

# **Pronóstico de tasas de cambio de divisas a través de modelos neuro-difusos Takagi-Sugeno**

Juan David Piedrahita López

Orientador:  
Danilo Bedoya Valencia

Trabajo de grado presentado como requisito parcial para optar por el título de:  
**Matemático**

Universidad de Antioquia  
Facultad de Ciencias Exactas y Naturales  
Instituto de Matemáticas  
El Carmen de Viboral (Ant.)  
Colombia  
2020



# Agradecimientos

*“ Si he visto más lejos, es porque estoy sentado sobre los hombros de gigantes...”*

*Sir Isaac Newton*

Agradezco inmensamente a mi madre, mi padre, mi hermano y hermanas y a todos y cada uno de los miembros de mi familia, ellos hacen parte de mi razón de ser. Agradezco a todos y cada uno de los docentes que hicieron parte de mi formación académica desde el inicio, y que mostraron su apoyo incondicional a pesar de las dificultades iniciales que tuve. También, agradezco a todos y cada uno de mis compañeros de pregrado, han sido fuente incansable de ánimos y de ayuda en mi desarrollo intelectual y personal. Un agradecimiento muy especial al docente Danilo Bedoya Valencia, por su perseverancia y acompañamiento durante la formulación, estructuración y culminación de este proyecto, por su amabilidad, paciencia y dedicación en este proyecto, por su gran disponibilidad. A ellos, y quienes faltaron por mencionar... Muchísimas gracias

# Resumen

En el presente trabajo, se pretende esbozar un sistema adaptativo de inferencia neuro-difusa tipo Takagi-Sugeno para la predicción de tasas de cambio del mercado de divisas. Estará dividido en tres capítulos: en el primero, se muestran los resultados teóricos relativos a teoría de la medida y análisis funcional, que permiten de manera rigurosa justificar predicciones por medio de regresiones en las series de tiempo. Adicionalmente, se muestran funciones y propiedades relativas a las series temporales. En el segundo capítulo, se aborda la teoría relativa a la lógica difusa, el principio de inferencia difusa, la definición de un modelo de inferencia difusa tipo Sugeno y su relación con un modelo ANFIS (Sistema Adaptativo de Inferencia Difusa). Se muestra además un resultado teórico que permite ver los modelos neurodifusos tipo Takagi-Sugeno de primer orden como aproximadores de funciones continuas. Finalmente, en el tercer capítulo, se bosqueja un modelo ANFIS relacionado a modelos de predicción de series temporales aplicado a un conjunto de pares de divisas, se muestran resultados, conclusiones y posible trabajo a futuro.

# Introducción

A menudo, la forma y el motivo de las decisiones del ser humano está mediada por la subjetividad. En gran variedad de aspectos de la vida cotidiana, mediante dicha subjetividad buscamos tomar decisiones que generen algún tipo de beneficio, y que en la medida de lo posible, las mismas conlleven el menor riesgo posible, en pro de nuestro bienestar o algún tipo de meta u objetivo particular. Por ello, es importante que se puedan crear y usar herramientas que permitan tratar dicha subjetividad de un modo más cuantificable, que brinden un mayor y mejor análisis sobre las relaciones de riesgo, pérdida y ganancia que existen a la hora de evaluar nuestras acciones. Este es el objetivo de la lógica difusa; brindar las herramientas necesarias que modelan la imprecisión y complementar conjuntos de información incompletos, aspectos que son consecuencia de la subjetividad inherente en la toma de decisiones.

Existen gran variedad de áreas donde herramientas derivadas de la lógica difusa se ven involucradas, particularmente en donde la imprecisión, la constante toma de decisiones y el análisis de datos son aspectos importantes. Tal es el caso de los mercados financieros: dichos mercados funcionan según una gran variedad de factores, y proveen conjuntos de información para los cuales se hace indispensable tener herramientas para su análisis y modelado. [Cavalcante et al., 2016] relatan que, gracias a la necesidad práctica de estos mercados, su estudio toma gran importancia en campos como estadística matemática, economía y finanzas. Sumado a esto, mecanismos computacionales se tornan relevantes a la hora de manejar grandes cantidades de datos.

Como es bien sabido, gran cantidad de mercados financieros existen y operan actualmente de forma global, y a menudo descentralizada. Según el tipo de activo con el que comercian, dichos mercados varían desde el comercio de materias primas, el intercambio de divisas (Foreign Exchange), y la inversión y flujo de distintos tipos de capital. Gracias a su funcionamiento, dichos mercados a menudo permiten una interacción global de distintos conjuntos de personas, y debido a su alta complejidad, a menudo se requiere profundizar en áreas que van mucho más allá del manejo y análisis de datos. Es el caso del mercado Forex, que juega un papel importante en lo referente al manejo e inversión de monedas internacionales, predicción de valores de cambio y modos de interacción de quienes lo componen, lo que permite definirlo como uno de los mercados económicos más importantes y con mayor influencia a nivel global.

El mercado Forex (Foreign Exchange, por sus siglas en inglés) es un mercado que permite

la interacción a nivel global e inmediata de sus participantes, y que como se mencionó anteriormente, tiene como principal objeto el intercambio y comercio de pares de monedas relativas a un país de referencia. Dicho mercado llega a tener un flujo activo de más de 5 billones de dólares por día según el Banco de Pagos Internacionales, y a menudo es considerado como uno de los mercados financieros más accesibles a nivel general dada la tecnologización de dicho mercado, a pesar de ser concebido inicialmente como un mercado destinado a inversores netamente institucionales. [Contreras V. et.al, 2018, ]

El comercio en el mercado Forex funciona en base a una serie de cambios en los precios de una moneda respecto a otra dada como referencia (pares de monedas). El rango de precios en la moneda inicial alcanzado en un intervalo de tiempo se denomina serie de precios. Dicha serie permite entender el comportamiento del mercado en dicho lapso de tiempo, mediante distintas técnicas y métodos de análisis de información. Al considerar dicha serie de precios como una serie temporal, se puede entonces establecer un vínculo entre las herramientas estadísticas de series de tiempo y su aplicación al conjunto de series de precios mencionadas anteriormente. Así, el pronóstico de series de tiempo da información adicional respecto a la toma de decisiones en el mercado de divisas, particularmente en la compra y venta de las mismas. Los pronósticos se basan en una serie de análisis de la información que provee la serie de tiempo inicialmente y un conjunto de medios adicionales que ayudan a predecir características del comportamiento futuro de las mismas. Modelos estadísticos, en su mayoría computacionales, permiten un uso relativamente óptimo de los datos que se disponen y brindan información adicional sobre el comportamiento de dicha serie. Aquellos que pronostican el comportamiento y analizan la volatilidad del mercado son bastante usados, y a menudo basados en distintas herramientas y que abordan distintos enfoques: modelos basados en redes neuronales artificiales, algoritmos genéticos para su optimización, modelos ARIMA, modelos basados en métodos de inferencia difusa (Mamdani, Sugeno, Tsukamoto), entre otros. Aunque como se mencionó antes, dichos modelos dan una serie de aspectos clave de dicha serie, lograr una aproximación buena de la misma es bastante complicado, dada la volatilidad del mercado en el que se busca su aplicación.

Uno de los modelos estadísticos que tiene como finalidad la predicción de series de datos (particularmente variaciones de precios en el mercado Forex) parte de la concepción de un modo de razonamiento difuso, conformado por una serie de reglas de la forma if-then dadas por una relación determinada entre los datos de entrada y de salida del modelo. El modelo de inferencia de Sugeno fue inicialmente propuesto por Michio Sugeno y Takago Tomohiro, y actualmente es ampliamente usado en problemas complejos en altas dimensiones, y gracias al conjunto de reglas de las que parte dicho modelo de inferencia difusa, se obtiene una mayor flexibilidad en el uso y el manejo de datos [Pradeepkumar and Ravi, 2018].

Tanto el análisis funcional como la teoría de la medida permiten abordar y entender de manera más formal todo lo concerniente a la teoría de aproximación de funciones y la teoría de series de tiempo. Dichas teorías brindan además una serie de métodos y herramientas que permiten soluciones formales a los problemas concernientes a la aproximación y optimización. Por ejemplo, mediante el teorema de Stone-Weierstrass, un modelo de inferencia

neuro-difusa puede verse como la aproximación de una función continua definida sobre un conjunto compacto. Gracias a la teoría referente a productos interiores de un espacio lineal, la mejor predicción lineal de una serie de tiempo está dada por la combinación lineal de las variables de referencia, dicha serie vista como la realización de un proceso estocástico definido sobre un espacio de medida dado.

Es por todo lo anterior que este trabajo tiene como motivación entender y profundizar la relación entre el análisis funcional, teoría de la medida y los resultados que permiten la construcción de herramientas que tengan como campo de aplicación el modelado y pronóstico de series temporales, mediante un modelo de desiciones cuyas bases están dadas por la teoría de razonamiento difuso desarrollado por Lofti Zadeh. Dichas predicciones se basan en una serie de métodos que muestran una serie de condiciones necesarias para entender el comportamiento de la serie. Para ello, se muestra además el funcionamiento del mercado Forex, sus modos de transacción y las formas en que los participanetes interactúan entre ellos. La aplicación del modelo de inferencia neuro-difusa se hará mediante el software libre Metatrader 5, que además permite la construcción de variedad de modelos con aplicación al mercado Forex.

Más específicamente, este trabajo tiene como objetivo principal, En el marco del espacio de Lebesgue  $L_2(\Omega, \Sigma, \mu)$  de funciones real-valuadas de cuadrado integrable, estudiar las propiedades de la familia de modelos neuro-difusos tipo Takagi-Sugeno que justifican su aplicación en tareas de pronóstico de series temporales estacionarias no lineales. Adicionalmente, emplear dicha familia en el pronóstico de series de tasas de cambio de divisas del mercado Forex y evaluar su capacidad predictiva en este contexto. Para lograrlo, se plantean principalmente tres objetivos específicos, a saber: plantear el marco teórico para el pronóstico de series temporales estacionarias no-lineales en el contexto de las funciones real-valuadas de cuadrado integrable, asociadas al espacio de Lebesgue  $L_2(\Omega, \Sigma, \mu)$ , mediante el uso de las propiedades derivadas de su condición como espacio de Hilbert, a saber, el cálculo de proyecciones y el teorema de la mejor aproximación, estudiar la propiedad de aproximación universal de la familia de modelos neuro-difusos Takagi-Sugeno, específicamente, estudiar las propiedades que hacen que esta familia sea un conjunto denso dentro de un espacio de funciones adecuado y su relación con el Teorema de Stone-Weierstrass y finalmente identificar y modelar las características estocásticas de un conjunto representativo de series temporales de tasas de cambio dentro del mercado Forex y evaluar su desempeño en pronóstico por medio de las medidas de exactitud usualmente empleadas en análisis de series temporales.

El presente trabajo, estará dividido en tres capítulos, que de forma secuencial establecen las bases para la aplicación del modelo: en el primero, se abordan definiciones y resultados teóricos importantes del análisis funcional, aplicados particularmente al espacio de funciones cuadrado-integrables realvaluadas, así como abordar las propiedades de la esperanza condicional de funciones vista como operador. Se plantea, además, un resultado teórico que permite la predicción lineal y no lineal de una serie de tiempo dado un conjunto de  $n$  observaciones. Se dan las bases generales de una serie de modelos de series temporales que permiten esbozar un comportamiento a diferentes plazos de las mismas. En el segundo capítulo, se aborda la

teoría, conceptos y resultados de lógica difusa, se definen los modelos de inferencia difusa, su estructura y funcionamiento y se muestra la relación entre los modelos de inferencia de Sugeno de orden cero y el teorema de Stone-Weierstrass, un resultado importante concerniente a la aproximación de funciones. Finalmente, en el tercer capítulo, se definirán conceptos puntuales relativos al funcionamiento del software, se plantearán algunas herramientas adicionales y se plantearán las reglas del modelo difuso, se establecerán una serie de modelos de series derivados de los capítulos anteriores y que se usarán en conjunto con el modelo neuro-difuso en cuestión, se examinará su aplicación y los resultados posteriores para un conjunto de pares de divisas.

# Índice general

## Agradecimientos

<b>Resumen</b>	<b>I</b>
<b>Introducción</b>	<b>II</b>
<b>1. Preliminares</b>	<b>1</b>
1.1. Funciones en espacios $\mathcal{L}_2$ y espacios de Hilbert . . . . .	1
1.2. Variables Aleatorias cuadrado-integrables . . . . .	4
1.2.1. Valor esperado condicional . . . . .	4
1.3. Predicción . . . . .	6
1.3.1. Predicción Lineal . . . . .	6
1.3.2. Predicción no lineal . . . . .	8
1.3.3. Series de Fourier . . . . .	9
1.4. Series de tiempo . . . . .	10
1.4.1. Propiedades de Series de Tiempo . . . . .	14
1.4.2. Modelos ARIMA . . . . .	18
<b>2. Aproximación universal</b>	<b>24</b>
2.1. Lógica difusa . . . . .	24
2.1.1. Razonamiento difuso . . . . .	28
2.2. Modelo de inferencia de Sugeno (Modelo TSK) . . . . .	30
2.3. Aproximación universal . . . . .	33
2.4. Sistemas adaptativos de inferencia neuro-difusa (ANFIS) . . . . .	36
<b>3. Construcción y Aplicación</b>	<b>39</b>
3.1. Mercado de Divisas . . . . .	40
3.2. Planteamiento del Modelo . . . . .	42
3.2.1. Predictibilidad . . . . .	43
3.2.2. Tamaño de Inversión Variable . . . . .	46
3.3. Aplicación y Resultados . . . . .	50
3.4. Discusión . . . . .	51
<b>Conclusiones y Trabajos a Futuro</b>	<b>53</b>

**Bibliografía**



# Capítulo 1

## Preliminares

### 1.1. Funciones en espacios $\mathcal{L}_2$ y espacios de Hilbert

Se consideran inicialmente los espacios de medida de funciones cuadrado-integrables, las definiciones básicas y los resultados importantes que permiten, bajo un determinado producto interior, inferir un resultado útil para la predicción de datos.

Dado un espacio de medida  $(\Omega, \mathcal{U}, \mu)$ , se define  $\mathcal{L}_2 := \mathcal{L}_2(\Omega, \mathcal{U}, \mu)$  como el conjunto de todas las funciones medibles  $f : \Omega \rightarrow \mathbb{C}$  (o en  $\mathbb{R}$ , según sea el caso) tal que  $\int |f|^2 d\mu < \infty$ . Se entiende que una función de valores complejos es medible si sus partes real y compleja son respectivamente medibles. En cuyo caso, se tiene por definición que

$$\int f d\mu = \int \operatorname{Re} f d\mu + i \int \operatorname{Im} f d\mu \quad (1.1.1)$$

donde la integral del lado izquierdo existe si y solo si ambas integrales del lado derecho existen y son finitas. Además, se define

$$\langle f_1, f_2 \rangle = \int f_1 \bar{f}_2 d\mu \quad (1.1.2)$$

$$\|f\| = \sqrt{\int |f|^2 d\mu} \quad (1.1.3)$$

$$d(f_1, f_2) = \|f_1 - f_2\| = \sqrt{\int |f_1 - f_2|^2 d\mu} \quad (1.1.4)$$

como un semiproducto, una seminorma, y una semidistancia respectivamente, todas definidas sobre  $\mathcal{L}_2$ . Es importante aclarar que el uso del prefijo semi hace alusión a que, aunque se cumplen la mayoría de las propiedades de distancia, norma y producto interior, las propiedades que implican una igualdad al elemento nulo (por ejemplo  $\|f\| = 0$ ) no necesariamente implican que  $f = 0$ , solo implicaría que  $f = 0 \mu - c.t.p.$

Para poder dotar a  $\mathcal{L}_2$  de una estructura de producto interior (resp. estructura métrica, estructura de norma), se hace la identificación de funciones definida por la siguiente relación de equivalencia: dos funciones  $f, g \in \mathcal{L}_2$  están relacionadas si y solo si

$$f = g \quad \mu - c.t.p.$$

A dicha relación se le denomina  $\mu$ -equivalencia. La clase de equivalencia determinada por  $f$  en  $\mathcal{L}_2$  es denotada por  $[f]$ . A partir de lo anterior, se definen entonces las clases de funciones cuadrado-integrables. Vale aclarar que a partir de este punto, se denotará la clase de un elemento por su representante, esto con el fin de simplificar el manejo de notación.

**Definición 1.1.1 (Clases de funciones cuadrado-integrables)** El espacio de funciones  $L_2 := L_2(\Omega, \mathcal{U}, \mu)$  es el conjunto de todas las clases de equivalencia en  $\mathcal{L}_2$  bajo la relación de  $\mu$ -equivalencia.

**Proposición 1.1.1 (Completez de  $L_2$ )** El espacio métrico  $L_2$  es completo bajo la métrica  $d$ .

Para ver una demostración completa sobre este resultado, cunsúltese [Bartle, 1995].

**Proposición 1.1.2 (Desigualdad de Cauchy-Schwarz)** Para cualquier par de elementos  $f, g$  en  $L_2$  se cumple que  $|\langle f, g \rangle| \leq \|f\| \|g\|$ .

**Definición 1.1.2 (Espacio de Hilbert)** Un espacio de Hilbert es un conjunto equipado con un producto interior que es completo bajo la correspondiente métrica.

El espacio  $L_2$  es un ejemplo de un espacio de Hilbert. Se puede probar que un espacio de Hilbert es isométricamente isomorfo a un espacio  $L_2(\Omega, \mathcal{U}, \mu)$  para algún espacio medible  $(\Omega, \mathcal{U}, \mu)$

**Definición 1.1.3 (Ortogonalidad)** Dos elementos  $f, g \in L_2$  son ortogonales si  $\langle f, g \rangle = 0$ , y se denota por  $f \perp g$ . Dos subconjuntos  $\mathcal{F}, \mathcal{G}$  de  $L_2$  son ortogonales si  $f \perp g$  para toda  $f \in \mathcal{F}$ ,  $g \in \mathcal{G}$  y se denota por  $\mathcal{F} \perp \mathcal{G}$ .

Como un importante resultado, obtenemos el siguiente teorema. Este permitirá obtener las condiciones necesarias para la predicción de datos.

**Teorema 1.1.1 (Mejor aproximación)** Sea  $L \subseteq L_2$  un subespacio lineal cerrado. para cada  $f \in L_2$  existe un único elemento  $\Pi f \in L$  que minimiza  $l \mapsto \|f - l\|$  sobre  $l \in L$ . Este elemento está determinado de manera única por las condiciones  $\Pi f \in L$  y  $f - \Pi f \perp L$ .

A la función  $\Pi f$  del teorema anterior se le denomina proyección (ortogonal) de  $f$  sobre el subespacio  $L$ .

**Justificación.** Al tener que  $\|f - l\|$  es no negativa para  $l \in L$  se tiene que dicho conjunto de reales no negativos está acotado inferiormente, con lo cual se obtiene la existencia de  $d = \inf_{l \in L} \|f - l\|$ . Dado de  $d$  es punto de acumulación del conjunto de elementos de la forma

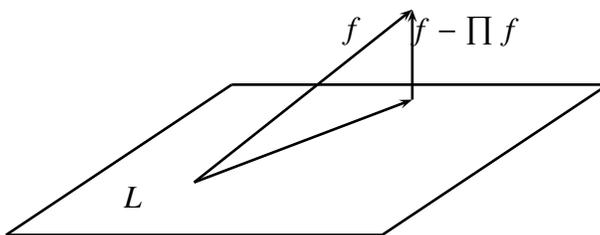


Figura 1.1: Proyección de  $f$  sobre el espacio lineal  $L$ . El vector resultante  $f - \Pi f$  es ortogonal a  $L$ .

$\|f - l\|$ ,  $l \in L$ , se toma una sucesión  $(l_n)_{n \in \mathbb{N}}$  de elementos de  $L$  de tal manera que  $\|f - l_n\| \rightarrow d$  y se muestra que  $(l_n)_{n \in \mathbb{N}}$  es de Cauchy en  $L$ , y mediante la condición de que el subespacio  $L$  sea un subespacio cerrado del espacio completo  $L_2$  se llega a la conclusión de que el límite  $l$  de la sucesión es el que minimiza a  $\|f - l\|$ , que es precisamente  $d$ . Para mostrar la unicidad, se suponen dos elementos  $\Pi_1 f$ ,  $\Pi_2 f$  que cumplen que minimizan  $\|f - l\|$ ,  $l \in L$ . Al definir la sucesión

$$\begin{aligned} l_{2k} &= \Pi_1 f \\ l_{2k+1} &= \Pi_2 f \end{aligned}$$

para  $k \in \mathbb{N}$ , y siguiendo un razonamiento similar al usado anteriormente, se llega a la igualdad.

Finalmente, para mostrar la caracterización, primeramente se debe mostrar que  $\langle f - \Pi f, l \rangle = 0$  para  $l \in L$ . Se considera entonces la igualdad

$$\|f - (\Pi f + al)\|^2 = \|f - \Pi f\|^2 - 2\operatorname{Re} \langle f - \Pi f, al \rangle + \|al\|^2$$

como función de  $a$  (ya sea que  $a \in \mathbb{R}$  o  $a = ci$ ,  $c \in \mathbb{R}$ ), esto con el fin de poder concluir que tanto  $\operatorname{Re} \langle f - \Pi f, al \rangle$  como  $\operatorname{Im} \langle f - \Pi f, al \rangle$  son iguales a cero. La segunda parte de la demostración de dicha caracterización es inmediata, pues al asumir el hecho de que  $\langle f - \Pi f, l \rangle = 0$ , dado que  $f - \Pi f \in L$ , por Teorema de Pitágoras se concluye el resultado.

A partir de la caracterización de ortogonalidad de  $\Pi f$  y de la proposición 1.1.2, se puede ver que el mapeo  $f \mapsto \Pi f$  es lineal, es decir, dicho mapeo cumple las siguientes propiedades:

$$\begin{aligned} \Pi(f + g) &= \Pi f + \Pi g \\ \Pi(\alpha f) &= \alpha \Pi f \\ \|\Pi f\| &\leq \|f\| \end{aligned} \tag{1.1.5}$$

Se puede verificar a partir de las propiedades anteriores de linealidad que dicho mapeo es Lipschitz continuo, es decir

$$\|\Pi f - \Pi g\| \leq \|f - g\|$$

Una aplicación bastante importante, gracias al hecho de ortogonalidad se refiere al hecho de proyecciones iteradas. Si  $\Pi_L f$  denota la proyección de  $f$  sobre el espacio  $L$ , y  $L_1$  y  $L_2$  son subespacios lineales cerrados, entonces

$$\Pi_{L_1} (\Pi_{L_2} f) = \Pi_{L_1} f \quad \text{si } L_1 \subseteq L_2 \quad (1.1.6)$$

Se puede encontrar, por tanto, la proyección por pasos, encontrando la proyección ( $\Pi_{L_2} f$ ) sobre un subespacio  $L_2$  mas grande y luego hallando la proyección de esta sobre  $L_1$ , usando la relación de ortogonalidad.

## 1.2. Variables Aleatorias cuadrado-integrables

Para  $(\Omega, \mathcal{U}, \mu)$  un espacio de medida tal que  $\mu(\Omega) = 1$ , el espacio de Hilbert  $_2(\Omega, \mathcal{U}, \mu)$  es equivale al conjunto de todas las variables aleatorias complejas (o reales)  $X$  con segundo momento finito  $E|X|^2$ . El producto interior se define como la esperanza del producto, es decir  $\langle X, Y \rangle = E(X\bar{Y})$  y el producto interior entre variables centradas es la covarianza:

$$\langle X - EX, Y - EY \rangle = cov(X, Y) \quad (1.2.1)$$

De lo anterior, la desigualdad de Cauchy-Schwarz toma la forma

$$|E(X\bar{Y})|^2 \leq E|X|^2 E|Y|^2$$

Y por tanto se llega a que  $|cov(X, Y)|^2 \leq var(X)var(Y)$ . La convergencia  $X_n \rightarrow X$  relativa a la norma significa que  $E|X_n - X|^2 \rightarrow 0$  pues  $E|X| \leq \sqrt{E|X|^2}$  por la desigualdad de Cauchy - Schwarz. La continuidad del producto interno permite inferir que

$$E|X_n - X|^2 \rightarrow 0, E|Y_n - Y|^2 \rightarrow 0 \Rightarrow cov(X_n, Y_n) \rightarrow cov(X, Y)$$

### 1.2.1. Valor esperado condicional

Sea  $\mathcal{U}_0 \subseteq \mathcal{U}$  una sub  $\sigma$ -álgebra de la  $\sigma$ -álgebra  $\mathcal{U}$ . Entonces, la colección  $L$  de todas las variables  $\mathcal{U}_0$ -medibles  $Y \in L_2(\Omega, \mathcal{U}, \mu)$  es un subespacio lineal cerrado de  $L_2(\Omega, \mathcal{U}, \mu)$  (denotado por  $L_2(\Omega, \mathcal{U}_0, \mu)$ ). Por el teorema 1.1.1, toda variable cuadrado-integrable  $X$  posee una proyección sobre  $L$ . Dicha proyección resulta ser de gran importancia, y será abordada con más detalle.

**Definición 1.2.1 (Valor esperado condicional como proyección)** La proyección de  $X \in L_2(\Omega, \mathcal{U}, \mu)$  sobre el conjunto de todas las variables aleatorias cuadrado-integrables que son  $\mathcal{U}_0$ -medibles es llamada la esperanza condicional de  $X$  dado  $\mathcal{U}_0$ , que es denotada por  $E(X|\mathcal{U}_0)$

Puede darse el caso en el que la  $\sigma$ -álgebra  $\mathcal{U}_0$  sea generada por un mapeo medible  $Y : \Omega \rightarrow \mathbb{D}$  con valores en un espacio medible  $(\mathbb{D}, \mathcal{D})$ , de la siguiente manera: La  $\sigma$ -álgebra generada por  $Y$  se define como  $\sigma(Y) = \{Y^{-1}(\mathcal{D}) | \mathcal{D} \in \mathcal{D}\}$ . La notación  $E(X|Y)$  es una abreviación de  $E(X|\sigma(Y))$ , y es llamada la esperanza condicional de  $X$  dado  $Y$ .

Existe otra interpretación del concepto de esperanza condicional, dada por la siguiente definición.

**Definición 1.2.2 (Valor esperado condicional como variable aleatoria)** *La esperanza condicional dado  $\mathcal{U}_0$  de una variable aleatoria  $X$  la cual es no negativa o integrable es definida como una variable aleatoria  $\mathcal{U}_0$ -medible  $X'$  tal que  $E(X1_A) = E(X'1_A)$  para todo  $A \in \mathcal{U}_0$*

Se tiene que, como una variable aleatoria  $X \in L_2(\Omega, \mathcal{U}, \mu)$  es automáticamente integrable, la definición 1.2.2 da una esperanza condicional para una clase mayor de variables que la definición 1.2.1. Si  $E|X|^2 < \infty$ , de modo que ambas definiciones apliquen, entonces las dos definiciones concuerdan. Para demostrar lo anterior será suficiente probar que una proyección  $E(X|\mathcal{U}_0)$  como en la primera definición es la esperanza condicional  $X'$  de la segunda definición. Ahora, se tiene que  $E(X|\mathcal{U}_0)$  es  $\mathcal{U}_0$ -medible por definición y satisface la igualdad  $E(X - E(X|\mathcal{U}_0))1_A = 0$  para todo  $A \in \mathcal{U}_0$ , dada la relación de ortogonalidad de una proyección. Entonces  $X' = E(X|\mathcal{U}_0)$  satisface los requerimientos de la definición 1.2.2.

La condición de medibilidad requerida en en la  $\sigma$ -álgebra  $\mathcal{U}_0$  mas pequeña dice que la esperanza condicional  $X' = E(X|\mathcal{U}_0)$  es un “engrosamiento” de la variable original  $X$ : está basada en menos información. La definición 1.2.2 muestra que las dos variables tienen los mismos valores promedio  $EX1_A/P(A)$  y  $EX'1_A/P(A)$  sobre todo conjunto medible  $A \in \mathcal{U}_0$ . Se consideran a continuación algunos ejemplos para dar una idea mas clara sobre esperanza condicional.

### Ejemplo (Valor esperado ordinario)

La esperanza  $EX$  de una variable aleatoria  $X$  es un número, y puede ser considerado como una variable aleatoria degenerada. Dicha esperanza es también la esperanza condicional relativa a la  $\sigma$ -álgebra trivial:  $E(X|\{\phi, \Omega\}) = EX$ , lo cual resulta nuevamente de la relación de ortogonalidad. Más generalmente, se tiene que  $E(X|\mathcal{U}_0)$  si  $X$  y  $\mathcal{U}_0$  son independientes, lo que es intuitivamente claro, pues una  $\sigma$ -álgebra independiente  $\mathcal{U}_0$  no brinda información sobre  $X$  y así la esperanza dado  $\mathcal{U}_0$  es la esperanza no condicional.

De las definiciones anteriores se puede establecer una forma de hallar la esperanza condicional de dos variables aleatorias  $X, Y$  dado que  $Y$  tome un valor (es decir,  $Y = y$ ), como se muestra a continuación.

### Ejemplo (Función de densidad de probabilidad condicional)

Al definir  $(X, Y) : \Omega \rightarrow \mathbb{R} \times \mathbb{R}^k$  medible, de tal manera que posea una función de densidad  $f(x, y)$  relativa a una medida producto  $\sigma$ -finita  $\mu \times \nu$  sobre  $\mathbb{R} \times \mathbb{R}^k$  (por ejemplo, la medida de Lebesgue sobre  $\mathbb{R}^{k+1}$ ), se torna entonces natural definir una densidad condicional de  $X$  dado  $Y = y$  por

$$f(x|y) = \frac{f(x, y)}{\int f(x, y) d\mu(x)}$$

Lo anterior está definido para todo  $y$  para el cual el denominador es positivo. Como el denominador es por definición la función de densidad marginal  $f_Y$  de  $Y$  evaluada en  $y$ , para todo  $y$  en un conjunto de medida uno bajo la distribución de  $Y$ , se obtiene que la esperanza condicional está dada por la fórmula

$$E(X|Y) = \int x f(x|Y) d\mu(x)$$

Considerando entonces las definiciones 1.2.1, 1.2.2, se tiene el siguiente lema:

**Lema 1.2.1 (Propiedades)** Para  $\alpha, \beta \in \mathbb{C}$ ,  $\mathcal{U}_0 \subseteq \mathcal{U}$ ,  $X, Y \in \mathcal{L}_2$ , se cumple que:

- i)  $E(E(X|\mathcal{U}_0)) = E(X)$
- ii) Si  $Z$  es  $\mathcal{U}_0$ -medible, entonces  $E(ZX|\mathcal{U}_0) = ZE(X|\mathcal{U}_0)$  (Es necesario, además, que  $X \in L_p(\Omega, \mathcal{U}, \mu)$ ,  $Z \in L_q(\Omega, \mathcal{U}, \mu)$  para  $1 \leq p \leq \infty$  y  $p^{-1} + q^{-1} = 1$ )
- iii) (linealidad)  $E(\alpha X + \beta Y|\mathcal{U}_0) = \alpha E(X|\mathcal{U}_0) + \beta E(Y|\mathcal{U}_0)$
- iv) (positividad) si  $X \geq 0$  Entonces  $E(X|\mathcal{U}_0) \geq 0$
- v) si  $\mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \mathcal{U}$  entonces  $E(E(X|\mathcal{U}_1)|\mathcal{U}_0) = E(X|\mathcal{U}_0)$

## 1.3. Predicción

### 1.3.1. Predicción Lineal

Se parte ahora del hecho de que se tienen los valores  $X_1, \dots, X_n$  de una serie de tiempo estacionaria (éste y conceptos relacionados se verán en la próxima sección) de media cero  $X_t$ , donde  $t$  pertenece a algún conjunto indicador. A partir de dichos valores, se busca entonces encontrar la combinación lineal de las variables dadas que mejor prediga las variables futuras. Lo anterior se conoce como *el problema de predicción lineal*.

**Definición 1.3.1 (Mejor predictor lineal)** Dada una serie de tiempo<sup>1</sup> con media cero  $X_t$ , el mejor predictor lineal de  $X_{n+1}$  es la combinación lineal  $\phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_n X_1$  que minimiza  $E|X_{n+1} - Y|^2$  sobre todas las combinaciones lineales  $Y$  de  $X_1, \dots, X_n$ . El valor mínimo  $E|X_{n+1} - \phi_1 X_n - \phi_2 X_{n-1} - \dots - \phi_n X_1|^2$  es llamado el error cuadrático de predicción.

<sup>1</sup>Conjunto de observaciones hechas en un tiempo  $t$ . Se definirá de mejor manera este concepto mas adelante

En la terminología de la sección anterior, se tiene que el mejor predictor lineal de  $X_{n+1}$  es la proyección de  $X_{n+1}$  sobre el subespacio  $\text{lin}(X_1, \dots, X_n)$  generado por  $X_1, \dots, X_n$ . Un tipo común de notación para dicha proyección es  $\prod_n X_{n+1}$ , para  $\prod_n$  la proyección lineal sobre  $\text{lin}(X_1, \dots, X_n)$ . Los mejores predictores lineales de otras variables aleatorias se definen de manera similar.

Los coeficientes  $\phi_1, \dots, \phi_n$  en la fórmula  $\prod_n X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_n X_1$  dependen de  $n$ , incluso cuando se suprime la dependencia de dicha notación.

Por el Teorema 1.1.1, el mejor predictor lineal se puede encontrar mediante las *ecuaciones de predicción* dadas por

$$\langle X_{n+1} - \phi_1 X_n - \phi_2 X_{n-1} - \dots - \phi_n X_1, X_t \rangle = 0 \quad t = 1, \dots, n \quad (1.3.1)$$

donde  $\langle \cdot, \cdot \rangle$  denota el producto interior en  $L_2(\Omega, \mathcal{U}, \mu)$  con  $\mu(\Omega) = 1$ . Para una serie de tiempo estacionaria (se abordará de manera mas formal el concepto de estacionariedad más adelante) realvaluada, este sistema puede ser escrito de la forma

$$\begin{pmatrix} \gamma_X(0) & \gamma_X(1) & \dots & \gamma_X(n-1) \\ \gamma_X(1) & \gamma_X(0) & \dots & \gamma_X(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_X(n-1) & \gamma_X(n-2) & \dots & \gamma_X(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_n \end{pmatrix} = \begin{pmatrix} \gamma_X(1) \\ \vdots \\ \gamma_X(n) \end{pmatrix} \quad (1.3.2)$$

donde las entradas  $\gamma_X(i), i \in \{1, \dots, n\}$  de la matriz y el vector anteriores están dadas por la funcion de autocovarianza definida como

$$\gamma_X(h) = \text{Cov}(X_t, X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))], \quad h \in \mathbb{N}$$

Si la matriz  $n \times n$  de la izquierda es no singular, entonces los coeficientes  $\phi_1, \dots, \phi_n$  se pueden hallar de manera única. De no ser así, entonces existen múltiples soluciones para el vector  $(\phi_1, \dots, \phi_n)$ , pero cualquier solución dará el mejor predictor lineal  $\prod_n X_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_n X_1$ , debido a que está únicamente determinada por el teorema 1.1.1. Las ecuaciones expresan  $\phi_1, \dots, \phi_n$  en función de la autocovarianza  $\gamma_X$ . Dado que en situaciones prácticas no se conoce dicha función, se utilizan los datos disponibles para hallar los estimados, que se usan después para hallar los correspondientes estimados de  $\phi_1, \dots, \phi_n$  y poder así hallar el predictor.

El error cuadrático de predicción se puede expresar en términos de los coeficientes mediante el Teorema de Pitágoras, donde para una serie de tiempo estacionaria  $X_t$

$$\begin{aligned} E|X_{n+1} - \prod_n X_{n+1}|^2 &= E|x_{n+1}|^2 - E|\prod_n X_{n+1}|^2 \\ &= \gamma_X(0) - (\phi_1, \dots, \phi_n) \Gamma_n(\phi_1, \dots, \phi_n)^T \end{aligned} \quad (1.3.3)$$

siendo  $\Gamma_n$  la matriz de covarianza del vector  $(X_1, \dots, X_n)$ , es decir, la matriz del lado derecho de la ecuación 1.3.2.

Para calcular predicciones de elementos  $X_{n+h}$ ,  $h \geq 1$  se puede proceder de forma similar a la anterior. Sin embargo, para cada  $h$  se tendría que resolver un sistema lineal que puede volverse tedioso si  $n$  se torna grande. Existen algoritmos que ayudan a simplificar en gran medida dichos cálculos, que pueden ser vistos en [der Vaart, 2010].

Si una serie de tiempo  $X_t$  no está centrada en cero, entonces se puede añadir un término constante en el predictor. Se escribe 1 para la variable aleatoria que es equivalente a  $1 - \mu - c.t.p.$  (casi seguramente).

**Definición 1.3.2 (Mejor predictor lineal como proyección)** *El mejor predictor lineal de  $X_{n+1}$  basado en  $X_1, \dots, X_n$  es la proyección de  $X_{n+1}$  sobre el espacio lineal generado por  $1, X_1, \dots, X_n$ .*

Si la serie de tiempo  $X_t$  tiene media cero, entonces la introducción del término constante 1 no es de mucha ayuda. De hecho, la relación  $E(X_t) = 0$  es equivalente por definición a que  $X_t \perp 1$ , esto implica que  $1 \perp \text{lin}(X_1, \dots, X_n)$  y que la proyección de  $X_{n+1}$  sobre  $\text{lin}(1)$  es cero. Se tiene por tanto que la proyección de  $X_{n+1}$  sobre  $\text{lin}(1, X_1, \dots, X_n)$  es la suma de sus proyecciones sobre  $\text{lin}(1)$  y  $\text{lin}(X_1, \dots, X_n)$ . Como la primera proyección es cero, entonces la proyección de  $X_{n+1}$  sobre  $\text{lin}(1, X_1, \dots, X_n)$  es coincide con la proyección de  $X_{n+1}$  sobre  $\text{lin}(X_1, \dots, X_n)$ .

Si la media de la serie de tiempo no es cero, entonces al añadir una constante al predictor se reduce el error de predicción. Mediante un argumento similar al usado en el párrafo anterior, se tiene que para una serie de tiempo con media  $\mu = E(X_t)$  posiblemente no nula, se tiene que

$$\prod_{\text{lin}(1, X_1, \dots, X_n)} = \mu + \prod_{\text{lin}(X_1 - \mu, \dots, X_n - \mu)}(X_{n+1} - \mu) \quad (1.3.4)$$

La igualdad anterior se prueba teniendo en cuenta que para  $L_1, L_2$  subespacios lineales cerrados, se cumple que

$$\prod_{L_1 + L_2} f = \prod_{L_1} f + \prod_{L_2} f \quad \text{si } L_1 \perp L_2 \quad (1.3.5)$$

donde  $L_1 + L_2 := \{l_1 + l_2 | l_i \in L_i\}, i = 1, 2$ .

Así, para calcular la predicción de una serie de tiempo no centrada, se procede de la siguiente manera: se resta la media de todo  $X_t$ , se calcula la proyección de la serie de tiempo centrada  $X_t - \mu$ , y al final se suma la media. Como la función de autocovarianza  $\gamma_X$  se define como el producto interior de los proceso centrados, los coeficientes  $\phi_1, \dots, \phi_n$  de  $X_n - \mu, \dots, X_1 - \mu$  están dados por las ecuaciones de predicción dadas en la igualdad 1.3.2.

### 1.3.2. Predicción no lineal

El método de predicción lineal es bastante usado en el análisis de series de tiempo, debido a que su principal ventaja es su simplicidad: el predictor lineal depende solo de la media

y la función de autocovarianza, de una manera sencilla. Por otro lado, el uso de funciones generales  $f(X_1, \dots, X_n)$  de las observaciones como predictores puede ayudar a decrecer el error de predicción.

**Definición 1.3.3 (Mejor predictor)** *El mejor predictor de  $X_{n+1}$  basado en  $X_1, \dots, X_n$  es la función  $f_n(X_1, \dots, X_n)$  que minimiza  $E|X_{n+1} - f(X_1, \dots, X_n)|^2$  sobre todas las funciones medibles  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .*

Teniendo en cuenta lo visto en la sección 1.2, se tiene entonces que el mejor predictor es la esperanza condicional  $E(X_{n+1}|X_1, \dots, X_n)$  de  $X_{n+1}$  dadas la variables  $X_1, \dots, X_n$ , donde la  $\sigma$ -álgebra definida para dicho conjunto de variables es la menor  $\sigma$ -álgebra tal que cada  $X_i$  es medible, para  $i \in 1, \dots, n$ . Los mejores predictores de otras variables se definen de manera similar como esperanzas condicionales.

### 1.3.3. Series de Fourier

Como un apartado importante y como ilustración de la teoría previa, se definen las series de Fourier y se muestran algunas de sus importantes propiedades. El análisis de dichas series permite, bajo ciertas circunstancias, definir y hallar determinadas propiedades de series de tiempo que presentan un comportamiento cíclico. Dichas series pueden expresarse entonces como sumas de funciones periódicas (particularmente senos y cosenos) producto determinadas amplitudes, teniendo en cuenta además un desfase (desplazamiento) de las mismas.

Considere el espacio de Hilbert  $L_2[\pi, \pi] := L_2([-\pi, \pi], \mathcal{U}, P)$  donde  $\mathcal{U}$  denota la  $\sigma$ -álgebra de Borel en  $[-\pi, \pi]$  y  $P$  la medida de probabilidad uniforme. Para  $f, g \in L_2[-\pi, \pi]$ , se define el producto interior de  $f$  y  $g$  como

$$\begin{aligned} \langle f, g \rangle &= E(f\bar{g}) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx \end{aligned}$$

Para  $n \in \mathbb{Z}$ , se define  $e_n$  como la función complejo-evaluada  $e_n(x) = e^{inx}$ . Dichas funciones tienen la propiedad de ser ortonormales, es decir, se cumple que

$$\langle e_m, e_n \rangle = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases}$$

A partir de lo anterior se obtiene la siguiente definición:

**Definición 1.3.4 (Aproximaciones de Fourier y coeficientes)** *La aproximación de Fourier de  $n$ -ésimo orden a cualquier función  $f \in L_2[-\pi, \pi]$  se define como la proyección de  $f$  sobre el subespacio  $\text{lin}\{e_j \mid |j| < n\}$ , es decir, si  $S_n f$  denota la aproximación de Fourier de orden  $n$ , entonces*

$$S_n f = \Pi_{\text{lin}\{e_j \mid |j| < n\}} f$$

Para un conjunto  $e_1, \dots, e_n$  de un espacio de Hilbert  $\mathcal{H}$  y para  $x \in \mathcal{H}$  se cumple que

$$\Pi_{\text{lin}\{e_1, \dots, e_n\}} = \sum_{i=1}^n \langle x, e_i \rangle e_i$$

Así, la aproximación de Fourier de orden  $n$  toma la forma

$$\begin{aligned} S_n f &= \sum_{j=-n}^n \langle f, e_j \rangle e_j \\ &= \sum_{j=-n}^n \langle f, e_j \rangle e^{ijx}, \quad x \in [-\pi, \pi] \end{aligned} \tag{1.3.6}$$

Se puede demostrar que  $S_n f \rightarrow f$  en media cuadrática, es decir  $\|S_n f - f\|^2 \rightarrow 0$  cuando  $n \rightarrow \infty$ , teniendo en cuenta la desigualdad de Bessel <sup>2</sup> y el hecho de que en una serie convergente, sus colas tienden a cero.

Se tienen algunas propiedades de dichas series (su demostración puede ser consultada en [Brockwell and Davis, 1997]).

**Lema 1.3.1 (Propiedades)**

i)  $L_2[-\pi, \pi] = \overline{\text{span}} \{e_j \mid j \in \mathbb{Z}\}$ , donde  $\overline{\text{span}} \{e_j \mid j \in \mathbb{Z}\}$  es el menor subespacio cerrado de  $L_2[-\pi, \pi]$  que contiene a  $e_j, \forall j \in \mathbb{Z}$ .

ii)  $\|f\|^2 = \sum_{j=-\infty}^{\infty} |\langle f, e_j \rangle|^2$

iii)  $\langle f, g \rangle = \sum_{j=-\infty}^{\infty} \langle f, e_j \rangle \overline{\langle g, e_j \rangle}$

## 1.4. Series de tiempo

Para poder dar un mejor entendimiento a la teoría de las secciones previas, se abordarán conceptos referentes a las series de tiempo, definiciones y ejemplos relacionados que permitan una comprensión mas clara de todo lo desarrollado anteriormente.

**Definición 1.4.1 (Serie de Tiempo)** Se define una serie de tiempo como un conjunto de observaciones  $x_t$  hechas en un tiempo específico  $t$ .

---

<sup>2</sup>Desigualdad de Bessel: Para  $x$  un elemento de un espacio de Hilbert y  $\{e_1, \dots, e_n\}$  un subconjunto ortonormal, se tiene que  $\sum_{i=1}^n |\langle x, e_i \rangle|^2 \leq \|x\|^2$

Dependiendo del tipo del conjunto indizante (el decir, el conjunto a cual  $t$  pertenece), se define el tipo de la serie. Una serie de tiempo es discreta si el conjunto indizante (denotado como  $T_0$ ) es un conjunto contable. Por otro lado, una serie de tiempo continua es aquella en la cual  $t$  corre de manera continua sobre algún intervalo de tiempo  $[a, b]$ . A menudo se usa la notación  $x(t)$  para indicar que la serie es del segundo tipo.

Ahora, para el análisis de una serie de tiempo se debe seleccionar un modelo plausible que permita el modelado de los datos, es decir, se parte del hecho de que el valor  $x_t$  es el resultado de una variable aleatoria  $X_t$  evaluada en  $\omega \in \Omega$  fijo.

**Definición 1.4.2 (Proceso estocástico)** A Una familia de variables aleatorias  $\{X_t | t \in T\}$ , con  $T_0 \subseteq T$ , definidas sobre el espacio de probabilidad  $(\Omega, \mathcal{U}_0, \mu)$  se le denomina proceso estocástico.

**Definición 1.4.3 (Realización de un proceso estocástico)** Las funciones  $\{X_t(\omega) | \omega \in \Omega\}$  definidas sobre  $T$  son conocidas como las realizaciones del proceso  $\{X_t | t \in T\}$

Frecuentemente se usa el término “serie de tiempo” para hacer referencia tanto al conjunto de datos como al proceso del cual son realizaciones. Se debe tener en cuenta además que, de la definición de una variable aleatoria, para cada  $t \in T$  fijo  $X_t$  es de hecho una función  $X_t(\cdot)$  definida sobre  $\Omega$ , y para cada  $\omega \in \Omega$  fijo  $X_t(\omega)$  es una función definida sobre  $T$ .

### Ejemplo (Ruido blanco)

Una secuencia doble infinita de variables aleatorias independientes idénticamente distribuidas  $x_t$  es una serie estrictamente estacionaria. Su función de autocovarianza es, con  $\sigma^2 = \text{var}(X_{x_t})$ ,

$$\gamma_X(h) = \begin{cases} \sigma^2, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Se denomina ruido blanco a una serie estacionaria  $x_t$  con media cero y función de autocovarianza como la dada anteriormente. Cualquier secuencia de variables aleatorias independientes idénticamente distribuidas con varianza finita es un ruido blanco.

### Ejemplo (Autorregresión)

Dada una serie ruido blanco  $\omega_t$  con varianza  $\sigma^2$  y  $\phi \in \mathbb{R}$ , considere las ecuaciones

$$x_t = \theta x_{t-1} + \omega_t$$

A dicha serie se le denomina serie autorregresiva de orden 1, y se denota por AR(1). Se tiene que, para  $|\phi| < 1$ ,  $x_t$  se puede expresar de la forma

$$x_t = \sum_{j=1}^{\infty} \phi^j \omega^{t-j}$$

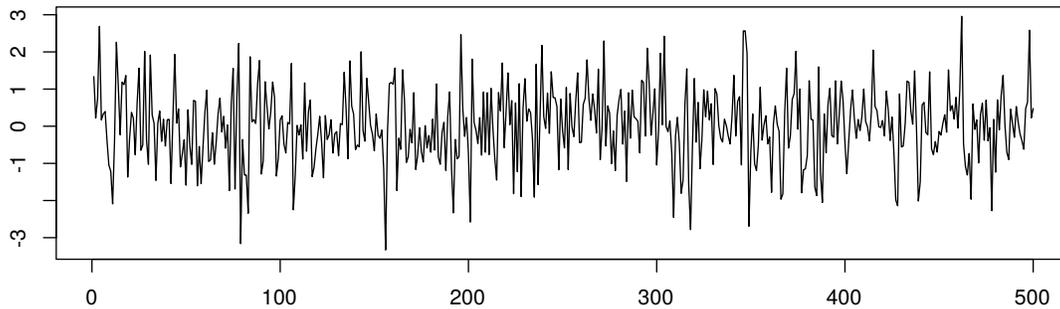


Figura 1.2: Ruido blanco Gaussiano (las variables  $\omega_t$  siguen una distribución normal estándar) generado a partir de una muestra de datos de tamaño 500, con media cero y varianza 1

Se puede demostrar además que dicha serie es convergente  $\mu - c.t.p.$  Al sustituir la expresión anterior en la recursión inicial, se puede verificar que  $x_t$  cumple la igualdad.

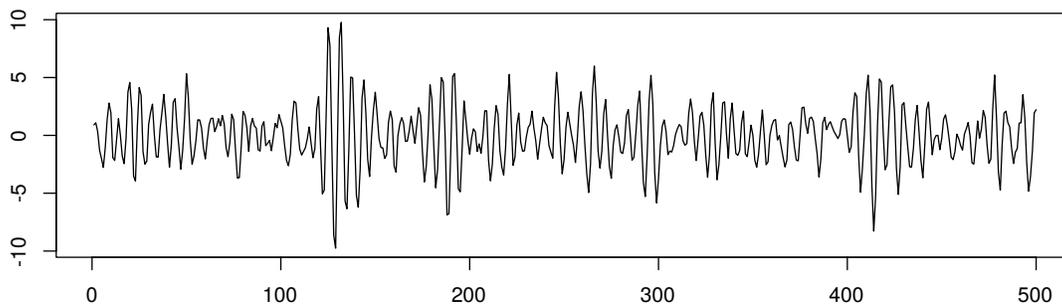


Figura 1.3: Serie autorregresiva dada por la ecuación  $x_t = x_{t-1} - 0.9x_{t-2} + \omega_t$ , para  $\omega_t$  el ruido blanco dado en la figura ??

### Ejemplo (Media móvil)

Para  $\omega_t$  un ruido blanco dado con varianza  $\sigma^2$  y  $\theta$  un escalar, sea

$$x_t = \omega_t + \theta\omega_{t-1}$$

A la serie  $x_t$  se le denomina media móvil de orden 1. Se puede verificar que dicha serie es estacionaria con  $EX_t = 0$ . Se puede verificar que la función de autocovarianza de la media móvil es de la forma

$$\gamma_X(h) = \text{Cov}(\omega_{t+h} + \theta\omega_{t+h-1}, \omega_t + \theta\omega_{t-1}) = \begin{cases} (1 + \theta^2)\sigma^2, & h = 0 \\ \theta\sigma^2, & h = \pm 1 \\ 0, & \text{en otro caso} \end{cases}$$

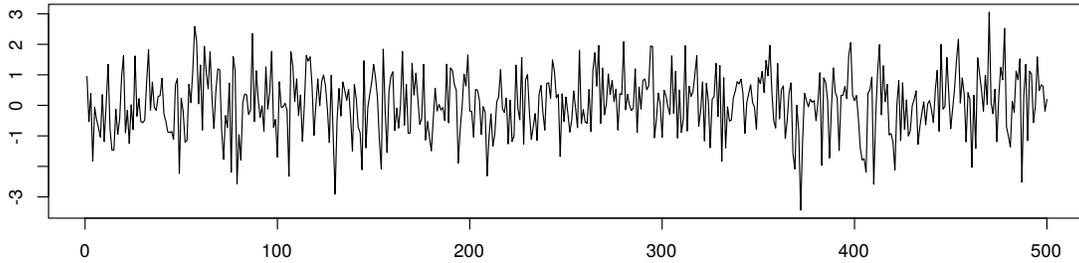


Figura 1.4: Serie de media móvil dada por la ecuación  $x_t = \omega_t + -0.5\omega_{t-1}$ , siendo  $\omega_t$  un ruido blanco Gaussiano

Al tener que  $X_t$  es una variable aleatoria, se puede considerar entonces el problema de hallar el mejor predictor lineal de una serie de tiempo  $X_t$ , como lo muestra el siguiente ejemplo:

#### **Ejemplo (Mejor predictor lineal de una serie autorregresiva)**

Para una serie de tiempo autorregresiva como la dada en el ejemplo 1.4 con  $|\phi| < 1$ , se puede ver que el mejor predictor lineal de  $x_{n+1}$  dadas las variables  $x_1, \dots, x_n$  es  $\phi x_n$ . Por la relación de recurrencia, se tiene que:

$$E(x_{n+1}|x_1, \dots, x_n) = E(\theta x_n + \omega_{n+1}|x_1, \dots, x_n)$$

por el lema 1.2.1 junto con del hecho de que  $X_n \in \text{lin}(x_1, \dots, x_n)$ , se tiene que

$$\begin{aligned} E(\phi x_n + \omega_{n+1}|x_1, \dots, x_n) &= \phi E(x_n|x_1, \dots, x_n) + E(\omega_{n+1}|x_1, \dots, x_n) \\ &= \phi x_n + E(\omega_{n+1}|x_1, \dots, x_n) \end{aligned}$$

dado que por definición  $\omega$  es una secuencia de variables aleatorias independientes con media cero, se obtiene que  $\omega_{n+1}$  es independiente de las variables  $\omega_i$ ,  $i \in 1, \dots, n$ , lo anterior implica que  $E(\omega_{n+1}|x_1, \dots, x_n) = 0$  y por tanto

$$E(x_{n+1}|x_1, \dots, x_n) = \phi x_n$$

Así, para predecir  $x_{n+1} = \theta x_n + \omega_{n+1}$ , basta tomar  $\omega_{n+1}$  e igualarla al valor de la media, que es cero.

### 1.4.1. Propiedades de Series de Tiempo

A menudo, una serie de tiempo puede describirse mediante una suma de componentes, tales como una componente que refleja el comportamiento temporal de la misma, otra que refleja su patrón de tendencia, mas una componente residual. Para poder realizar inferencias estadísticas sobre dichas componentes, es necesario simplificar la estructura que rige el comportamiento de la serie. Una de dichas asunciones, y quizá de las más importantes, es que las propiedades probabilísticas de la serie sean independientes de la variación del tiempo, es decir, que dicha serie posee una especie de *equilibrio probabilístico*. Esto es conocido normalmente bajo el concepto de **estacionariedad**. formalmente, se define como sigue:

**Definición 1.4.4 (Serie estacionaria)** *Un proceso estocástico  $\{X_t\}$  se dice que es estacionario (o estrictamente estacionario) si, para un conjunto de variables aleatorias  $X_{t_1}, \dots, X_{t_n}$ , se cumple que la función de distribución conjunta de dichas variables coincide con la función de distribución conjunta de las variables  $X_{t_1-k}, \dots, X_{t_n-k}$ , para cualquier tupla de coordenadas temporales  $t_1, \dots, t_n$  y cualquier  $k \in \mathbb{Z}$  [Cryer and Chan, 2008]. De manera similar, un proceso es estrictamente estacionario si se cumple que*

$$P(X_{t_1} \leq c_1, \dots, X_{t_n} \leq c_n) = P(X_{t_1-k} \leq c_1, \dots, X_{t_n-k} \leq c_n)$$

para cualquier  $n \in \mathbb{N}$ , para cualquier escogencia de puntos  $t_1, \dots, t_n$  y para cualquier  $k \in \mathbb{Z}$  [Shumway and Stoffer, 2010]

Teniendo en cuenta el hecho de que casi nunca se posee el proceso del que proviene una serie de tiempo, la definición anterior de estacionariedad es muy difícil de aplicar, aunque muy fuerte. Es por ello que se trabaja con una version mas debil de la misma, que defines ciertas condiciones para los dos primeros momentos de la misma. Para definir dicha nocion de estacionariedad, primero se definen una serie de funciones, que serán bastante útiles más adelante:

**Definición 1.4.5 (Medidas de dependencia)** *Sea  $X_t$  una serie de variables aleatorias. Entonces:*

- la **función media** de  $X_t$  se define como

$$\mu_{xt} = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx$$

dado el hecho de que exista.  $E()$  denota el operador de valor esperado (o esperanza), y  $f_t(x)$  denota la función de densidad de la variable  $X_t$ . cuando se sobreentiende la serie, dicha función se denota como  $\mu_t$

- la **función de autocovarianza** se define como

$$\gamma_x(s, t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

para cualquier  $s, t \in \mathbb{Z}$ . Cuando no hay confusión sobre la serie con la que se trabaja, dicha función se denota como  $\gamma(s, t)$

- la **función de autocorrelación (ACF)** de  $X_t$  se define como

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

La función de autocovarianza mide la dependencia lineal entre dos observaciones de la serie  $X_t$ . A pesar de que entre dos puntos cualesquiera  $s, t$  de la serie se cumpla que  $\gamma(s, t) = 0$ , esto no implica que los puntos  $X_s, X_t$  sean independientes, dado que puede existir un tipo de relación entre ellas diferente a la lineal. Por su parte, la ACF se encarga de medir qué tan bien se puede predecir una serie en el tiempo  $t$  usando sólo la variable  $s$ . Puede mostrarse que la ACF cumple que  $-1 \leq \rho(s, t) \leq 1$ . Si se puede predecir el valor  $t$  mediante una relación lineal  $x_t = \beta_0 + \beta_1 x_s$ , entonces el valor de la ACF será cercano a 1 cuando  $\beta_1 > 0$  y cercano a  $-1$  cuando  $\beta < 0$ . De la ACF se obtiene entonces una primera medida de predicción de una serie.

Nuevamente, se debe tener en cuenta que casi nunca se poseen las funciones de densidad de las variables que conforman la serie, dado que *a priori* solo se trabaja con un conjunto de datos. Es por ello que para calcular dichas funciones, se usan estimadores que trabajan con base a los datos, y de cuyos detalles puede verse en [Shumway and Stoffer, 2010]. Para ilustrar mejor la definición 1.4.5, se tienen los siguientes ejemplos:

### **Ejemplo (Funciones de media, autocovarianza y ACF)**

Se tienen los siguientes ejemplos:

- la función media definida para una media móvil dada por

$$v_t = \frac{1}{3} (\omega_{t-1} + \omega_t + \omega_{t+1})$$

siendo  $\{\omega_t\}_{t \in \mathbb{Z}}$  un ruido blanco Gaussiano, es

$$\mu_{v_t} = E(v_t) = \frac{1}{3} [E(\omega_{t-1}) + E(\omega_t) + E(\omega_{t+1})] = 0$$

dado que por definición  $E(\omega_t) = 0$

- la función de autocovarianza de la media móvil definida anteriormente está dada por

$$\begin{aligned} \gamma_v &= E[(v_s - 0)(v_t - 0)] \\ &= \frac{1}{9} E[(\omega_{s-1} + \omega_s + \omega_{s+1})(\omega_{t-1} + \omega_t + \omega_{t+1})] \end{aligned}$$

Es preferible calcular dicha función de autocovarianza como relación de la separación entre  $t$  y  $s$ : si  $h = s - t$  entonces se calculan los valores de dicha función para  $h \in \mathbb{Z}$ .

Para  $h = 0$

$$\begin{aligned}\gamma_v(t, t) &= \frac{1}{9} E [(\omega_{t-1} + \omega_t + \omega_{t+1})(\omega_{t-1} + \omega_t + \omega_{t+1})] \\ &= \frac{1}{9} [E(\omega_{t-1}\omega_{t-1}) + E(\omega_t\omega_t) + E(\omega_{t+1}\omega_{t+1})] \\ &= \frac{3}{9}\end{aligned}$$

Para  $h = 1$

$$\begin{aligned}\gamma_v(t + 1, t) &= \frac{1}{9} E [(\omega_t + \omega_{t+1} + \omega_{t+2})(\omega_{t-1} + \omega_t + \omega_{t+1})] \\ &= \frac{1}{9} E [E(\omega_t\omega_t) + E(\omega_{t+1})(\omega_{t+1})] \\ &= \frac{2}{9}\end{aligned}$$

De manera similar se obtienen resultados para los valores de  $\gamma_v(t - 1, t) = 2/9$ ,  $\gamma_v(t + 2, t) = \gamma_v(t - 2, t) = 1/9$  y 0 para una separación mayor de los valores  $s, t$ . Así, la función de autocovarianza para  $v_t$  está dada por

$$\gamma_v(t, s) = \begin{cases} 1/3, & s = t \\ 2/9, & |s - t| = 1 \\ 1/9, & |s - t| = 2 \\ 0, & |s - t| \geq 3 \end{cases}$$

Con base en las funciones anteriores, se puede definir la noción de estacionariedad débil (de ahora en adelante denotada simplemente como estacionariedad):

**Definición 1.4.6 (Serie débilmente estacionaria)** Una serie débilmente estacionaria  $\{X_t\}_{t \in \mathbb{Z}}$  es un proceso de varianza finita que además cumple que:

- la función media, definida en 1.4.5 es constante y no depende del tiempo  $t$  y
- la función de autocovarianza definida en 1.4.5 solo depende de la separación entre los índices  $s, t$ , es decir, solo depende de  $|s - t|$ .

Puede probarse que cualquier serie que sea estrictamente estacionaria es débilmente estacionaria. Sin embargo, el recíproco no es cierto. Con base a la definición anterior, las funciones de media, autocovarianza y ACF definidas antes pueden ser simplificadas en su notación: para una serie estacionaria  $X_t$ , como la función de media es independiente de  $t$ , puede reescribirse simplemente como  $\mu_t = \mu$ . Además, por la segunda parte de la definición anterior la función

de autocovarianza solo depende de  $|s - t|$ , puede simplificarse como se muestra: si  $s = t + h$ , donde  $h$  representa la diferencia de tiempo (retroceso), entonces se tiene que

$$\begin{aligned}\gamma(t + h, t) &= E [(x_{t+h} - \mu)(x_t - \mu)] \\ &= E [(x_h - \mu)(x_0 - \mu)] \gamma(h, 0)\end{aligned}$$

donde se ha asumido que  $Var(x_t) = \gamma(0, 0) < \infty$ . Dado que  $\gamma(h, 0)$  no depende del argumento de tiempo  $t$ , se denotará la función de autocovarianza solamente como la separación entre los índices  $s, t$ , es decir, se escribirá como  $\gamma(h)$ . Así, de lo anterior se obtienen las definiciones alternas de las funciones de autocovarianza y ACF de una serie estacionaria  $x_t$ :

$$\begin{aligned}\gamma(h) &= E (x_{t+h} - \mu)(x_t - \mu) \\ \rho(h) &= \frac{\gamma(t + h, t)}{\sqrt{\gamma(t + h, t + h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)}\end{aligned}$$

Un hecho importante que no se mencionó sobre las funciones de autocovarianza y ACF de series estacionarias es que cumplen varias propiedades, entre ellas una de las más útiles es que cummplen que son simétricas alrededor del origen. Así, para calcular sus valores, solo se usan las separaciones  $h$  (como se mostró en el ejemplo anterior) para los cuales  $h \geq 0$ .

### Ejemplo (Función de autocovarianza de la media móvil)

Se vió anteriormente que la función de autocovarianza definida para la media móvil

$$v_t = \frac{1}{3} [\omega_{t-1} + \omega_t + \omega_{t+1}]$$

con  $\{\omega_t\}_{t \in \mathbb{Z}}$  un ruido blanco Gaussiano está dada por

$$\gamma_v(t, s) = \begin{cases} 1/3, & s = t \\ 2/9, & |s - t| = 1 \\ 1/9, & |s - t| = 2 \\ 0, & |s - t| \geq 3 \end{cases}$$

Puede apreciarse entonces que dicha función solo depende de la serapación de puntos de la serie en los tiempos  $s$  y  $t$ , con lo cual, sumado al hecho de que tiene media constante cero por ser combinación de un ruido blanco Gaussiano, se obtiene que la serie de medias móviles  $\{v_t\}_{t \in \mathbb{Z}}$  es una serie estacionaria.

Anteriormente, distintos tipos de series de tiempo se han mostrado como ejemplo, debido a su amplia utilidida y su relativa facilidad de aplicación a la hora de modelar el comportamiento de una serie de tiempo. Estos modelos, junto con determinadas herramientas mostradas más adelante permiten la construcción de modelos con una serie de propiedades importante, y que mediante un método esencialmente igual al planteado en secciones anteriores (regresión), permite a dichos modelos dar una serie de predicciones sobre los valores futuros sobre una serie.

### 1.4.2. Modelos ARIMA

Los modelos ARIMA permiten el modelado de series de tiempo que cumplen determinadas condiciones. En esencia, dichos modelos permiten entender el comportamiento y la predicción de series no estacionarias; dichos modelos son particularmente útiles en el mercado de divisas, dado que muchas herramientas que se usan para describir el comportamiento del mercado se basan en dichos modelos. Anteriormente, se vieron ejemplos de modelos autorregresivos y de medias móviles, que hacen parte de su construcción. Para dar una definición formal sobre los modelos ARIMA, se abordarán de manera un poco más general las partes mencionadas.

**Definición 1.4.7 (Modelo Autoregresivo de orden  $p$ )** Para una serie estacionaria  $x_t$ , un modelo autoregresivo de orden  $p$  (denotado  $AR(p)$ ) se define como

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \omega_t$$

donde  $\phi_1, \dots, \phi_p$  son constantes con  $\phi_p \neq 0$ ,  $\omega_t$  es un ruido blanco Gaussiano con media 0 y varianza  $\sigma_\omega^2$ .

Generalmente la media de  $x_t$  es cero. De no serlo, se reemplaza  $x_t$  por  $x_t - \mu$  en la igualdad anterior, y se llega a un modelo  $AR(p)$  de la forma

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \omega_t$$

donde  $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$ . Una forma alterna de escribir un modelo  $AR(p)$  tiene que ver con el operador de retroceso, definido como sigue:

**Definición 1.4.8 (Operador de Retroceso)** Para una serie de tiempo  $x_t$ , se define el operador de retroceso como

$$Bx_t = x_{t-1}$$

Puede extenderse a potencias naturales, por ejemplo  $B^2 x_t = x_{t-2}$ . De manera general,  $B^k x_t = x_{t-k}$

Un concepto importante en el análisis de series de tiempo es la diferenciación. Para una serie de tiempo  $x_t$ , se define la serie diferenciada como

$$\nabla x_t = x_t - x_{t-1}$$

puede probarse que, al diferenciar una serie, la serie resultante es estacionaria. Una relación con el operador de retroceso y la diferenciación de una serie está dada por

$$\nabla x_t = (1 - B)x_t$$

el orden de diferenciación se define como

$$\nabla^d = (1 - B)^d$$

La técnica de diferenciación es una componente importante de los modelos ARIMA de Box y Jenkins [Shumway and Stoffer, 2010] Mediante el operador de retroceso, el modelo AR(p) puede ser descrito como

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = \omega_t$$

De lo anterior, puede considerarse el operador de autorregresión, definido como

$$\Phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$$

Como se mencionó anteriormente, un proceso AR(1) de la forma

$$x_t = \phi x_{t-1} + \omega_t$$

puede describirse como un modelo de media móvil infinito, teniendo en cuenta que para ello se debe cumplir que  $|\phi| < 1$ , es decir,  $x_t$  se puede expresar de la forma

$$x_t = \sum_{j=0}^{\infty} \phi^j \omega_{t-j}$$

De lo anterior pueden derivarse propiedades de las funciones de autocovarianza y ACF de  $x_t$ , que pueden verse en [Shumway and Stoffer, 2010]

**Definición 1.4.9 (Modelo de media móvil de orden q)** El modelo de media móvil de orden q (denotado por MA(q)) se define como

$$x_t = \omega_t + \theta_1 \omega_{t-1} + \dots + \theta_q \omega_{t-q}$$

con  $\theta_q \neq 0$ , y  $\theta_1, \dots, \theta_q$  son parámetros. Se asume además que  $\omega_t$  es un ruido blanco Gaussiano.

Al igual que con el modelo AR(p), se puede establecer una relación entre el operador de retroceso y proceso de media móvil, considerando el operador de media móvil definido como

$$\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

un hecho importante es que el proceso de media móvil es estacionario para cualquier conjunto de parámetros  $\theta_1, \dots, \theta_q$ . Los operadores autoregresivos y de media móvil son bastante útiles para mostrar propiedades generales de los modelos ARMA, definidos a continuación

**Definición 1.4.10 (Modelo Autorregresivo de Media Móvil - ARMA)** Sea  $\{x_t\}_{t \in \mathbb{Z}}$  una serie de tiempo estacionaria. Dicha serie es un modelo ARMA de órdenes p, q (denotado por ARMA(p,q)) si se cumple que

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \dots + \theta_q \omega_{t-q}$$

donde  $\phi_p \neq 0, \theta_q \neq 0$  y  $\sigma_\omega^2 > 0$ . Los parámetros p y q son llamados los órdenes autoregresivos y de media móvil, respectivamente. Si  $x_t$  tiene media no nula, entonces considerando un proceso similar al visto en la definición 1.4.7, se fija  $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$  y el modelo se reescribe como

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \omega_t + \theta_1 \omega_{t-1} + \dots + \theta_q \omega_{t-q}$$

Claramente, los modelos AR( $p$ ) y MA( $q$ ) son casos particulares del modelo ARMA. Teniendo en cuenta lo mencionado sobre los operadores de retroceso y de media móvil, un modelo ARMA puede expresarse mediante dichos operadores como

$$\Phi(B)x_t = \Theta(B)\omega_t$$

existen una serie de problemas subyacentes en los modelos ARIMA que pueden derivarse de la definición de los mismos (ver ejemplos 3.2, 3.4 y 3.5 de [Shumway and Stoffer, 2010]) Resumidamente, son los siguientes:

- El modelo posee redundancia de parámetros
- Modelos estacionarios AR pueden depender de observaciones futuras
- No unicidad de los modelos MA

Para solucionar dichos problemas, se requieren restricciones adicionales sobre los parámetros del modelo, para ello se consideran los siguientes polinomiales:

**Definición 1.4.11** (*Polinomiales AR y MA*) *Respectivamente, se definen los polinomiales AR y MA como*

$$\begin{aligned}\phi(z) &= 1 - \phi_1 z - \cdots - \phi_p z^p, & \phi_p &\neq 0 \\ \theta(z) &= 1 + \theta_1 z + \cdots + \theta_q z^q, & \theta_q &\neq 0\end{aligned}$$

para  $z \in \mathbb{C}$

Para solucionar los problemas de posibles dependencias futuras y no unicidad de los modelos MA, se definirán las nociones de causalidad e invertibilidad de un modelo ARMA. Para efectos prácticos, se verán propiedades equivalentes de dichos conceptos. Las definiciones respectivas pueden verse en [Shumway and Stoffer, 2010]

**Definición 1.4.12** *Causalidad de un modelo ARMA( $p,q$ )* *Un modelo ARMA( $p,q$ ) es causal si y solo si  $\phi(z) \neq 0$  para  $|z| \leq 1$ , es decir, dicho modelo es causal solo cuando las raíces del polinomial  $\phi(z)$  caen por fuera del círculo unitario.*

**Definición 1.4.13** *Invertibilidad de un modelo ARMA( $p,q$ )* *Un modelo ARMA( $p,q$ ) es invertible si y solo si  $\theta(z) \neq 0$  para  $|z| \leq 1$ , es decir, un modelo ARMA( $p,q$ ) es invertible solo si las raíces de  $\theta(z)$  caen por fuera del círculo unitario.*

Intuitivamente, la noción de causalidad de un modelo ARMA es que el valor de la serie en un punto  $t$  dependerá (si ese es el caso) solo de las observaciones pasadas, mientras que un modelo es invertible si el modelo MA que conforma la serie puede expresarse de manera única. Para el problema final de la redundancia de parámetros, lo que se hace es partir de la representación del modelo dada por  $\Phi(B)x_t = \Theta(B)\omega_t$ . Al considerar los operadores  $\Phi, \Theta$

dependientes de la variable  $z$ , se obtienen los polinomiales AR, MA definidos anteriormente. se obtiene entonces de lo anterior que

$$\phi(z) = \theta(z)$$

Al factorizar dichos polinomiales, dado el caso de que hayan parametros redundantes, estos se verán expresados como factores de las expresiones polinomiales de la igualdad anterior. Lo que se hace, simplemente, es cancelar dichos factores que estén a ambos lados de la igualdad. Se examinará a continuación un ejemplo sobre cómo proceder respecto a la redundancia de parámetros, causalidad e invertibilidad

### **Ejemplo (Redundancia de parámetros)**

Considere el proceso

$$x_t = 0.4x_{t-1} + 0.45x_{t-2} + \omega_t + \omega_{t-1} + 0.25\omega_{t-2}$$

que en su forma de operador se puede expresar como

$$(1 - 0.4B - 0.45B^2)x_t = (1 + B + 0.25B^2)\omega_t$$

Podría darse la impresión de que, en la forma inicialmente mostrada,  $x_t$  sigue un proceso ARMA(2,2). Sin embargo, al considerar los polinomiales asociados

$$\begin{aligned}\phi(z) &= (1 - 0.4z - 0.45z^2) = (1 + 0.5z)(1 - 0.9z) \\ \theta(z) &= (1 + z + 0.25z^2) = (1 + 0.5z)^2\end{aligned}$$

se observa que ambos polinomiales tienen un factor asociado. Después de operar, los polinomiales adquieren la forma  $\phi(z) = (1 - 0.9z)$  y  $\theta(z) = (1 + 0.5z)$ , así que el modelo puede ser descrito como

$$x_t = 0.9x_{t-1} + 0.5\omega_{t-1} + \omega_t$$

Lo cual, permite concluir que  $x_t$  sigue un proceso ARMA(1,1). Puede verse además que, al calcular las raíces de los polinomiales  $\phi(z)$ ,  $\theta(z)$ , se observa que los módulos de dichas raíces caen por fuera del círculo unitario.

Los conceptos y definiciones desarrollados anteriormente permiten su aplicabilidad a series estacionarias. Sin embargo, lo que se encuentra comúnmente es que las series de tiempo, sin importar su índole, presentan comportamientos o patrones que indica que la serie sea no estacionaria. Es por ello que se hace necesario el desarrollar medios para los cuales herramientas desarrolladas anteriormente puedan ser aplicables en este contexto particular.

Generalmente, las series no estacionarias pueden ser vistas como una suma de dos componentes: una componente no estacionaria (vista como componente de tendencia) mas otra

componente estacionaria de media cero. Como se mencionó antes, el diferenciar una serie permite transformarla en una serie estacionaria, eliminando dichos componentes que generan algún tipo de temporalidad o tendencia. Así, la clase de modelos integrados ARMA o ARIMA, es aquella clase de modelos de series no estacionarios, que involucra un proceso de diferenciación para obtener como resultado otra serie estacionaria.

**Definición 1.4.14 (Modelos Autoregresivos Integrados de Media Móvil - ARIMA)** un proceso  $x_t$  se dice que es ARIMA de órdenes  $p, d, q$  (denotado por  $ARIMA(p, d, q)$ ) si

$$\nabla^d x_t = (1 - B)^d x_t$$

es  $ARMA(p, q)$ . En general, un modelo  $ARIMA(p, d, q)$  se denota como

$$\Phi(B)(1 - B)^d x_t = \Theta(B)\omega_t$$

Si se tiene que  $E(\nabla^d x_t) = \mu$  (valor esperado de la serie diferenciada), entonces se denota el modelo como

$$\Phi(B)(1 - B)^d x_t = \alpha + \Theta(B)\omega_t$$

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

El índice  $d$  del modelo ARIMA indica las veces que se debe diferenciar la serie de tal manera que la serie diferenciada  $d$  veces (denotado por  $\nabla^d x_t$ ) cumpla que es un modelo  $ARMA(p, q)$ .

Algunos modelos ARIMA son bastante importantes debido al hecho de que varias series financieras pueden modelarse exitosamente mediante estos. De ellos, se verá un ejemplo a continuación.

### Ejemplo (Modelos IMA(1,1) y EWMA)

El modelo  $ARIMA(0, 1, 1)$  (denotado como  $IMA(1, 1)$ ) es usado en serie financieras tanto para entender como para predecir el comportamiento de dicha serie. Para la predicción de una serie, tiende a usarse como un método llamado *Media Móvil Exponencial Ponderada* (EWMA por sus siglas en inglés). Inicialmente, dicho modelo es descrito de la forma

$$x_t = x_{t-1} + \omega_t - \lambda\omega_{t-1}$$

Puede verse que cuando  $|\lambda| < 1$ , dicho modelo tiene una representación invertible de la forma

$$x_t = \sum_{j=1}^{\infty} (1 - \lambda) \lambda^{j-1} x_{t-j} + \omega_t$$

Ahora, se asume que se conocen todos los valores de la serie menores o iguales a un punto  $n$  dado. Sea  $\tilde{x}_{n+1}$  el valor esperado condicional del valor  $x_{n+1}$  dados los valores de la serie con índice menor que  $n$ , esto es

$$\tilde{x}_{n+1} = E(x_{n+1} | x_n, x_{n-1}, \dots)$$

(Acá dicha esperanza condicional significa el valor<sup>3</sup> de la serie que minimiza el error cuadrático medio entre la observación real y la predicha). De lo anterior y mediante una serie de manipulaciones, se obtiene que

$$\tilde{x}_{n+1} = (1 - \lambda)x_n + \lambda\tilde{x}_n$$

para  $n \in \mathbb{N}$  y sujeta a la condición inicial de que  $\tilde{x}_0 = 0$ . Así, se obtiene la predicción del valor  $n + 1$  de la serie sería una combinación lineal de la anterior predicción (la predicción en el momento  $n$ ) y el valor de la serie en el momento  $n$ . Al parámetro  $\lambda$  generalmente se le conoce como constante de suavizado y se restringe a valores entre 0 y 1.

Para una serie  $x_t$ , puede calcularse un modelo ARIMA(p,d,q) que permita modelar el comportamiento bajo ciertos criterios (puede verse más en [Shumway and Stoffer, 2010]). Dichos modelos pueden usarse en conjunto con otros tipos de modelos predictivos aplicados al contexto de series financieras que permiten la mejora significativa de los valores de predicción del modelo. El modelo que compete en este trabajo (modelo ANFIS) puede usarse en complemento modelos de series de tiempo para para lograr un mejor comportamiento modelando ciertas variables. La estructura y funcionamiento de dichos modelos se mostrará en el siguiente capítulo.

---

<sup>3</sup>Dicho valor puede ser calculado mediante una serie de algoritmos. Dichos métodos de cálculo se muestran en [Shumway and Stoffer, 2010]

# Capítulo 2

## Aproximación universal

En la lógica clásica, se dan afirmaciones generalmente de dos tipos, donde su valor de verdad se ve comprendido solo entre dos valores. Sin embargo, dada la naturaleza del pensamiento humano, existen afirmaciones cuyo valor de verdad tiende a ser algo subjetivo, es decir afirmaciones donde su validez cambia según la percepción. Por ejemplo, si se considera bajo la lógica clásica un conjunto de personas donde su altura es mayor que 2 mt, una persona cuya altura sea de 1.99 mt no sería una persona alta. Sin embargo, si se considera dicha afirmación bajo la lógica difusa, el valor de verdad se verá comprendido entre un rango de valores mayor al doblete  $\{0, 1\}$ . La lógica difusa permite dotar de significado matemático a dichas afirmaciones, generalizar la teoría de deducción de lógica clásica para un determinado conjunto de afirmaciones.

Este capítulo tiene entonces como objeto abordar dicha teoría, dar una comprensión sobre los conjuntos difusos, definir una estructura sobre dichos conjuntos y posteriormente trabajar sobre un conjunto de reglas para abordar así el teorema de aproximación universal, que brinda una de tantas formas de aproximación de funciones mediante ANFIS.

### 2.1. Lógica difusa

De manera intuitiva, se define un conjunto difuso sin una frontera clara, es decir, donde la pertenencia de los elementos a dicho conjunto se ve de manera parcial, tomando un conjunto de valores distintos a los dados la pertenencia clásica. Dicha pertenencia se ve caracterizada por una función (función indicadora, denotada por  $1_A$ ) definida para  $A \subseteq X$ , como se muestra en 2.1.1. Se busca, por ende, que la pertenencia para un conjunto difuso se vea reflejada de manera similar sobre un determinado rango de valores. La siguiente definición mostrará de manera formal la definición de un conjunto difuso.

$$1_A : X \rightarrow \{0, 1\}$$
$$x \mapsto 1_A(x) : \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \quad (2.1.1)$$

**Definición 2.1.1 (Conjunto difuso)** Sea  $X$  un conjunto. Un conjunto difuso  $A$  en  $X$  se define como el conjunto de pares ordenados

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

donde  $\mu_A : X \rightarrow T$ , se denomina función de pertenencia (o de manera abreviada MF) del conjunto  $A$ , para  $T$  un subconjunto del intervalo  $[0, 1]$ .

Se puede ver que la definición anterior es una extensión de la definición de función indicadora de un conjunto clásico  $A$ , tomando  $T = \{0, 1\}$ . La definición anterior indica que a cada elemento del conjunto  $A$  se le asocia un valor comprendido entre 0 y 1, denominado el valor de pertenencia del elemento al conjunto  $A$ . Al conjunto  $X$  se le conoce como el universo relativo al conjunto  $A$ . Dicho universo se conoce como discreto si su cardinal es contable. de lo contrario, se dice que  $X$  es continuo.

#### **Ejemplo (Conjunto difuso con un universo discreto)**

Sea  $X = \{1, 2, 3, 4, 5, 6\}$  el número de hijos que una familia puede escoger tener. Entonces el conjunto  $A$  definido como “*número sensible de niños en una familia*” puede tener la forma

$$A = \{(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.7), (5, 0.3), (6, 0.1)\}$$

De manera gráfica, dicho conjunto tiene la forma mostrada en la figura 2.1 (a).

#### **Ejemplo (Conjunto difuso con un universo continuo)**

Sea  $X = \mathbb{R}^+$  el conjunto de posibles edades para seres humanos. Entonces el conjunto difuso  $B =$  “*aproximadamente 50 años*” se puede expresar de la forma

$$B = \{(x, \mu_B(x)) \mid x \in X\}$$

donde

$$\mu_B(x) = \frac{1}{1 + \left(\frac{x-50}{10}\right)^4}$$

De manera gráfica, el conjunto difuso  $B$  tiene la forma mostrada en la figura 2.1 (b)

Es importante aclarar que la especificación de funciones de pertenencia puede ser de manera subjetiva, es decir, las funciones especificadas por el mismo concepto pueden variar considerablemente, esta subjetividad viene de diferencias individuales al percibir o expresar

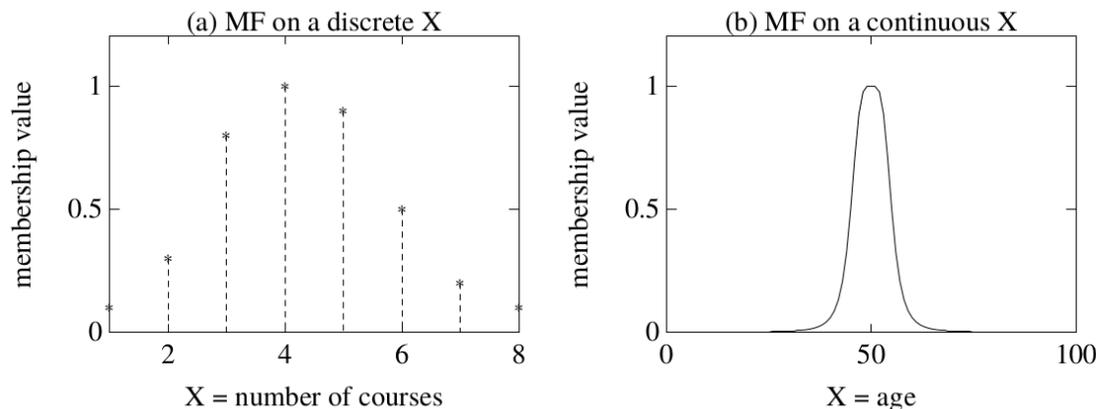


Figura 2.1: La parte a) muestra de manera el conjunto difuso  $A$  definido sobre un universo discreto. La parte b) muestra conjunto difuso  $B$  definido sobre un universo continuo. A cada elemento de  $X$  se le asocia un grado de pertenencia mediante  $\mu$ . Tomado de [Jang and Sun, 1995]

conceptos abstractos y tiene que ver con la aleatoriedad. Por tanto, la subjetividad y la no aleatoriedad de los conjuntos difusos es la principal diferencia entre el estudio de los conjuntos difusos y la teoría de la probabilidad.

Con base a la función de pertenencia de un conjunto se pueden definir relaciones para conjuntos difusos  $A, B$  tal como se hace en la teoría de conjuntos clásica.

**Definición 2.1.2 (Relaciones entre conjuntos difusos)** Sean  $A, B$  conjuntos difusos. Se definen las relaciones de contención, intersección, unión y complemento de dichos conjuntos como sigue:

i) **Contención.** Se dice que  $A \subseteq B$  sí  $\forall_{x \in X} \mu_A(x) \leq \mu_B(x)$

ii) **Unión.** La unión de los conjunto difusos  $A, B$  es nuevamente un conjunto difuso  $C$ , caracterizado por la MF dada por

$$\mu_C(x) = \max\{\mu_A(x), \mu_B(x)\} = \mu_A(x) \vee \mu_B(x)$$

Dicha unión se denota por  $A \cup B$ .

iii) **Intersección.** Para  $A, B$  conjuntos difusos, la intersección de dichos conjuntos difusos es un conjunto difuso  $C$  caracterizado por la MF dada por

$$\mu_C(x) = \min\{\mu_A(x), \mu_B(x)\} = \mu_A(x) \wedge \mu_B(x)$$

Lo anterior se denota por  $A \cap B$ .

iv) **Complemento.** El complemento del conjunto difuso  $A$  está dado por la MF

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x)$$

Donde  $\bar{A}$  denota el complemento del conjunto difuso  $A$ . También se suele denotar por  $\sim A$

Se pueden ilustrar las relaciones anteriores de manera gráfica, como se muestra a continuación

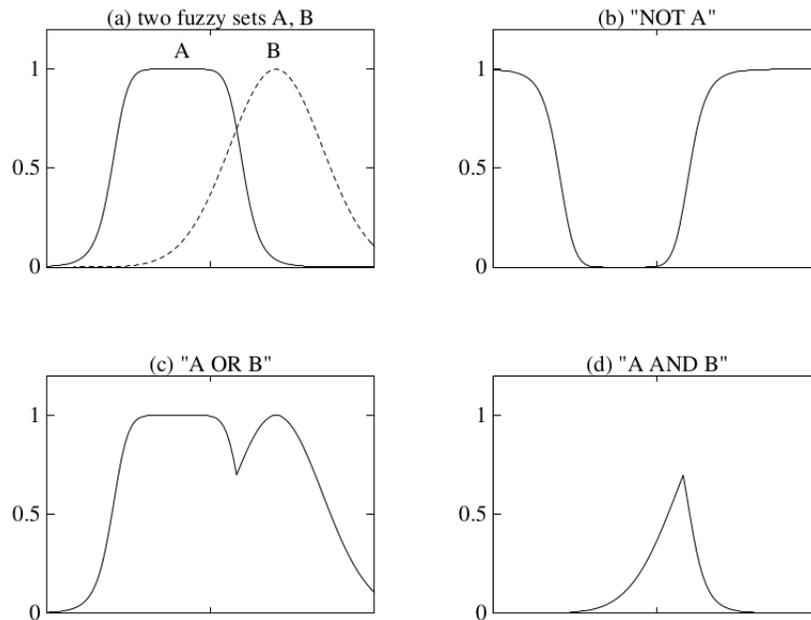


Figura 2.2: La parte (a) muestra dos conjuntos difusos  $A, B$ . La parte (b) muestra el complemento del conjunto difuso  $A$ . las partes (c), (d) muestran las operaciones de unión e intersección de los conjuntos difusos  $A, B$  respectivamente. Tomado de [Jang and Sun, 1995]

Para dos conjuntos difusos  $A, B$  se puede definir una relación que haga las veces de producto cartesiano entre dichos conjuntos.

**Definición 2.1.3 (Producto cartesiano)** Sean  $A, B$  conjuntos difusos en  $X$  y  $Y$  respectivamente. El producto  $A \times B$  es un conjunto difuso en el conjunto  $X \times Y$  caracterizado por la función de pertenencia

$$\mu_{A \times B}(x, y) = \min\{\mu_A(x), \mu_B(y)\}$$

De la definición 2.1.1 se infiere que los conjuntos difusos están caracterizados por la función de pertenencia. Se busca, por tanto, dar una expresión matemática que permita la simplicidad a la hora de ver la representación del conjunto difuso. Algunos tipos de funciones de pertenencia se pueden ver en [Jang and Sun, 1995].

### 2.1.1. Razonamiento difuso

La inferencia difusa permite desarrollar técnicas de deducción similares a las usadas en la lógica clásica; a partir de un conjunto de reglas y de hechos conocidos, se pueden derivar determinandas conclusiones. Dicho conjunto de reglas se denomina conjunto de reglas difusas, donde cada una es de la forma “si  $x$  es  $A$  entonces  $y$  es  $B$ ”, con  $A$  y  $B$  conjuntos difusos en universos  $X, Y$  respectivamente. dicha regla suele ser denotada por  $A \rightarrow B$ .

En el razonamiento difuso, el *modus ponens* se usa de una manera “aproximada”(se le conoce como *Modus Ponens generalizado - GMP*. Cuando los conjuntos son nitidos, es decir, clásicos, se obtiene el modus ponens clásico). Por ejemplo, si se tiene la afirmación “si el tomate es rojo entonces el tomate está maduro” y se tiene que el tomate en efecto es rojo, entonces se concluye que el tomate está maduro. Si se sabe que el tomate está más o menos rojo, entonces se infiere que el tomate está “más o menos maduro”. Si se denota  $A$  como ser rojo,  $B$  como ser maduro,  $A'$  como ser mas o menos rojo y  $B'$  como más o menos maduro, lo anterior toma la siguiente forma: si  $x$  es  $A$  entonces  $y$  es  $B$  y se tiene que  $x$  es  $A'$ , entonces se infiere que  $y$  es  $B'$ . Se puede ver de lo anterior que  $A', B'$  son conjuntos cercanos a  $A, B$  respectivamente, es decir, conjuntos donde su intersección es más de un elemento. De manera general, se define el razonamiento difuso como sigue.

**Definición 2.1.4 (Razonamiento difuso - Razonamiento aproximado)** Sean  $A, A'$  conjuntos difusos de  $X$  y  $B$  conjunto difuso de  $Y$ . Si la implicación difusa  $A \rightarrow B$  se expresa como una relación en  $X \times Y$ . Entonces el conjunto difuso  $B'$  inducido por la afirmación  $x$  es  $A'$  y la regla difusa<sup>1</sup> si  $x$  es  $A$  entonces  $y$  es  $B$  está caracterizado por la función de pertenencia

$$\begin{aligned}\mu_{B'}(y) &= \max_x \min\{\mu_{A'}(x), \mu_R(x, y)\} \\ &= \vee_x (\mu_{A'}(x) \wedge \mu_R(x, y))\end{aligned}$$

donde  $\vee_x$  y  $\wedge$  denotan máximo y mínimo, respectivamente y  $\mu_R(x, y)$  se define como una relación binaria difusa en el producto cartesiano  $X \times Y$ .

A partir de lo anterior, se consideran entonces casos donde la forma y el número de reglas difusas varíe.

#### Una regla con un antecedente

La definición anterior expresa de manera formal una regla difusa con un antecedente: dado que  $\mu_{B'}(y) = \vee_x (\mu_{A'}(x) \wedge \mu_R(x, y))$ , se obtiene la ecuación<sup>2</sup>

$$\mu_{B'}(y) = \vee_x (\mu_{A'}(x) \wedge \mu_A(x)) \wedge \mu_B(y) = \omega \wedge \mu_B(y).$$

<sup>1</sup>Referida también como regla difusa si-entonces.

<sup>2</sup>Hay que tener en cuenta que la implicación difusa  $A \rightarrow B$  se expresa como una relación en  $X \times Y$ , donde la MF de dicha relación está dada por  $\mu_R(x, y) = \mu_A(x) \wedge \mu_B(y)$ . A lo anterior se le conoce como la implicación difusa de Mamdani. Ver [Jang and Sun, 1995]

A dicho  $\omega$  se le conoce como el grado de coincidencia entre  $A$  y  $A'$ . De manera gráfica, se tiene que dicha regla toma la forma mostrada en la siguiente figura. La MF del conjunto  $B'$  sería equivalente a la MF de  $B$  cortada por  $\omega$ .

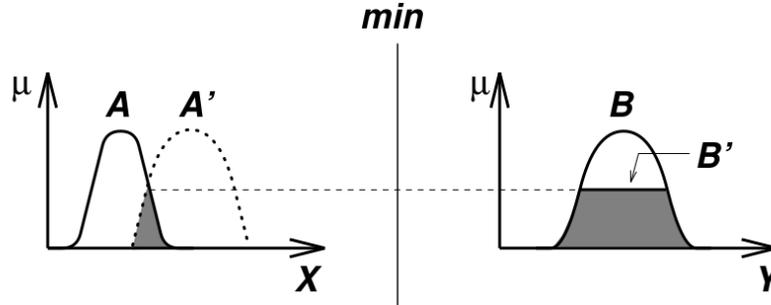


Figura 2.3: Razonamiento difuso para una regla con un antecedente. Tomado de [Jang and Sun, 1995]

### Una regla con múltiples antecedentes

Una regla difusa con dos antecedentes es usualmente escrita como *si  $x$  es  $A$  y  $y$  es  $B$  entonces  $z$  es  $C$* . El correspondiente problema para el GMP es expresado como sigue: si se tiene que *si  $x$  es  $A$  y  $y$  es  $B$  entonces  $z$  es  $C$*  y se cumple además que  *$x$  es  $A'$  y  $y$  es  $B'$*  se concluye que  *$z$  es  $C'$* .

La MF del conjunto difuso  $C'$  está dada, para  $(x, y, z) \in X \times Y \times Z$ <sup>3</sup>, por

$$\begin{aligned}\mu_{C'}(z) &= [\vee_x(\mu_A(x) \wedge \mu_{A'}(x))] \wedge [\vee_y(\mu_B(y) \wedge \mu_{B'}(y))] \wedge \mu_C(z) \\ &= (\omega_1 \wedge \omega_2) \wedge \mu_C(z)\end{aligned}$$

Se puede generalizar para una regla con mas de dos antecedentes. De manera gráfica, puede apreciarse esto en la figura 2.5

### Múltiples reglas con múltiples antecedentes

Si se tienen los hechos  *$x$  es  $A'$  y  $y$  es  $B'$*  y las reglas *si  $x$  es  $A_1$  y  $y$  es  $B_1$  entonces  $z$  es  $C_1$*  y *si  $x$  es  $A_2$  y  $y$  es  $B_2$  entonces  $z$  es  $C_2$* , se concluye que  *$z$  es  $C'$* , donde  $C'$  es el máximo de los  $C'_i$  dados al tener las reglas con múltiples antecedentes (se puede ver lo anterior de manera gráfica en la siguiente figura). La MF de dicho conjunto  $C'$  es de la forma

$$\mu_{C'}(z) = \mu_{C'_1}(z) \vee \mu_{C'_2}(z)$$

El razonamiento difuso permite crear determinados modelos que funcionan acorde a una forma y definición particular del tipo de reglas de inferencia, dependiendo de su definición en la parte consecuente. A continuación, se definirá y se profundizará en un modo de formulación de reglas que permite definir el modelo de inferencia de Sugeno.

<sup>3</sup> $C, C'$  conjuntos difusos del universo  $Z$

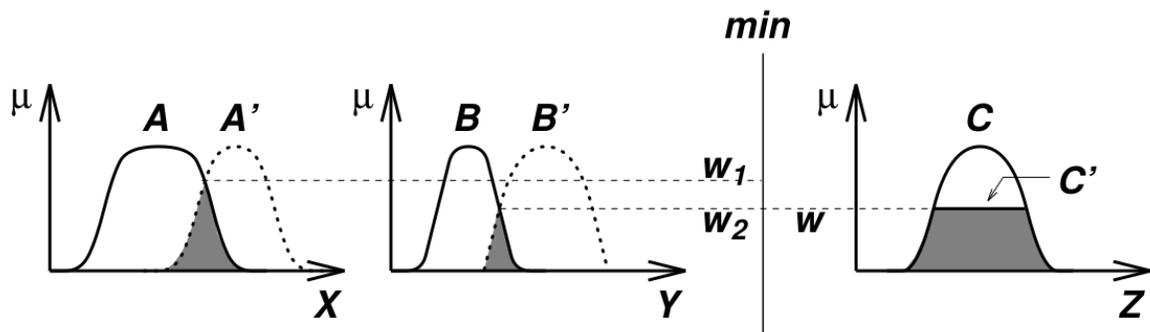


Figura 2.4: Razonamiento difuso para una regla con múltiples antecedentes. Tomado de [Jang and Sun, 1995]

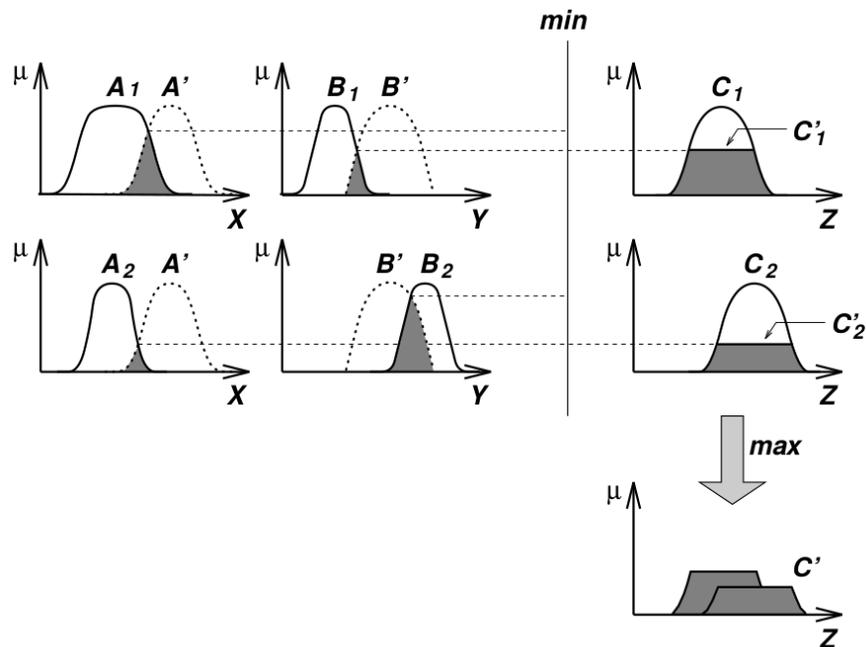


Figura 2.5: Razonamiento difuso para múltiples reglas con múltiples antecedentes. Tomado de [Jang and Sun, 1995]

## 2.2. Modelo de inferencia de Sugeno (Modelo TSK)

Los modelos de inferencia difusa son modelos computacionales basados en teoría difusa de conjuntos, reglas difusas y razonamiento difuso. Sus aplicaciones van desde control automático, clasificación de datos, visión computacional, etc.

la estructura clásica de un sistema de inferencia difusa consiste de 3 componentes con-

ceptuales:

- i) Una regla base, que contiene una selección de reglas difusas.
- ii) Una base de datos (denominada diccionario), que define las funciones de pertenencia usadas en las reglas difusas.
- iii) Un mecanismo de razonamiento, que realiza el proceso de inferencia sobre las reglas y hechos dados para dar una salida o conclusión.

Un sistema de inferencia difusa puede tomar entradas tanto nítidas como difusas. Dicho sistema implementa un mapeo no lineal de su espacio de entrada a su espacio de salida, realizado por medio de un número de reglas difusas que describen de manera local el comportamiento del modelo. De manera particular, el antecedente de una regla define una región difusa en el espacio de entrada, mientras que el consecuente define una región difusa en el espacio de salida.

**Definición 2.2.1 (Modelo difuso de Sugeno)** las reglas de un modelo difuso de Sugeno (también conocido como modelo Takagi-Sugeno-Kang) tienen la siguiente forma: si  $x$  es  $A$  y  $y$  es  $B$  entonces  $z = f(x, y)$ , donde  $A, B$  son conjuntos difusos en el antecedente y  $z = f(x, y)$  es una función en el consecuente.

Usualmente  $f(x, y)$  es un polinomio en las variables de entrada  $x, y$ , pero puede ser cualquier función que describa de manera apropiada la salida del modelo con la región difusa especificada por el antecedente de la regla. Si  $f(x, y)$  es un polinomio de primer orden, entonces el modelo difuso es llamado modelo de Sugeno de primer orden. Si  $f$  es constante, entonces el modelo es llamado modelo de Sugeno de orden cero.

### **Ejemplo (Modelo de Sugeno con una entrada - comparación.)**

Para ilustrar un poco la definición, un ejemplo de una sola entrada de un modelo difuso de Sugeno es dado por

$$\begin{cases} \text{Si } X \text{ es pequeño entonces } Y = 0.1X + 6.4 \\ \text{Si } X \text{ es mediano entonces } Y = -0.5X + 4 \\ \text{Si } X \text{ es grande entonces } Y = X - 2 \end{cases}$$

Donde “pequeño, mediano, grande” son conjuntos con MF dadas en la figura 2.7. Si dichos conjuntos se consideran como conjuntos clásicos, sus funciones de pertenencia son constantes, y la respectiva función de salida es lineal. Si por el contrario dichos conjuntos se consideran como conjuntos difusos, con funciones de pertenencia dadas por la figura c) entonces se obtiene como salida una función suave como la vista en d).

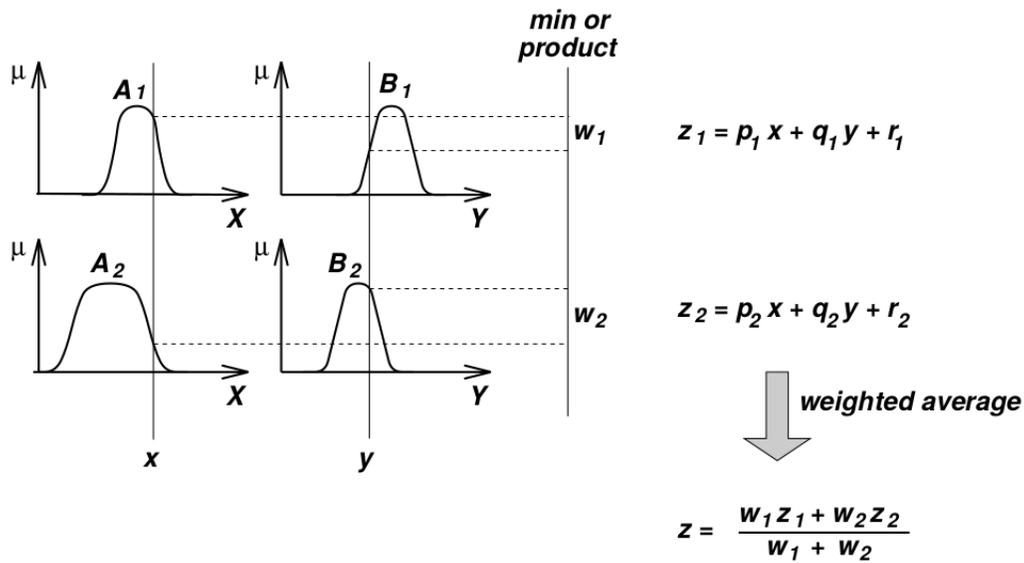


Figura 2.6: Razonamiento difuso para un modelo de Sugeno de primer orden; la salida general de dicho modelo es un promedio ponderado. Tomado de [Jang and Sun, 1995]

Puede darse el caso que para un modelo de Sugeno de orden lineal relativamente simple, la salida pueda tener un comportamiento complejo. El siguiente ejemplo servirá como ilustración.

### Ejemplo (Modelo de Sugeno con dos entradas)

Un modelo de Sugeno de dos entradas con una sola salida puede ser expresado mediante las reglas

$$\begin{cases} \text{Si } X \text{ es pequeño y } Y \text{ es pequeño entonces } z = x + y + 1 \\ \text{Si } X \text{ es pequeño y } Y \text{ es grande entonces } z = 3 - y \\ \text{Si } X \text{ es grande y } Y \text{ es pequeño entonces } z = 3 - x \\ \text{Si } X \text{ es grande y } Y \text{ es grande entonces } z = x + y + 2 \end{cases} \quad (2.2.1)$$

El resultado del modelo inferencia, aunque puede parecer algo compleja dada la forma de la superficie del conjunto de salida (como se ve en la figura 2.9), muestra como las reglas difusas logran una transición suave entre las superficies, viendose de manera relativamente clara como están dadas por las ecuaciones de los planos de las partes consecuentes de las reglas.

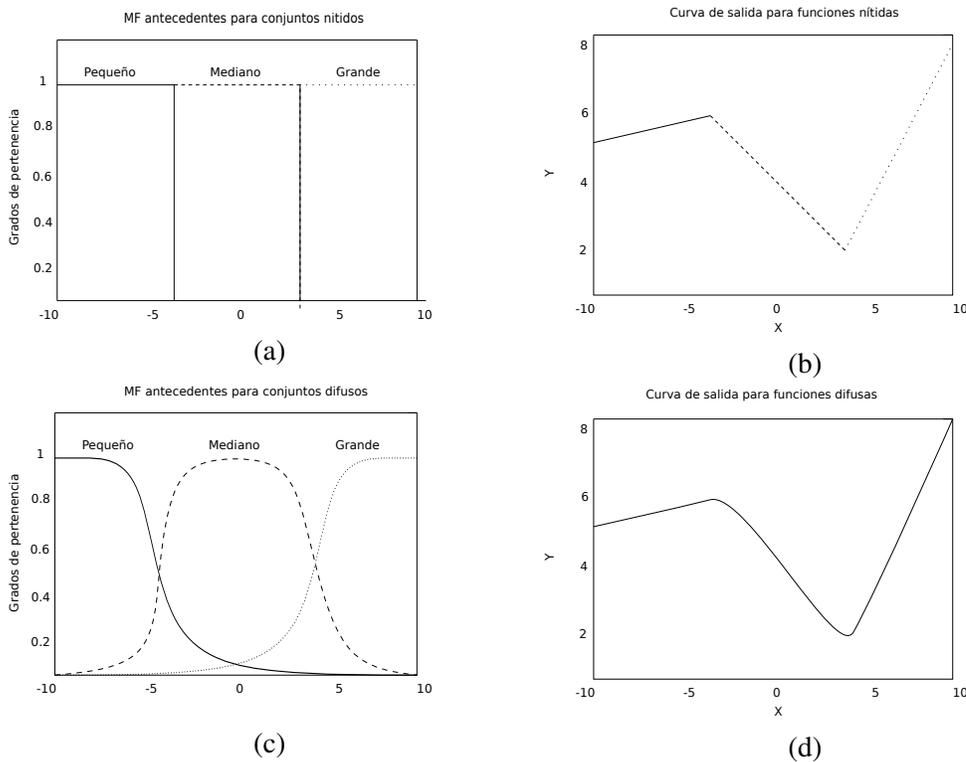


Figura 2.7: Comparación de las salidas c), d) del modelo de Sugeno dado para entradas a), b)

## 2.3. Aproximación universal

Existen distintas formas de aproximar funciones continuas definidas sobre un conjunto compacto  $K$ . Por ejemplo, el Teorema de aproximación de Weierstrass permite aproximar funciones continuas con polinomios de determinado grado, mientras que las series de Fourier permiten expresar una función continua como suma de senos y cosenos. Desde la parte computacional, se han demostrado resultados que permiten de igual manera dicha aproximación. Por ejemplo, se demostró que una red neuronal con funciones de activación sigmoideas y tres capas puede aproximar cualquier función  $f \in C(I)$ . Se verá un resultado que permite la aproximación de funciones continuas en  $K$  mediante un modelo difuso de Sugeno de orden cero, esto como consecuencia del teorema de Stone-Weierstrass.

**Teorema 2.3.1 (Stone-Weierstrass)** Sea  $D$  un espacio compacto en  $\mathbb{R}^n$ , y  $\mathcal{F}$  un conjunto de funciones continuas realvaluadas tales que:

- i) La función  $f = 1 \in \mathcal{F}$ .
- ii) Para cualesquier  $x_1, x_2$  en  $D$  existe una función  $f \in \mathcal{F}$  tal que  $f(x_1) \neq f(x_2)$ .
- iii) si  $f, g \in \mathcal{F}$  entonces  $fg$  y  $af + bg \in \mathcal{F}$  para  $a, b \in \mathbb{R}$ .

Entonces  $\mathcal{F}$  es denso en  $C(D)$ , es decir, si  $g \in C(D)$  y  $\epsilon > 0$ , existe  $f \in \mathcal{F}$  tal que para cualquier  $x \in D$  se cumple que  $|f(x) - g(x)| < \epsilon$ .

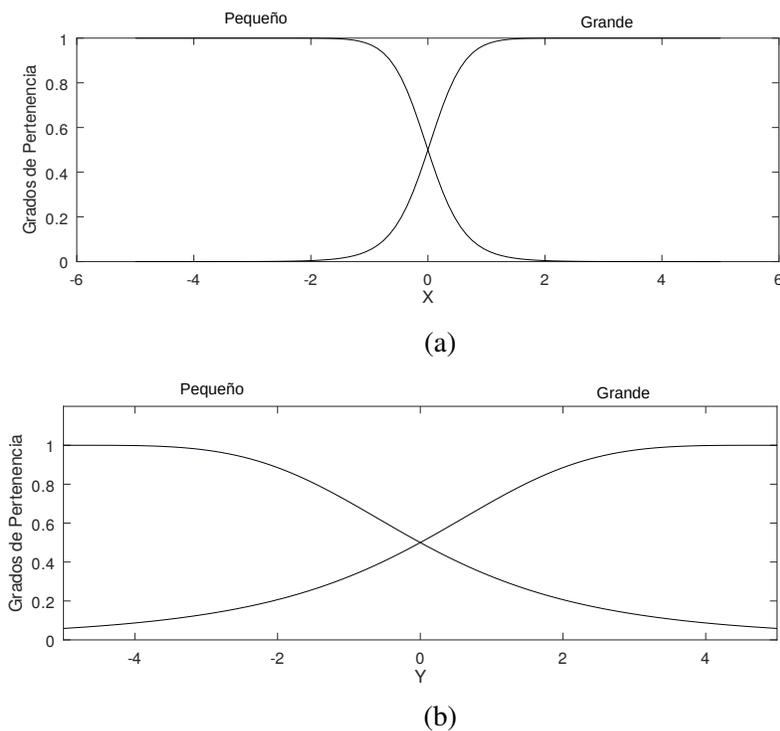


Figura 2.8: Las partes a) y b) muestran las funciones de pertenencia de las variables difusas  $X$  y  $Y$

Se verificarán las condiciones anteriores mediante modelo de inferencia difusa de Sugeno adecuados. Si  $\mathcal{F}$  denota el conjunto de modelos de Sugeno de orden cero, entonces:

### Función unitaria

Para verificar la primera condición se puede construir un modelo de Sugeno de orden cero que conste de una regla, con el consecuente definido como la función  $f = 1$ .

### Separabilidad

Se requiere, para verificar la segunda condición, que para puntos  $x_1, x_2 \in D$  con  $x_1 \neq x_2$  se cumple que existe un modelo de Sugeno  $S$  de tal manera que  $S(x_1) \neq S(x_2)$ <sup>4</sup>. Para ver lo anterior, se puede considerar el siguiente modelo:

$$\left\{ \begin{array}{l} \text{Si } x \text{ es } A_1 \text{ entonces } y = a_1 \\ \text{Si } x \text{ es } A_2 \text{ entonces } y = a_2 \end{array} \right.$$

con  $a_1, a_2 \in \mathbb{R}$  y  $a_1 \neq a_2$ . Si  $z_1 = S(x_1)$  y  $z_2 = S(x_2)$ , entonces se puede ver que  $z_1 \neq z_2$ .

<sup>4</sup>Se debe tener en cuenta que al usar la notación  $S(x)$  para un modelo  $S$ , se está indicando que dicho modelo recibe como dato (o datos) de entrada el valor (o los valores)  $x$

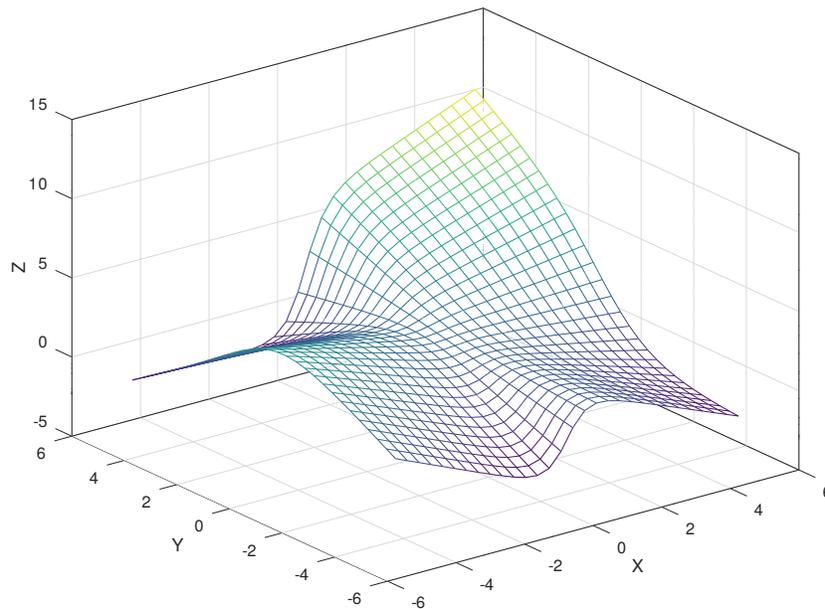


Figura 2.9: Superficie de salida del modelo de inferencia 2.2.1

**Aditividad**

Sean  $S, \hat{S} \in \mathcal{F}$ . Se quiere verificar que para constantes  $a, b$ ,  $aS + b\hat{S} \in \mathcal{F}$ . Si

$$S = z := \frac{\omega_1 f_1 + \omega_2 f_2}{\omega_1 + \omega_2}$$

y

$$\hat{S} = \hat{z} := \frac{\hat{\omega}_1 \hat{f}_1 + \hat{\omega}_2 \hat{f}_2}{\hat{\omega}_1 + \hat{\omega}_2}$$

Se tiene que

$$z + \hat{z} = \frac{\omega_1 \hat{\omega}_1 (a f_1 + b \hat{f}_1) + \omega_1 \hat{\omega}_2 (a f_1 + b \hat{f}_2) + \omega_2 \hat{\omega}_1 (a f_2 + b \hat{f}_1) + \omega_2 \hat{\omega}_2 (a f_2 + b \hat{f}_2)}{\omega_1 \hat{\omega}_1 + \omega_1 \hat{\omega}_2 + \omega_2 \hat{\omega}_1 + \omega_2 \hat{\omega}_2}$$

Dicha salida se puede verificar mediante el sistema de inferencia difusa formado por el conjunto de reglas

$$\left\{ \begin{array}{l} \text{Si } X \text{ es } A_1 \text{ y } Y \text{ es } B_1 \text{ entonces } z = a f_1 + b \hat{f}_1 \\ \text{Si } X \text{ es } A_2 \text{ y } Y \text{ es } B_2 \text{ entonces } z = a f_1 + b \hat{f}_2 \\ \text{Si } X \text{ es } A_3 \text{ y } Y \text{ es } B_3 \text{ entonces } z = a f_2 + b \hat{f}_1 \\ \text{Si } X \text{ es } A_4 \text{ y } Y \text{ es } B_4 \text{ entonces } z = a f_2 + b \hat{f}_2 \end{array} \right.$$

Para verificar la cerradura bajo multiplicación, se puede modelar un modelo de 4 reglas, de manera similar a la anterior [Jang and Sun, 1995].

Así, bajo una clase apropiada de funciones de pertenencia, un modelo de Sugeno de orden cero cumple las cuatro propiedades del teorema 2.3.1. Por tanto, para  $\epsilon > 0$  dado y cualquier función  $g \in C(D)$ , existe un modelo de Sugeno de orden cero  $S$  tal que para todo  $x \in D$ , se cumple que  $|g(x) - S(x)| < \epsilon$ .

## 2.4. Sistemas adaptativos de inferencia neuro-difusa (ANFIS)

Se puede establecer una analogía entre los modelos de inferencia difusa de Sugeno y las redes neuronales artificiales<sup>5</sup>. Dichas redes brindan una interpretación gráfica (hasta cierto grado) del funcionamiento de dichos modelos y permiten el uso de herramientas relacionadas con redes artificiales que ayudan a la optimización de los mismos.

Gracias a su arquitectura, los sistemas difusos y las redes neuronales son herramientas complementarias en la construcción de sistemas inteligentes. Mientras que las redes neuronales son estructuras de bajo nivel computacional que se desempeñan de buena manera cuando se trabaja con datos en bruto, la lógica difusa se ocupa de un razonamiento a un mayor nivel. Sin embargo, los sistemas difusos carecen de la capacidad de aprender y no pueden adaptarse por sí mismos. La fusión de una red neuronal con un sistema difuso es un sistema integrado que permite la construcción de modelos de predicción a corto plazo [Negnevitsky and Potter, 2006].

### Arquitectura ANFIS

Un sistema adaptativo de inferencia neuro-difusa (ANFIS) es un tipo de red adaptativa feed-forward, donde cada uno de sus nodos consta de un tipo de funciones (ya sea que tengan o no parámetros adaptables) que reciben los datos de salida de los nodos anteriores, los procesa y los pasa a los nodos de la siguiente capa. Dicho sistema está conformado por 5 capas, dispuestas de la forma mostrada en 2.10.

Dado que el manejo de la información de un ANFIS se maneja por capas, se verá cómo funciona cada capa. Dicha explicación puede ser vista a mayor profundidad en [Jang and Sun, 1995], [Ata and Kocyigit, 2010]. Un ANFIS incorpora las reglas vistas en la sección 2.1.1 y permite pequeños ajustes en los parámetros de las funciones de pertenencia de acuerdo a un par de datos de entrada-salida.

---

<sup>5</sup>Puede verse a mayor profundidad la teoría de redes neuronales en [Russell et al., 1995]

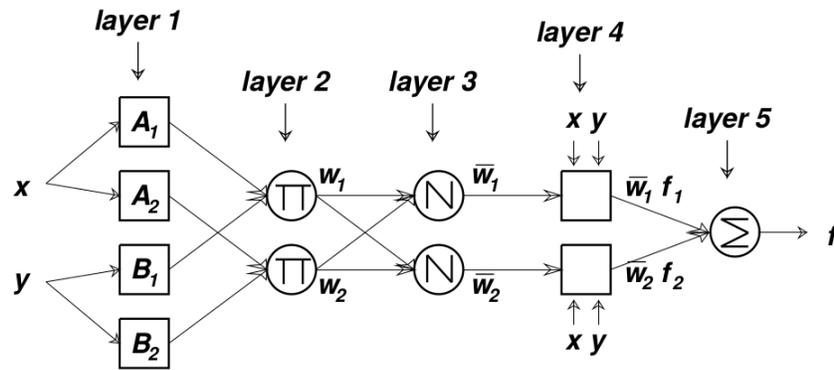


Figura 2.10: Arquitectura de un ANFIS. Cada capa puede tener un conjunto de nodos dependientes de parámetros modificables, conectados de tal modo que el flujo de datos sea hacia adelante (red adaptativa multicapa feed-forward). Los nodos adaptativos son representados por cuadrados, mientras que los nodos circulares representan nodos fijos. Tomado de [Jang and Sun, 1995]

Supóngase por simplicidad un sistema de inferencia difusa formado por reglas que constan de dos entradas  $x, y$  y una salida  $z$ . Para un modelo difuso de Sugeno de primer orden, un conjunto de reglas está dado de la forma

$$\begin{cases} \text{Si } x \text{ es } A_1 \text{ y } y \text{ es } B_1 \text{ entonces } z = p_1x + q_1y + r_1 \\ \text{Si } x \text{ es } A_2 \text{ y } y \text{ es } B_2 \text{ entonces } z = p_2x + q_2y + r_2 \end{cases}$$

La salida del  $i$ -ésimo nodo en la capa  $l$  es denotado por  $O_{l,i}$ . El proceso de razonamiento que sigue dicho sistema, para un conjunto de parámetros  $\{\theta, p_i, q_i, r_i\}, i = 1, 2$  siendo  $\theta$  el conjunto de valores de la función de pertenencia es análogo al modelo de inferencia difusa de Sugeno.

### Capa 1

Todo nodo  $i$  en esta capa es un nodo adaptativo con una función nodo de la forma

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x), \quad i = 1, 2 \\ O_{1,i} &= \mu_{B_{i-2}}(y), \quad i = 3, 4 \end{aligned}$$

donde  $x$  (o  $y$ ) es la entrada en el nodo  $i$  y  $A_i$  (o  $B_i$ ) es un conjunto difuso. En otras palabras,  $O_{1,i}$  es el grado de pertenencia de un conjunto difuso  $A$  ( $A = A_1, A_2, B_1, B_2$ ). La función de pertenencia puede ser cualquiera de las vistas anteriormente, al cambiar los parámetros se obtienen distintas MF.

### Capa 2

Todo nodo en esta capa es un nodo fijo  $\Pi$ , donde cada salida es el producto de las señales entrantes  $O_{2,i} = \omega_i = \mu_{A_i}(x)\mu_{B_i}(y), i = 1, 2$ .

**Capa 3**

Todo nodo en esta capa es un nodo fijo  $N$ . el  $i$ -ésimo nodo calcula la razón entre cada  $\omega_i$  y la suma de los  $\omega_i$ , es decir, las salidas de estos nodos son de la forma

$$O_{3,i} = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, i = 1, 2$$

**Capa 4**

En esta capa, los nodos que la conforman son nodos adaptativos con una función nodo

$$\begin{aligned} O_{4,i} &= \bar{\omega}_i(f_i) \\ &= \bar{\omega}_i(p_i x + q_i y + r_i) \end{aligned}$$

siendo  $\{p_i, q_i, r_i\}$  el conjunto de parámetros de esta capa.

**Capa 5**

EL único nodo de esta capa es un nodo etiquetado  $\Sigma$  que computa la salida de las señales entrantes

$$\begin{aligned} O_{5,1} &= \sum_i \bar{\omega}_i f_i \\ &= \frac{\sum_i \omega_i f_i}{\sum_i \omega_i} \end{aligned}$$

Al igual que una red neuronal, un ANFIS tiene dos momentos: el paso hacia adelante, y el paso hacia atrás. El paso hacia adelante parte de un conjunto de datos de entrada y el modelo calcula las salidas capa por capa, dando como resultado una salida general. Los parámetros consecuentes de la regla son optimizados mediante el método de mínimos cuadrados. Una vez se ha concluido este proceso, se realiza el paso hacia atrás; dicho paso consiste en la variación de los parámetros de los antecedentes a medida que los consecuentes se mantienen constantes. En un ANFIS, ambos pasos se realizan mediante un algoritmo de aprendizaje híbrido, basado en el método del descenso del gradiente, y la dependencia implícita de las capas respecto a las capas anteriores.

# Capítulo 3

## Construcción y Aplicación

La complejidad implícita del mercado de divisas ha forzado a los inversores a desarrollar una serie de herramientas que permitan mejorar sus ganancias en las operaciones financieras. Recientemente, se han adoptado técnicas provenientes del campo de la inteligencia artificial que complementan el desarrollo y funcionamiento de los sistemas financieros que implican la toma de decisiones.

Según cuenta [Naranjo and Santos, 2019], los sistemas que involucran la lógica difusa han tenido una gran variedad de éxitos gracias a su capacidad subyacente para emular el proceso de pensamiento humano. Esto la convierte en una herramienta supremamente útil a la hora de modelar procesos con aplicabilidad al mercado de divisas, dado que la complejidad del mismo no solo radica en el volumen inmenso de datos de los que se dispone: su principal dificultad está en el entendimiento del comportamiento humano, siendo éste el que guía de manera determinante las fluctuaciones del mercado.

Gracias a su simplicidad y a su interpretabilidad, los modelos difusos brindan un enfoque que facilita la toma de decisiones que tienen como objeto algún beneficio financiero, hecho que los hace especialmente útiles en el mercado de divisas. Sin embargo, es necesario destacar que para un buen desempeño en dicho mercado se requiere no solo una buena capacidad y análisis de datos; se requiere además contar con una forma efectiva y eficiente de manejo de capital y gestión monetaria, siendo todo esto estructurado mediante una estrategia de inversión.

Un problema importante comúnmente enfrentado por inversores individuales e institucionales es el de determinar el momento idóneo para poder comprar o vender un par. Al observar los precios predominantes en el mercado, un inversor especulativo puede elegir entrar en el mercado inmediatamente o esperar por una oportunidad en el futuro. Al completarla, el inversor debe determinar cuando es el mejor momento para cerrar la posición. Es por esto que se busca una secuencia de tiempos óptimos para las operaciones.

Este capítulo provee un acercamiento sistemático sobre el tiempo óptimo de negociación en los mercados de precios con dinámica de reversión a la media. Se presenta un enfoque

aplicado que destila cuestiones matemáticas fundamentales desde los diferentes problemas de comercio (de capítulos anteriores), y también incorpora aspectos prácticos de negociación tales como manejo del riesgo. Aunque el núcleo principal de esta tesis es matemático, varios problemas básicos de trading son discutidos en este capítulo guiados por estudios analíticos y numéricos. A partir de la reversión a la media en el largo plazo y su predictibilidad dependiendo de los datos anteriores en el corto plazo, en presencia de heterocedasticidad condicional, proponemos un procedimiento de reversión a la media condicional junto a un modelo ANFIS de inversión variable. Los resultados indican que los instrumentos analizados revierten a la media a lo largo del periodo 2015 a 2019 exhibiendo dinámicas explosivas en diferentes periodos.

### 3.1. Mercado de Divisas

Se define una divisa como una moneda internacional respecto a un país de referencia. Forex (siglas para Foreign Exchange o mercado de divisas) es un mercado internacional que permite el comercio de determinado conjunto de divisas, predicción de valores y variaciones en tasas de cambio, lo que le da la característica de ser uno de los mercados financieros más importantes a nivel mundial. Forex funciona de manera descentralizada, es decir, permite un intercambio directo entre compradores y vendedores. Gracias a tal funcionamiento, dicho mercado es uno de los mercados con mayor liquidez (es decir, tiene la cualidad de que los activos con los que se comercia pueden ser transformados en dinero fácilmente) del mundo, teniendo la capacidad de mover más de 5 billones de dólares por día según las estadísticas del Banco de Pagos Internacionales. Además, la forma descentralizada en la que funciona y el amplio acceso tecnológico permiten una participación a un nivel bastante general.

En lo que respecta a transacciones, el funcionamiento Forex se basa en el intercambio de pares de divisas: la primera, denominada la divisa base, es la divisa sobre la cual se quieren hacer las operaciones financieras. La segunda, denominada la divisa cotizada, es con la cual se hacen las operaciones. Dicho par a menudo se suele denotar por divisa 1/divisa 2 o por  $\text{div1div2}$ . Por ejemplo, considérense las divisas EUR (euro) y USD (dólar estadounidense). El par EURUSD indica que el activo sobre el cual se hacen las operaciones (sean compra o venta) es el euro (las primeras tres letras) y que el medio por el cual se hace dicho intercambio es el dólar (últimas tres). Cuando se quiere comprar un par de divisas, se hace alusión a que se espera una mejora del precio de la divisa base respecto a la divisa cotizada, y cuando se quiere vender, ocurre lo contrario, es decir, se espera una mejora de la divisa cotizada respecto a la divisa base. Cuando se compra un par de divisas, lo que ocurre es que se compra la divisa base al tiempo que se vende la divisa cotizada, y cuando se vende un par, es viceversa; se vende la divisa base y se compra la cotizada.

## Candlestick o Velas Japonesas

El precio de una divisa dada puede verse ampliamente afectado dados distintos factores. Es por ello que se torna importante una forma de manejo de la información que sea eficiente y que proporcione el máximo de información posible. Los Candlestick (o velas japonesas), son una serie de estructuras que indican distintos valores que el precio de dicha divisa alcanzó en determinado intervalo de tiempo.

En los mercados financieros, los candlestick se usan para resumir la evolución del precio (en este caso, de una divisa) en un determinado intervalo de tiempo. Dichos candlestick son definidos principalmente con base a 4 parámetros: precio de apertura (precio que tiene la divisa en el inicio del intervalo de tiempo), precio de cierre (precio que tiene la divisa al término de dicho lapso de tiempo) y precios máximo y mínimo. Según la disposición de los precios de apertura y de cierre, los candlestick se pueden clasificar en dos tipos: alcistas y bajistas. El alcista se identifica por el hecho de que el precio de cierre fue mayor que el precio de apertura. Por su parte, el bajista muestra que el precio de apertura fue mayor que el precio de cierre. Ambos pueden apreciarse en la siguiente figura

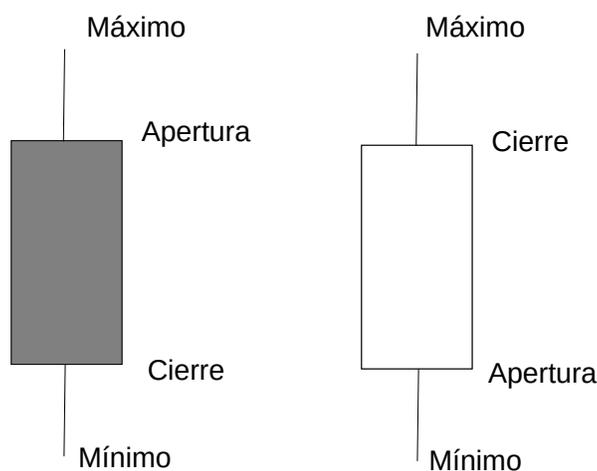


Figura 3.1: Candlestick bajista y alcista (en dicho orden)

## Series de Precios

Los datos, cuando están ordenados de la manera que muestran los candlestick, brindan no solo información que no se obtendría de manera inmediata con otro tipo de estructuras; además, dichos gráficos dan una ayuda visual al estar la información ordenada de manera conveniente, lo que permite apreciar a mayor profundidad el comportamiento de un precio en un lapso de tiempo. A partir de la disposición de la información mostrada en dichos candlesticks, pueden extraerse un conjunto de datos, útiles para entender dicho comportamiento.



Figura 3.2: Serie de precios del par EURUSD desde 2019.03.11 hasta 2020.03.11

Los Candlestick se forman cada cierto intervalo de tiempo (ya sea cada minuto, cada hora, cada día, semana o mes). En cada uno de dichos candlestick, se encuentra condensada la información del comportamiento del precio en dicho intervalo de tiempo. De cada uno, se pueden extraer los cuatro precios mencionados anteriormente, y con cada uno de estos precios, se forman distintas series de tiempo. En este trabajo, particularmente, se utilizarán los precios de cierre de cada intervalo de tiempo, y con base a ellos se formulará el modelo.

## 3.2. Planteamiento del Modelo

El comportamiento del precio de un par de divisas puede ser bastante complejo, y por ello a menudo suele ser extremadamente difícil determinar el punto ideal para el cual se puede comprar o vender dicho par. Junto a esto, se suma el hecho de que se requiere un buen manejo del capital: saber qué cantidad es la adecuada para invertir a partir de las condiciones determina en gran medida el porcentaje de beneficio que se puede esperar de una acción de compra o venta. Debido a esto, los modelos aplicados al mercado de divisas suelen tener componentes tanto de predictibilidad de la serie como de manejo de riesgo o de porcentaje de inversión, con el fin de minimizar pérdidas y lograr obtener un beneficio en el transcurso de la operación.

Con base a lo anterior, en lo que sigue se plantea un modelo que tome en cuenta tanto una componente predictiva como una componente de manejo del tamaño de la inversión, ambos teniendo en cuenta la teoría desarrollada de las secciones previas. El diseño y aplicación de este modelo se hará en el software libre Metatrader 5, equipado con un editor de código y un lenguaje propios para el trabajo concerniente al mercado de divisas.

### 3.2.1. Predictibilidad

La predictibilidad del precio de un activo financiero es un aspecto importante que siempre ha recibido un gran interés dentro de la modelación de la dinámica de un instrumento financiero. Mediante diferentes enfoques, se busca capturar una serie de hechos de las series de precios, y con base a ellos mostrar que en algún grado estas son predecibles. Entre los muchos resultados de investigación obtenidos al respecto, ha sido ampliamente observado que muchas de estas dinámicas provenientes de los mercados financieros exhiben una característica de reversión a la media.

El análisis de reversión a la media en mercados financieros atrae el interés de muchos inversores, profesionales del mercado e investigadores. Dicho análisis tiene implicaciones fundamentales en las decisiones de inversión, desde la selección de un portafolio hasta los precios de opciones financieras; por ejemplo, un mercado con reversión a la media sugiere que sus instrumentos presentan menor riesgo a largo plazo. Los hallazgos al día de hoy indican que los precios con reversión a la media son más comunes que lo que se encontraba anteriormente.

### Reversión a la Media

Con el fin de determinar el valor esperado condicional del proceso estocástico que genera el conjunto de datos de la serie, se usarán los siguientes modelos (casos particulares de los modelos de series temporales desarrollados anteriormente), que constituyen una serie de medios diversos que describen adecuadamente la dinámica de las tasas de cambio.

- Media móvil simple (SMA, Simple Moving Average)

$$E[y_t | \mathbf{x}_{t-1}] = \text{SMA}_p(t) = \frac{1}{p} \sum_{i=1}^p y_{t-i} \quad (3.2.1)$$

- Media móvil exponencial (EMA, Exponential Moving Average)

$$E[y_t | \mathbf{x}_{t-1}] = \text{EMA}_p(t) = \alpha y_{t-1} + (1 - \alpha) \text{EMA}_p(t-1) \quad (3.2.2)$$

donde  $\alpha$  es una constante que determina el porcentaje de uso del valor del precio.

- Media móvil suavizada (SMMA, Smoothed Moving Average)

$$E[y_t | \mathbf{x}_{t-1}] = \text{SMMA}_p(t) = (t-1) \text{SMMA}_p(t-1) + \frac{y_{t-1}}{n} \quad (3.2.3)$$

- Media móvil ponderada linealmente (LWMA, Linear Weighted Moving Average)

$$E[y_t | \mathbf{x}_{t-1}] = \text{LWMA}_p(t) = \frac{1}{p} \sum_{i=1}^p i y_{t-i} \quad (3.2.4)$$

Cada tipo de media móvil refleja características particulares del mercado: por su estructura, la media móvil simple asigna igual importancia a cada observación del mercado. Principalmente, este tipo de media móvil permite, dependiendo del número de precios usados, dar un pronóstico más o menos acertado de la tendencia de los precios a corto y mediano plazo. La media móvil ponderada linealmente se basa en las afirmaciones de que los precios más recientes del mercado pueden tener información sobre el comportamiento futuro de la serie. Por esto, a cada precio se le asigna un peso creciente relativo a su temporalidad: cuanto más reciente es la observación, mayor peso tendrá dicho precio, lo que contribuirá a que dicha media sea más sensible a los cambios relacionados con los últimos precios. Aunque la media móvil linealmente ponderada permite destacar (mediante los pesos) la influencia que pueden tener los últimos precios respecto a un valor futuro, la misma queda relegada a la asignación de los pesos a dichos precios, lo cual puede llegar a limitar el comportamiento de la media, en tanto que afecta la sensibilidad de los cambios de los valores de los posibles precios. Por su parte, la media móvil exponencial permite una asignación más variable de valores de los pesos dependiendo del valor de la constante  $\alpha$ ; mediante el ajuste de este valor (tomado como un valor en  $[0, 1]$ ), puede darse mayor peso a los precios en determinado rango, ya sea a los más recientes o a los más antiguos. Debido a la ponderación de los precios en las medias anteriores, estos modelos adquieren una especie de sensibilidad a los cambios relativos a los últimos precios. Un inconveniente derivado de lo anterior es que, se aumenta la probabilidad de falsas señales de tendencia. Respecto a la media móvil suavizada, permite reducir comportamiento volátiles de una serie, proporcionando una forma más clara y definida de la dirección de la tendencia. El funcionamiento de una media móvil suavizada se basa en la teoría de series abordada anteriormente, y el aspecto de suavizado hace referencia al proceso de diferenciar la serie. [Zakamulin, 2017]

Las fórmulas (3.2.1), (3.2.2), (3.2.3) y (3.2.4) son usadas con el propósito de acelerar los cálculos de cómputo, disminuyendo el número de operaciones hechas en cada iteración. Adicionalmente a lo anterior, existen propiedades relativas al tiempo de adaptación de cada modelo a la actualización de los datos (retroceso), y la forma en que pueden calculados y reducidos de acuerdo al modelo. Estas pueden ser vistas en [Zakamulin, 2017]

### **Acumulación y Distribución**

Cada instrumento financiero está regido principalmente por dos dinámicas: una de corto plazo y la otra de largo plazo. La media condicional de largo plazo descrita por alguno de los modelos mencionados anteriormente rige el comportamiento general, mientras que una media condicional de corto plazo rige las desviaciones del precio desde la media de largo plazo, para finalmente retornar a la media de largo plazo de manera reiterativa. Al comportamiento anterior, se le denomina un comportamiento de reversión a la media.

Así, los tiempos de entrada y salida del mercado quedan determinados por el inicio y el final de cada reversión a la media por parte del precio, es decir, por los cruces el modelo condicional de largo plazo y el de corto plazo. Tras el inicio de un nuevo cruce entre ambos



Figura 3.3: Un proceso de reversión a la media condicional determinado por un modelo SSMA de largo plazo y un modelo EMA de corto plazo.

modelos condicionales, una operación de venta (resp. compra) termina y una nueva operación de compra (resp. venta) inicia, dando lugar a un comportamiento cíclico propio de este tipo de mercados. La implementación de dicho comportamiento se hará como se muestra a continuación en `mql5`.<sup>1</sup>

El algoritmo 1 muestra el funcionamiento general del modelo: se determinan los periodos para la media móvil de posición larga y la de posición corta. Seguidamente se determinan unos arreglos dinámicos donde se guardarán los resultados de las medias de corto y largo plazo, con el fin de acceder a ellos posteriormente. Se determinan el número de velas al inicio de la ejecución y se crea una variable que guarda la cantidad de velas dadas en un intervalo de tiempo. Al verificar si el número total de velas es diferente al de previas, se indica que se formó una nueva vela, con lo cual se procede a verificar las condiciones de la operación. Para la venta, se verifica si el precio de apertura de la vela es mayor que el de cierre, y se verifica si el precio de cierre está acotado por las dos medias de corto y largo plazo. Para la parte de compra, se verifica si el precio de cierre es menor que la media a corto plazo, y ésta a su vez es menor que la media a largo plazo. Si se cumple además que el precio de cierre es mayor que el precio de apertura, se procede a ejecutar la operación. La idea es que el precio siempre esté acotado (ya sea por abajo o por arriba) por ambas medias, y que la operación se mantenga mientras que dicha propiedad de acotamiento persista.

En esta parte, sólo se determinaron las condiciones para las cuales el modelo ejecutará una orden en el mercado, lo cual es una parte del modelo final. Para la segunda parte, se

<sup>1</sup>`mql5` es un lenguaje para la plataforma de trading Metatrader. Es un derivado de C, con funciones especializadas para este caso.

**Algorithm 1** Abrir posiciones de compra y venta

---

```

Input Periodo Corto, periodo largo, tipos de media móvil.
Output Posiciones de compra, venta
MA-1 Period  $\leftarrow$  Periodo Corto
MA-2 Period  $\leftarrow$  Periodo Largo
MA-Fast[ ]  $\leftarrow$  Datos de MA corto (Dinámico)
MA-Long[ ]  $\leftarrow$  Datos de MA largo (Dinámico)
Total-Candle  $\leftarrow$  Total de velas al inicio
Prev-Candle  $\leftarrow$  velas previas a una operación
if Total-Candle  $\neq$  Prev-Candle then
    Total-Candle = Prev-Candle
    if Price-Close > MA-FAST[1] y MA-Fast[1]  $\leq$  MA-Long[1] y Price-Close < Price-
    Open then
        Vender (Tasa de inversión: a determinar)
    end if
    if Price-Close < MA-Long[1] y MA-Fast[1]  $\geq$  Ma-Long y Price-Close > Price-Open
    then
        Comprar (Tasa de inversión: a determinar)
    end if
end if

```

---

determina la tasa de inversión mediante un modelo sencillo de inferencia neuro-difusa, que calcule la tasa de inversión con base a la intensidad con la que se ve reflejada la tendencia del mercado.

### 3.2.2. Tamaño de Inversión Variable

Existen variedad de formas en las que la lógica difusa ayuda a modelar cuestiones de amplia vaguedad en variedad de situaciones. En el mercado de divisas, esto es particularmente útil, dada la infinidad de posiciones y criterios que los inversores tienen a la hora de efectuar sus operaciones. Como ejemplos de aplicación, en [Naranjo and Santos, 2019], se propone un modelo de decisión basado en lógica difusa mediante la recopilación de algunos patrones formados por conjuntos de velas, definir una serie de reglas difusas con base a ellos y posteriormente elaborar un sistema de inferencia de tipo Mamdani, junto a una estrategia de inversión dada.

Para el manejo de la tasa de inversión, se propone un modelo neuro-difuso básico que provea porcentajes de inversión con base a un medidor de la intensidad de la tendencia. Dicha intensidad será abordada con base a una de las herramientas que brinda Metatrader 5, para posteriormente formular las reglas de inferencia.

Una de las tantas herramientas que trae Metatrader para modelar y entender el com-

portamiento de las series de precios son los indicadores. Dichos indicadores funcionan con base a un conjunto de fórmulas y modelos y permiten concluir gran variedad de aspectos y características de la serie. Es el caso del indicador *índice de fuerza direccional* (Average Movement Index - ADX). La idea intuitiva de dicho indicador es que, con base a un conjunto de velas (tomadas desde la última vela formada hacia atrás) se concluye la intensidad de la tendencia actual y su posible dirección en un rango de valores de 0 a 100. lo que funciona además como un complemento para los modelos de medias móviles mencionados en la parte de predictibilidad (puede consultarse más información sobre el funcionamiento y el cálculo de dichos valores en [Naranjo et al., 2015]). Respecto a la salida de dicho indicador, existen convenciones para indicar la intensidad de la tendencia del mercado en un instante dado. De los resultados del ADX, puede catalogarse la tendencia principalmente en tres fases: débil, promedio y fuerte. Para esta separación, no hay cantidades fijas que indiquen cuándo se pasa de un tipo de tendencia a otra, lo que hace que definir un rango adecuado para el manejo de la misma sea muy variable. Es por ello que la lógica difusa funciona bien en estos casos, pues permite modelar regiones de manera flexible, que vaya mas acorde a la experiencia del usuario, o en este caso, de inversor.

A continuación, se mostrará la estructura del modelo ANFIS considerados. Los valores del ADX se tomarán en el rango de 30 a 100, con la siguiente disposición de los conjuntos *débil*, *promedio* y *fuerte*, que indican la intensidad de la tendencia en desarrollo:

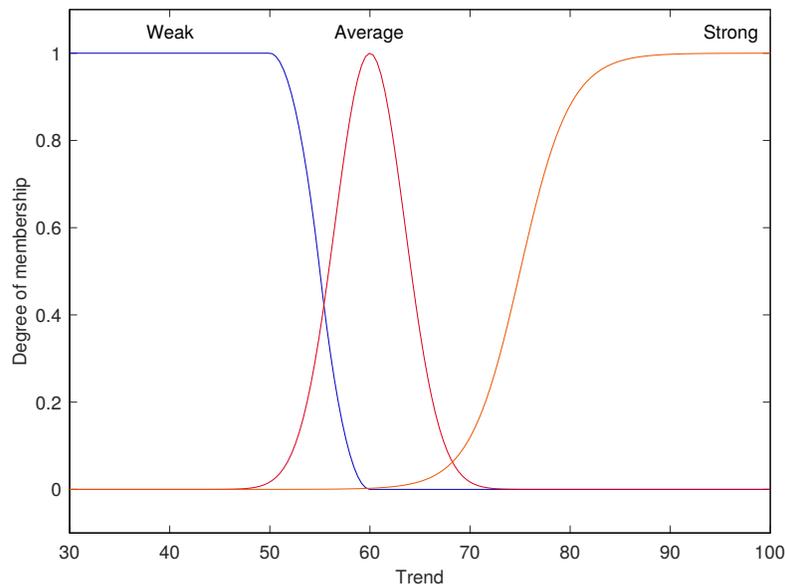


Figura 3.4: Funciones de pertenencia para los términos de tendencia *débil*, *promedio* y *fuerte*

De forma analítica, las funciones de pertenencia se expresan de la siguiente forma:

- *Débil:*

$$\mu_1(x; a, b) = \begin{cases} 1, & x \leq a; \\ 1 - 2 \left( \frac{x-a}{b-a} \right)^2, & a \leq x \leq \frac{a+b}{2}; \\ 2 \left( \frac{x-b}{b-a} \right)^2, & \frac{a+b}{2} \leq x \leq b; \\ 0, & x \geq b; \end{cases}$$

donde  $a = 50, b = 60$ .

- *Promedio:*

$$\mu_2(x; a, b) = e^{-\frac{(x-b)^2}{2a^2}}$$

para  $a = 3.5, b = 60$ .

- *Fuerte:*

$$\mu_3(x; a, b) = \frac{1}{1 + e^{-a(x-b)}}$$

En este caso,  $a = 0.4, b = 75$

Se definen las constantes  $c_1, c_2, c_3$  que funcionarán como parámetros de la parte consecuente de las reglas.

### Forma de las reglas y Funcionamiento

Como se ha mencionado anteriormente, el modelo dependerá del nivel del indicador ADX y dará como resultado sugerencia de porcentajes de riesgo (de inversión) que serán aplicados a un capital inicial. La estructura de las reglas del modelo es la siguiente: si  $x$  es el valor del ADX, bajo el hecho de que se toma  $x \geq 30$ , se consideran las siguientes reglas:

$$\begin{aligned} \text{If } x \text{ is Weak} &\Rightarrow y = c_1 \\ \text{If } x \text{ is Average} &\Rightarrow y = c_2 \\ \text{If } x \text{ es Strong} &\Rightarrow y = c_3 \end{aligned}$$

La forma de las reglas muestra que el proceso de inferencia difusa está dado por uno del tipo Sugeno de orden cero; cada una de las constantes de la parte consecuente hace referencia a un nivel de riesgo de acuerdo al valor  $x$  de entrada. Para valores de  $x$  entre 0 y 30, se toma una tasa de inversión de 0.01 del capital. La salida de dicho modelo varía entre 2 y 10, con lo cual, para tener un nivel de inversión aceptable, se multiplica el resultado del modelo por 0.01.

El algoritmo 2 muestra de forma general el funcionamiento de dicho sistema de inferencia. Bosqueja el funcionamiento y la organización de las capas vistas a partir de la figura 2.10

---

**Algorithm 2** Tasa de inversión Variable

---

**Input** constantes  $c_1, c_2, c_3$ , valor del ADX (denotado por  $x$ )  
**Output** porcentaje de inversión  $r$   
**for**  $i \in \{1, 2, 3\}$  **do**  
     $y_i \leftarrow c_i$   
     $\omega_i \leftarrow \mu_i(x)$  (Definidas anteriormente)  
     $\bar{\omega}_i \leftarrow \frac{\omega_i}{\omega_1 + \omega_2 + \omega_3}$   
**end for**  
 $R \leftarrow \text{sum}(c_i \bar{\omega}_i, i \in \{1, 2, 3\})$  (suma las componentes  $c_i \bar{\omega}_i$ )  
 $r = (0.01)R$

---

Ahora, con base a los procedimientos mencionados, se tiene entonces una forma general del modelo basado en los tipos de medias móviles y la tasa de inversión. Dicho modelo funciona de manera dinámica; al tener como base los precios de apertura y de cierre, se proveen los parámetros de entrada tanto de la parte predictiva como de la parte difusa. Después, se realiza el procedimiento determinado en el algoritmo 1, para examinar las condiciones de compra y venta. Al darse dichas condiciones, se procede a ejecutar la operación con los niveles de inversión dados por el algoritmo 2. Es de resaltar nuevamente que lo anterior se realiza solo cada vez que se detecta un nuevo precio de cierre, y no cada variación del precio en un tiempo dado.

---

**Algorithm 3** Posiciones de compra y venta - Inversión Variable

---

**Input** Periodo Corto, periodo largo, tipos de media móvil,  $c_1, c_2, c_3$   
**Output** Posiciones de compra, venta, inversión  $r$  variable (Dada por el algoritmo 2)  
MA-1 Period  $\leftarrow$  Periodo Corto  
MA-2 Period  $\leftarrow$  Periodo Largo  
MA-Fast[ ]  $\leftarrow$  Datos de MA corto (Dinámico)  
MA-Long[ ]  $\leftarrow$  Datos de MA largo (Dinámico)  
Total-Candle  $\leftarrow$  Total de velas al inicio  
Prev-Candle  $\leftarrow$  velas previas a una operación  
**if** Total-Candle  $\neq$  Prev-Candle **then**  
    Total-Candle = Prev-Candle  
    **if** Price-Close  $>$  MA-FAST[1] y MA-Fast[1]  $\leq$  MA-Long[1] y Price-Close  $<$  Price-Open **then**  
        Vender (Lote de tamaño  $r$ )  
    **end if**  
    **if** Price-Close  $<$  MA-Long[1] y MA-Fast[1]  $\geq$  Ma-Long y Price-Close  $>$  Price-Open **then**  
        Comprar (Lote de tamaño  $r$ )  
    **end if**  
**end if**

---

Algo que no se mencionó anteriormente es que dicho algoritmo puede abrir varias operaciones (compra/venta) de manera sucesiva, y que dadas ciertas condiciones el mismo puede cerrar una operación abierta o modificar una existente. El modelo da entonces una forma de entender y pronosticar el comportamiento de una serie con una forma de cálculo de inversión variable.

En lo que sigue, se muestra la aplicación del modelo descrito por el algoritmo 3, a un conjunto de 6 pares de divisas, y se hace un pequeño análisis de los resultados.

### 3.3. Aplicación y Resultados

Antes de proceder con la aplicación del modelo, se definirán dos términos, importantes para entender los resultados:

- **Balance:** hace referencia al capital inicial de una cuenta. Dicho capital se ve afectado solo cuando se ven consolidadas las operaciones de compra o venta, y se ve afectado sólo cuando existen pérdidas o ganancias fijas.
- **Patrimonio:** es el resultado del balance de la cuenta junto con las potenciales ganancias o pérdidas que pueden generar las operaciones de compra o venta, tanto consolidadas como sin consolidar.

La aplicación de los modelos propuestos y la estrategia mencionada se llevó a cabo sobre el conjunto de 6 instrumentos financieros: EURUSD, GBPUSD, USDCHF, USDJPY, EURJPY y CADCHF, en un rango de tiempo de 3 años; desde 2017-01-01 hasta 2020-05-15. La ejecución se llevó a acabo partiendo de un capital inicial de 10000 USD sobre los precios de apertura, cierre, máximo y mínimos en la temporalidad D1 (diaria). El procesador empleado para la ejecución fue un Intel(R) Core(TM) i5-5200U Quad Core, 4TH GEN, 2.20 GHz, 4G RAM, con tiempos variables en los procesos de optimización. La técnica usada para la optimización de los tipos de modelos, los periodos de los mismos y las constantes  $c_1, c_2, c_3$  fue a base de algoritmos genéticos, una técnica incorporada en el software Metatrader.

Instrumento	Modelo		Constantes		
	Largo Plazo	Corto Plazo	$c_1$	$c_2$	$c_3$
EURUSD	SMA(63)	SSMA(33)	3.4	4.8	9.8
GBPUSD	LWMA(100)	SMA(30)	3.8	4.8	9.0
USDCHF	LWMA(90)	EMA(10)	4.0	4.8	6.8
USDJPY	LWMA(130)	SSMA(100)	4.0	5.0	9.2
EURJPY	EMA(190)	SMA(120)	3.0	5.0	9.0
CADCHF	LWMA(160)	SSMA(140)	3.6	5.0	9.0

Cuadro 3.1: Estadísticas del desempeño del modelo propuesto balance inicial de 10000 USD sobre el periodo de tiempo de 3 años: desde 2017-01-01 hasta 2020-05-15.

A partir de la tabla anterior, se hacen simulaciones del comportamiento del modelo en los diferentes instrumentos. Se obtienen los siguientes resultados:

<b>Instrumento</b>	<b>Rentabilidad</b> (USD)	<b>Operaciones</b>
EURUSD	11422	86
GBPUSD	11642	122
USDCHF	10366	74
USDJPY	22764	298
EURJPY	16266	166
CADCHF	16508	180

Cuadro 3.2: Desempeño de los modelos óptimos con un balance inicial de 10000 USD sobre el periodo de tiempo de 3 años: desde desde 2017-01-01 hasta 2020-05-15

La figura 3.5 muestra el resultado de la aplicación de ambas partes del modelo de manera conjunta: por un lado, la curva de balance (la línea azul) presenta un aumento progresivo a lo largo del tiempo. Esto indica que, al menos para la temporalidad establecida, el comportamiento de la serie está dado por una dinámica reversión a la media condicional, donde cada modelo de reversión está dado de manera particular según el instrumento financiero (ver tabla 3.1).

Por otro lado, la curva de patrimonio (la línea verde) presenta un comportamiento muy variable: esto indica que, durante el tiempo en que las operaciones estuvieron sin consolidar, el margen de ganancias y pérdidas se vió ampliamente influenciado por la cantidad de capital invertida, siendo dicha cantidad dictada por el modelo de inversión variable. A pesar de su amplia variabilidad a lo largo del tiempo, a menudo presentando un comportamiento similar al del balance, y al igual que este, muestra un comportamiento al alza: esto que indica que el comportamiento del mismo se vió afectado por la estrategia de predictibilidad dada por el modelo de reversión a la media.

En la mayoría de los casos, en los tramos finales de las simulaciones el patrimonio suele mostrar un comportamiento, como ya se dijo, casi análogo al del balace, esto podria indicar como el modelo de tasa variable se adapta a al comportamiento del capital de inversión a largo plazo, de manera un poco más uniforme.

### 3.4. Discusión

Aunque la familia de modelos elegidos destaca por su baja complejidad frente a los modelo completos expuestos a lo largo de esta tesis, los resultados obtenidos muestran que las tasas de cambio elegidas tienen a exhibir reversión a la media condicional unas en mayor medida que otras, al igual que tienden a ser mas viables ciertos instrumentos a tasas de inversión variable

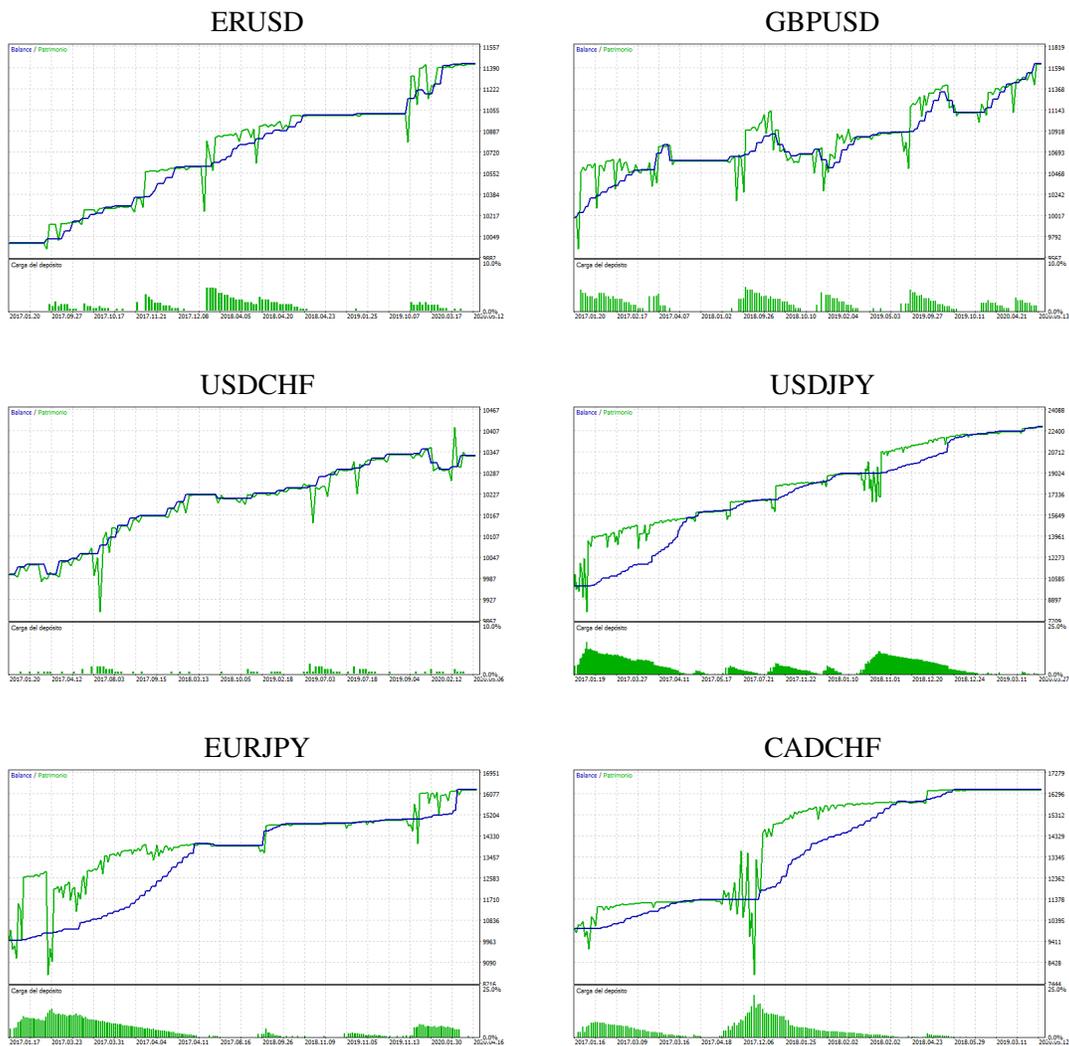


Figura 3.5: Curvas de balance y patrimonio para la temporalidad D1, muestran la dinámica del precio durante los últimos 3 años.

que otros. Por ejemplo, en los casos USDCHF, EURUSD y GBPUSD, presentan un margen de ganancia casi nulo relativo al capital inicial y al tiempo de aplicación. Esto los deja, en principio, como opciones inviables en la ejecución real del modelo.

Es de resaltar el hecho de que el modo de inversión se hizo, de cierta manera a priori, debido a la falta de experiencia en este campo. Sin embargo, en principio resulta viable para maximizar las posibles ganancias en tanto la parte predictiva del modelo permita ejecutar una operación. Esto podría entonces inspirar estudios rigurosos sobre gestión de riesgo que permitan un margen variable, para lograr una implementación más adecuada de modelos de inferencia difusa junto a dichas técnicas de gestión.

Aunque con este tipo de modelos se ilustra de forma relativamente clara un modo más integral de aplicación de conceptos de áreas diferentes, se hace importante el hecho de que se han dejado factores que afectan el comportamiento a distintos tiempos de la serie, como lo es el estudio de la volatilidad de la misma. Nuevamente, esto implica que se pueden abordar modos de vincular dicha volatilidad al estado actual del modelo, permitiendo que este pueda tener más factores en cuenta, y posiblemente brindar mejores resultados.

## Conclusiones y Trabajos a Futuro

En este trabajo, se ilustra un modelo matemático flexible a través de la media condicional para los tiempos óptimos de comercio en el mercado de divisas para diferentes instrumentos, sobre la base de modelos para la media condicional, una hipótesis de reversión a la media y una parte de adaptabilidad al manejo de riesgo en determinadas condiciones del mercado. Se hizo sobre un rango de tiempo de 3 años, considerando los distintos tipos de precios de la serie formados cada día, los tiempos óptimos de apertura y cierre de las operaciones dados por las dinámicas de reversión a la media ayudan al inversor a cuantificar su visión del riesgo y a desarrollar una gestión de mismo más adecuada, y desarrollar modos en los que puede hacerse de manera dinámica la gestión del mismo adaptándose a las condiciones del mercado. Una dirección natural de trabajo futuro involucra el uso de gestión monetaria; es importante tener medios adecuados de manejo del dinero que permitan una pérdida mínima y que, junto con las diversas estrategias, maximizar los beneficios de una operación de trading.

La estructura de dependencia entre los múltiples factores de riesgo es crucial en el desempeño de la estrategia sobre la dinámica reversión a la media. Las reglas de comercio de los participantes del mercado pueden variar dependiendo de sus preferencias de riesgo, lo cual lleva a la pregunta interesante de como la aversión o propensión al riesgo puede generar variaciones en los tiempos y la distribuciones de las entradas al mercado. Así, podría ser útil incorporar dinámicas complementarias de comportamiento que complementen la toma de decisiones del mercado, profundizar y comprender un poco sobre los diversos factores que se ven involucrados en las fluctuaciones de precios, de manera que se pueda tener un modo más integral de participación en el mercado de divisas.

Como se ha mencionado antes, una serie de posibles continuaciones de este trabajo podrían ser la gestión de riesgo y de la inversión derivada de este, usando medios que permitan tanto flexibilidad como rigurosidad a la hora de calcular la inversión. La volatilidad es un aspecto clave a tener en cuenta; la misma ayuda a determinar de mejor manera rangos de comportamiento a futuro de la serie, que en complemento con técnicas de gestión monetaria permiten un modelo integral, adaptable a diversas condiciones del mercado. Un enfoque diferente respecto a la continuación de este trabajo podría ser en profundizar sobre el mejor modo de integrar modelos de predicción de series de tiempo con modelos ANFIS, de tal manera que se brinde un modelo integral que proporcione formas de predicción de las series como adaptabilidad a diversas condiciones.

# Bibliografía

- [Contreras V. et.al, 2018, ] Antonio V. Contreras, Antonio Llanes, Alberto Pérez-Bernabeu, Sergio Navarro, Horacio Pérez-Sánchez, Jose J. López-Espín José M. Cecilia, Y.(2018) ENMX: An elastic network model to predict the FOREX market evolution *Simulation Modelling Practice and Theory*, 86:1–10
- [Ata and Kocyigit, 2010] Ata, R. and Kocyigit, Y. (2010). An adaptive neuro-fuzzy inference system approach for prediction of tip speed ratio in wind turbines. *Expert Systems with Applications*, 37(7):5454–5460.
- [Bartle, 1995] Bartle, R. G. (1995). *The elements of integration and Lebesgue measure*. Wiley.
- [Brockwell and Davis, 1997] Brockwell, P. J. and Davis, R. A. (1997). *Springer Series in Statistics*.
- [Cavalcante et al., 2016] Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. P., Nobrega, J. P., and Oliveira, A. L. I. (2016). Computational Intelligence and Financial Markets: A Survey and Future Directions. *Expert Systems With Applications*, 55:194–211.
- [Cryer and Chan, 2008] Cryer, J. and Chan, K. (2008). *Time Series Analysis: With Applications in R*. Springer Texts in Statistics. Springer.
- [der Vaart, 2010] der Vaart, A. V. (2010). *Time Series*.
- [Jang and Sun, 1995] Jang, J. S. R. and Sun, C. T. (1995). Neuro-Fuzzy Modeling and Control. *Proceedings of the IEEE*, 83(3):378–406.
- [Naranjo et al., 2015] Naranjo, R., Meco, A., Arroyo, J., and Santos Peñas, M. (2015). An intelligent trading system with fuzzy rules and fuzzy capital management. *International Journal of Intelligent Systems*, 30.
- [Naranjo and Santos, 2019] Naranjo, R. and Santos, M. (2019). A fuzzy decision system for money investment in stock markets based on fuzzy candlesticks pattern recognition. *Expert Systems with Applications*, 133:34 – 48.
- [Negnevitsky and Potter, 2006] Negnevitsky, M. and Potter, C. W. (2006). Innovative short-term wind generation prediction techniques. *2006 IEEE PES Power Systems Conference and Exposition, PSCE 2006 - Proceedings*, pages 60–65.

- [Pradeepkumar and Ravi, 2018] Pradeepkumar, D. and Ravi, V. (2018). Soft computing hybrids for FOREX rate prediction: A comprehensive review. *Computers and Operations Research*, 99:262–284.
- [Russell et al., 1995] Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., and Edwards, D. D. (1995). *Artificial Intelligence A Modern Approach*. Technical report.
- [Shumway and Stoffer, 2010] Shumway, R. and Stoffer, D. (2010). *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Springer New York.
- [Zakamulin, 2017] Zakamulin, V. (2017). *Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules*. New Developments in Quantitative Trading and Investment. Springer International Publishing.