, f + g, f - g, fg, f/g$ (donde el dominio de f/g es $\{x \in A : g(x) \neq 0\}$) son medibles.

Lema 5.2. [3] Sea (X, \mathcal{A}) un espacio medible, sea A un conjunto de X perteneciente a \mathcal{A} y $f, g : A \rightarrow [-\infty, +\infty]$ dos funciones medibles. Los conjuntos $\{x \in A : f(x) < g(x)\}, \{x \in A : f(x) = g(x)\}$ y $\{x \in A : f(x) \leq g(x)\}$ pertenecen a \mathcal{A} .

Haciendo uso de los lemas anteriores se demostrará ahora un resultado necesario para que el modelo sea estacionario.

Teorema 5.3. La integral de Choquet discreta $C_m : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función Borel-medible.

Demostración. Para probar que la función es Borel-medible, basta tomar $a \in \mathbb{R}$ cualquiera y demostrar que el conjunto $C_m^{-1}((-\infty, a))$ es medible. Para ello, sea $a \in \mathbb{R}$ cualquiera pero fijo. Sean $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $f_i(\mathbf{x}) = x_i$ con $i \in \{1, \dots, n\}$, que son medibles. Sea $B_{\sigma_i} = \{x \in B : f_{\sigma(i)} \leq f_{\sigma(i+1)}\}$ para $i = 1, \dots, n - 1$. Estos conjuntos son medibles gracias al lema 5.2. Se define $B_\sigma = \bigcap_{i=1}^n B_{\sigma_i}$ medible por ser intersección finita de conjuntos medibles.

Se tiene ahora un conjunto medible B_σ tal que si $\mathbf{x} \in B_\sigma$, $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$. Sea tal B_σ cualquiera pero fijo, se tiene entonces que las coordenadas de los puntos en ese conjunto están ordenadas según la permutación σ . Observando cómo serían los A_i definidos en la ecuación (4.3), $A_i = \{a \in [n] : x_a \geq x_i\}$, restringidos a B_σ , estos son fijos, independientemente del vector \mathbf{x} considerado.

Por lo tanto $C_m|_{B_\sigma}(\mathbf{x}) = \sum_{i=1}^n (a_i - a_{i-1})m(A_i)$ con los A_i constantes y por tanto independientes de \mathbf{x} y consecuentemente $m(A_i)$ también para todo $i \in \{1, \dots, n\}$. Así, en la expresión se tiene la resta de componentes multiplicadas por constantes. Por el lema 5.1 la resta de funciones medibles y la multiplicación de una función medible por una constante son medibles, por lo que la integral de Choquet finita restringida a B_σ es medible, es decir, $C_m^{-1}((-\infty, a)) \cap B_\sigma$ es medible $\forall \sigma \in \Sigma$ donde Σ es el conjunto de todas las permutaciones del conjunto $[n]$. Por tanto

$$\begin{aligned} \bigcup_{\sigma \in \Sigma} (B_\sigma \cap C_m^{-1}((-\infty, a))) &= \left(\bigcup_{\sigma \in \Sigma} B_\sigma \right) \cap C_m^{-1}((-\infty, a)) = \\ &= \mathbb{R}^n \cap C_m^{-1}((-\infty, a)) = C_m^{-1}((-\infty, a)) \end{aligned} \quad (5.1)$$

es medible por ser unión finita de conjuntos medibles. \square

Una vez probada la medibilidad de la integral, queda concluir que el método es estrictamente estacionario. Para ello se usará el siguiente lema auxiliar.

Lema 5.4. [8] *Dado un modelo autorregresivo no lineal de orden p :*

$$Z_t = \phi(Z_{t-1}, \dots, Z_{t-p}) + a_t \quad (5.2)$$

donde $\phi : \mathbb{R}^p \rightarrow \mathbb{R}$ es una función medible, es estrictamente estacionario si existe $0 < \lambda < 1$ tal que:

$$|\phi(\mathbf{Z})| \leq \lambda \max\{|Z_1|, \dots, |Z_p|\}, \quad \mathbf{Z} = (Z_1, \dots, Z_p)' \quad (5.3)$$

En el lema aparece un $0 < \lambda < 1$ por lo que ha sido necesaria la introducción de una constante $|\Phi| < 1$ para el cumplimiento de las condiciones de dicho lema.

Teorema 5.5. *El proceso estocástico dado por:*

$$Z_t = \Phi C_m(Z_{t-1}, \dots, Z_{t-k}) + a_t \quad (5.4)$$

donde $0 < |\Phi| < 1$, m es la capacidad normalizada que mejor se ajusta a los datos y a_t es un ruido blanco aleatorio, es estrictamente estacionario.

Demostración. El objetivo de la prueba es apoyarse en el lema 5.4 para demostrar la estacionariedad del método, por lo que es necesario encontrar λ , $0 < \lambda < 1$ tal que, para $\mathbf{Z} = (Z_1, \dots, Z_k)'$:

$$|C_m(\mathbf{Z})| \leq \lambda \max\{|Z_1|, \dots, |Z_k|\} \quad (5.5)$$

Dado que m es una capacidad normalizada, según el punto 8 de la proposición 4.1, C_m es una función interna, luego $\min(\mathbf{Z}) \leq C_m(\mathbf{Z}) \leq \max(\mathbf{Z})$, lo que implica que $|C_m(\mathbf{Z})| \leq |\max(\mathbf{Z})| \leq \max(|Z_1|, \dots, |Z_k|)$. Ahora bien, aún es necesario encontrar la constante λ . Para ello, se modifica el modelo de la siguiente manera:

$$Z_t = \Phi C_m(Z_{t-1}, \dots, Z_{t-k}) + a_t \quad (5.6)$$

Considerando de nuevo la proposición 4.1, $|C_m(\mathbf{Z})| \leq |\max(\mathbf{Z})|$, con lo que $\Phi|C_m(\mathbf{Z})| \leq \Phi|\max(\mathbf{Z})|$ y, por tanto, $|\Phi C_m(\mathbf{Z})| \leq \Phi|\max(\mathbf{Z})|$ si $\Phi > 0$. Basta tomar $0 < |\Phi| < 1$ y $\lambda = |\Phi|$ para tener que existe $0 < \lambda < 1$ tal que, para $\mathbf{Z} = (Z_1, \dots, Z_k)'$:

$$|\Phi C_m(\mathbf{Z})| \leq \lambda \max\{|Z_1|, \dots, |Z_k|\} \quad (5.7)$$

Según el lema 5.1 una función medible por una constante también es medible, por lo que $\Phi C_m(Z_{t-1}, \dots, Z_{t-k})$ es medible, por lo tanto se cumplen las condiciones del lema 5.4 con $\phi = C_m$ y $\lambda = |\Phi|$. \square

Por lo tanto, queda un modelo con varios fenómenos a estudiar:

- k que será el número entero de realizaciones previas utilizadas para predecir el valor de la serie.
- Φ , una constante entre -1 y 1 , que indicará la correlación entre el valor t y el valor de la integral de Choquet de los valores $t-1, \dots, t-k$ pasados.

- m , que será una capacidad normalizada que indicará cómo es la relación entre los valores anteriores dentro de la integral de Choquet.

Antes de comenzar con la aplicación del modelo a bases de datos reales, hecho en el siguiente capítulo, se hará un pequeño estudio del efecto de la constante Φ en unos datos simulados, siendo $k = 3$, $a_t \sim \mathcal{N}(0, 1)$ y $m = (0, 0.3, 0.4, 0.5, 0.2, 0.8, 0.9, 1)$, ordenada en binario (000, 001, 010, 011, 100, 101, 110, 111) donde 1 denota que el elemento está en el conjunto sobre el que se está aplicando la capacidad y un 0 que no. Para mejor comprensión se ha hecho el diagrama de Hasse de esta capacidad en la figura 5.1 Se han simulado 1000 datos, tomando los 3 primeros con una $\mathcal{N}(0, 1)$ y descartando posteriormente el 10% inicial de los datos.

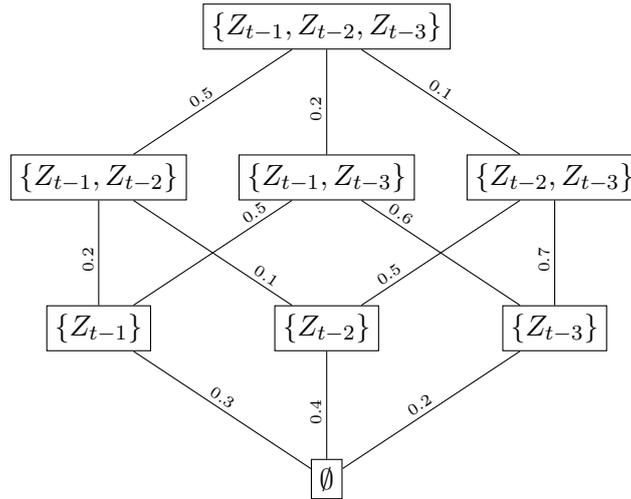


Figura 5.1: Diagrama de Hasse para la capacidad utilizada.

Como se muestra en la figura 5.2 a simple vista se observa que los valores de Φ tales que $|\Phi| < 1$ son estacionarios. Para comprobarlo se ha utilizado la función *adf.test* del paquete **R**. Para las simulaciones en las que $|\Phi| = 2$, la varianza, vista en función de t , tiende a infinito a medida que t crece. Cabe destacar los casos límite en el que $|\Phi| = 1$. En el lado negativo ($\Phi = -1$) se obtiene una serie estacionaria, así comprobado con la función antes

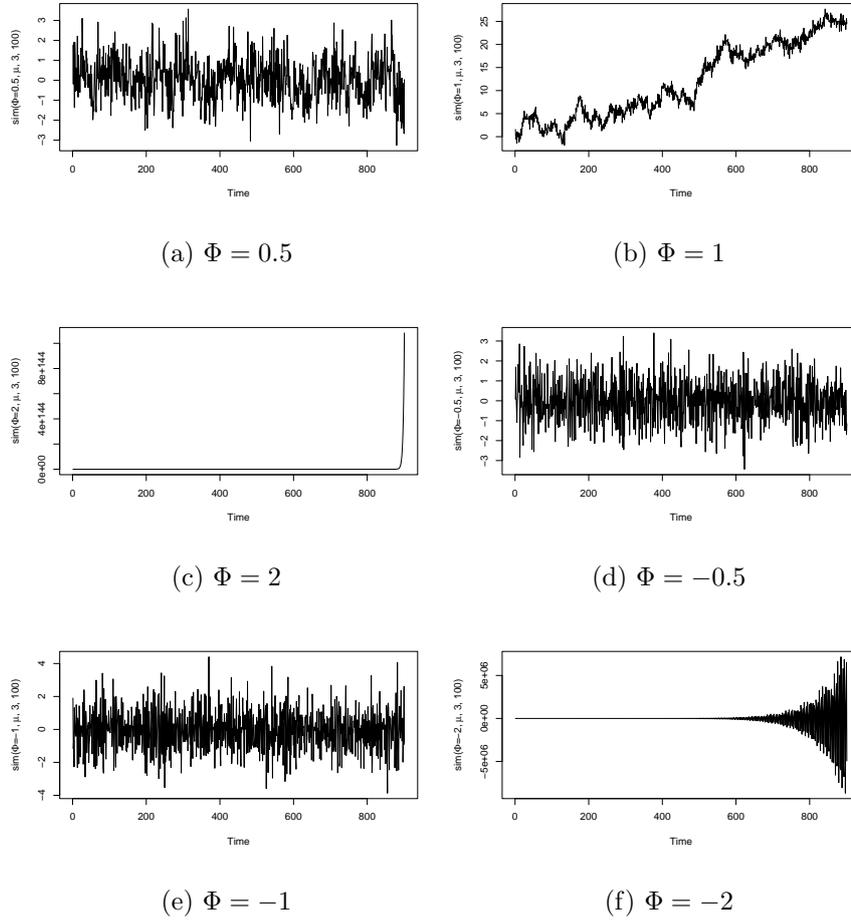


Figura 5.2: Simulación del modelo $Z_t = \Phi C_m(Z_{t-1}, \dots, Z_{t-k}) + a_t$ para distintos valores de Φ .

mencionada. Nótese que en el desarrollo teórico no se ha demostrado que para valores en los que $|\Phi| \geq 1$ la serie no sea estacionaria. Por su parte, la simulación en la que $\Phi = 1$ es un caso singular. Se observa una tendencia positiva y lineal.

Capítulo 6

Estudio del modelo con series temporales reales

El objetivo de este capítulo es intentar predecir valores para series temporales reales utilizando el procedimiento definido en el capítulo anterior y comparar los valores obtenidos con la serie real y con el modelo SARIMA, estudiando como de próximos son. Para ello es necesario conocer los parámetros introducidos: Φ , m y k . Con el fin de conseguir los mejores parámetros se han optimizado los tres. Para ello, la base de datos se divide en dos secciones. Si n es la longitud de la serie temporal, la primera sección consta de los primeros $n - k$ valores, es decir, desde el inicio de la serie hasta el valor $n - k$. Esta será la ventana de entrenamiento, es decir, los valores que se utilizarán para optimizar los parámetros. La ventana comprendida desde $n - k + 1$ hasta n serán los k valores utilizados para predecir.

6.1. El número de realizaciones previas k

Fijar el número k óptimo es crucial para hacer un modelo eficiente. Tomando valores de más no se conseguirá una mejor predicción, si no que el programa creado emplee más tiempo en conseguir el mismo valor a predecir,

por lo que será en vano.

Para obtener el valor de k se hará un estudio de la función de autocorrelación parcial (PACF). Con \mathbb{R} se puede obtener fácilmente representada. Para ilustrar un ejemplo se ha tomado la serie temporal *airmiles* de \mathbb{R} , correspondiente a los miles de pasajeros que vuelan en aerolíneas comerciales en los Estados Unidos para cada año desde 1937 hasta 1960. La serie viene dada por los 24 valores en la tabla 6.1.

	Año		1937	1938	1939	1940	1941	1942	1943	1944
Miles de pasajeros			412	480	683	1052	1385	1418	1634	2178
1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	
3362	5948	6109	5981	6753	8003	10566	12528	14760	16769	
		1955	1956	1957	1958	1959	1960			
		19819	22362	25340	25343	29269	30514			

Tabla 6.1: Valores de los miles de pasajeros que vuelan en aerolíneas comerciales en los Estados Unidos para cada año desde 1937 hasta 1960

En la figura 6.1 se muestra dicha representación con el intervalo de confianza del 95% asociado a que la autocorrelación parcial es 0, en cuyo caso no tiene sentido considerar esas variables en un modelo autorregresivo, siendo estos la banda entre líneas azules. Los valores superiores a esa banda son valores significativos. Si un valor fuese 0, tendría una probabilidad de 0.95 de estar en dicha banda. Para los valores fuera de la banda se rechaza la hipótesis de que la autocorrelación es 0, luego se acepta que no es 0 y tiene sentido considerarlo en el modelo. Por tanto el valor de k será el último valor del retraso que sea significativo, en caso de la figura 6.1 será uno.

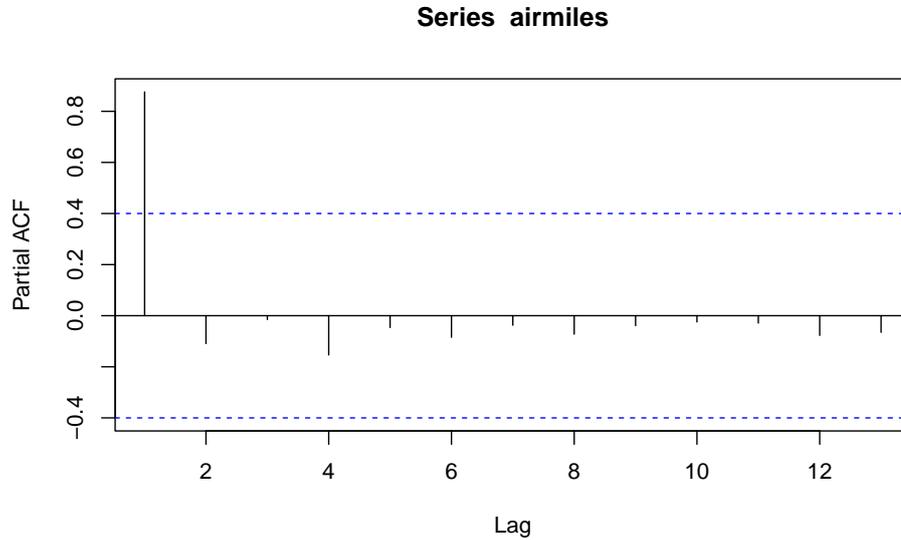


Figura 6.1: Figura que representa la función de autocorrelación parcial (PACF) con respecto a la distancia entre valores.

6.2. La capacidad m

La capacidad es lo que determinará como de importantes son los valores pasados. El valor de la capacidad asociado a cierta realización pasada determina el peso de esa realización en la integral de Choquet. Cuanto más cercano sea ese valor a 1 más importante será. Además, la capacidad tiene en cuenta la relación cuando se toman realizaciones de forma conjunta.

Para implementar la mejor capacidad en el modelo se ha utilizado el paquete `Rfmtool`, que permite calcular una capacidad dando valores minimizando la suma de errores absolutos [2].

6.3. La constante Φ

Por último resta optimizar Φ . Para ello se ha creado un programa que calcula la media de los errores cuadráticos comparando los valores de entrenamiento con la integral de Choquet de esa realización utilizando los k

valores previos.

Para calcular la integral de Choquet y compararla es necesaria una capacidad, por lo que se utilizará otra vez el paquete `Rfmtool`, pero esta vez introduciendo Φ dentro de la capacidad, es decir:

$$Z_t = \Phi C_m(Z_{t-1}, \dots, Z_{t-k}) = C_m(\Phi Z_{t-1}, \dots, \Phi Z_{t-k}) \quad (6.1)$$

Esto es posible gracias a la propiedad de invarianza en escala de la integral de Choquet, vista en la proposición 4.1. Una vez calculados el error cuadrático medio se optimiza la función `optim` con el método `BFGS`.

6.4. Primera comparación

Para hacer una primera predicción se ha utilizado la serie temporal estacional antes mencionada en la figura 3.5. Esta serie no es estacionaria, por lo que es necesario diferenciarla primero para que lo sea. Así, aplicando dos procesos de diferenciación, primero uno que quite la tendencia lineal con retraso 1 y luego otro que quite la componente estacional con retraso 12, se obtiene una serie estacionaria. Como se ha mencionado antes se ha concluido que el mejor k para esta serie es 2. Se muestra en la figura 6.2 se ve una comparación entre el modelo con la integral de Choquet, un modelo SARIMA y la serie original.

Para hacer un análisis numérico y no solo gráfico se ha calculado el MSE en ambos métodos, obteniendo:

$$MSE_{Choquet} = 510.4472$$

y

$$MSE_{SARIMA} = 613.5189$$

Como se puede apreciar, el error cuadrático medio es ligeramente mejor para el modelo hecho con la integral de Choquet, pero es necesario un estudio más amplio para poder sacar conclusiones.

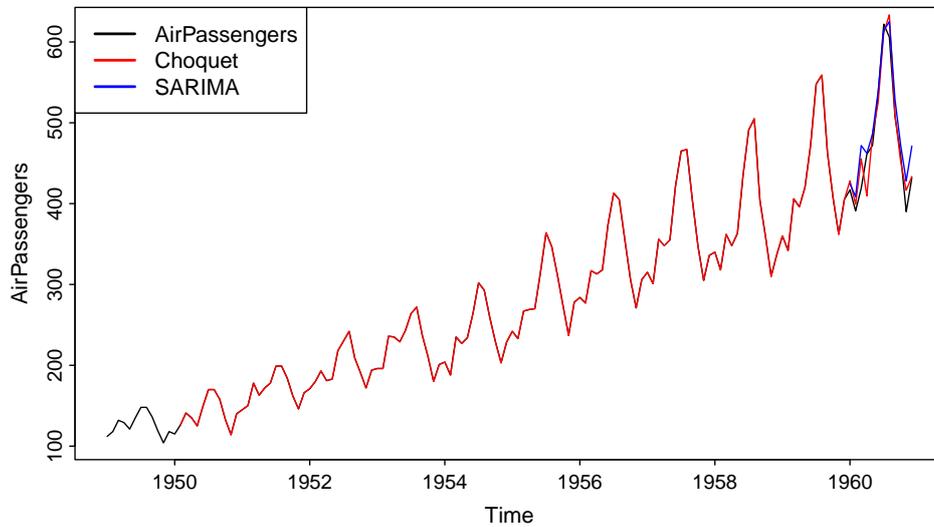


Figura 6.2: Comparación entre SARIMA, modelo con Choquet y serie original.

6.5. Análisis comparativo para varias series

Con el fin de comparar ambos modelos, se han predicho 50 series y se han comparado los errores. Dado que los valores de las series no tienen por que tener el mismo rango, el MSE deja de ser tan útil, por lo que se ha elegido el $MAPE$ para poder comparar bien las series. Para todas las series se ha utilizado una ventana de predicción de 12 valores y una k igual en ambos modelos, pero variando la óptima para cada serie.

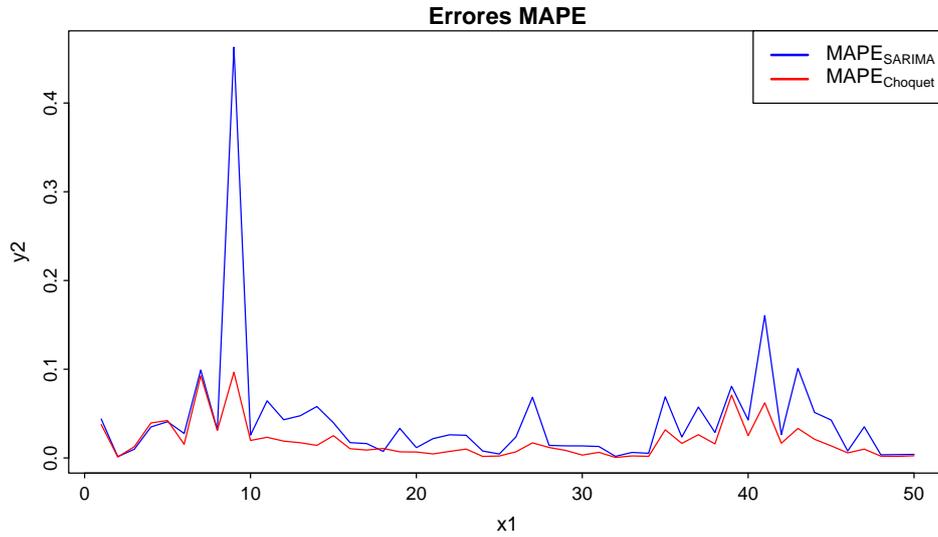


Figura 6.3: Gráfica que muestra los errores $MAPE$ para cada serie en el mismo orden de la tabla 6.4

Nombre de la serie	$MAPE_{Choquet}$	$MAPE_{SARIMA}$	Nombre de la serie	$MAPE_{Choquet}$	$MAPE_{SARIMA}$
<i>AirPassengers</i>	0.0377167	0.04394699	<i>Acciones Iberdrola</i>	0.007390865	0.02611784
<i>co2</i>	0.001129259	0.001429323	<i>Acciones Mapfre</i>	0.009982343	0.02555839
<i>nhtemp</i>	0.01274801	0.009906919	<i>Acciones Grupo Catalana Occidente</i>	0.01179371	0.01418655
<i>nottem</i>	0.03946891	0.03495694	<i>Acciones banco Santander</i>	0.01640713	0.02358033
<i>USAccDeaths</i>	0.04222263	0.0408705	<i>Acciones banco Sabadell</i>	0.02623899	0.05738916
<i>WWUsage</i>	0.0154425	0.0154425	<i>Acciones Caixabank</i>	0.01582815	0.02872357
<i>Johnson.Johnson</i>	0.09254986	0.09916061	<i>Acciones Mediaset</i>	0.02522638	0.04275764
<i>birth</i>	0.03114437	0.03159152	<i>Acciones DIA</i>	0.06216884	0.1603747
<i>UKDriverDeaths</i>	0.09665242	0.4628988	<i>Acciones BBVA</i>	0.01661546	0.02630021
<i>EuStockMarkets\$DAX</i>	0.01976111	0.02581615	<i>Acciones Unicaja</i>	0.03326858	0.1010246
<i>EuStockMarkets\$SMI</i>	0.02330737	0.06440679	<i>Acciones REPSOL</i>	0.02106847	0.05132973
<i>EuStockMarkets\$CAC</i>	0.01899637	0.04308983	<i>Acciones Bankinter</i>	0.01364488	0.04280128
<i>EuStockMarkets\$FTSE</i>	0.01715217	0.01715217	<i>Precio del euro en dólares</i>	0.001744542	0.007809525
<i>UKgas</i>	0.07085635	0.08076566	<i>Precio del euro en libras</i>	0.002202661	0.004475573
<i>Acciones Apple</i>	0.01419437	0.05800407	<i>Precio del euro en francos suizos</i>	0.0005739688	0.001836605
<i>Acciones AMD</i>	0.02511445	0.03959743	<i>Precio del euro en yenes japoneses</i>	0.002290293	0.006233359
<i>Acciones P&S 500</i>	0.01045278	0.01729815	<i>Precio del euro en dólares canadienses</i>	0.001800796	0.005267518
<i>Acciones IBM</i>	0.008920448	0.01621846	<i>Precio del euro en coronas suecas</i>	0.001827024	0.003793551
<i>Acciones Nvidia</i>	0.0106938	0.00739304	<i>Precio del euro en yuanes</i>	0.002473489	0.004032005
<i>Acciones Oracle</i>	0.006901524	0.03333042	<i>Precio del dólar en euros</i>	0.002006264	0.003599334
<i>Acciones META</i>	0.0170356	0.068466	<i>Precio del oro en dólares</i>	0.00689238	0.0236548
<i>Acciones IBEX 35</i>	0.006696305	0.01160301	<i>Precio de la plata en dólares</i>	0.009950476	0.03512711
<i>Acciones Endesa</i>	0.004547334	0.02176098	<i>Precio del petróleo en dólares</i>	0.008480486	0.01362982
			<i>Precio del cobre en dólares</i>	0.006342721	0.01291869
			<i>IPC México</i>	0.003243379	0.01358663
			<i>Índice Crypto 200</i>	0.03172266	0.06895993
			<i>Promedio Industrial Dow Jones</i>	0.005708392	0.007837826
			<i>Media de las observaciones</i>	0.019411959	0.041160251

Figura 6.4: Tablas de errores $MAPE$

Capítulo 7

Conclusiones

Este trabajo fin de grado se ha centrado en la predicción de series temporales. Para ello, se ha hecho una breve introducción sobre los conceptos básicos necesarios para desarrollar los temas importantes, hecho en el capítulo 2. Seguidamente, en el capítulo 3 se ha hecho un breve estudio sobre series temporales y algunos de los métodos más habituales para predecirlas, con énfasis en los modelos ARIMA y SARIMA. En el capítulo 4 se ha investigado en el tema de la integral de Choquet tanto continua como discreta, viendo esta última como función de agregación. A continuación se han unido ambos temas proponiendo un modelo de predicción de series temporales haciendo uso de la integral de Choquet discreta, haciendo un estudio de los parámetros existentes, lo que ha constituido el capítulo 5, un capítulo de resultados novedosos demostrados expresamente para este trabajo. Por último, en el capítulo 6, se ha aplicado el modelo a series temporales reales y se ha comparado con el método SARIMA, obteniendo algunas interesantes conclusiones.

Como norma general el modelo de Choquet es ligeramente mejor que el clásico método SARIMA. Se ha obtenido un error medio porcentual de $MAPE_{Choquet} = 1,94\%$ que, comparado con $MAPE_{SARIMA} = 4,12\%$ es ligeramente inferior.

Las series que tenían componente estacional (*AirPassengers*, *co2*, *nottem*, *USAccDeaths*, *JohnsonJohnson*, *birth*, *UKgas*) tienen los errores porcentuales más altos, obteniendo el mayor error en *JohnsonJohnson* de un 9.25 % quizá debido a esta componente estacional. Para estas mismas series no se puede concluir que el modelo *SARIMA* sea mejor dado que los errores son aproximados, oscilando entre cual de los dos modelos obtiene mejor error.

De todo lo anterior se deduce que en este trabajo se ha presentado una nueva metodología en el análisis de series temporales, que presenta un buen comportamiento respecto a los métodos clásico, pudiendo dar lugar a mejores predicciones.

Los resultados obtenidos son prometedores y hacen que continuar investigación en esta dirección parezca un siguiente paso natural. Así, una primera futura línea de trabajo sería intentar combinar el modelo de Choquet con un modelo de media móvil y comparar tanto con este modelo como con el modelo *SARIMA* para concluir si se mejora o no.

Bibliografía

- [1] N. Beck and J. N. Katz. Modeling dynamics in time-series–cross-section political economy data. *Annual Review of Political Science*, 14:331–352, 2011.
- [2] G. Beliakov, S. James, and J.-Z. Wu. *Discrete fuzzy measures*. Springer, 2020.
- [3] D. L. Cohn. *Measure theory*, volume 1. Springer, 2013.
- [4] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*, volume 127. Cambridge University Press, 2009.
- [5] J. A. Gras. Uso de los modelos de series temporales como técnica de análisis de los diseños conductuales. *Anuario de psicología/The UB Journal of psychology*, pages 17–34, 1981.
- [6] J. T. Hepworth and S. G. West. Lynchings and the economy: A time-series reanalysis of hovland and sears. *Journal of Personality and Social Psychology*, 55(2):239–247, 1940.
- [7] R. A. Johnson, I. Miller, and J. E. Freund. *Probability and statistics for engineers*. Pearson Education London, 2000.
- [8] O.-S. Lee and M.-H. Kim. On stationarity of nonlinear ar processes with nonlinear arch errors. *Communications of the Korean Mathematical Society*, 17(2):309–319, 2002.

- [9] J. R. B. Martínez. *Vigilancia de la resistencia bacteriana en pediatría y su relación con el uso de antibióticos por medio de análisis de series temporales*. PhD thesis, Universitat de Valencia, 2005.
- [10] D. C. Montgomery, C. L. Jennings, and M. Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
- [11] T. Murofushi, M. Sugeno, et al. Fuzzy measures and fuzzy integrals. *Fuzzy measures and integrals: theory and applications*, pages 3–41, 2000.
- [12] J. Neggers and H. S. Kim. *Basic posets*. World Scientific, 1998.
- [13] L. Paolini. Análisis de la respuesta de la temperatura de superficie al crecimiento urbano utilizando series temporales modis. *Revista de Teledetección*, 38:28–35, 2012.
- [14] V. K. Rohatgi and A. M. E. Saleh. *An introduction to probability and statistics*. John Wiley & Sons, 2015.
- [15] C. Q. Ruiz. *Elementos de inferencia estadística*. Editorial Universidad de Costa Rica, 1993.
- [16] W. W. Wei. Time Series Analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis*. Oxford University Press, 2013.