

**TRABAJO FINAL
LICENCIATURA EN
CIENCIAS DE LA COMPUTACIÓN**



**MODELOS OCULTOS DE MARKOV
APLICADOS AL
RECONOCIMIENTO DE PATRONES DEL
ANÁLISIS TÉCNICO BURSÁTIL**

CRISTIÁN F. FERNÁNDEZ

Director: Dr. OSCAR BUSTOS



**FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA
UNIVERSIDAD NACIONAL DE CÓRDOBA
CÓRDOBA, MARZO DE 2008**

Resumen

El análisis técnico bursátil desde hace más de un siglo se ha utilizado para intentar predecir el comportamiento futuro de los precios de un activo o índice bursátil. Dicho análisis se basa en la detección de patrones en series temporales financieras. Esta detección o *reconocimiento* ha sido, y es hasta nuestros días, realizada en forma *manual* y de manera *artesanal*. En el presente trabajo se introduce un novedoso sistema automático para el reconocimiento de patrones del análisis técnico bursátil, para el cual, los modelos ocultos de Markov constituyen una pieza fundamental.

Primero se desarrolla en detalle la teoría necesaria aplicada al caso específico, la cual abarca a las series temporales financieras, los modelos ocultos de Markov, el análisis técnico bursátil y los sistemas de reconocimiento de patrones. Como corolario, se aplica la teoría desarrollada a un caso de estudio. Los resultados de la aplicación demuestran la efectividad de la teoría.

Palabras clave: MOM, modelos ocultos de Markov, reconocimiento de patrones, análisis técnico bursátil, series temporales financieras, algoritmo forward, algoritmo backward, algoritmo de Viterbi, algoritmo de Baum-Welch.

Abstract

The technical analysis of stock market for more than a century has been used to try to predict the future price of an asset or stock index. This analysis is based on the detection of patterns in financial time series. This detection has been and is to this day, done by hand and in an artisanal manner. In this work introduces a new automatic system for patterns recognition of technical analysis of stock market, for which, the hidden Markov models are a key element.

First, developed in detail the necessary theory applied to the specific case, which includes financial time series, hidden Markov models, the technical analysis of stock market and pattern recognition systems. As a corollary, applies the theory developed to a case study. The results of the application demonstrate the effectiveness of the theory.

Keywords: HMM, hidden Markov models, pattern recognition, technical analysis of stock market, financial time series, forward algorithm, backward algorithm, Viterbi algorithm, Baum-Welch algorithm.

Agradecimientos

ÍNDICE GENERAL

Capítulo 1 INTRODUCCIÓN Y OBJETIVOS	1
1.1 Introducción	1
1.2 Motivación	3
1.3 Objetivos	4
1.4 Organización de la exposición	5
Capítulo 2 MODELOS DE SERIES TEMPORALES FINANCIERAS	7
2.1 Introducción	7
2.2 Serie Temporal y proceso estocástico	8
2.2.1 Análisis de series temporales	8
2.2.2 Serie temporal como realización de un proceso estocástico	10
2.2.3 Proceso estocástico estacionario	12
2.3 Características de las series temporales financieras	15
2.4 Modelos de series temporales financieras	18
2.4.1 Modelos estacionarios	20
2.4.2 Modelos no estacionarios	23
2.4.3 Modelos basados en el análisis técnico bursátil	27
Capítulo 3 MODELOS OCULTOS DE MARKOV (HMM)	28
3.1 Introducción	28
3.2 Definiciones	30
3.2.1 Proceso de Markov	30
3.2.2 Cadena de Markov	32
3.3 Modelos ocultos de Markov (HMM)	36
3.3.1 Historia de los HMM	38
3.3.2 Arquitectura de un HMM	39
3.3.3 Formalización de un HMM	40
3.3.4 Topología de un HMM	43
3.3.5 Problemas fundamentales de los HMM	44
3.3.5.1 Problema 1: Probabilidad de una observación	45
3.3.5.2 Problema 2: Secuencia de estados más probable	45
3.3.5.3 Problema 3: Estimación de los parámetros del modelo ...	46

Capítulo 4 SOLUCIÓN A LOS PROBLEMAS FUNDAMENTALES DE LOS HMM	47
4.1 Introducción	47
4.2 Solución al problema 1: Probabilidad de una observación	47
4.2.1 Algoritmos de avance – retroceso (forward – backward)	49
4.2.1.1 Procedimiento de avance	50
4.2.1.2 Procedimiento de retroceso	54
4.3 Solución al problema 2: Secuencia de estados más probable	56
4.3.1 El algoritmo de Viterbi	58
4.3.2 Implementación alternativa del algoritmo de Viterbi	63
4.4 Solución al problema 3: Estimación de los parámetros del modelo	65
4.4.1 Algoritmo de Baum-Welch	66
4.4.2 Limitaciones del algoritmo de Baum-Welch	70
4.5 Problemas de implementación de los HMM	71
4.5.1 Estimación inicial de los parámetros del HMM	71
4.5.2 Múltiples secuencias de observaciones	73
4.5.3 Uso de coeficientes de escalado	74
4.5.4 Insuficientes datos para el entrenamiento	75
4.5.5 Elección del modelo	76
Capítulo 5 ANÁLISIS TÉCNICO BURSÁTIL	78
5.1 Introducción	78
5.1.1 Análisis técnico vs. análisis fundamental	78
5.1.2 Historia del análisis técnico	81
5.2 Filosofía y fundamento	82
5.2.1 Hipótesis del análisis técnico	82
5.2.1.1 Los movimientos del mercado descuentan toda información	83
5.2.1.2 Los precios se mueven siguiendo tendencias	83
5.2.1.3 La historia se repite	84
5.2.2 Hipótesis del mercado eficiente	84
5.3 Gráficos	86
5.3.1 Gráficos lineales	86
5.3.2 Gráficos de barras	86
5.3.3 Escala simétrica vs. escala logarítmica	87
5.4 Tendencia	90

5.5	Volumen	92
5.6	Medias móviles	93
5.6.1	Simples	94
5.6.2	Ponderadas	95
5.6.3	Exponenciales	96
5.7	Patrones del análisis técnico	97
5.7.1	Patrones de cambio de tendencia	97
5.7.1.1	Cabeza y hombros	98
5.7.1.2	Cabeza y hombros invertidos	99
5.7.1.3	Doble cresta y doble valle	100
5.7.1.4	Triple cresta y triple valle	102
5.7.1.5	Fondos redondeados y crestas redondeadas	103
5.7.2	Patrones de continuación de tendencia	105
5.7.2.1	Triángulos	105
5.7.2.2	Banderas y banderines	108
5.7.2.3	Cuñas	109
5.7.3	Etapas del mercado	111
5.7.3.1	Etapa 1: la base	112
5.7.3.2	Etapa 2: alcista	113
5.7.3.3	Etapa 3: el techo	115
5.7.3.4	Etapa 4: bajista	116
Capítulo 6 RECONOCIMIENTO DE PATRONES		118
6.1	Introducción	118
6.2	Conceptos y definiciones	119
6.2.1	Clases	121
6.2.2	Patrones	122
6.2.3	Características	122
6.3	Sistemas de reconocimiento de patrones	125
6.3.1	Adquisición de datos	126
6.3.2	Extracción de características	127
6.3.2.1	Discretización de las características	128
6.3.2.2	Árboles de decisión	130
6.3.2.3	Reglas IF-THEN	132
6.3.2.4	Regresión lineal simple	135
6.3.3	Clasificación de patrones	139

6.3.3.1	Clasificación por mínima distancia	139
6.3.3.2	Enfoque estadístico	140
6.3.3.3	Clasificación usando lógica difusa	141
6.3.3.4	Clasificación estructural	141
6.3.3.5	Clasificación usando redes neuronales	142
6.4	Sistemas de reconocimiento de patrones basados en HMM	143
Capítulo 7	CASO DE ESTUDIO: IDENTIFICACIÓN DE LAS ETAPAS DEL	
	MERCADO	146
7.1	Introducción	146
7.2	Objetivos	147
7.3	Adquisición de datos	147
7.4	Extracción de características	149
7.4.1	Pendiente de la recta tangente a la <i>MMEP</i>	150
7.4.2	Precio de cierre con respecto a la <i>MMEP</i>	152
7.4.3	Volumen con respecto a la <i>MMEV</i>	153
7.4.4	Árbol de decisión	153
7.5	Clasificación / Reconocimiento	154
7.6	Iniciación de los parámetros del modelo	156
7.7	Resultados	159
7.8	Detalles de implementación	167
Capítulo 8	CONCLUSIONES Y FUTUROS DESARROLLOS	174
8.1	Conclusiones	174
8.2	Aportes de esta tesis	177
8.3	Líneas de continuación	178
APÉNDICE A	DEMOSTRACIONES DEL CAPÍTULO 4	179
A.1	Demostración de la proposición 4.2	179
A.2	Demostración de la proposición 4.3	180
A.3	Demostración de la proposición 4.4	182
A.4	Demostración de la proposición 4.6	185
A.5	Demostración de la proposición 4.7	188
APÉNDICE B	190
B.1	Archivo de ejemplo para el ingreso manual de un HMM	190
BIBLIOGRAFÍA	192

Capítulo 1

INTRODUCCIÓN Y OBJETIVOS

“Casi todo lo que realice será insignificante, pero es muy importante que lo haga”.

Mahatma Gandhi.

1.1 Introducción

En los mercados bursátiles, los agentes participantes necesitan desarrollar estrategias para acudir al mercado en el momento de negociar la compra o la venta de un activo. El problema de la toma de decisiones que enfrentan dichos agentes es muy complejo, pues son muchos los factores que influyen en el comportamiento de los precios de mercado.

Con la finalidad de intentar *predecir* dicho comportamiento se han desarrollado a lo largo del tiempo diferentes modelos de análisis. Tal vez el más antiguo de estos modelos, es el *análisis técnico bursátil*, cuyo basamento está en el reconocimiento de *patrones* en las series temporales de precios, denominadas series temporales financieras.

La hipótesis del análisis técnico bursátil en concordancia con [Mur00, AP97, Nis96, Wei98], es que si un determinado patrón determinó a posteriori, un comportamiento particular de los precios en el *pasado*, el hallazgo de dicho patrón en el presente, indica que debe esperarse el mismo comportamiento en la serie de precios en el futuro cercano. En realidad, la hipótesis indica que existe una alta *probabilidad* de que se repita el mismo comportamiento registrado en el pasado.

Además los patrones del análisis técnico bursátil, al igual que los patrones de la voz humana, o de imágenes digitalizadas entre otros, no siempre aparecen de igual manera. Por el contrario, se encuentran patrones *parecidos* a los patrones de

referencia, donde parecido es un concepto al que naturalmente se puede asociar una probabilidad (*existe una cierta probabilidad de que un patrón pertenezca a la “clase” de otro patrón de referencia, en función de su “parecido” con éste*).

Por lo tanto, un sistema de reconocimiento de patrones del análisis técnico bursátil debería contar con un *modelo estocástico* (probabilístico) en su *etapa de clasificación* (reconocimiento) de patrones.

Se pueden vislumbrar dos características importantes del análisis técnico bursátil a saber:

- Se apoya en la tesis de que todo lo que pueda afectar en el futuro los precios de un activo en particular, se refleja con anticipación en la misma serie de precios de mercado, y por ende sólo es necesario analizar las series temporales de precios para predecir futuros comportamientos.
- No persigue el objetivo de acertar en forma exacta los precios futuros de un determinado activo, sino determinar la *tendencia*¹ subyacente en los mismos, ya sea a corto, mediano o largo plazo, y/o detectar posibles cambios de tendencia con cierto grado de certeza.

Por lo tanto las series temporales de precios de mercado constituyen un insumo importante para esta tesis, y contar con una provisión de datos reales fidedignos que conformen las mismas también. A lo largo de este trabajo y en diferentes capítulos, se mostrarán ejemplos reales de series temporales financieras. Todas las series temporales financieras utilizadas en dichos ejemplos fueron series de *precios y volúmenes operados de acciones cotizantes en la Bolsa de Comercio de Buenos Aires*. La fuente de dichos datos fue precisamente la Bolsa de Comercio de Buenos Aires, a través de su portal en internet: *www.bolsar.com*.

Es importante recalcar, que las series temporales financieras serán objeto de análisis en este trabajo, pero a los efectos del desarrollo de la teoría no participan de manera directa, sino a través de las *características* extraídas de las mismas. Los conjuntos de valores de dichas características conformarán los patrones del análisis técnico bursátil que luego podrán ser clasificados de acuerdo a patrones de referencia.

¹ El concepto de tendencia en el contexto del análisis técnico bursátil se analizará en el capítulo 5.

1.2 Motivación

Los patrones del análisis técnico bursátil han sido identificados en series temporales financieras, mediante la utilización de gráficos (*charts* – en inglés –)². En un comienzo dichos gráficos se elaboraban en forma manual, con el advenimiento de la tecnología y los sistemas de cómputo se empezaron a elaborar en forma automática a partir de una serie temporal financiera dada. Sin embargo, la identificación de patrones continuó ejecutándose en forma manual (considerándose incluso un proceso *artesanal*).

Ahora bien, si se establecen claramente los parámetros para la identificación de un patrón en particular y se utiliza el sistema adecuado para procesar la serie temporal subyacente, no existe razón para que el proceso de detección del mismo deba ser manual.

Por lo tanto *la primera motivación de este trabajo surge de comprobar que el tema de la detección automática de patrones del análisis técnico bursátil, es un tema que posee un tratamiento casi nulo.*

Luego de una vasta investigación, sólo pudo encontrarse un trabajo relacionado al tema, que le corresponde a John Murphy, autor de [Mur00] y que es un accesorio (*plug-in*) para el programa comercial MetaStock® de Equis International y está desarrollado en el lenguaje de fórmulas propio de MetaStock. Con respecto a este trabajo es importante señalar que no se trata de una solución estocástica con las limitaciones que esto implica, que está basada en la teoría del análisis técnico bursátil, pero no incluye a la teoría necesaria sobre reconocimiento de patrones y por último, al ser un producto comercial, no se incluye una descripción detallada de su desarrollo.

Por otro lado, los *modelos ocultos de Markov*³ son modelos estocásticos que han sido aplicados con éxito en numerosas áreas, como se verá en los capítulos correspondientes de esta tesis, pero a pesar de que fueron utilizados con y sin modificaciones para el análisis de series temporales financieras y se aplicaron con importante éxito al reconocimiento de patrones en diversos ámbitos, como patrones de voz humana, patrones en imágenes digitales, etc., no se encontraron trabajos en

² Los gráficos y sus implicaciones en el ámbito del análisis técnico bursátil se estudiarán en el capítulo 5.

³ Los modelos ocultos de Markov se verán en detalle en los capítulos 3 y 4 de este trabajo.

donde se apliquen los modelos ocultos de Markov al reconocimiento de patrones del análisis técnico bursátil.

Por lo tanto, *la segunda motivación de esta tesis surge de comprobar que los modelos ocultos de Markov, a pesar de constituirse actualmente en herramientas clave para muchos sistemas de reconocimiento de patrones, no han sido utilizados con ese fin para los patrones del análisis técnico bursátil.*

1.3 Objetivos

Los objetivos primarios perseguidos por esta tesis son los siguientes:

- *Desarrollar la teoría necesaria que sirva como base para elaborar sistemas de reconocimiento de patrones del análisis técnico bursátil.*
- *Aplicar la teoría desarrollada a un caso de estudio concreto y lograr demostrar de esta forma su efectividad.*

Además de los objetivos primarios antes descritos y aunque este trabajo se centre en el estudio de los patrones del análisis técnico bursátil, un objetivo secundario de esta tesis, pero no de menor importancia, es que la teoría desarrollada tenga cierto carácter general, de manera que sirva de base para su aplicación en el reconocimiento de patrones en otros ámbitos. Ya que la utilización de los modelos ocultos de Markov en el desarrollo abre un abanico de posibilidades en ese sentido, pues han demostrado su efectividad en los más variados ámbitos.

1.4 Organización de la exposición

Este trabajo posee la siguiente organización:

- En el capítulo 2 se efectúa una introducción a las series temporales, con especial énfasis en las series temporales financieras. Se realiza un repaso de los distintos modelos existentes para el análisis de series temporales financieras y se propone una taxonomía novedosa para los mismos, en la cual se podrá apreciar el lugar que ocupa este trabajo.
- En el capítulo 3 se introducen definiciones básicas que serán usadas en el resto de la tesis y se efectúa la presentación de los modelos ocultos de Markov que serán usados en el sistema de reconocimiento de patrones.
- En el capítulo 4 se analizan en detalle los problemas fundamentales de los modelos ocultos de Markov y los algoritmos utilizados para su resolución. También se muestran problemas de implementación de los mismos.
- En el capítulo 5 se introduce la teoría del análisis técnico bursátil, ofreciendo una completa introducción a temas propios del mismo, como los gráficos, la tendencia, el volumen y las medias móviles, todos necesarios para el entendimiento de los patrones del análisis técnico bursátil que se muestran como conclusión del capítulo.
- En el capítulo 6 se presentan los lineamientos fundamentales de los sistemas de reconocimiento de patrones, mostrando cada etapa de los mismos, con énfasis en los temas aplicables a este trabajo. Finalmente se muestra la metodología propuesta, fundamentada en los sistemas de reconocimiento de patrones basados en modelos ocultos de Markov.
- Como corolario, en el capítulo 7 se aplica la teoría desarrollada en los capítulos anteriores al caso de estudio: *Identificación del patrón de las etapas del mercado*. Luego se utiliza el sistema de reconocimiento de la forma del patrón de las etapas del mercado con series temporales financieras obtenidas

del mundo real, mostrando los resultados obtenidos. Finalmente se muestra la implementación del sistema desarrollado.

- En el capítulo 8 se recogen las conclusiones más relevantes de este trabajo, así como también los aportes realizados por el mismo. Se especifican además las líneas de investigación que se abren a partir de esta tesis y/o futuros desarrollos derivados.

En realidad los capítulos 2 al 6, están diseñados con una doble finalidad, la primera finalidad es contribuir al desarrollo de la teoría que permita la elaboración de un sistema de reconocimiento de patrones del análisis técnico bursátil (que constituye uno de los objetivos primarios de esta tesis); la segunda finalidad es ofrecer una guía del tema tratado, con el suficiente detalle para contribuir al objetivo secundario, que es dotar a la teoría de cierto carácter general para su posible aplicación a la detección de patrones en otros ámbitos. Es importante destacar que en el caso del capítulo 5, se tuvo la intención de que el mismo sirviera como una guía resumida para este tipo de análisis, por lo cual, dicho capítulo puede estudiarse antes de cualquier otro, o bien en el orden que se eligió para él, lo importante es que sea leído antes del capítulo 7.

Capítulo 2

MODELOS DE SERIES TEMPORALES FINANCIERAS

*“Un camino de mil millas comienza
con un paso”.*

Benjamin Franklin.

2.1 Introducción

Si bien en este trabajo, el estudio estará centrado en el reconocimiento de patrones del análisis técnico bursátil, la serie temporal *subyacente* debe ser considerada, pues será desde donde se extraigan los valores de las *características*⁴, que luego conformarán los patrones a reconocer, y las peculiaridades de dicha serie en mayor o menor medida, directa o indirectamente, se trasladan a los valores de dichas características. Por otro lado, el capítulo servirá para mostrar las técnicas utilizadas en el ámbito de la predicción en series temporales financieras.

En la actualidad, el análisis de series de datos ordenados en el tiempo, es imprescindible en numerosos campos como la medicina, la ingeniería, la economía y muchos más. Investigar cómo ha evolucionado una variable hasta el presente puede ser de gran utilidad en el momento de predecir su comportamiento futuro. De manera similar, establecer qué otras series de datos tuvieron un comportamiento análogo puede ayudar en la toma de decisión concerniente a los pasos a seguir, ya sea para conservar la dirección actual o para modificarla substancialmente.

En este capítulo, primero se muestran algunas definiciones y conceptos básicos que serán de utilidad en el desarrollo de esta tesis. A continuación se especifican las propiedades estadísticas que son propias de las series temporales financieras.

⁴ El tema de extracción de características se tratará en detalle en el capítulo 6 de este trabajo, pues comprende una de las etapas del reconocimiento de patrones.

Además se propone una clasificación en cuanto al modelado de dichas series, pudiéndose apreciar el lugar que ocupa este trabajo en la taxonomía propuesta, en ese contexto se da una breve descripción de los modelos utilizados para analizar series temporales financieras de acuerdo a las referencias consultadas.

2.2 Serie Temporal y proceso estocástico

En esta sección, además de las definiciones de serie temporal, proceso estocástico y proceso estocástico estacionario, se mostrarán las formas de análisis de las series temporales, dentro de lo cual se hará especial hincapié en las series temporales como realización de un proceso estocástico dado.

2.2.1 Análisis de series temporales

Una serie temporal es una *sucesión de observaciones* que se distribuyen de acuerdo al tiempo de su resultado. El objetivo fundamental del análisis de series temporales es desarrollar *modelos* matemáticos que provean fieles descripciones de los datos [SS06].

El análisis de series temporales es una herramienta especialmente adecuada cuando se registran las siguientes circunstancias:

- Se dispone de una buena cantidad de *información muestral*.
- No existe la posibilidad de conseguir una precisa *información teórica* (es decir, no es posible realizar un *análisis fundamental*).
- El objetivo primordial es *predecir*.

Está claro que una cantidad importante de datos muestrales es necesaria para que el análisis sea representativo en la población general a la que pertenece la serie. Por otro lado, si se conocen suficientes datos teóricos que fundamenten firmemente las

causas del comportamiento de una serie, el análisis de las mismas se torna accesorio, pero no indispensable.

En general, mediante el análisis de series temporales se persigue la posibilidad de efectuar una predicción, donde por predicción se entiende la estimación de los valores futuros de una serie, en función del comportamiento pasado de la misma.

Si a pesar de que se conocen los valores anteriores de una serie, no se puede predecir con precisión el valor futuro de la variable, se dice que la serie es no determinística o aleatoria. El análisis de series temporales se centra en el estudio de dichas series. Por supuesto, como podrá inferirse, las series temporales financieras, objeto de este trabajo en forma indirecta, poseen un comportamiento no determinista.

Los métodos tradicionales para el análisis de series temporales [Cha96] implican la descomposición de las mismas en varias partes, como tendencia, variaciones estacionales y otras irregularidades.

- La *tendencia*⁵ es la dirección de la variable en el periodo de observación, es decir, el cambio de la media de la serie en el periodo respectivo.
- La *estacionalidad* se corresponde con las periódicas fluctuaciones de la variable, en determinados periodos de tiempo.
- Luego de extraer de la serie las variaciones estacionales y la tendencia, quedarán unas *fluctuaciones irregulares* generadas por valores residuales, que pueden ser o no aleatorios.

Cuando una serie contiene tendencias, como es el caso de las series temporales financieras⁶, se puede analizar dicha tendencia a través de diferentes métodos. La tendencia más simple es la *tendencia lineal con ruido*, que se puede analizar utilizando técnicas de regresión [Cha96, Dev02].

Siendo β_0 y β_1 los parámetros a estimar y u_t la componente aleatoria en el tiempo t (el ruido), (que en la mayoría de los casos se supone una variable independiente e idénticamente distribuida con distribución normal de media 0 y varianza σ^2), la

⁵ La tendencia, en el ámbito específico de las series temporales financieras y en particular del análisis técnico bursátil, es de gran importancia para este trabajo y se trata en detalle en el capítulo 5, sección 5.4

⁶ Según la hipótesis del mercado eficiente (Eugene Fama, 1965) esto no es cierto, pero en la actualidad la opinión generalizada es contraria a la tesis de Fama (ver capítulo 5, sección 5.2.2)

tendencia lineal con ruido puede modelarse con el modelo lineal o de regresión lineal simple⁷, donde:

$$y_t = \beta_0 + \beta_1 t + u_t \quad (2.1)$$

El *nivel medio* en el tiempo t está representado por $m_t = \beta_0 + \beta_1 t$. Los cambios de tendencia estarán representados por los *cambios* en el nivel medio por unidad de tiempo, aunque algunos autores utilizan la pendiente β_1 como la tendencia y por ende los cambios de tendencia se referirán a los cambios en dicho parámetro.

El lector que desee ampliar conocimientos sobre el tema encontrará una excelente guía de análisis de series temporales con métodos lineales en [Cha96, Ham94], y con métodos no lineales en [KS04].

2.2.2 Serie temporal como realización de un proceso estocástico

“Un proceso estocástico es la abstracción matemática de un proceso empírico, cuyo desarrollo está gobernado por las leyes de la probabilidad” [Doo90]. Desde un punto de vista más matemático un proceso estocástico es una colección de variables aleatorias $\{z_t, t \in T\}$. El proceso se dirá de parámetro discreto o simplemente discreto si T es un conjunto discreto y se dirá de parámetro continuo o sólo continuo si T es un intervalo. Nótese que T es un conjunto no necesariamente limitado al tiempo, pero en la práctica z_t se corresponderá con la observación en el tiempo t , y T es el rango de tiempo involucrado. Además, en la mayoría de los problemas, sólo un conjunto finito de variables aleatorias serán consideradas, es decir, T será un conjunto finito.

Para ejemplificarlo, supóngase la existencia de un conjunto de T observaciones ordenadas en el tiempo $\{p_1, \dots, p_T\}$ donde $p_t \in \mathbb{R}^r \forall t = 1, \dots, T$. Supóngase además que el valor observado p_t de la serie en un momento de tiempo t se corresponde con el

⁷ En el capítulo 6, sección 6.3.2 se verá en detalle el método de la regresión lineal simple en el contexto de extracción de características de un sistema de reconocimiento de patrones.

valor de una variable aleatoria z_t definida en dicho momento. Se denomina proceso estocástico [Nel95, Ros96] al conjunto definido de variables aleatorias $\{z_t\}, t=1, \dots, T$ ordenadas en el tiempo. La serie observada $\{p_1, \dots, p_T\}$ es por consiguiente una muestra de un vector de T variables aleatorias $\{z_1, z_2, \dots, z_T\}$ a la que se denomina trayectoria o realización del proceso estocástico.

Las variables aleatorias $\{z_1, z_2, \dots, z_T\}$ correspondientes al proceso estocástico, tienen asociadas las siguientes funciones de densidad:

$$\left. \begin{array}{l} z_1 \sim f_1 : \mathbb{R}^r \rightarrow \mathbb{R} \\ \dots \\ z_t \sim f_t : \mathbb{R}^r \rightarrow \mathbb{R} \\ \dots \\ z_T \sim f_T : \mathbb{R}^r \rightarrow \mathbb{R} \end{array} \right\} \quad (2.2)$$

En la Figura 2.1 se representa la noción de proceso estocástico y al mismo tiempo se pueden apreciar varias realizaciones del proceso. En la misma puede verse cómo en cada lapso temporal, las observaciones son generadas por la función de densidad de la variable correspondiente.

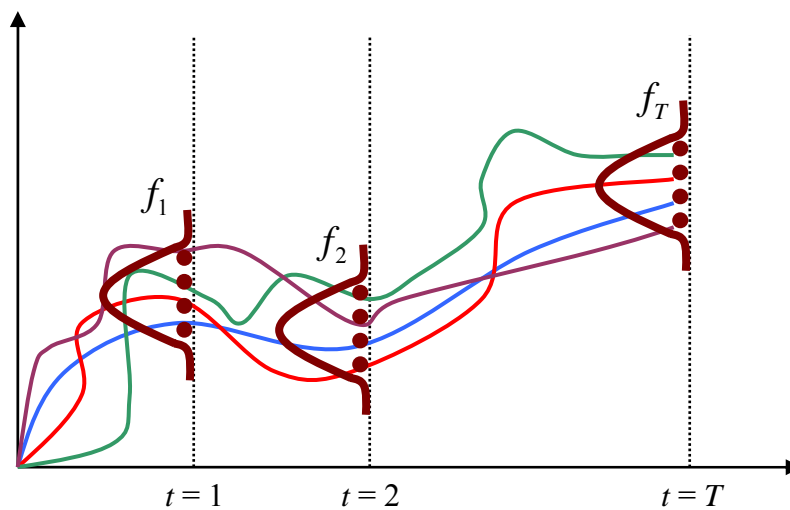


Figura 2.1: Proceso estocástico.

Dado un proceso estocástico $\{z_t\}$, una serie temporal $\{p_1, p_2, \dots, p_T\}$ como se desprende de lo expuesto anteriormente, no es otra cosa que una realización de dicho proceso, el que puede definirse tanto en un espacio temporal discreto como en un espacio temporal continuo, con respecto al valor del parámetro t .

Es conveniente dejar en claro que a diferencia del ejemplo correspondiente a la figura 2.1, en su mayoría, los procesos que se encuentran en el mundo real cuentan solamente con una realización del proceso estocástico.

2.2.3 Proceso estocástico estacionario

Un proceso estocástico $z = \{z_t\}_{t=1}^T$ será catalogado como *estrictamente estacionario* [SS06] si las funciones de distribución multidimensionales (de dimensión n) son idénticas para cada par de conjuntos de variables cualesquiera del tipo $[z_{t_1}, z_{t_2}, \dots, z_{t_n}]$ y $[z_{t_1+l}, z_{t_2+l}, \dots, z_{t_n+l}]$ con $n = 1, 2, \dots, T$. Formulando de otra manera:

$$P[z_{t_1} \leq x_1; z_{t_2} \leq x_2; \dots; z_{t_n} \leq x_n] = P[z_{t_1+l} \leq x_1; z_{t_2+l} \leq x_2; \dots; z_{t_n+l} \leq x_n] \quad (2.3)$$

o en términos de la función de distribución acumulada:

$$F[x_1; x_2; \dots; x_n; t_1 \dots t_n] = F[x_1; x_2; \dots; x_n; t_1 + l \dots t_n + l] \quad (2.4)$$

para todo t y para todo l .

La definición anterior de proceso estacionario es excesivamente estricta y resulta muy difícil encontrar ejemplos de procesos reales, ya sea en la naturaleza o como producto de la acción humana, que confirmen la condición impuesta por la ecuación (2.4). Dicha condición implica en particular que las distribuciones unidimensionales deben ser iguales para todo t . Por esta razón, además de los procesos estacionarios expuestos, se considera también un tipo de procesos débilmente estacionarios definidos de la siguiente manera:

Un proceso estocástico $z = \{z_t\}_{t=1}^T$ es débilmente estacionario [SS06], cuando para todo valor de t se tiene que:

- $E[z(t)]$ es constante y finito.
- $E[z(t+l) \cdot z(t)]$ existe y solamente depende de l .

Si un proceso estocástico es estrictamente estacionario entonces también es débilmente estacionario.

En la Figura 2.2 se puede apreciar un ejemplo de serie temporal estacionaria. Aunque para tener la certeza se deberían utilizar procedimientos para garanticen que realmente la serie es estacionaria, intuitivamente se puede verificar que no existen cambios sistemáticos ni en la media de la serie, ni tampoco en la varianza de la serie, lo cual de acuerdo a [Cha96] permite afirmar que la serie es estacionaria, al menos en sentido débil.

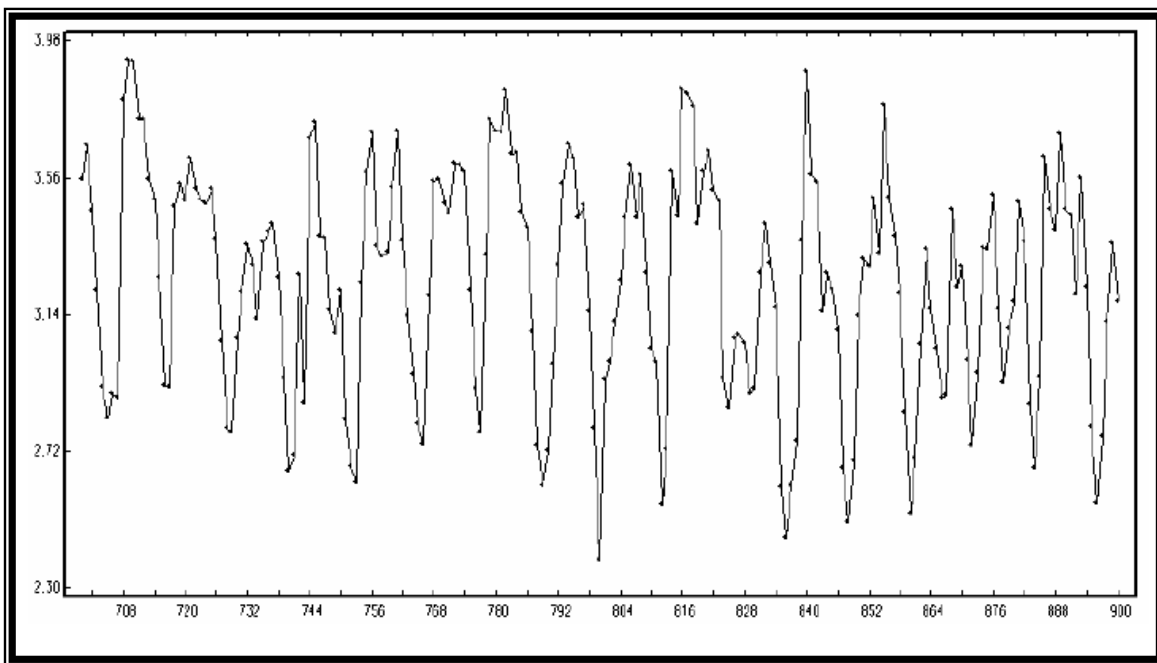


Figura 2.2: Proceso estocástico débilmente estacionario.

En la Figura 2.3 se puede ver un ejemplo claro de serie no estacionaria en donde se aprecia que la media de la serie tiene una variación significativa. Se desprende de la figura que la media aumenta con el tiempo. Este es un ejemplo de una serie en donde se verifica una *tendencia creciente*.

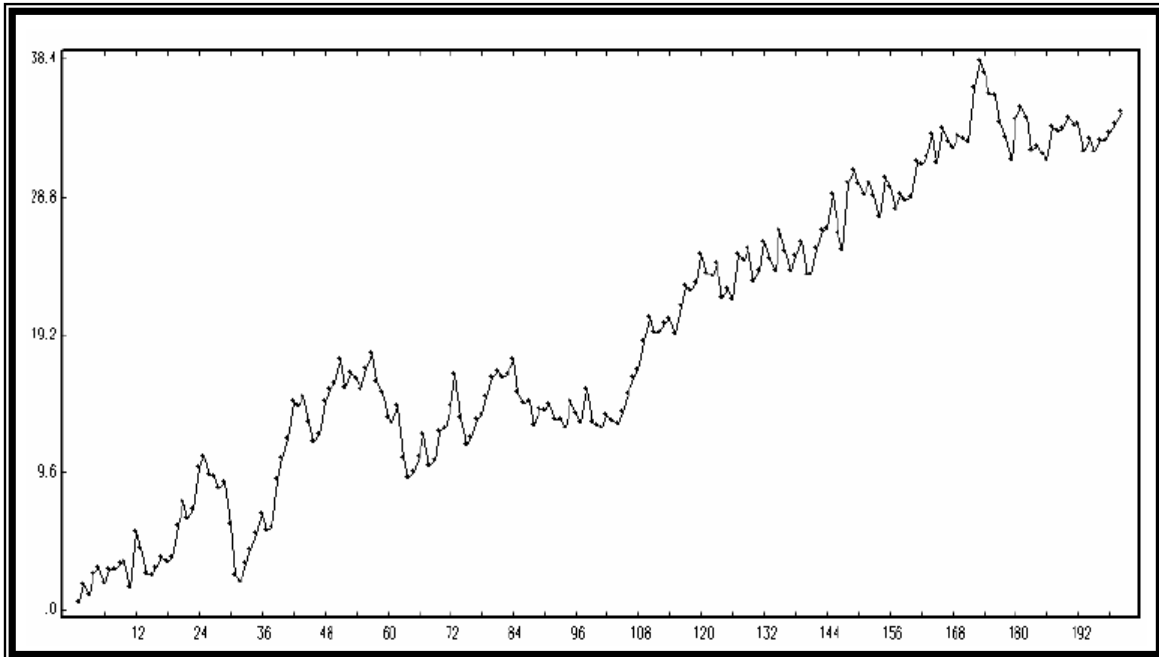


Figura 2.3: Proceso estocástico no estacionario (con variación de la media).

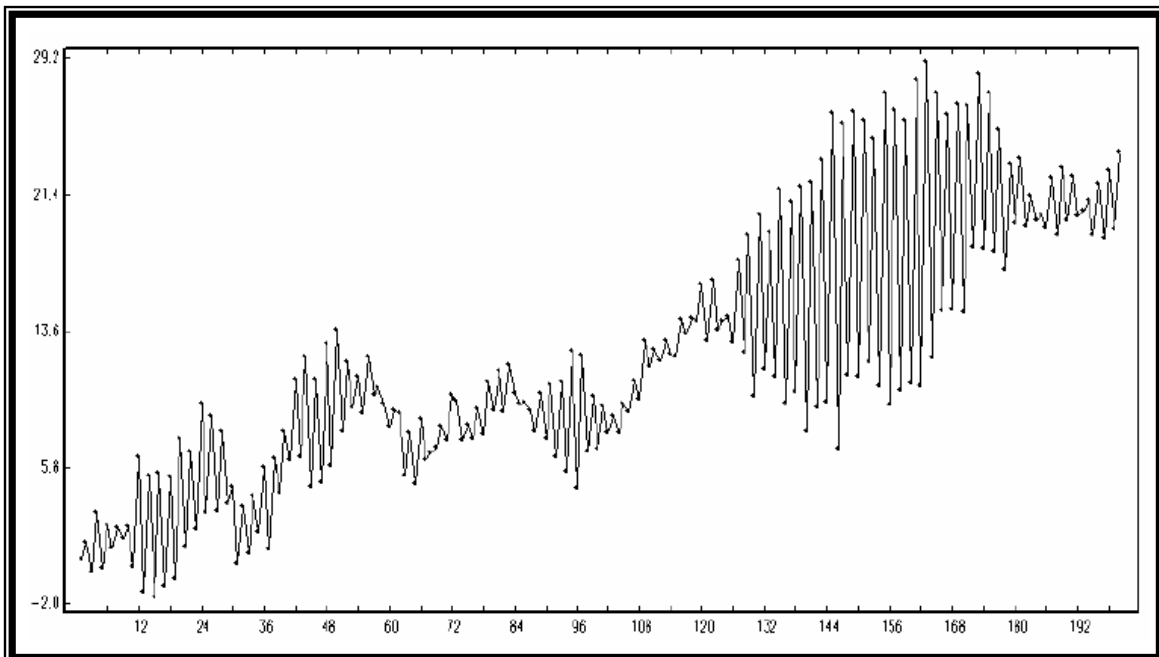


Figura 2.4: Proceso estocástico no estacionario (con variación de la media y de la varianza).

Finalmente en la Figura 2.4 se exhibe otro ejemplo de serie temporal no estacionaria. En la misma es posible observar que existe, al igual que en el ejemplo anterior, una tendencia creciente, puesto que la media de la serie se incrementa con el tiempo. Pero en este ejemplo se verifica también que la varianza de la serie fluctúa con el tiempo.

2.3 Características de las series temporales financieras

Las series temporales financieras son secuencias de precios de activos o índices financieros en un determinado periodo de tiempo, por ende “el análisis de series temporales financieras es la disciplina que concierne a la teoría y práctica sobre la valoración de activos a través del tiempo” [Tsa02]. Es una disciplina altamente *empírica*, pero al igual que otros campos de la teoría científica es la base para hacer *inferencias*. Existe un rasgo característico que distingue al análisis de series temporales financieras de otros análisis de series temporales. Tanto la teoría financiera como las series temporales empíricas contienen un elemento de *incertidumbre*. Como ejemplo, en la figura 2.5 se puede observar el gráfico de un activo financiero (la acción de una empresa en este caso) con su respectiva serie de retornos. Donde el retorno del periodo t (R_t) se define en función del precio del activo subyacente del periodo t (P_t) y del periodo $t - 1$ (P_{t-1}) como sigue:

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (2.5)$$

(al resultado se lo puede multiplicar por cien para tenerlo en términos de porcentaje), hay varias definiciones de la volatilidad de los activos, y en una serie de valores de retorno como puede verse en la figura 2.5, la volatilidad no es directamente observable [Tsa02].

Como resultado de la incertidumbre implícita, la teoría y los métodos estadísticos juegan un papel importante en el análisis de series temporales financieras.

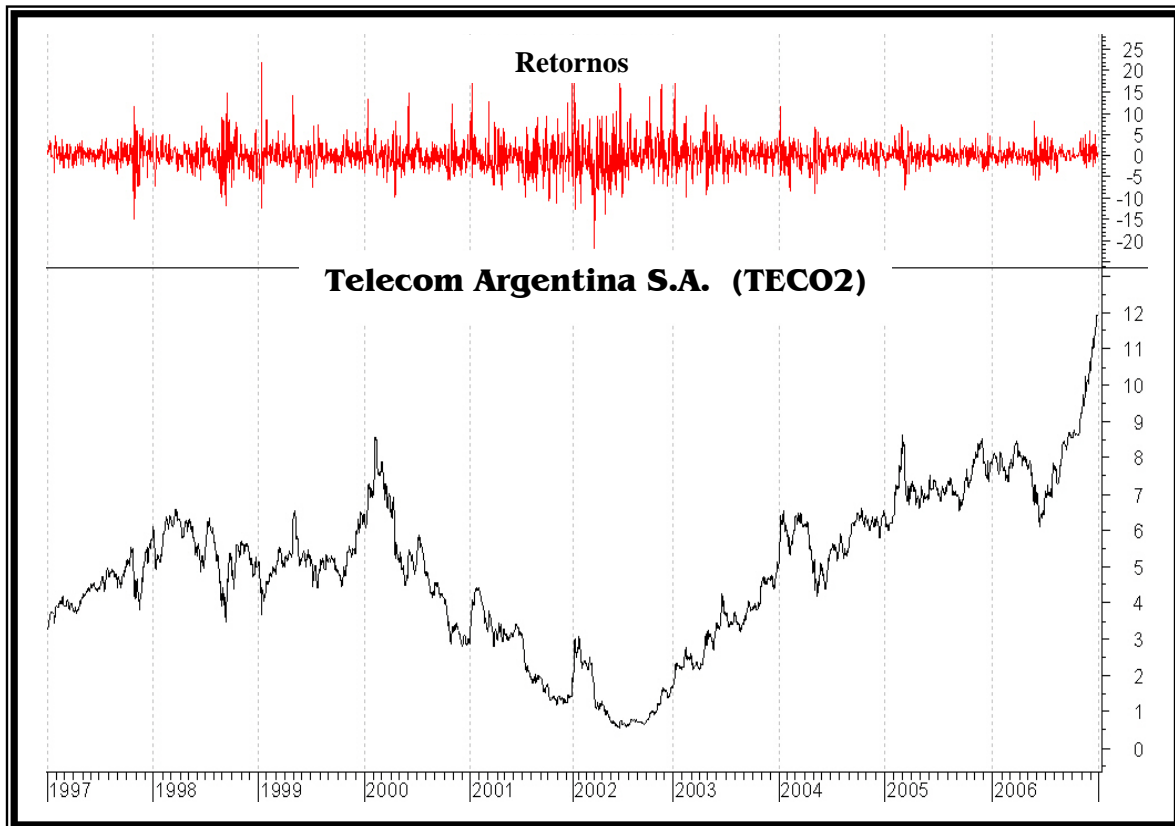


Figura 2.5: Gráfico del precio de cierre de Telecom Argentina S.A. desde enero de 1997 hasta diciembre de 2006 con retornos diarios.

Es usual en el análisis de series temporales financieras, que se analicen los retornos y no los precios del activo subyacente. En este trabajo no se analizarán los retornos más allá de su uso en ejemplos, por la sencilla razón de que los patrones del análisis técnico han sido desarrollados en base a series de precios y volúmenes de operación (como se verá en el capítulo 5) y no sobre series de retornos.

Existe consenso entre los autores consultados [Tsa02, FD03, WS98, ME07] de que las series temporales financieras presentan comportamientos no estacionarios y altamente volátiles.

En la figura 2.6 se puede ver un ejemplo de dichos comportamientos. Como se puede observar en una simple inspección del gráfico, la serie presenta variabilidad tanto en la media como en la varianza. Un método muy útil que permitirá identificar la componente no estacionaria de la serie consiste en agrupar observaciones consecutivas y hacer el cálculo de la media, varianza y desvío estándar sobre cada grupo por separado. Finalmente los datos obtenidos se representan en un gráfico que se denomina gráfico de media y desvío estándar.

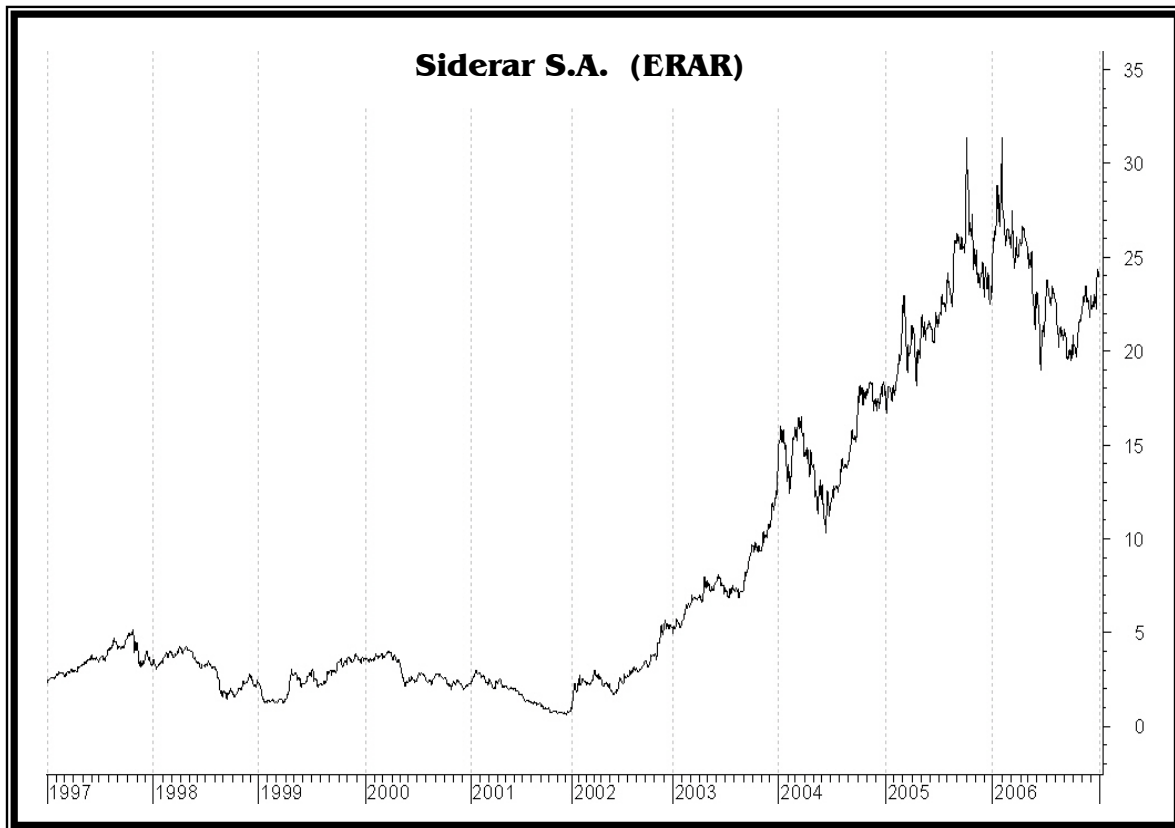


Figura 2.6: Gráfico del precio de cierre de Siderar S.A. desde enero de 1997 hasta diciembre de 2006.

La representación en el gráfico de media y desvío estándar mostrará la estacionariedad o no de la media y la varianza, y además hará relucir la relación existente entre ambas.

En la figura 2.7 se puede observar el gráfico de media y desviación estándar para el ejemplo de la figura 2.6. Como los diez años de observaciones conformaban un total de 2.462 datos diarios de cierre del valor de la acción, se utilizaron 98 intervalos de 25 días cada uno (descartando los últimos 12 datos), sobre estos intervalos se calculó la media, la varianza y la desviación estándar y luego se representó la relación entre la media (eje de las abscisas) y la desviación estándar (eje de las ordenadas). La información revelada por el gráfico es concluyente en cuanto a la no estacionalidad de la media y también de la varianza, puesto que en un proceso estacionario estricto el gráfico debería estar representado por un único punto, y como se puede observar, está muy lejos de eso. Además es clara la alta volatilidad puesto que se registran para valores similares de media, valores de desvíos estándar altamente diferentes en

magnitud (por ejemplo, para valores de media de 25, se registran desvíos de 0,30 aproximadamente y también de más de 1,80). Por último, es también visible una cierta relación lineal entre la media y la varianza.

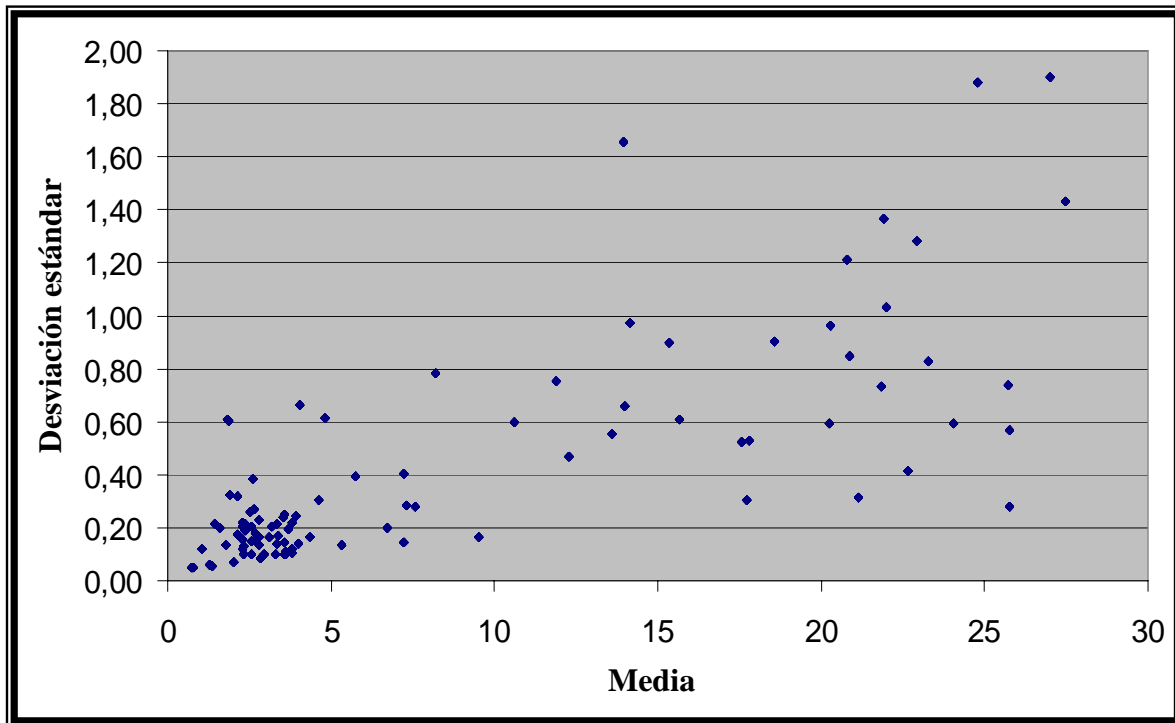


Figura 2.7: Gráfico de media y desviación estándar para los datos de la figura 2.6 utilizando intervalos de 25 días.

2.4 Modelos de series temporales financieras.

En esta sección se presenta una revisión de los diferentes enfoques utilizados para el análisis de series temporales financieras. Como resultado de dicha revisión, en la figura 2.8 se presenta una clasificación de los modelos usados, con la aclaración que se clasifican sólo los más relevantes encontrados, pudiendo existir más modelos con dicho fin. La taxonomía propuesta permite además que se pueda apreciar el lugar que ocupa este trabajo en dicha clasificación.

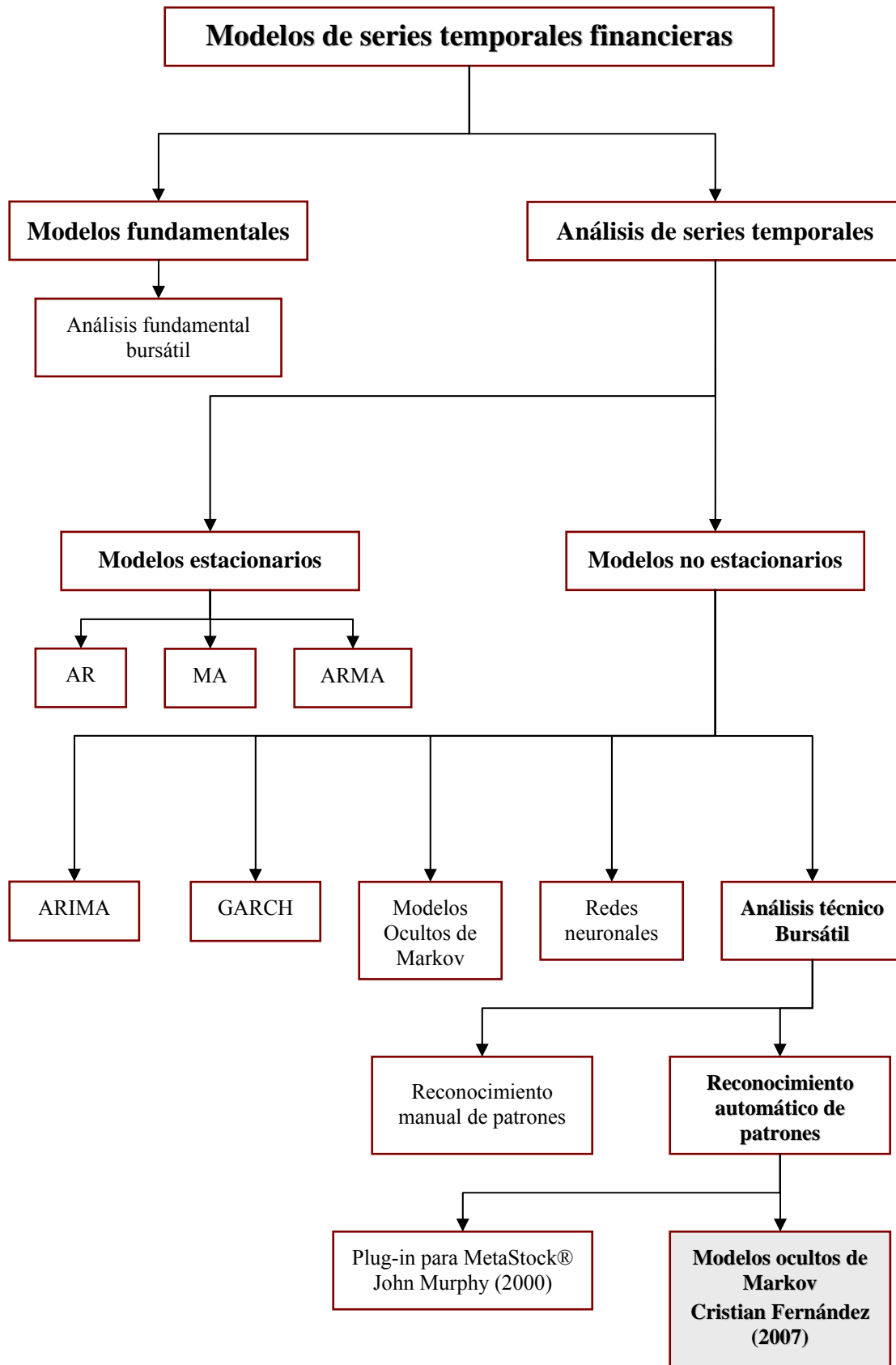


Figura 2.8: Clasificación de los modelos de series temporales financieras.

En la clasificación anterior se puede ver una división inicial entre modelos fundamentales y modelos basados en el análisis de series temporales. Con respecto a los primeros se puede decir que analizan la *causa* de los movimientos registrados en las series temporales y no el *efecto*, es decir, se interesan por los motivos fundamentales que conllevaron a dicho movimiento. Este trabajo se ubica en el enfoque técnico, como se desprende de la taxonomía, por lo que no se darán mayores precisiones sobre el modelo fundamental, de todos modos en el capítulo 5 se provee una breve descripción del análisis fundamental bursátil en pos de su comparación con al análisis técnico bursátil, en dicho capítulo también se citan las referencias pertinentes para el lector que desee profundizar sobre este tipo de análisis.

Con respecto al análisis basado en series temporales, se pueden dividir en dos grandes grupos, los modelos basados en series estacionarias y los basados en series no estacionarias.

2.4.1 Modelos estacionarios

A pesar de que, como se dijo anteriormente, las series temporales financieras poseen una componente claramente no estacionaria, existen varios modelos puramente estacionarios que aprovechan que la mayoría de la teoría de la probabilidad referente a series temporales concierne a series estacionarias,

Ahora bien, para poder aplicar estos métodos se deben aplicar *transformaciones o filtros* a la serie original para *convertirla* en estacionaria, es decir, se deben remover tendencias, variaciones estacionarias, etc. Algunos lineamientos para estas transformaciones [Cha96] son:

- *estabilizar la varianza*, si se observa que la desviación estándar es directamente proporcional a la media entonces una transformación logarítmica es adecuada. (observar que este es el caso del ejemplo de la figura 2.6)
- *convertir el efecto estacionario en aditivo*, si se observa que dicho efecto es directamente proporcional a la media, entonces el efecto es multiplicativo y una transformación logarítmica será adecuada para convertirlo en aditivo.

- *hacer que los datos estén normalmente distribuidos*, en los modelos usualmente se asume que los datos están normalmente distribuidos. Si es necesario, pueden utilizarse con dicho fin transformaciones llamadas de Box-Cox. Dada una serie temporal observada $\{x_t\}$ y un parámetro de transformación λ , la serie transformada está dada por:

$$y_t = \begin{cases} (x_t^\lambda - 1) / \lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases} \quad (2.6)$$

También pueden ser usados *filtros lineales* para convertir una serie $\{x_t\}$ en otra $\{y_t\}$ a través de una operación lineal, un ejemplo de filtro es el de *media móvil* [Cha96]. Además, como muestra la figura 2.9, estos filtros pueden ser aplicados en cadena.

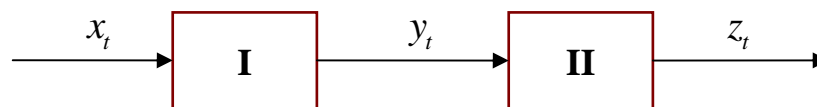


Figura 2.9: Filtros en cadena.

Una excelente guía sobre éstas y otras transformaciones y/o filtros puede ser encontrada en [Cha96].

A continuación se citan los modelos estacionarios relevantes que se utilizan en el modelado de series temporales financieras, es importante aclarar que se aplican sobre series temporales de retornos (el retorno del periodo t es r_t).

- *Modelos autoregresivos* (AR) (autoregressive models): El más simple es el AR(1) que se define a través de la siguiente ecuación:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \quad (2.7)$$

donde $\{a_t\}$ son variables independientes e idénticamente distribuidas con media 0 y varianza σ^2 (ruido blanco – *white noise*), ϕ_0 y ϕ_1 son los parámetros (a estimar) del modelo. La generalización de este modelo AR(p) está dada por la siguiente ecuación:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + a_t \quad (2.8)$$

donde p es un entero mayor o igual a 1 y ϕ_i $i = 0, 1, \dots, p$ son los parámetros del modelo.

- *Modelos de media móvil (MA)* (moving-average models): El más sencillo es el MA(1) que está definido por la siguiente ecuación:

$$r_t = c_0 + a_t - \theta_1 a_{t-1} \quad (2.9)$$

donde c_0 es una constante, θ_1 es el parámetro del modelo y $\{a_t\}$ es una serie de ruido blanco al igual que en el anterior. La generalización del mismo MA(q), está dada por la siguiente igualdad:

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.10)$$

donde q es un entero mayor o igual a 1 y θ_i $i = 0, 1, \dots, q$ son los parámetros del modelo.

- *Modelos autoregresivos de media móvil (ARMA)* (autoregressive moving-average models): Es básicamente una combinación de las ideas de los dos métodos anteriores, logrando una forma compacta que reduce parámetros. El más común es el ARMA(1,1), una serie temporal r_t sigue un modelo ARMA(1,1) si satisface la siguiente ecuación:

$$r_t - \phi_1 r_{t-1} = \phi_0 + a_t - \theta_1 a_{t-1} \quad (2.11)$$

donde la parte izquierda de la ecuación 2.11 se corresponde con la componente AR y la parte derecha con la componente MA. El término constante es en este caso ϕ_0 . Este modelo requiere que $\phi_1 \neq \theta_1$, de otra forma la ecuación se cancelaría y el proceso se reduciría a una serie de ruido blanco. La generalización de este modelo ARMA(p, q), se puede visualizar en la siguiente ecuación:

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{i=1}^q \theta_i a_{t-i} \quad (2.12)$$

donde p y q son enteros mayores a 1.

Una descripción detallada de estos modelos en el ámbito financiero puede ser encontrada en [Tsa02] y en series temporales en general en [Cha96, Ham94]. Es importante recalcar que estos modelos tienen la ventaja de que las herramientas para manejarlos son accesibles, y debido a que como se dijo, la gran mayoría de la teoría de la probabilidad se aplica a series estacionarias, estas herramientas están muy extendidas. Sin embargo, su uso en series extensas es dificultoso, pues se torna complicado convertirlas a estacionarias, a lo sumo debería hacérselo en tramos, con la pérdida de información que puede conllevar.

2.4.2 Modelos no estacionarios

Estos modelos permiten trabajar con series temporales financieras no estacionarias. De acuerdo a la clasificación propuesta, a continuación se verán los más relevantes.

- *Modelos autoregresivos integrados con media móvil (ARIMA)* (autoregressive integrated moving-average models): Se basa en los modelos ARMA ya vistos. La ecuación 2.12 si se tiene que $Br_t = r_{t-1}$, se convierte en:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) r_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) a_t \quad (2.13)$$

Un modelo ARIMA (p, d, q) se obtiene mediante el proceso de *integración* de un modelo ARMA (p, q), de acuerdo a la siguiente ecuación:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) (1 - B)^d r_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) a_t \quad (2.14)$$

donde d es un entero positivo. (Si $d = 0$, el modelo se convierte en ARMA, pues no hay ninguna integración). Nótese que un proceso ARIMA(0, 1, 0) es una caminata aleatoria (random walk) ($r_t = r_{t-1} + a_t$), que es una cadena de Markov, las cadenas de Markov se verán en el próximo capítulo. Es importante señalar, en línea con [Tsa02] que $r_t - r_{t-1} = (1 - B)r_t$ debe conformar un modelo ARMA(p, q) estacionario, por lo que este modelo no es indicado para series de precios financieros, aunque su uso si es posible en series de retornos logarítmicos $r_t = \ln(p_t) - \ln(p_{t-1})$. El lector que desee ver en detalle los modelos ARIMA, en [Tsa02] se encuentran aplicados a series de retornos financieros y en [Cha96, SS06] a series temporales en general.

- *Modelos autoregresivos condicionales heterocedásticos generalizados* (GARCH) (generalized autoregressive conditional heteroscedastic models): Los modelos anteriores tienen dificultades para modelar series con alta volatilidad, con el objeto de solucionar este problema, en la literatura se encuentran modelos heterocedásticos que son modelos en donde la varianza de la serie temporal es dependiente del tiempo. La versión más simple ARCH(m) (no generalizada) indica que el precio de mercado se puede modelar según la siguiente ecuación:

$$p_t = f(t) + u_t, \quad u_t = \sigma_t a_t, \quad \sigma_t^2 = \beta_0 + \beta_1 u_{t-1}^2 + \dots + \beta_m u_{t-m}^2 \quad (2.15)$$

donde $f(t)$ es una función determinista del precio en el momento t , u_t constituye la perturbación del precio en el momento t , σ_t es la desviación

estándar, y $\{a_t\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza 1. Nótese que la varianza condicional del momento t se modela de acuerdo a un proceso de regresión dinámica de orden m del término de perturbación, donde β_i $i=1, \dots, m$ son los parámetros a estimar.

La generalización de este modelo (GARCH) impone que la varianza no solo dependa de los pasados valores de perturbación, sino también de los pasados valores de la misma varianza, así en GARCH(m, n) el precio se modela igual que en la ecuación 2.15, solo que la varianza se modela como indica la siguiente igualdad:

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^m \beta_i u_{t-i}^2 + \sum_{j=1}^n \alpha_j \sigma_{t-j}^2 \quad (2.16)$$

El lector que quiera profundizar sobre los modelos GARCH, en [Tsa02, FD03] encontrará buenas opciones aplicadas a series financieras, y en [SS06] a series temporales en general.

- *Modelos basados en redes neuronales* (neural networks): El enfoque de redes neuronales, es un enfoque no lineal del análisis de series temporales. Al igual que los modelos ocultos de Markov, estos modelos también son útiles en los sistemas de reconocimiento de patrones, (en la etapa de clasificación de los mismos), en el capítulo 6, sección 6.3.3 se dará un vistazo de dicha aplicación. En términos generales una red neuronal puede ser vista como un sistema que conecta un conjunto de entradas (inputs) a un conjunto de salidas (outputs) de una manera (posiblemente) no lineal. Estas conexiones se producen a través de una o más *capas ocultas* como puede apreciarse en la figura 2.10.

De acuerdo al problema específico, se deberá elegir la estructura de la red adecuada, que consiste en determinar la cantidad de capas y la cantidad de neuronas por cada capa.

La principal razón de la popularidad de las redes neuronales en el análisis de series temporales, es que estos modelos han demostrado ser capaces de aproximar casi cualquier función no lineal arbitrariamente cerrada. Por lo tanto, cuando se aplica a una serie temporal que se caracteriza por relaciones

dinámicas verdaderamente no lineales, la red neuronal detectará las mismas y proporcionará un ajuste superior en comparación con modelos lineales de series de tiempo, sin la necesidad de construir un modelo paramétrico no lineal específico.

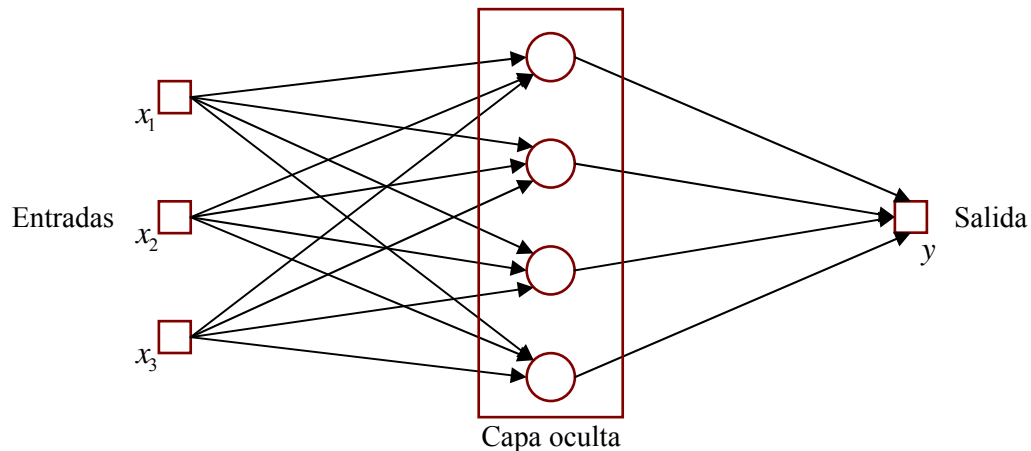


Figura 2.10: Ejemplo de red neuronal con 3 entradas, una capa oculta de 4 neuronas, y una salida.

Si se desea una guía detallada sobre redes neuronales aplicadas a series temporales, en [FD03] se encontrará una excelente guía desde el enfoque de series financieras, y en [Cha96] otra con un enfoque general.

- *Modelos ocultos de Markov*: Los modelos ocultos de Markov son modelos probabilísticos especialmente útiles para el análisis de procesos que tienen diferentes periodos de comportamiento, y con distintas variantes se han aplicado con éxito al modelado de series temporales financieras. Como se expresó anteriormente, en este trabajo no se aplicarán los modelos ocultos de Markov directamente a las series temporales financieras, sino en el contexto del análisis técnico bursátil, dentro de la etapa de clasificación de un sistema de reconocimiento de patrones. Los modelos ocultos de Markov discretos y los sistemas de reconocimiento de patrones serán tratados en detalle en próximos capítulos.

El lector que desee ampliar sobre el tema específico de la aplicación directa de los modelos ocultos de Markov a series temporales financieras (en particular series temporales de retornos), en [ME07] y en [BH04] encontrará dos excelentes opciones.

2.4.3 Modelos basados en el análisis técnico bursátil

El análisis técnico bursátil se basa fundamentalmente en la detección y reconocimiento de patrones en *gráficos* de series temporales financieras con el propósito de pronosticar *tendencias futuras* [Mur00, AP97]. Con ese objetivo se buscan conjuntos de características en la serie temporal, y dichos conjuntos conformarán los patrones a reconocer. Por lo tanto, sin dudas es un modelo de análisis no estacionario (ya que no es necesario estacionalizar la serie pues las características no estacionarias no son un escollo para el análisis, es más, muchas veces son parte del conjunto de características buscado).

El reconocimiento de patrones en series temporales financieras se ha realizado desde hace más de un siglo en forma manual, sólo muy recientemente y con el auge de los sistemas de reconocimiento de patrones se ha avanzado en el reconocimiento automático. En realidad los propios sistemas de reconocimiento de patrones, con excepción de muy pocos ejemplos como el reconocimiento del habla, conforman una disciplina todavía incipiente.

Con respecto al reconocimiento automático de los patrones del análisis técnico bursátil, como ya se dijo en el capítulo 1, se ha encontrado solamente un trabajo en ese sentido, de parte de John Murphy, autor de [Mur00]. Dicho trabajo consiste en un accesorio (plug-in) para las versiones 7 o superiores del programa comercial MetaStock® de la firma Equis International, el plug-in tiene el nombre de “*Chart Pattern Recognition For MetaStock*” desarrollado con el lenguaje de fórmulas propio del programa MetaStock. El programa tiene la seria desventaja de no ser una solución estocástica, y ésta es una limitación importante, ya que si por ejemplo hubiera una alta probabilidad de que en un gráfico se haya formado un patrón en particular, pero no hubiera la total certeza, este plug-in no lo detectaría.

Por el contrario, los modelos ocultos de Markov, introducidos por este trabajo, son modelos probabilísticos de probada eficacia en sistemas de reconocimiento de patrones, dichos modelos participan en la etapa de clasificación de los sistemas nombrados. En cuanto al ejemplo antes citado, el sistema de reconocimiento de patrones bursátiles basado en modelos ocultos de Markov, a diferencia del plug-in comentado, retornaría el patrón con su respectiva probabilidad. En los siguientes dos capítulos se tratarán en detalle los modelos ocultos de Markov, y los sistemas de reconocimiento de patrones se tratarán en detalle en el capítulo 6.

Capítulo 3

MODELOS OCULTOS DE MARKOV (HMM)

“No se puede desatar un nudo sin saber cómo está hecho”.

Aristóteles.

3.1 Introducción

Los procesos que se pueden encontrar en el mundo real [Rab89] producen generalmente “salidas observables” (outputs) que pueden ser caracterizados como “señales”. Estas señales pueden ser *discretas* (caracteres de un alfabeto, stock de mercadería, etc.) o *continuas* (muestras de voz, medidas de temperatura, música, etc.). El origen de la señal puede ser *estacionaria* u *homogénea* (sus propiedades no varían con el tiempo) o *no-estacionarias* (las propiedades de la señal varían con el tiempo), la señal puede ser *pura* (de un solo origen) o *corrupta* por otras señales de orígenes distintos (ej.: ruido) o por distorsiones en la transmisión, reverberación, etc.

Un problema fundamental es el de diseñar modelos que interpreten estas señales del mundo real. Hay muchas razones para querer modelar dichas señales, no se citarán todas, pero la razón más importante es que estos modelos funcionan muy bien en la práctica, habilitando la eficiente realización de importantes sistemas como pueden ser: sistemas de predicción, sistemas de reconocimiento, sistemas de identificación, etc.

Existen dos tipos de modelos a saber:

- *Modelos determinísticos*: recalcan en propiedades específicas conocidas de la señal, todo lo que se requiere es determinar (o estimar) los parámetros del modelo (ej.: amplitud, frecuencia, fase de una onda, etc.)
- *Modelos estadísticos*: caracterizan solamente las propiedades estadísticas de la señal. Asumen que la señal puede ser caracterizada como un *proceso aleatorio parametrizado* y que los parámetros pueden ser determinados (estimados) de forma precisa y bien definida. Algunos modelos estadísticos son: procesos Gaussianos, procesos de Poisson, procesos de decisión de Markov, modelos ocultos de Markov, etc.

En este capítulo se exhibirá una introducción general a los modelos ocultos de Markov (HMM⁸ – por sus siglas en inglés –). Los HMM son modelos estadísticos que resultan especialmente útiles para modelar procesos que se desarrollan a través de diferentes periodos o regímenes de comportamiento. De esta forma, los HMM están diseñados para capturar la dinámica de la serie, reconociendo por un lado un conjunto finito de estados discretos por los que se desarrolla el proceso, y por otro identificando las probabilidades de que el sistema evolucione de un estado a otro.

En general son muy útiles para el *reconocimiento de patrones*⁹ y en particular estos modelos son substancialmente efectivos para modelar el comportamiento del hombre [PL99, HOZ03] o procesos en los que existe la intervención del hombre. En el caso específico que concierne a este trabajo, cuando los agentes que participan de un mercado bursátil acuden al mismo se origina una interacción entre ellos, que junto con las situaciones concretas de oferta y demanda del mercado dan lugar a un proceso que a lo largo del tiempo evoluciona de forma similar a un modelo oculto de Markov.

Los HMM han demostrado su eficacia para modelar sistemas dinámicos que evolucionan a través de un conjunto finito de estados, prueba de ello son los trabajos de otros autores detallados en la sección 3.3.1 de este capítulo. El análisis técnico bursátil modela las series temporales financieras considerando que evolucionan a través de un conjunto finito de *estados de mercado*. Estos estados *dependerán del patrón* que dicho análisis desee reconocer en las series.

⁸ Hidden Markov Models (La sigla HMM se usará en este trabajo para referirse a los Modelos ocultos de Markov – en plural y en singular –).

⁹ Se tratará en detalle en el Capítulo 6.

3.2 Definiciones

Las nociones básicas sobre procesos de Markov introducidas en esta sección son una herramienta necesaria para abordar en la siguiente sección (3.3) el tratamiento de los Modelos ocultos de Markov.

3.2.1 Proceso de Markov

Un proceso de Markov¹⁰ es un tipo especial de proceso estocástico en donde se verifica la “propiedad de Markov”, dicha propiedad puede verbalmente enunciarse como: “la probabilidad de que ocurra un evento, conociendo los estados de muchos momentos previos, es la misma que conociendo solamente el último estado dado” o más informalmente significa que “el pasado no tiene influencia en el futuro excepto a través del presente” [Chu67] Esta condición es también conocida como “principio de Markov” [Din65] en el cual “el futuro es independiente del pasado para un presente conocido”.

Específicamente, se puede decir [Nel95] que un proceso estocástico $\{s_t : t \in T\}$ es un proceso de Markov si al escoger cualquier conjunto $\{t_1, \dots, t_n\}$ de n valores del conjunto parámetro T , la probabilidad condicional del evento $(s_{t_n} \leq S_n)$ ¹¹ cuando han acontecido los eventos $s_{t_1} = S_1; s_{t_2} = S_2; \dots; s_{t_{n-1}} = S_{n-1}$ sólo depende de $s_{t_{n-1}}$, o de otro modo:

$$P\left(s_{t_n} \leq S_n \mid s_{t_1} = S_1; s_{t_2} = S_2; \dots; s_{t_{n-1}} = S_{n-1}\right) = P\left(s_{t_n} \leq S_n \mid s_{t_{n-1}} = S_{n-1}\right) \quad (3.1)$$

La equivalencia enunciada en la ecuación (3.1) se conoce normalmente como “propiedad de Markov” de orden uno. Por lo expuesto se desprende que en un

¹⁰ En honor a Andrei A. Markov, una completa descripción de su vida y obra puede encontrarse en [BLN03]

¹¹ Se refiere a $S_i, i = 1, \dots, n$ como los elementos del espacio de estados S .

proceso de Markov es preciso saber la siguiente probabilidad, llamada probabilidad de transición:

$$P\left[s_t \in S^* \mid s_{t_0} = S_j\right] \quad (3.2)$$

que se interpreta como la probabilidad de que en el momento t el proceso sea parte del conjunto S^* encontrándose en el estado S_j en el momento t_0 .

Como puede inferirse por lo visto anteriormente, los procesos de Markov de orden uno se caracterizan porque el futuro $s_{t_{n+1}} = S_{n+1}$ depende exclusivamente del presente $s_{t_n} = S_n$ y no del pasado $s_{t_1} = S_1; s_{t_2} = S_2; \dots; s_{t_{n-1}} = S_{n-1}$. Además es factible la definición de procesos de Markov de orden k ($k > 1$), es decir, donde el futuro $s_{t_{n+1}} = S_{n+1}$ depende del presente $s_{t_n} = S_n$ y de los $(k-1)$ valores previos $s_{t_{n-k+1}} = S_{n-k+1}; s_{t_{n-k+2}} = S_{n-k+2}; \dots; s_{t_{n-1}} = S_{n-1}$.

Los procesos de Markov se clasifican de acuerdo a la condición de los conjuntos T y S , dichos conjuntos pueden ser discretos o continuos. Existe una leve discordancia entre los autores, algunos clasifican los procesos de Markov de acuerdo al espacio de estados S , denominando “Cadena de Markov” si el espacio de estados es discreto y “Proceso de Markov” para referirse a un espacio de estados continuo (ver [Chu67, Ros96, Ros97, HPS72, Law95]), esta nomenclatura da lugar a la necesidad de referirse a “Cadenas de Markov de tiempo discreto (o de parámetro discreto)” y “Cadenas de Markov de tiempo continuo (o de parámetro continuo)”. Otros autores en cambio clasifican los procesos de Markov de acuerdo al conjunto parámetro T , designando como “Cadena de Markov” a aquella cuyo conjunto parámetro T sea discreto y “Proceso de Markov” si el conjunto parámetro T es continuo (ver [Nel95]).

3.2.2 Cadena de Markov

En este trabajo una “cadena de Markov” será un proceso de Markov donde el espacio de estados S es discreto y el conjunto parámetro T también es discreto¹².

Una cadena de Markov constituye de esta manera un sistema dinámico en el que para cada valor del parámetro¹³ de tiempo t el sistema está en uno de los estados determinados por el conjunto discreto de estados $S = \{S_1, S_2, \dots, S_n\}$.

Se puede pensar esta dinámica de la siguiente manera [Nel95]:

El sistema comienza en el momento de tiempo t_0 en uno de los estados $\{S_1, S_2, \dots, S_n\}$. En ese periodo t_0 el vector que otorga las probabilidades iniciales $\{\pi_1, \pi_2, \dots, \pi_n\}$ establece la probabilidad de que el sistema empiece en cada uno de los estados determinados. El progreso del sistema del instante t al instante $t + 1$ está gobernado por la matriz que establece las probabilidades de transición A . En dicha matriz el elemento a_{ij} simboliza la probabilidad de que el sistema esté en el estado S_i en el momento t y evolucione al estado S_j en el momento $t + 1$ (se puede apreciar un ejemplo de cadena de Markov a través de un *diagrama de transición* en la figura 3.1).

Adaptando la ecuación (3.1) al caso específico de las cadenas de Markov se tiene:

$$P\left(s_{t_n} = S_n \mid s_{t_1} = S_1; s_{t_2} = S_2; \dots; s_{t_{n-1}} = S_{n-1}\right) = P\left(s_{t_n} = S_n \mid s_{t_{n-1}} = S_{n-1}\right) \quad (3.3)$$

¹² “Cadena de Markov” será análogo de “Cadena de Markov de tiempo discreto”

¹³ El conjunto parámetro T , de una cadena de Markov, en general se refiere al tiempo, pero no hay que perder de vista la generalidad de la definición.

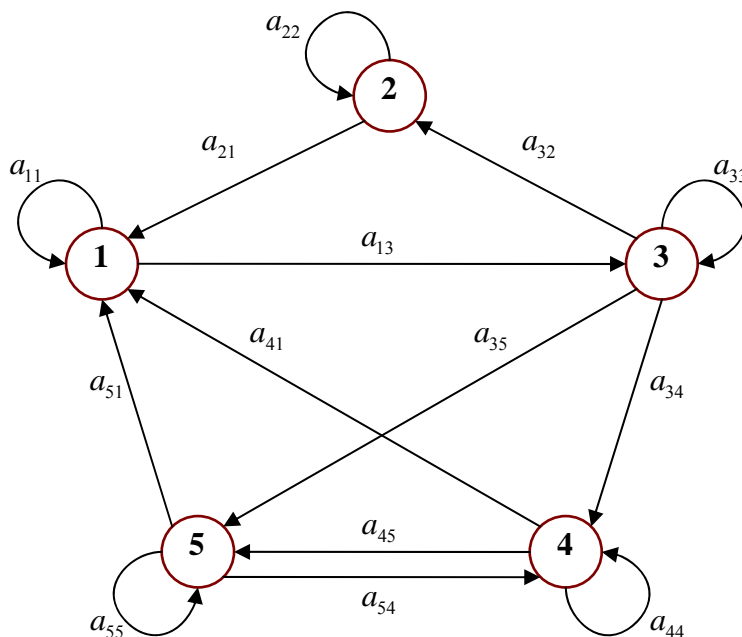


Figura 3.1: Cadena de Markov de 5 estados con transiciones entre ellos.

Formalmente, una cadena de Markov se define como una 3-tupla $C = \{\Pi, S, A\}$ en donde:

- S es un conjunto finito de estados. $S = \{S_1, S_2, \dots, S_n\}$
- Π es el vector de probabilidades iniciales. $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$, con $\sum_{i=1}^n \pi_i = 1$
- A es la matriz de probabilidades de transición. $A = \{(a_{i,j})\}$ $i, j = 1, \dots, n$ siendo a_{ij} la probabilidad condicionada $P(s_{t+1} = S_j | s_t = S_i)$, por tanto $\sum_{i=1}^n a_{ij} = 1$, $\forall j = 1, \dots, n$

Una cadena de Markov recibe el apelativo de “homogénea” [Nel95, Ros96, Ros97, Law95, HPS72] cuando la matriz de transición de probabilidades que corresponde a dicha cadena es independiente del tiempo¹⁴, esto es,

$$P(s_{t+1} = S_j | s_t = S_i) = P(s_{t+k+1} = S_j | s_{t+k} = S_i), \forall k \quad (3.4)$$

Supóngase que se tiene la siguiente secuencia de estados (s_1, s_2, \dots, s_T) . Teniendo en cuenta la propiedad de Markov, la probabilidad de que dicha secuencia de estados haya sido concebida por una cadena de Markov vendrá dada por¹⁵:

$$P(s_1, s_2, \dots, s_T) = P(s_1, s_2, \dots, s_{T-1})P(s_T/s_{T-1}) = P(s_0) \prod_{t=1}^T P(s_t/s_{t-1}) \quad (3.5)$$

Esto indica que una cadena de Markov está unívocamente determinada en primer lugar por la matriz de probabilidades iniciales y en segundo lugar por la matriz de probabilidades de transición [Nel95].

Ejemplo 3.1 (Basado en un ejemplo de [Rab89, RJ93] y adaptado) Supóngase que se tiene una cadena de Markov homogénea con tres estados proveniente de la observación del clima: (1) lluvia, (2) nublado ó (3) sol, presúmase también que el día t se identifica por solo uno de esos tres estados y sea la matriz A de las probabilidades de transición para los estados como sigue:

$$A = \{a_{ij}\} = \begin{bmatrix} 0,4 & 0,3 & 0,3 \\ 0,2 & 0,6 & 0,2 \\ 0,1 & 0,1 & 0,8 \end{bmatrix}$$

y además todos los estados poseen idéntica posibilidad de convertirse en el estado inicial, esto es, $\pi_i = \frac{1}{3}, i = 1, 2, 3$. (Se puede ver este modelo en la figura 3.2)

¹⁴ Notar que si no fuera homogénea, se debería tener una matriz para cada momento de tiempo.

¹⁵ Aquí $P(s_i)$ significa $P(s_i = S_i)$

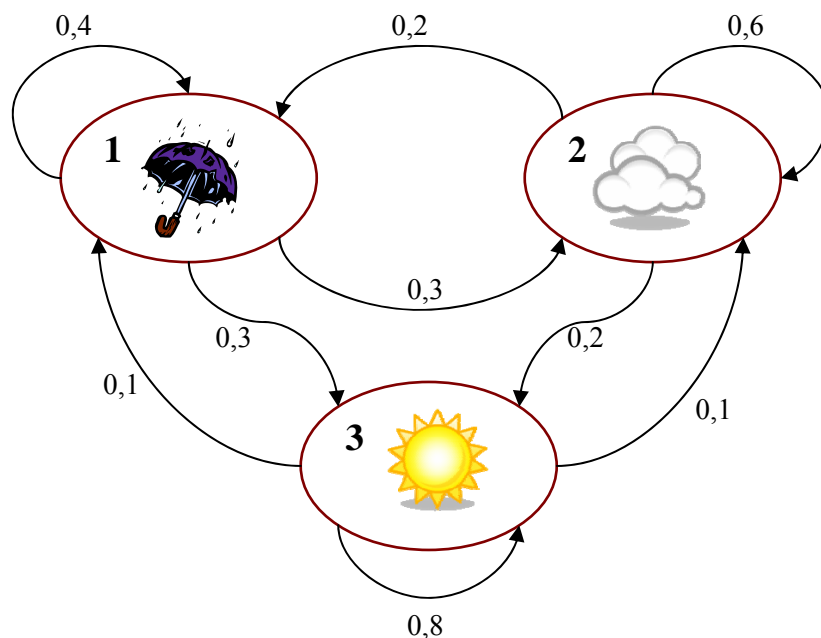


Figura 3.2: Cadena de Markov de 3 estados (aplicado al clima – ejemplo 3.1 –).

La probabilidad de que se produzca una secuencia de estados en particular es fácil de calcular. Sea $s_i \in \{1,2,3\} \quad \forall i$, entonces para una secuencia finita de variables aleatorias s_1, s_2, \dots, s_T se tiene que:

$$P(s_1, s_2, \dots, s_T) = P(s_1 = S_1)P(s_2 = S_2 | s_1 = S_1)P(s_3 = S_3 | s_2 = S_2) \dots P(s_T = S_T | s_{T-1} = S_{T-1}) = \pi_{s_1} \prod_{t=1}^{T-1} a_{s_t, s_{t+1}}$$

Ahora se puede calcular un caso específico, como es la probabilidad que se produzca la secuencia de estados *sol-sol-nublado-lluvia* y entonces aplicando lo anterior se tiene:

$$P(3,3,2,1) = P(3)P(3|3)P(2|3)P(1|2) = \pi_3 a_{33} a_{32} a_{21} = \frac{1}{3} \times 0,8 \times 0,1 \times 0,2 \approx 5,33 \times 10^{-3}$$

□

3.3 Modelos ocultos de Markov (HMM)

Un problema con los procesos de Markov de orden k (ver 3.2.1) es que se tornan inmanejables para un k “grande. [Ben99]. Si se tiene un proceso de Markov con n estados, el número de parámetros requeridos para representar las probabilidades de transición está en $O(n^{k+1})$, por lo que es necesario usar un valor pequeño de k . El inconveniente es que la mayoría de los datos secuenciales observados de interés no satisfacen la propiedad de Markov para un k pequeño (entre los que se encuentran las series temporales financieras). Como una solución a este problema nacen los Modelos ocultos de Markov, en un HMM no se asume que la secuencia de observaciones cumple la propiedad de Markov, (la única condición es la de ser un proceso estocástico), sin embargo en un HMM se tiene otro proceso estocástico no observado (oculto), y dicho proceso en un HMM básico¹⁶, es una Cadena de Markov *homogénea* con espacio de estados *finito*, (es decir: cumple la propiedad de Markov y posee una matriz de transición de probabilidades constante e independiente del tiempo).

Un HMM es un proceso estocástico bivariado $\{X_k, Y_k\}_{k \geq 0}$, donde $\{X_k\}$ es una cadena de Markov homogénea con espacio de estados finito y $\{Y_k\}$ es un proceso estocástico cuya distribución condicional sólo depende de $\{X_k\}$. [CMR07]

Los *estados* de un HMM se corresponden a los estados de la cadena de Markov oculta, y por lo tanto, a diferencia de un proceso de Markov común, son *ocultos*, sólo son visibles las variables (*observaciones*) condicionadas por los estados [BH94].

Para ejemplificar los conceptos básicos de un HMM se introduce el siguiente ejemplo clásico de urnas y bolas. [Rab89, RJ86, RJ93]

Ejemplo 3.2 Supóngase que se dispone de un conjunto de N urnas y cada urna tiene en su interior una cantidad de bolas de distintos colores, habiendo M distintos colores de bolas. De este modo, cada urna posee una cantidad de bolas, la que puede diferir de una urna a otra. Por lo tanto en cada urna se tendrá una distribución de

¹⁶ Existen otros HMM's ampliados como los IOHMM (*input-output HMM*) donde la cadena de Markov correspondiente al proceso estocástico oculto inherente a un HMM puede no ser homogénea, pero su tratamiento escapa al alcance de este trabajo (ver Futuros desarrollos Cáp. 8)

probabilidad diferente, vinculada al color de las bolas ($b_i(k)$ será la probabilidad de obtener una bola de color k de la urna i) (ver figura 3.3).

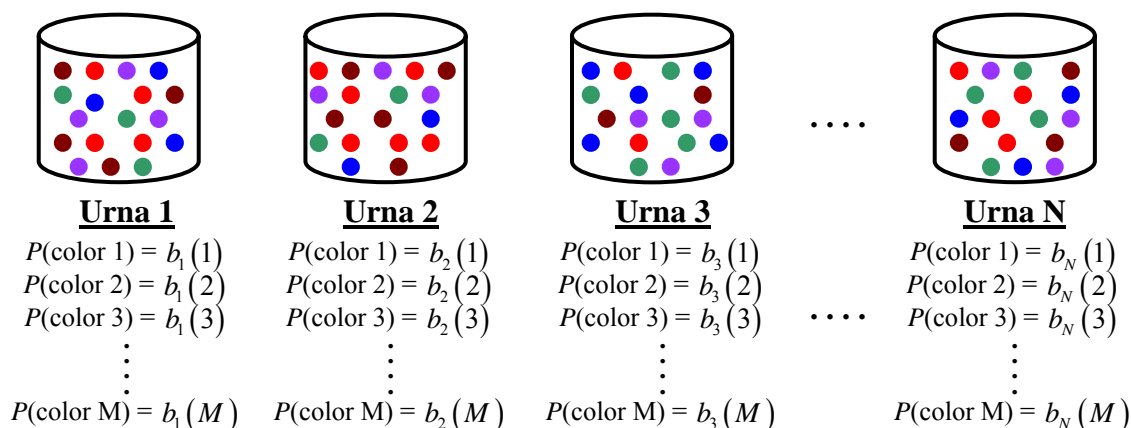


Figura 3.3: HMM para N urnas con bolas de M colores.

El procedimiento físico para obtener una secuencia de observaciones compatible con un HMM sería el que se muestra a continuación:

Una persona se encuentra frente a las urnas y elige al azar una de ellas para luego extraer una bola, entonces se registra el color observado y se devuelve la bola a la urna; a continuación, esta persona continúa el proceso aleatorio de elección de urnas y saca una nueva bola, de la que también se registra su color y se devuelve, este proceso se repite un número finito de veces, de manera que se registra una secuencia finita de colores.

Nótese que las urnas se corresponden con los estados del HMM, la elección de las urnas se realiza de acuerdo a la matriz de probabilidades de transición correspondiente, la secuencia de colores constituye la salida observable del modelo y cada estado posee una distribución de probabilidad perfectamente definida correspondiente a los colores. Es importante destacar que una observación aislada de un color determinado, no dice de qué urna viene dicha bola (los estados están *ocultos* al observador).

□

3.3.1 Historia de los HMM

La teoría básica concerniente a los HMM fue introducida al final de la década de 1960 y comienzos de la década de 1970 por Leonard E. Baum y sus colegas quienes le otorgaron el nombre de *funciones probabilísticas de cadenas de Markov*, recién en 1980 en un trabajo sobre reconocimiento del habla se otorga la denominación de *Hidden Markov Models* (HMM). Sin embargo a pesar de algunos trabajos sobre procesamiento del habla de mediados a fines de la década del 1970, la verdadera explosión de aplicaciones de los HMM ocurre a finales de la década de 1980 y continúa hasta nuestros días.

Existen varias razones para esto, pero se pueden citar en concordancia con [Rab89] las siguientes:

- La Teoría básica de los HMM fue dada a conocer en publicaciones matemáticas, las cuales no eran generalmente leídas por profesionales que trabajaban en diversos ámbitos donde los HMM podían ser aplicables.
- Las publicaciones originales no contenían suficiente material de apoyo (tutoriales) para que la mayoría de los lectores pudieran entender la teoría y estuvieran preparados para aplicarla a sus trabajos.

Entre 1986 y 1989, Lawrence R. Rabiner¹⁷ [Rab89,RJ86] y en parte como respuesta a esto último, describe la teoría básica de Baum, proveyendo detalles prácticos de los métodos de aplicación de la teoría, describiendo los problemas fundamentales y su solución, combinando múltiples fuentes y ayudando efectivamente a obtener una base para su aplicación y si bien estos trabajos son dirigidos al reconocimiento del habla, es muy recomendable su consulta para la obtención de directrices y referencias sobre el desarrollo y uso de los HMM's. Esta es la razón por la cual estos trabajos se constituyeron en una de las principales fuentes de consulta para la realización de esta tesis.

En la práctica los HMM han sido empleados en cuantiosos trabajos concernientes a disciplinas muy diferentes como Biociencia [KBM93], extracción de direcciones postales de una imagen [TCPN05], reconocimiento del habla [Rab89, RJ86, RJ93],

¹⁷ Lawrence R. Rabiner Ph.D MIT y AT&T Research es una referencia obligada en *Procesamiento de señales digitales y Reconocimiento del habla*.

clustering de series temporales [OFC04], reconocimiento de gestos [YX94], series temporales financieras [Bul06], interfase humano/computadora [Lan99], análisis del genoma humano [Edd97], reconocimiento de emociones en el habla [NFS03], etc.

3.3.2 Arquitectura de un HMM

En la figura (3.4) se puede apreciar la arquitectura general de un HMM [Bul06, Ben99]. No se debe confundir este diagrama con el diagrama de un HMM (que se puede ver en la figura (3.5)), Aquí se trata de una Red Bayesiana¹⁸ dinámica simple utilizada para entender la dinámica de la arquitectura de un HMM. En la figura se puede apreciar la cadena de Markov (oculta) cumpliendo la propiedad de Markov de orden 1 correspondiente y el proceso estocástico relativo a las observaciones y que es dependiente de cada estado oculto respectivo. Cada círculo simboliza una variable aleatoria que puede adquirir determinados valores. La variable aleatoria q_t representa el valor de la variable oculta en el momento de tiempo t . La variable aleatoria y_t representa el valor de la variable observada en el mismo momento de tiempo t . Las flechas representan dependencias condicionales.

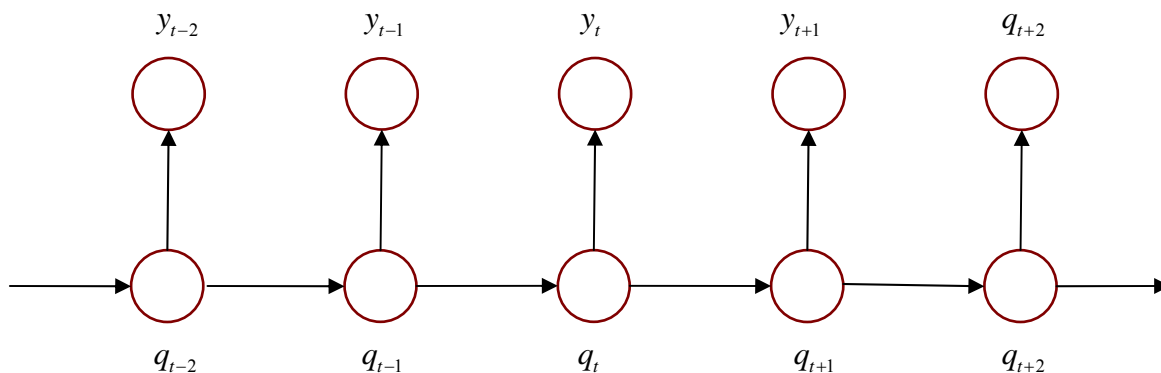


Figura 3.4: Red Bayesiana dinámica simple para analizar la arquitectura de un HMM.

¹⁸ Una **red bayesiana** es un modelo probabilístico gráfico que representa a un conjunto de variables y sus dependencias.

En el diagrama se puede notar claramente que el valor de la variable oculta q_t (en el instante t) depende solamente del valor de la variable oculta q_{t-1} (en el momento $t - 1$). De forma análoga, el valor de la variable observada y_t depende solamente del valor de la variable oculta q_t (ambas en el momento t).

3.3.3 Formalización de un HMM

Un HMM discreto¹⁹ se define formalmente [Rab89, RJ86, RJ93] como una 5-tupla (S, V, Π, A, B) , donde:

- S es el conjunto finito de estados (ocultos) del HMM $S = \{S_1, S_2, \dots, S_N\}$ donde N es la cantidad de estados en el modelo. El estado actual en el instante de tiempo t se denota como q_t .
- V es el conjunto de valores o símbolos diferentes que se pueden observar en cada uno de los estados, puede considerarse también un alfabeto finito. Cada uno de los símbolos que un estado puede emitir se denota como $\{v_1, v_2, \dots, v_M\}$ donde M es el número de símbolos del alfabeto y cada v_k se refiere a un símbolo diferente.
- $\Pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ es la distribución de probabilidad inicial de los estados. De esta forma,

$$\pi_i = P(q_1 = S_i), \quad \pi_i \geq 0, \quad 1 \leq i \leq N \quad \text{y} \quad \sum_{i=1}^N \pi_i = 1$$

Donde $\pi_i = P(q_1 = S_i)$ constituye la probabilidad de que el sistema inicie en el estado S_i .

¹⁹ En este trabajo HMM se usará como análogo de HMM discreto, escapa al alcance de este trabajo los HMM continuos (con observaciones continuas). (Recuerde el lector que los estados ocultos en un HMM siempre deben ser discretos, más aún, finitos).

- $A = \{a_{ij}\}$ es la distribución de probabilidad de las transiciones entre estados, por lo tanto,

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i), \quad 1 \leq i, j \leq N, \quad 2 \leq t \leq T, \quad a_{ij} \geq 0 \quad \forall i, j, \quad \sum_{j=1}^N a_{ij} = 1, \quad \forall i.$$

Donde $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$ constituye la probabilidad de que el sistema se encuentre en el estado S_i en el instante $t - 1$ y pase al estado S_j en el instante t .

- $B = \{b_i(k)\}$ es la distribución de probabilidad de las observaciones, es decir,

$$b_i(k) = P(o_t = v_k | q_t = S_i), \quad b_i(k) \geq 0, \quad 1 \leq i \leq N, \\ 1 \leq k \leq M, \quad 1 \leq t \leq T \quad \sum_{k=1}^M b_i(k) = 1, \quad \forall i.$$

Donde $b_i(k) = P(o_t = v_k | q_t = S_i)$ constituye la probabilidad de que el sistema, estando en el instante t en el estado S_i , genere la observación v_k .

Una secuencia de observaciones se denota como un conjunto $O = (o_1, o_2, \dots, o_T)$ donde cada observación o_i , $i = 1, \dots, T$ es un elemento de V , y T es la cantidad de observaciones en la secuencia.

La especificación anterior es rigurosa, pero en la práctica, es común ver una notación mas compacta para denotar un HMM, usando sus tres distribuciones de probabilidad [Rab89, RJ93]:

$$\lambda = (A, B, \Pi) \tag{3.6}$$

Todas las distribuciones de probabilidad vistas en la formalización (A, B, Π) , sólo tienen sentido, dado un modelo en particular, por ejemplo, en el caso de las transiciones, se tiene que $a_{ij} = P(q_t = S_j | q_{t-1} = S_i) = P(q_t = S_j | q_{t-1} = S_i, \lambda)$, donde λ es un modelo fijo.

Siguiendo esta formalización se puede ver un ejemplo de HMM en la figura 3.5

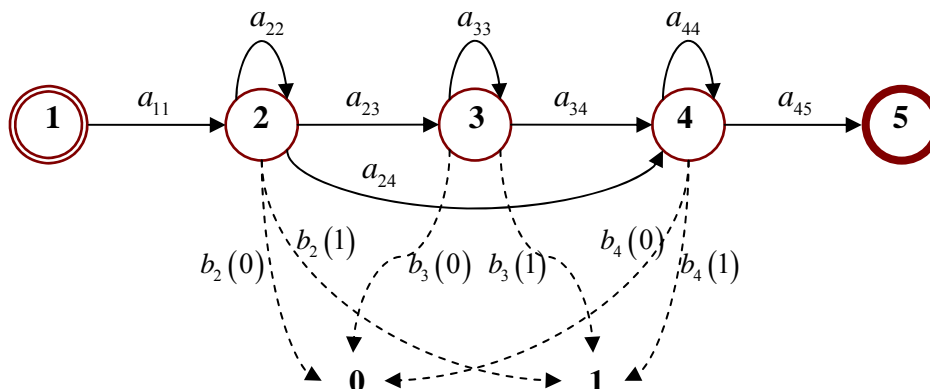


Figura 3.5: Diagrama de estados para un HMM con 5 estados y un alfabeto de observaciones con 2 elementos, las flechas continuas representan las transiciones entre estados ($a_{ij} > 0$). Y las flechas punteadas representan las probabilidades de observación de cada estado. Este es un caso particular de HMM denominado *de izquierda a derecha (left-right)* y en donde los estados 1 y 5 son *no emisores* (sin función de observación).

Otorgando valores apropiados a N , M , A , B y Π , el HMM puede ser usado para generar una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, el procedimiento sería como se detalla a continuación [Rab89, RJ93]:

Algoritmo 3.1 Generación de una secuencia de observaciones de tamaño T .

1. Elegir un estado inicial $q_1 = S_i$ de acuerdo a la distribución de probabilidad inicial de los estados Π .
2. Hacer $t = 1$.
3. Elegir $o_t = v_k$ de acuerdo a la distribución de probabilidad de las observaciones para el estado actual (es decir: $b_i(k)$).
4. Pasar a un nuevo estado $q_{t+1} = S_j$ de acuerdo a la distribución de probabilidad de las transiciones entre estados (es decir: a_{ij}).
5. Hacer $t = t + 1$; si $(t < T)$ ir al paso 3; en caso contrario, devolver $O = (o_1, o_2, \dots, o_T)$ y terminar.

3.3.4 Topología de un HMM

La topología de un HMM está determinada por la cantidad de estados que lo integran y las transiciones habilitadas entre dichos estados. La elección de la topología apropiada, obedecerá a la aplicación para la que sean elaborados los HMM, esto forma parte del problema de la elección del modelo (ver sección 4.5.4). Como ejemplo se citarán algunas de las topologías más comunes e importantes de HMM.

- *Topología de Bakis o de izquierda a derecha.* En los HMM con esta topología sólo son permitidas las transiciones *hacia delante*, por lo tanto el índice de los estados visitados, crecerá (no necesariamente en forma estricta) en el tiempo. Lo que indica que la matriz de transiciones A , será triangular superior:

$$a_{ij} \geq 0 \quad \forall i > j \quad (3.6)$$

Por lo tanto, si el proceso se encuentra en el estado i , entonces, o bien continuará en el mismo estado con probabilidad a_{ii} , o bien efectuará una transición hacia el estado j (con índice superior a i) con probabilidad a_{ij} . En la figura 3.5 se puede ver un ejemplo de esta topología en un HMM.

- *Topología de HMM ergódico.* De forma estricta, un HMM ergódico es aquel, en el cual todos sus estados (ocultos) son no nulos, no periódicos y recurrentes²⁰, pero en la práctica se denomina HMM con topología ergódica, a aquel en el cual todos sus estados se conectan entre sí, o lo que es lo mismo, en el que todas las posibles transiciones entre estados están habilitadas ($a_{ij} > 0 \quad \forall i, j$). En la figura 3.6 se puede apreciar un ejemplo de HMM ergódico.

²⁰ Una excelente guía sobre clasificación de los estados de una cadena de Markov, puede encontrarse en [Nel95], en [Ros96] y también en [Law95].

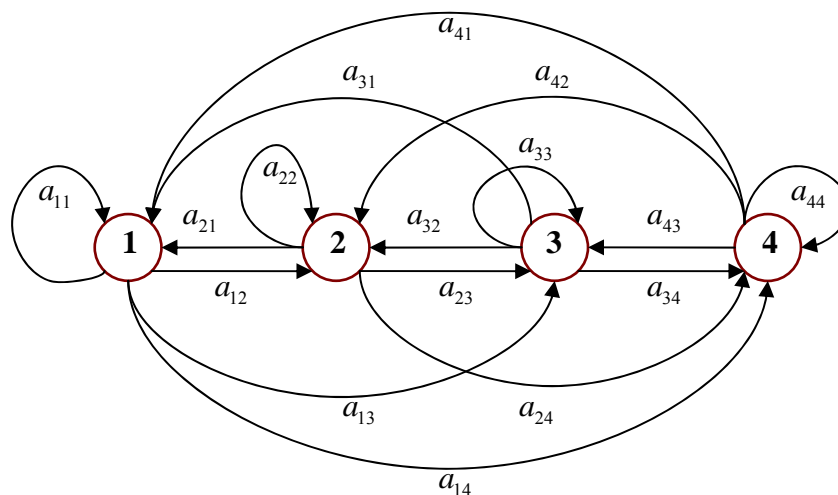


Figura 3.6: Ejemplo de HMM ergódico de 4 estados, (no están graficadas las distribuciones de las observaciones).

3.3.5 Problemas fundamentales de los HMM

Existen tres problemas fundamentales o canónicos asociados con los HMM [Rab89, RJ86, RJ93, Ste05] que deben estudiarse para que estos modelos puedan ser aplicados a sistemas reales. A continuación se verán cada uno de estos problemas:

1. Dada una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$ y un modelo $\lambda = (A, B, \Pi)$. ¿Cómo calcular de forma eficiente $P(O | \lambda)$?, o dicho de otro modo, ¿Cómo calcular la probabilidad de obtener dicha secuencia de observaciones dado un modelo fijo?
2. Dada una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$ y un modelo $\lambda = (A, B, \Pi)$. ¿Cómo encontrar la sucesión de estados ocultos más factible que pueda haber generado dicha secuencia de observaciones? (La que mejor explica la secuencia de observaciones dada).

3. Dada una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, ¿Cómo estimar los parámetros del modelo $\lambda = (A, B, \Pi)$ para maximizar $P(O | \lambda)$? (El modelo que mejor explica la secuencia de observaciones dada), o en otras palabras, entrénense a los parámetros del HMM dada una secuencia de observaciones.

3.3.5.1 Problema 1: Probabilidad de una observación

El primero constituye el problema de la evaluación de una secuencia de observaciones dado el modelo. La resolución de este problema provee la probabilidad de que la secuencia fuera generada por ese modelo. Un práctico ejemplo puede ser presuponer la presencia de varios HMM's pugnando entre sí. La evaluación de la secuencia en cada uno de ellos, dará la oportunidad elegir el modelo que mejor se ajusta con las observaciones.

Si se piensa en el caso del mercado bursátil, este problema se trasladará de la siguiente manera. Dada una secuencia o serie temporal de precios del mercado y la secuencia de volúmenes operados²¹ correspondiente, y un HMM ¿cómo se puede calcular de forma eficiente la probabilidad de que esa secuencia de precios-volúmenes²² haya sido generada con ese modelo particular?

3.3.5.2 Problema 2: Secuencia de estados más probable

El segundo se refiere al problema de encontrar lo oculto del modelo, es decir, trata sobre cómo descubrir el camino que siguió la cadena de Markov. Este camino oculto es de utilidad para la clasificación de las observaciones.

²¹ Para el lector que no tenga experiencia con el léxico bursátil, el volumen operado es la cantidad de activos financieros operado en una unidad del período analizado (si la serie es diaria, la unidad es un día) (Ver capítulo 5)

²² A los movimientos de precios conjuntamente con los movimientos de volumen, se los denomina “movimientos de mercado”, como se verá en el capítulo 5. Por ende una secuencia conjunta de movimientos de precio y volúmenes, será una *secuencia de movimientos de mercado*.

En el caso del mercado Bursátil, se puede establecer la siguiente correspondencia: dada una secuencia de precios del mercado bursátil y la secuencia de volúmenes operados correspondiente, y un HMM ¿qué sucesión de *estados del mercado*, describe mejor la secuencia de precios-volúmenes observada?

3.3.5.3 Problema 3: Estimación de los parámetros del modelo

El tercero de los problemas, es el problema crítico de los HMM puesto que permite adecuar en forma óptima los parámetros de un modelo a una secuencia de observaciones. También resulta el más complejo de resolver (ver capítulo 4). Notar que la solución a este problema permite la efectiva aplicación de los HMM a una cantidad de casos reales, a los que sería inviable su aplicación, pues determinar a priori las distribuciones correspondientes no es una opción en la mayoría de los casos. En lugar de eso, se dispone de una *secuencia de entrenamiento* con el fin de *enseñar* al HMM cuales son las distribuciones que mejor se adaptan.

En el trabajo de buscar una analogía con el mercado bursátil, este problema se traduce a ¿cómo diseñar el modelo HMM para que la probabilidad de una secuencia de precios del mercado bursátil con su correspondiente secuencia de volúmenes de operación sea máxima?

Capítulo 4

SOLUCION A LOS PROBLEMAS FUNDAMENTALES DE LOS HMM

“Todas las teorías son legítimas y ninguna tiene importancia. Lo que importa es lo que se hace con ellas”.

Jorge Luis Borges.

4.1 Introducción

De acuerdo a lo visto en la última sección del capítulo anterior, existen tres problemas fundamentales o canónicos referidos a los HMM. En el presente capítulo se abordará la resolución eficiente a dichos problemas y se tratarán en detalle los algoritmos necesarios para ello. La mayor parte de lo expuesto en este capítulo se basa en [RJ86, Rab89, RJ93, Ste05].

4.2 Solución al problema 1: Probabilidad de una observación

Se quiere calcular de manera eficiente la probabilidad de una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, dado un modelo $\lambda = (A, B, \Pi)$, es decir $P(O | \lambda)$.

Recordando del capítulo anterior que cada o_t depende solamente de q_t , (la observación en el tiempo t depende únicamente del estado en el tiempo t), entonces para producir la secuencia O se debe tener una secuencia de estados

$Q = (q_1, q_2, \dots, q_T)$. Luego esto permite inferir un sencillo método de resolución al problema en cuestión, simplemente se podría *calcular la probabilidad de obtener O , con cada una de todas las posibles secuencias de estados de longitud T que son posibles con el modelo, y finalmente sumar dichas probabilidades*. Notar que si N es el número de estados del modelo, existen N^T posibles secuencias de estados de longitud T .

Recuérdese que de acuerdo a la ecuación (3.5) del capítulo anterior, la probabilidad de que ocurra la secuencia de estados $Q = (q_1, q_2, \dots, q_T)$ en el modelo $\lambda = (A, B, \Pi)$ está dada por:

$$P(Q | \lambda) = \pi_{q_1} \prod_{t=1}^{T-1} a_{q_t q_{t+1}} \quad (4.1)$$

La probabilidad de O , dada la secuencia Q y el modelo $\lambda = (A, B, \Pi)$ es:

$$P(O | Q, \lambda) = \prod_{t=1}^T P(o_t | q_t, \lambda) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdots b_{q_T}(o_T) \quad (4.2)$$

Donde se asume independencia entre observaciones, recordar la estructura de un HMM (sección 3.3.2).

La probabilidad conjunta de O y Q (la probabilidad de obtener la secuencia de observaciones O simultáneamente con la secuencia particular de estados Q) dado el modelo, es el producto de (4.1) y (4.2) (usando la definición de probabilidad condicional en conjunto con la regla de la multiplicación [Dev01]), es decir:

$$P(O, Q | \lambda) = P(O | Q, \lambda) P(Q | \lambda) \quad (4.3)$$

Y como se comento al principio, para obtener $P(O | \lambda)$ se debe calcular (4.3) para cada secuencia posible de estados de tamaño T del modelo y luego calcular la suma de las mismas, por lo tanto:

$$P(O | \lambda) = \sum_{\forall Q} P(O | Q, \lambda) P(Q | \lambda) \quad (4.4)$$

O lo que es lo mismo:

$$P(O | \lambda) = \sum_{\forall Q} \left(\pi_{q_1} b_{q_1}(o_1) \prod_{t=1}^{T-1} a_{q_t q_{t+1}} b_{q_{t+1}}(o_{t+1}) \right) \quad (4.5)$$

4.2.1 Algoritmos de avance - retroceso (forward – backward)

Se puede interpretar la ecuación (4.5) de la siguiente forma. Al inicio (tiempo $t = 1$), el proceso se encuentra en el estado q_1 con probabilidad π_{q_1} y es generado el símbolo o_1 con probabilidad $b_{q_1}(o_1)$, en $t = 2$, se produce una transición del estado q_1 a estado q_2 con probabilidad $a_{q_1 q_2}$, y se genera el símbolo o_2 con probabilidad $b_{q_2}(o_2)$. El proceso continúa hasta que se llega al tiempo $t = T$, con una transición del estado q_{T-1} al estado q_T con probabilidad $a_{q_{T-1} q_T}$ y generando el símbolo o_T con probabilidad $b_{q_T}(o_T)$. En la figura 4.1 se puede ver una representación gráfica de lo anterior.

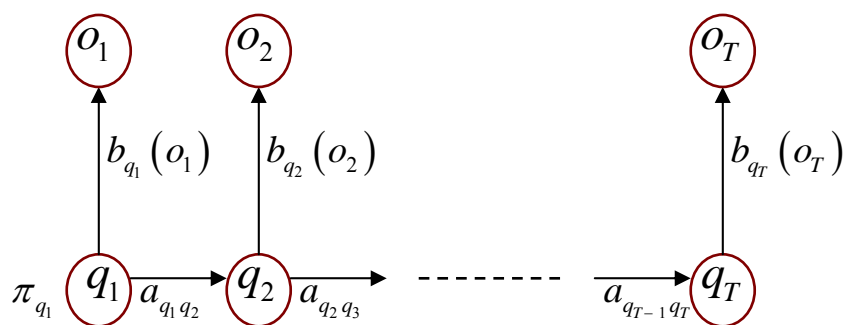


Figura 4.1: Proceso representativo de la ecuación (4.5)

Como ya se expresó en la sección anterior, si la cantidad de estados del modelo es N , la cantidad de secuencias de estados posibles de longitud T es N^T . Analizando la ecuación (4.5) se puede ver que para cada una de las N^T secuencias de la sumatoria se deben realizar $(2T - 1)$ multiplicaciones ($T - 1$ factores correspondientes a las probabilidades de transición y T factores correspondientes a las probabilidades de las observaciones, todos multiplicando la probabilidad inicial, en total $(2T - 1)N^T$ multiplicaciones) y $N^T - 1$ sumas (ya que son N^T sumandos). Por lo tanto esta forma de calcular $P(O|\lambda)$ está en el $O(2TN^T)$ lo que la convierte en impracticable.

Para solucionar este problema debe hacerse uso de técnicas de programación dinámica, con el objetivo de recordar soluciones parciales, en vez de recalcularlas. A continuación se mostrarán un par de técnicas para ese fin, denominados *procedimiento de avance* (forward) y *procedimiento de retroceso* (backward).

4.2.1.1 Procedimiento de avance

Considérese la siguiente variable $\alpha_t(i)$, definida de la siguiente manera:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \lambda) \quad (4.6)$$

Que representa la probabilidad conjunta de obtener la secuencia de observaciones parcial (o_1, o_2, \dots, o_t) hasta el tiempo t , y que el proceso se encuentre en el estado i en el tiempo t , dado el modelo $\lambda = (A, B, \Pi)$. Los resultados de $\alpha_t(i)$, para los estados del modelo en los diferentes momentos de tiempo, se pueden calcular de forma iterativa, y luego pueden ser usados para determinar $P(O|\lambda)$. El proceso es descrito en el siguiente algoritmo.

Algoritmo 4.1 Cálculo de $P(O|\lambda)$. Procedimiento de avance.

1. Inicialización:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (4.7)$$

2. Inducción:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad t = 1, 2, \dots, T-1; \quad 1 \leq j \leq N \quad (4.8)$$

3. Terminación:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (4.9)$$

□

El primer paso (4.7), inicializa todas las potenciales probabilidades de inicio, es decir, una para cada estado (N en total), como las probabilidades conjuntas de que inicie en el estado i y dicho estado genere la observación o_1 .

El paso de inducción (4.8) es el más complejo y es el núcleo del algoritmo de avance. Para entenderlo, primero nótese que $\alpha_t(i)$ representa la probabilidad conjunta de que el proceso genere la sucesión de observaciones (o_1, o_2, \dots, o_t) , estando en el estado i en el tiempo t , luego $\alpha_t(i) a_{ij}$ representa la probabilidad de que en el tiempo $t + 1$ exista transición desde el estado i hasta el estado j , simultáneamente a la probabilidad conjunta anterior para el tiempo t . Si se suman los productos para cada uno de los N estados posibles en el instante t , y se multiplica esa suma por $b_j(o_{t+1})$, (que es la probabilidad de que el símbolo o_{t+1} sea emitido por el estado j), se obtiene $\alpha_{t+1}(j)$, que será la probabilidad conjunta de obtener la secuencia parcial de observaciones $(o_1, o_2, \dots, o_{t+1})$ y de encontrarse en el estado j en

el instante de tiempo $t + 1$. Se realiza el cálculo anterior para cada uno de los N estados y para los instantes de tiempo $t = 1, 2, \dots, T - 1$. En la figura 4.2a y 4.2b se puede apreciar una representación del paso de inducción.

El paso de Terminación (4.9) obtiene $P(O | \lambda)$ sumando el valor de las N variables $\alpha_T(i)$. Nótese que por definición,

$$\alpha_T(i) = P(o_1, o_2, \dots, o_{T-1}, o_T, q_T = i | \lambda) \tag{4.10}$$

Es decir, la probabilidad conjunta de obtener la secuencia de observaciones (o_1, o_2, \dots, o_T) y que en el tiempo T el proceso se encuentre en el estado i , si se hace extensivo eso a todos los N estados, y se calcula la suma, se tendrá precisamente $P(O | \lambda)$, obteniendo la ecuación (4.9) el mismo resultado que la ecuación (4.5).

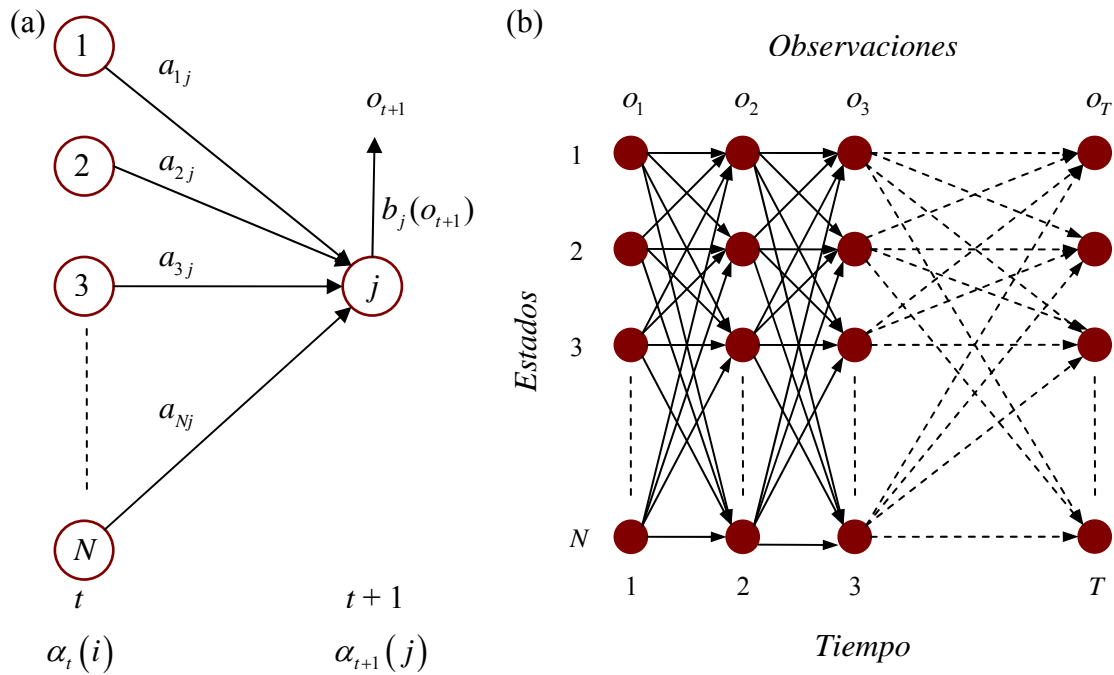


Figura 4.2: (a) Detalle del cálculo de $\alpha_{t+1}(j)$ con el procedimiento de avance y (b) representación de la implementación del cálculo del procedimiento de avance mediante un enrejado de N estados y T observaciones.

Proposición 4.1

La complejidad del algoritmo de avance es del $O(N^2T)$

Demostración: Para analizar el coste computacional de este algoritmo, se debe examinar el cálculo de $\alpha_t(i)$, $1 \leq t \leq T$, $1 \leq j \leq N$.

Primero se analizará la cantidad de multiplicaciones que se necesitan, nótese que en el paso 2, en la sumatoria, se tienen N multiplicaciones ($1 \leq i \leq N$), más una cuando al resultado de la sumatoria se le multiplica $b_j(o_{t+1})$, en total $N + 1$, y como varía t y j , ($1 \leq t \leq T - 1$), ($1 \leq j \leq N$), para el paso 2 se tienen en total $N(N + 1)(T - 1)$ multiplicaciones, a las que hay que sumarles N multiplicaciones más del paso 1.

Ahora se analizarán la cantidad de sumas necesarias, en la sumatoria del paso 2, se tienen $(N - 1)$ sumas, como varían t y j , ($1 \leq t \leq T - 1$), ($1 \leq j \leq N$), para el paso 2 se tienen en total $N(N - 1)(T - 1)$ sumas, a las que se debe sumar las $(N - 1)$ sumas del paso 3.

Este análisis muestra claramente que el algoritmo está en el $O(N^2T)$.

□

Esta complejidad representa una significativa mejora con respecto al cálculo directo visto al principio. Para entender semejante mejora, nótese que en cada momento de tiempo (como muestra la figura 4.2a), sólo se utilizan los N estados del tiempo inmediato anterior para el cálculo (cada $\alpha_t(i)$ se calcula solamente una vez); el algoritmo utiliza programación dinámica y justamente una de las propiedades de la misma es la de recordar las soluciones anteriores, necesita por supuesto más memoria que la versión directa, pero no es representativo en comparación a la cantidad de ordenes de magnitud ganado en eficiencia. Como existe un $\alpha_t(i)$ para cada estado y para cada momento de tiempo, la cantidad de memoria requerida está en el $O(NT)$.

Se debe señalar que este procedimiento presenta un grave problema de pérdida de precisión en su implementación en sistemas de cómputo debido a la tendencia a cero de los resultados de los productos para un t suficientemente grande. Este problema se tratará en la sección 4.5.

4.2.1.2 Procedimiento de retroceso

En forma análoga al procedimiento de avance, se puede considerar la variable $\beta_t(i)$ definida de la siguiente forma:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda) \quad (4.11)$$

Es decir, la probabilidad de obtener la secuencia de observación parcial $(o_{t+1}, o_{t+2}, \dots, o_T)$ desde el momento de tiempo $t + 1$ hasta el momento de tiempo T (final), dado que el estado es i en el momento de tiempo t y dado el modelo $\lambda = (A, B, \Pi)$. Luego, se pueden resolver las variables $\beta_t(i)$ nuevamente en forma inductiva y calcular el valor de $P(O | \lambda)$ a través del siguiente algoritmo.

Algoritmo 4.2 Cálculo de $P(O | \lambda)$. Procedimiento de retroceso.

1. Inicialización:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (4.12)$$

2. Inducción:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} \beta_{t+1}(j) b_j(o_{t+1}) \quad t = T-1, T-2, \dots, 1; \quad 1 \leq i \leq N \quad (4.13)$$

3. Terminación:

$$P(O | \lambda) = \sum_{i=1}^N \beta_1(i) \pi_i b_i(o_1) \quad (4.14)$$

□

El primer paso (4.12), otorga el valor 1 a las N variables $\beta_T(i)$.

El segundo paso (4.13), es el paso inductivo, nótese que $\beta_{t+1}(j)$ es la probabilidad de obtener la secuencia de observación parcial $(o_{t+2}, o_{t+3}, \dots, o_T)$ dado que el estado en el momento $t + 1$ es j y dado el modelo, la misma se multiplica por la probabilidad de transición entre el estado i y el j “ α_{ij} ” y por la probabilidad de que el estado j , en el tiempo $t + 1$, emita el símbolo “ o_{t+1} ”, o sea, “ $b_j(o_{t+1})$ ”. Obteniendo dicho producto para cada uno de los N estados posibles y sumándolos, se obtiene $\beta_t(i)$, que es precisamente la probabilidad de obtener la secuencia de observaciones parcial $(o_{t+1}, o_{t+2}, \dots, o_T)$, dado que el estado en el momento t es i , y dado el modelo. En la figura 4.3 puede ver una representación de este paso.

El paso final (4.14) obtiene lo que se quería en la dirección inversa al algoritmo anterior, nótese que en este caso a $\beta_1(i)$ se le debe multiplicar la probabilidad de comenzar en el estado i “ π_i ” y la probabilidad de que el estado i emita el símbolo o_1 “ $b_i(o_1)$ ”; esto último se consideraba en el algoritmo de avance (siguiendo la simetría) en el paso 1.

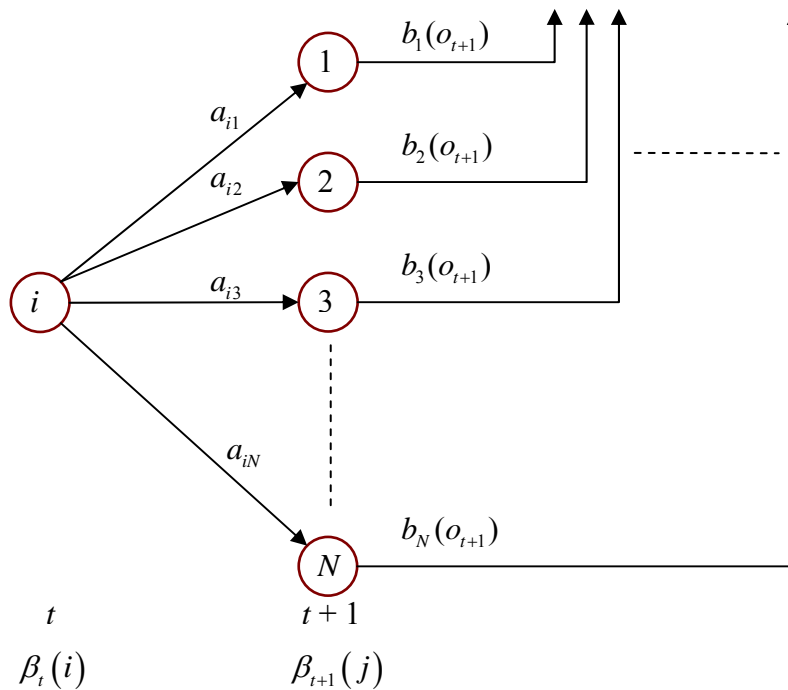


Figura 4.3: Cálculo de $\beta_t(i)$ con el procedimiento de retroceso.

El coste computacional del procedimiento de retroceso está, al igual que el procedimiento de avance, en el $O(N^2T)$, se deja al lector la demostración, que es similar al que se mostró para el caso del procedimiento de avance.

La razón para la introducción del procedimiento de retroceso es su utilidad en las soluciones a los problemas 2 y 3, que se verán en las siguientes secciones de este trabajo; al igual que el procedimiento de avance, presenta problemas de pérdida de precisión en su implementación (ver sección 4.5).

4.3 Solución al problema 2: Secuencia de estados más probable.

El problema 1 tenía una solución única y exacta, y se vio un algoritmo ineficiente y otro eficiente para encontrarla. El problema 2 se refiere a *encontrar la secuencia de estados "óptima", dada una secuencia de observaciones y un modelo*, en este caso existen varias maneras de solucionarlo; pues hay varios criterios de lo que sería *óptimo*.

Uno de esos criterios podría ser la selección de los *estados que son más probables individualmente en cada momento de tiempo*, Es decir, encontrar el estado en donde la generación de una observación concreta en el momento t , tenga la máxima probabilidad. Como implementación de este criterio, considérese la cantidad $\gamma_t(i)$ que se define como:

$$\gamma_t(i) = P(q_t = i | O, \lambda) \quad (4.15)$$

Esto es, la probabilidad de encontrarse en el estado i en el momento t , dada la secuencia de observaciones O y el modelo $\lambda = (A, B, \Pi)$.

Proposición 4.2

$$\gamma_t(i) = \frac{P(q_t = i, O | \lambda)}{\sum_{j=1}^N P(q_t = j, O | \lambda)} \quad (4.16)$$

Demostración: Ver apéndice A (A.1)

Proposición 4.3

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} \quad (4.17)$$

Demostración: Ver apéndice A (A.2)

Usando adecuadamente $\gamma_t(i)$, se puede resolver el problema antes citado, de encontrar el estado con máxima probabilidad de haber generado una observación concreta en el momento t , se llamará a dicho estado q_t^* , y será²³:

$$q_t^* = \arg \max_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T \quad (4.18)$$

La ecuación (4.18) hace que la cantidad de estados correctos obtenido sea máxima, eligiendo en cada momento t el estado más probable. Sin embargo, la secuencia de estados obtenida podría mostrar contradicciones con la matriz de transición de probabilidades, y exhibir estados contiguos no aceptados por dicha matriz.

Por ejemplo, en un HMM con probabilidad nula de transición entre algún par de estados ($a_{ij} = 0$ para algún i y algún j), podría suceder que dichos estados i y j figuren consecutivos en la secuencia óptima dada por la ecuación (4.18), siendo que esa secuencia no constituye una secuencia válida. Esto se debe a que la solución que provee la ecuación (4.18) establece solamente el estado con mayor probabilidad de generar una observación concreta en cada momento, pero nunca considera la probabilidad de la secuencia estados que está generando (es decir, puede generar una secuencia, cuya probabilidad sea nula).

Para intentar solucionar el problema anterior, se puede pensar en maximizar la cantidad de pares de estados correctos (q_t, q_{t+1}), o de triplas de estados correctos

²³ La función **arg max** encuentra el valor del argumento para el cual, la expresión dada logra su máximo valor, es decir: $\arg \max_x f(x) \in \{x \mid \forall y : f(x) \geq f(y)\}$

(q_t, q_{t+1}, q_{t+2}) , etc. Pero la solución más utilizada radica en descubrir la mejor secuencia considerando íntegramente todos los momentos de tiempo, es decir, la secuencia de estados $Q = (q_1, q_2, \dots, q_T)$ que maximiza $P(Q|O, \lambda)$. Para dicho fin, existe un procedimiento eficiente, que se basa, (al igual que en el caso del algoritmo de avance – retroceso) en técnicas de programación dinámica. Este procedimiento es el *algoritmo de Viterbi*.

4.3.1 El Algoritmo de Viterbi

Se quiere descubrir la secuencia de estados más probable $Q = (q_1, q_2, \dots, q_T)$, dada la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$ y el modelo $\lambda = (A, B, \Pi)$, para ello se considera la variable $\delta_t(i)$, que se define como:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t | \lambda) \quad (4.19)$$

Es decir, $\delta_t(i)$ es la probabilidad de la *mejor* secuencia de estados que finaliza en el estado i , teniendo en cuenta las primeras t observaciones.

Proposición 4.4

$$\delta_{t+1}(i) = \left[\max_{1 \leq j \leq N} \delta_t(j) a_{ji} \right] b_j(o_{t+1}) \quad (4.20)$$

Demostración: Ver apéndice A (A.3)

Se debe calcular $\delta_t(j)$ para cada momento de tiempo t y para cada estado j , para luego construir la secuencia de estados con un procedimiento en *retroceso* (*backtracking*), para ello se necesita recordar el argumento que maximizó la ecuación (4.20) para cada momento t y para cada estado j . Esto último se almacena

en las variables $\varphi_t(j)$. A continuación se provee la descripción completa del algoritmo de Viterbi.

Algoritmo 4.3 Cálculo de la secuencia $Q = (q_1^*, q_2^*, \dots, q_T^*)$ de estados que maximiza $P(Q|O, \lambda)$ (Algoritmo de Viterbi).

1. Iniciación:

$$\delta_1(i) = \pi_i b_i(o_1) \quad 1 \leq i \leq N \quad (4.21a)$$

$$\varphi_1(i) = 0 \quad 1 \leq i \leq N \quad (4.21b)$$

2. Inducción:

$$\delta_{t+1}(j) = \max_{1 \leq i \leq N} [\delta_t(i) a_{ij}] b_j(o_{t+1}) \quad 1 \leq t \leq T-1 \quad 1 \leq j \leq N \quad (4.22a)$$

$$\varphi_{t+1}(j) = \arg \max_{1 \leq i \leq N} [\delta_t(i) a_{ij}] \quad 1 \leq t \leq T-1 \quad 1 \leq j \leq N \quad (4.22b)$$

3. Terminación:

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (4.23)$$

4. Backtracking (construcción de la secuencia de estados):

$$q_t^* = \varphi_{t+1}(q_{t+1}^*) \quad t = T-1, T-2, \dots, 1 \quad (4.24)$$

□

Nótese que el Algoritmo de Viterbi tiene una gran analogía con el procedimiento de avance (Algoritmos de avance - retroceso), las diferencias que se pueden señalar, son: la sumatoria de las ecuaciones (4.8) y (4.9) cambiadas por la maximización en las ecuaciones (4.22a) y (4.23), y el cálculo en retroceso de la secuencia de estados (backtracking).

Proposición 4.5

La complejidad del algoritmo de Viterbi es del $O(N^2T)$

Demostración: Debe analizarse la cantidad de multiplicaciones necesarias para el cálculo de $\delta_t(j), 1 \leq t \leq T, 1 \leq j \leq N$.

En el paso 2 se tienen N multiplicaciones en $\delta_t(i)a_{ij}$ cuando se maximizan los estados ($1 \leq i \leq N$), más una cuando al resultado se le multiplica $b_j(o_{t+1})$, en total $N + 1$ multiplicaciones, y como t y j , varían ($1 \leq t \leq T - 1$), ($1 \leq j \leq N$), en total se tienen $N(N + 1)(T - 1)$ multiplicaciones para el paso 2, a las que se deben sumar N multiplicaciones más del paso 1.

Por lo que la cantidad total de multiplicaciones más asignaciones de algoritmo es de $N(N + 1)(T - 1) + N$. Las asignaciones de la ecuación (4.22b) y del paso 4 no son relevantes para el cálculo de la complejidad.

Este análisis muestra claramente que el algoritmo está en el $O(N^2T)$.

□

Ejemplo 4.1 Se presenta un ejemplo de ejecución del algoritmo de Viterbi. Supóngase un HMM que conste de 3 estados (ocultos) y tres símbolos observables (visibles), Se define el HMM $\lambda = (S, V, \Pi, A, B)$, donde:

$$S = \{1, 2, 3\}, \quad V = \{v_1, v_2, v_3\}, \quad \Pi = \begin{bmatrix} 0, 20 \\ 0, 30 \\ 0, 50 \end{bmatrix},$$

$$A = \begin{bmatrix} 0, 30 & 0, 20 & 0, 50 \\ 0, 50 & 0, 50 & 0 \\ 0, 50 & 0, 30 & 0, 20 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 0, 50 & 0, 20 & 0, 30 \\ 0, 50 & 0 & 0, 50 \\ 0, 20 & 0, 30 & 0, 50 \end{bmatrix}$$

Presúmase también que se quiere calcular la secuencia de estados con mayor probabilidad, dada la observación:

$$O = (v_3, v_1, v_3, v_2, v_2)$$

Los pasos detallados que realizaría el algoritmo de Viterbi para dicho cálculo, pueden verse a continuación:

Paso 1: Iniciación

$\delta_1(1) = \pi_1 b_1(v_3) = (0, 20)(0, 30)$	$\varphi_1(1) = 0$
$\delta_1(2) = \pi_2 b_2(v_3) = (0, 30)(0, 50)$	$\varphi_1(2) = 0$
$\delta_1(3) = \pi_3 b_3(v_3) = (0, 50)^2$	$\varphi_1(3) = 0$

Paso 2: Inducción

$\delta_2(1) = \max [\delta_1(1)a_{11}, \delta_1(2)a_{21}, \delta_1(3)a_{31}] b_1(v_1) = (0, 50)^4$	$\varphi_2(1) = 3$
$\delta_2(2) = \max [\delta_1(1)a_{12}, \delta_1(2)a_{22}, \delta_1(3)a_{32}] b_2(v_1) = (0, 30)(0, 50)^3$	$\varphi_2(2) = 2$
$\delta_2(3) = \max [\delta_1(1)a_{13}, \delta_1(2)a_{23}, \delta_1(3)a_{33}] b_3(v_1) = (0, 20)^2(0, 50)^2$	$\varphi_2(3) = 3$

$\delta_3(1) = \max [\delta_2(1)a_{11}, \delta_2(2)a_{21}, \delta_2(3)a_{31}] b_1(v_3) = (0, 30)^2(0, 50)^4$	$\varphi_3(1) = 1$
$\delta_3(2) = \max [\delta_2(1)a_{12}, \delta_2(2)a_{22}, \delta_2(3)a_{32}] b_2(v_3) = (0, 30)(0, 50)^5$	$\varphi_3(2) = 2$
$\delta_3(3) = \max [\delta_2(1)a_{13}, \delta_2(2)a_{23}, \delta_2(3)a_{33}] b_3(v_3) = (0, 50)^6$	$\varphi_3(3) = 1$

$\delta_4(1) = \max [\delta_3(1)a_{11}, \delta_3(2)a_{21}, \delta_3(3)a_{31}] b_1(v_2) = (0, 20)(0, 50)^7$	$\varphi_4(1) = 3$
$\delta_4(2) = \max [\delta_3(1)a_{12}, \delta_3(2)a_{22}, \delta_3(3)a_{32}] b_2(v_2) = 0$	$\varphi_4(2) = 3$
$\delta_4(3) = \max [\delta_3(1)a_{13}, \delta_3(2)a_{23}, \delta_3(3)a_{33}] b_3(v_2) = (0, 20)(0, 30)(0, 50)^6$	$\varphi_4(3) = 3$

$\delta_5(1) = \max \left[\underline{\delta_4(1)a_{11}}, \delta_4(2)a_{21}, \delta_4(3)a_{31} \right] b_1(v_2) = (0, 20)^2(0, 30)(0, 50)^7$	$\varphi_5(1) = 1$
$\delta_5(2) = \max \left[\delta_4(1)a_{12}, \delta_4(2)a_{22}, \delta_4(3)a_{32} \right] b_2(v_2) = 0$	$\varphi_5(2) = 1$
$\delta_5(3) = \max \left[\delta_4(1)a_{13}, \delta_4(2)a_{23}, \delta_4(3)a_{33} \right] b_3(v_2) = (0, 20)(0, 30)(0, 50)^8$	$\varphi_5(3) = 1$

Paso 3: Terminación

$$q_5^* = \arg \max_{1 \leq i \leq N} [\delta_5(i)] = 3$$

Paso 4: Backtracking

$q_4^* = \varphi_5(3) = 1$
$q_3^* = \varphi_4(1) = 3$
$q_2^* = \varphi_3(3) = 1$
$q_1^* = \varphi_2(1) = 3$

Nótese que en el paso 2 se encuentran subrayados los términos máximos. El valor del estado correspondiente en cada uno de los términos máximos (el valor de i), es el valor asignado a cada $\varphi_i(j)$. La máxima probabilidad para toda la secuencia observaciones se alcanza en $\delta_5(3)$, lo que implica que $q_5^* = 3$, y al construir en retroceso la secuencia de estados, el algoritmo obtuvo la siguiente secuencia en el paso 4:

$$Q = (3, 1, 3, 1, 3)$$

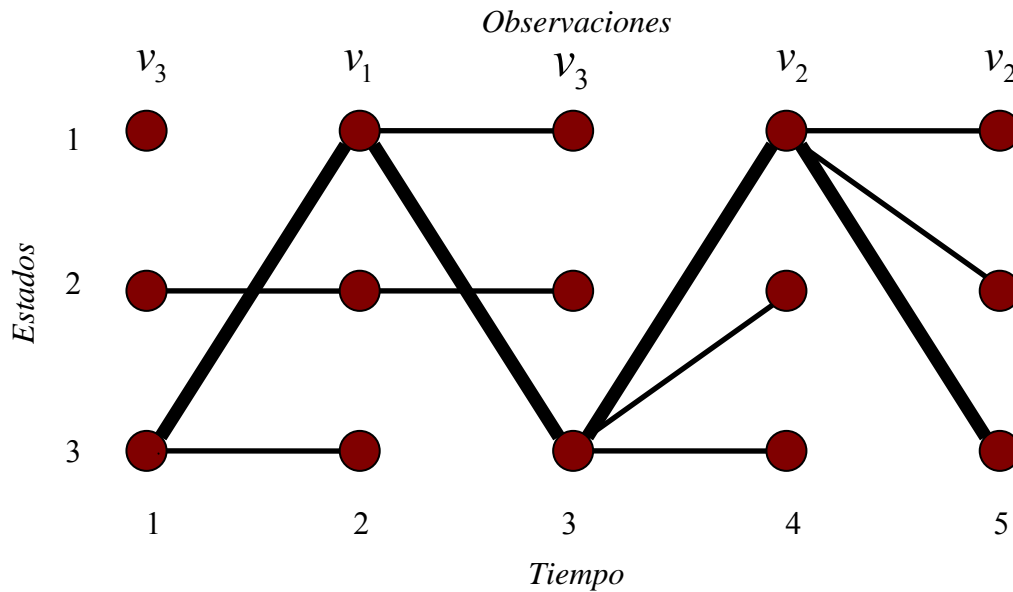


Figura 4.4: Traza de ejecución del algoritmo de Viterbi para el ejemplo 4.1

En la figura 4.4 se ha graficado con líneas para cada estado y para cada instante de tiempo, excepto el tiempo $t = 1$, de forma que se une cada estado j , con el estado del tiempo inmediato anterior que se indique de acuerdo a $\varphi_t(j)$. Por ejemplo el estado 2 del momento de tiempo 3 se une con el estado 2 del momento de tiempo 2, pues $\varphi_3(2) = 2$. La razón es que todavía no se sabe si el camino más probable, pasará por el estado 2 del momento de tiempo 3, pero si esto sucede, entonces luego pasará también por el estado 2 del momento de tiempo 2. Finalmente cuando en el paso de terminación del algoritmo se determina que el *mejor* estado en el momento 5 es el estado 3, entonces se construye en retroceso la secuencia de estados más probable, que puede visualizarse en la figura 4.4 con un trazo grueso.

4.3.2 Implementación alternativa del algoritmo de Viterbi

Como con el algoritmo de Viterbi no se persigue el objetivo de obtener una probabilidad exacta, sino una secuencia de estados, se podría aplicar logaritmos sobre los parámetros del modelo. La razón de hacerlo, es que se podría implementar

el algoritmo de Viterbi sólo con sumas, no necesitando ninguna multiplicación en el proceso. Lo que otorga dos ventajas, primero, las sumas se realizan más rápidamente en tiempo de ejecución que las multiplicaciones y segundo, se evita el problema de la pérdida de precisión a causa de las multiplicaciones, que es un problema que sucede rápido y es grave, recuérdese que se trabaja con números muy pequeños.

La eficiencia de este algoritmo es muy importante en la práctica, pues el proceso de estimación de parámetros (problema 3, próximo a tratar), se realiza inicialmente y de forma aislada, pero el problema de la secuencia de estados más probable, es central en la generalidad de las aplicaciones de los HMM.

A continuación se presenta una versión del algoritmo de Viterbi, que utiliza logaritmos sobre los parámetros del modelo con el fin antes expuesto.

Algoritmo 4.4 Cálculo de la secuencia $Q = (q_1^*, q_2^*, \dots, q_T^*)$ de estados que maximiza $P(Q|O, \lambda)$ (Algoritmo de Viterbi implementado con logaritmos y sumas).

0. Preproceso:

$$\begin{aligned}\tilde{\pi}_i &= \log(\pi_i), & 1 \leq i \leq N \\ \tilde{b}_i(o_t) &= \log[b_i(o_t)], & 1 \leq i \leq N, \quad 1 \leq t \leq T \\ \tilde{a}_{ij} &= \log(a_{ij}), & 1 \leq i, j \leq N\end{aligned}$$

1. Iniciación:

$$\begin{aligned}\tilde{\delta}_1(i) &= \log(\delta_1(i)) = \tilde{\pi}_i + \tilde{b}_i(o_1) & 1 \leq i \leq N \\ \varphi_1(i) &= 0 & 1 \leq i \leq N\end{aligned}$$

2. Inducción:

$$\begin{aligned}\tilde{\delta}_{t+1}(j) &= \log(\delta_{t+1}(j)) = \max_{1 \leq i \leq N} [\tilde{\delta}_t(i) + \tilde{a}_{ij}] + \tilde{b}_j(o_{t+1}), & 1 \leq t \leq T-1, \quad 1 \leq j \leq N \\ \varphi_{t+1}(j) &= \arg \max_{1 \leq i \leq N} [\tilde{\delta}_t(i) + \tilde{a}_{ij}] & 1 \leq t \leq T-1 \quad 1 \leq j \leq N\end{aligned}$$

3. Terminación:

$$q_T^* = \arg \max_{1 \leq i \leq N} [\tilde{\delta}_T(i)] \quad (4.23)$$

4. Backtracking (construcción de la secuencia de estados):

$$q_t^* = \varphi_{t+1}(q_{t+1}^*) \quad t = T-1, T-2, \dots, 1 \quad (4.24)$$

□

Esta implementación del algoritmo tiene una complejidad también en el $O(N^2T)$, la demostración es idéntica a la de la proposición 4.5, sólo debe cambiarse la palabra multiplicación, por suma. Aunque cabe aclarar que se debe agregar el coste del preproceso, pero dado que se realiza una sola vez, es despreciable.

4.4 Solución al problema 3: Estimación de los parámetros del modelo.

Este es el problema crucial de los HMM, por lo que se comentó en el capítulo anterior, determinar a priori las probabilidades correctas para un modelo en particular, es una tarea inviable en la mayoría de los problemas reales. De los tres problemas, es el más complicado de resolver y no se conoce una manera analítica de resolverlo, es decir, no se conoce la forma analítica de estimar un conjunto de parámetros para el modelo, que maximice la probabilidad de una secuencia de observaciones en particular, de una forma cerrada.

Por lo tanto es necesario determinar a través de una *secuencia de entrenamiento*, las probabilidades que *mejor se adaptan* al modelo, como se verá en esta sección, para dicha tarea se analizará un algoritmo, denominado algoritmo de Baum-Welch. La idea del mismo es maximizar *localmente* la probabilidad de una secuencia de observaciones, lo que consigue a través de un *procedimiento iterativo*.

4.4.1 Algoritmo de Baum-Welch.

Como ilustración, el algoritmo de Baum-Welch, intuitivamente, se podría describir de la siguiente manera. Inicialmente no se tiene conocimiento de los parámetros que mejor se ajustan a un modelo $\lambda = (A, B, \Pi)$, pero se dispone de una secuencia de observaciones (denominada secuencia de entrenamiento), que puede utilizarse para maximizar $P(O|\lambda)$. Como primer paso, se utilizan distribuciones de probabilidad preseleccionadas o se las eligen aleatoriamente²⁴, entonces se identificarán las transiciones y símbolos de observación más probables. Se incrementarán las probabilidades de dichas transiciones y símbolos, y de esta forma se tendrá un nuevo modelo revisado y *mejor* que el anterior, es decir, con una mayor probabilidad de haber generado la secuencia de entrenamiento dada. Este proceso, denominado *proceso de entrenamiento*, se repite hasta que el modelo generado por el mismo, no difiera del generado en el paso anterior, o sea, hasta que no pueda *mejorarse* el modelo.

Pasando a una descripción formal, primero se define $\xi_t(i, j)$, que será la probabilidad conjunta de que el proceso se encuentre en estado i , en el momento t , y que se produzca una transición hacia el estado j en el momento $t + 1$, dada la secuencia de observaciones O y dado el modelo $\lambda = (A, B, \Pi)$.

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda) \quad (4.25)$$

En la figura 4.5 se muestran esquemáticamente todos los caminos que satisfacen la ecuación anterior (4.25). De las definiciones de $\alpha_t(i)$ y $\beta_t(i)$ de la sección 4.2.1, se llega a la siguiente proposición:

Proposición 4.6

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^N \sum_{l=1}^N \alpha_t(k) a_{kl} b_l(o_{t+1}) \beta_{t+1}(l)} \quad (4.26)$$

²⁴ Más adelante en este capítulo se verá que *es conveniente* que algunas distribuciones iniciales para la estimación no sean aleatorias y también las opciones para conseguirlo.

Demostración: Ver apéndice A (A.4)

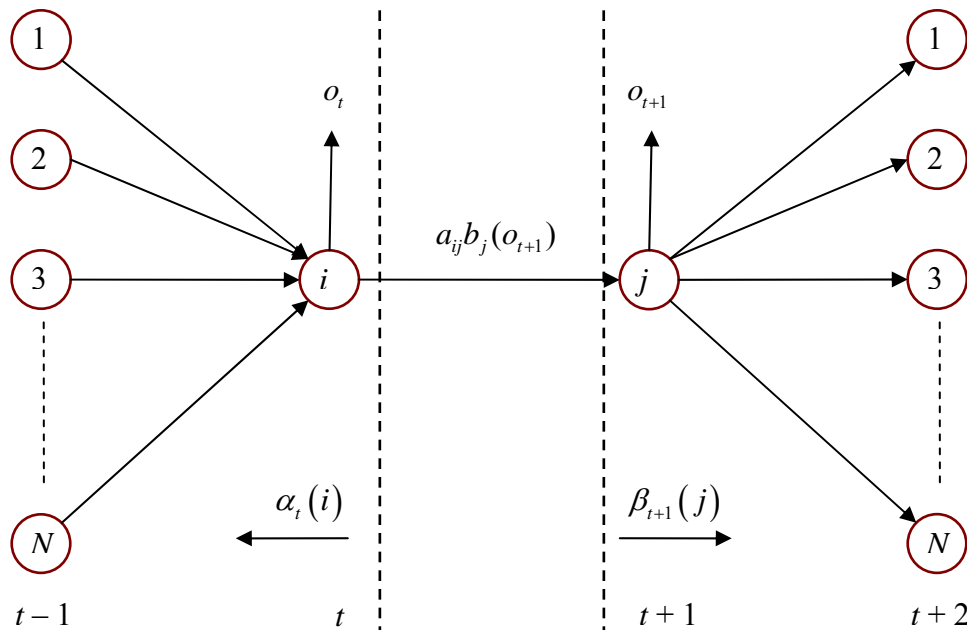


Figura 4.5: Todos los caminos correspondientes a la probabilidad conjunta de estar en el estado i en el momento t y luego estar en el estado j en el momento $t + 1$

Con antelación, en la sección 4.3 (ecuación 4.15), se definió $\gamma_t(i)$, como la probabilidad de que el proceso se encuentre en el estado i en el momento t , dada una secuencia de observaciones y dado el modelo. Razón por la cual, es posible relacionar $\gamma_t(i)$ con $\xi_t(i, j)$, notar que la diferencia, es que la última requiere que el proceso se encuentre en el estado j en el momento $t + 1$.

Proposición 4.7

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \tag{4.27}$$

Demostración: Ver apéndice A (A.5)

Si se suma $\gamma_t(i)$ a través del tiempo (desde $t = 1$ hasta $t = T - 1$), se obtendrá el *número esperado de veces que el proceso efectúa transiciones desde el estado i* , y si se efectúa el mismo procedimiento con $\xi_t(i, j)$, se obtiene el *número esperado de veces que el proceso produce transiciones desde el estado i al estado j* . O sea:

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{Número esperado de transiciones desde el estado } i \text{ para} \quad (4.28)$$

una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{Número esperado de transiciones desde el estado } i \quad (4.29)$$

hacia el estado j para $O = (o_1, o_2, \dots, o_T)$

Utilizando las ecuaciones (4.28) y (4.29), y la definición de $\gamma_t(i)$, se puede obtener un método para lograr la reestimación de los parámetros de un HMM. Sea $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\Pi})$ el modelo resultante de la reestimación de los parámetros, con $\bar{A} = \{\bar{a}_{ij}\} \quad 1 \leq i, j \leq N$, $\bar{B} = \{\bar{b}_j(k)\} \quad 1 \leq j \leq N, 1 \leq k \leq M$ y $\bar{\Pi} = \{\bar{\pi}_i\} \quad 1 \leq i \leq N$, entonces se pueden utilizar las siguientes fórmulas de reestimación:

$$\bar{\pi}_i = \text{cantidad de veces que el proceso se encuentra en} \quad = \gamma_1(i) \quad (4.30)$$

el estado i al momento ($t = 1$).

$$\bar{a}_{ij} = \text{número esperado de transiciones desde el estado } i \quad = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (4.31)$$

al j , sobre el número esperado de transiciones desde el estado i .

$$\bar{b}_j(k) = \text{número esperado de veces en las que el} \quad = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j) \text{ tal que } o_t = v_k} \quad (4.32)$$

proceso se encuentra en el estado j , observando el símbolo v_k sobre el número de veces en las que el proceso se encuentra en el estado j .

Si el modelo anterior a la aplicación de la reestimación es $\lambda = (A, B, \Pi)$, entonces Baum demostró (según [Rab89, RJ93]) que, o bien $\lambda = (A, B, \Pi)$ es un punto crítico de la función de probabilidad, y se verifica que $\bar{\lambda} = \lambda$, o bien, el modelo $\bar{\lambda}$ tiene una probabilidad mayor que λ donde la mayor probabilidad se refiere a que se verifica la desigualdad $P(O|\bar{\lambda}) > P(O|\lambda)$, lo que significa que se ha encontrado un modelo $\bar{\lambda}$, que posee mayor probabilidad de generar la secuencia de observaciones O . Esto asegura que si se repite el proceso tomando en cada paso el nuevo modelo como la base para el posterior cálculo, se producirá un paulatino acercamiento al punto crítico señalado. Lo anterior es la plataforma para el siguiente algoritmo.

Algoritmo 4.5 Estimación de los parámetros de un HMM (Baum-Welch)

Paso 1: Elegir distribuciones de probabilidad A, B, Π , tal que $\lambda = (A, B, \Pi)$.

Paso 2: Calcular $\bar{\pi}_i = \gamma_1(i)$, $\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$ y $\bar{b}_j(k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$ tal que $o_t = v_k$ $1 \leq i, j \leq N$, $1 \leq k \leq M$

Paso 3: Hacer $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\Pi})$.

Paso 4: Si $\bar{\lambda} = \lambda$ terminar y devolver $\bar{\lambda}$, sino hacer $\lambda = \bar{\lambda}$ y volver al paso 2.

□

4.4.2 Limitaciones del Algoritmo de Baum-Welch

A continuación se describirán algunas limitaciones del algoritmo de Baum-Welch al momento de la implementación del mismo, cuyas posibles soluciones se exhibirán en la sección 4.5.

El método de estimación de Baum-Welch para los parámetros de un HMM, tiene una fuerte dependencia de la *estimación inicial* de dichos parámetros, resulta que una de sus limitaciones, es que el algoritmo, de hecho, no asegura que se encuentre el mejor modelo si la iniciación de los parámetros no es la adecuada. Más precisamente, una iniciación aleatoria, podría causar que el algoritmo encontrara un *máximo local* de la función de probabilidad y que éste se encuentre lejos del *máximo global* buscado. Se debe tener en cuenta que en los problemas reales, con frecuencia, la función de probabilidad es no lineal, compleja y con muchos máximos locales, por lo tanto una buena iniciación de los parámetros es deseable para el éxito de este procedimiento.

Otra limitación de este método es su rápida pérdida de precisión al manejar números cada vez más pequeños, y en este caso, a diferencia del algoritmo de Viterbi, debido a las sumas involucradas en los cálculos, no es posible el uso de logaritmos de manera directa para solucionar este problema. Tan solo el cálculo de $\alpha_t(i)$ en la resolución de $\xi_t(i, j)$, implica la sumatoria de largos productos que rápidamente tienden a cero, y para un t suficientemente grande, la computación excedería la precisión de cualquier sistema de computo.

Finalmente se señalará como una limitación destacable, la necesidad del método de contar con suficientes datos de entrenamiento, ya que en caso contrario, se atentaría contra la buena estimación de los parámetros, y por ende, contra la eficiencia del procedimiento. Es importante señalar, que la secuencia de observaciones es *finita* por naturaleza, pero debe ser lo suficientemente extensa para que pueda realizarse una adecuada estimación de los parámetros del modelo. Como ejemplo, piénsese en una secuencia de observaciones reducida que no cuente con ninguna observación de un símbolo en particular v_k , entonces $b_j(k) = 0$ para todo estado j y las sucesivas reestimaciones del algoritmo no cambiarán este resultado.

4.5 Problemas de implementación de los HMM

En esta sección se tratarán las posibles soluciones a algunos de los problemas o limitaciones más importantes de los algoritmos vistos para la solución de los problemas fundamentales de los HMM.

Estos problemas surgen en la implementación de los mismos. Los conceptos vertidos en esta sección no son parte de la teoría básica de los HMM. Cabe aclarar que en algunos casos, las soluciones provistas, serán las que *empíricamente* han demostrado su eficacia.

4.5.1 Estimación inicial de los parámetros del HMM

Como se comentó en la sección 4.4.2, el método de reestimación (algoritmo de Baum-Welch), asegura solamente que se encontrará un *máximo local* de la función de probabilidad. Se necesita entonces elegir adecuadamente los parámetros para conseguir que dicho máximo local sea igual al máximo global o se acerque lo más posible. Una posible solución es intentar que la inicialización de los parámetros del modelo se produzca dentro del rango de parámetros cercanos a dicho máximo global.

Según [RJ93, Rab89], no existe una respuesta exacta e indiscutible para este problema, pero la experiencia empírica permite dar un par de lineamientos:

- Los parámetros de A y Π se pueden inicializar de manera *aleatoria* (respetando las restricciones estocásticas²⁵, y siendo distintos de cero) o de manera *uniforme*, esto produce una reestimación aceptable en la mayoría de los casos.
- Para los parámetros de B , se necesita una buena estimación inicial de los parámetros para obtener resultados aceptables. Esta estimación puede lograrse mediante varias maneras, algunas de ellas son:

²⁵ Restricciones estocásticas a saber: $\pi_i \geq 0$, $\sum_{i=1}^N \pi_i = 1$, $a_{ij} \geq 0$, $\sum_{j=1}^N a_{ij} = 1$, $\forall i, j$, $1 \leq i, j \leq N$

- a) Una segmentación manual de la secuencia de observaciones entre los estados, para luego promediar la ocurrencia de las mismas en dichos estados.
- b) Segmentación de máxima probabilidad de las observaciones, con promediado luego.
- c) Segmentación de la secuencia de observaciones mediante intervalos k proporcionales, usando el algoritmo *segmental K-means* [Rab89, RJ93] (ver algoritmo 4.6).

Algoritmo 4.6 Estimación *inicial* de los parámetros de B (algoritmo segmental K-means) para su reestimación posterior con el algoritmo Baum-Welch.

Paso 1: Inicializar el modelo $\lambda = (A, B, \Pi)$ de manera aleatoria o en forma manual.

Paso 2: Utilizar el algoritmo de Viterbi para obtener la secuencia de estados óptima dado el modelo $\lambda = (A, B, \Pi)$ y la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$.

Paso 3: $\hat{b}_j(k) = \frac{\text{cantidad de símbolos } v_k \text{ observados en el estado } j}{\text{cantidad de símbolos observados en el estado } j}$ $1 \leq j \leq N, 1 \leq k \leq M$

Nótese que la cantidad total de símbolos observados en el estado j es igual a la cantidad de ocurrencias de dicho estado en la secuencia devuelta por el algoritmo de Viterbi, y que la secuencia de observaciones utilizada es la misma que luego se usará en el proceso de reestimación de los parámetros a través del algoritmo de Baum-Welch.

□

En conclusión, cuanto mejor se puedan inicializar los parámetros, incluso los parámetros de A y Π , mejor será el resultado que se obtenga mediante el procedimiento de reestimación.

4.5.2 Múltiples secuencias de observaciones.

Puede suceder que se cuente con un HMM del tipo de izquierda a derecha (left-rigth) en donde el uso de una sola secuencia de observaciones para la estimación de los parámetros del modelo sea inadecuado, pues como no se retorna a un estado ya visitado, el conjunto de observaciones por estado puede ser limitado, por ende se necesitan varias observaciones para producir una estimación aceptable. Otro ejemplo son HMM en donde exista un probabilidad similar de inicio en muchos de sus estados, (o en todos), nótese que una sola secuencia de observaciones, indefectiblemente le otorgaría toda la probabilidad inicial al estado que eventualmente resultara en esa posición en dicha secuencia, pero esto no sería correcto.

Una manera de encarar este problema [Rab89, RJ93], es la utilización de varias secuencias de observaciones para la estimación, de esta forma se define:

$$O = \left[O^{(1)}, O^{(2)}, \dots, O^{(M)} \right] \quad (4.33)$$

Donde $O^{(k)} = (o_1^{(k)}, o_2^{(k)}, \dots, o_{T_k}^{(k)})$ es la secuencia de observaciones k-ésima. Con la hipótesis de que las secuencias de observaciones son independientes entre sí, para la estimación buscada, se debe encontrar el modelo $\lambda = (A, B, \Pi)$ que maximice:

$$P(O | \lambda) = \prod_{k=1}^M P(O^{(k)} | \lambda) \stackrel{def}{=} \prod_{k=1}^M P_k \quad (4.34)$$

Como la estimación de Baum-Welch se basa en frecuencias, sumando todas las frecuencias individuales y dividiéndolas por P_k se pueden modificar las fórmulas (4.30), (4.31) y (4.32), para el caso de múltiples secuencias de observaciones, como sigue:

$$\bar{\pi}_i = \sum_{k=1}^M \frac{1}{P_k} \alpha_1^k(i) \beta_1^k(i) \quad (4.35)$$

$$\bar{a}_{ij} = \frac{\sum_{k=1}^M \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i) a_{ij} b_j(o_{t+1}^{(k)}) \beta_{t+1}^k(j)}{\sum_{k=1}^M \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i) \beta_t^k(i)} \quad (4.36)$$

$$\bar{b}_j(r) = \frac{\sum_{k=1}^M \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(j) \beta_t^k(j)}{\sum_{k=1}^M \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(j) \beta_t^k(j)} \quad (4.37)$$

Es importante señalar que las probabilidades iniciales (fórmula (4.35)), pueden no estimarse cuando se conoce en forma fehaciente por cual estado inicia el modelo, como es el caso de los HMM con topología de izquierda a derecha, o también en los casos en los que existe la misma probabilidad de comenzar en cualquier estado, en este caso sería uniforme y no sería necesario estimar las probabilidades iniciales de los estados. Nótese que esto sólo puede suceder con las distribuciones de probabilidad iniciales de los estados, las demás deben ser estimadas.

4.5.3 Uso de coeficientes de escalado

El problema de la pérdida de precisión es un problema grave y ocurre rápidamente, todos los algoritmos vistos en este capítulo están afectados por este problema, con excepción de la implementación alternativa del algoritmo de Viterbi (ya que usa logaritmos para solucionarla), pero la aplicación directa de logaritmos no es posible para el caso de los algoritmos de avance, retroceso y Baum-Welch.

Una solución a este problema es el uso de coeficientes de escalado, la idea básica es la de intervenir en los cálculos de productos de los algoritmos, en cada iteración de los mismos, manteniéndolos dentro del rango de punto flotante para evitar exceder la precisión del sistema de cómputo usado. Al final de cada iteración se *cancelan*, pues se usan por igual en numerador y denominador.

Como ejemplo, considérese el cálculo de $\alpha_t(i)$ y $\beta_t(i)$, necesarios para los tres procedimientos citados, éstos se podrían multiplicar en cada iteración por un coeficiente de escalado que sea *independiente de i* , pero *dependiente de t* . Donde al decir “dependiente de t ”, intuitivamente se refiere al hecho que mientras mayor sea t , mayor será el coeficiente de escalado para mantener el cálculo dentro de la precisión del sistema de cómputo. En [Rab89, RJ93] se puede encontrar una excelente descripción detallada de este método.

4.5.4 Insuficientes datos para el entrenamiento

En el caso del procedimiento de reestimación (algoritmo de Baum-Welch), un problema que afecta al buen entrenamiento de los parámetros de un HMM, es una insuficiente cantidad de datos, es decir una secuencia de observaciones de entrenamiento limitada.

Se pueden citar algunas posibles soluciones a este problema:

- *Incrementar el tamaño de la secuencia de observaciones de entrenamiento.* Esto no es posible en todos los casos.
- *Reducir el tamaño del modelo* (reducir el número de estados, el número de símbolos de observaciones, el número de transiciones, etc.), esto siempre es posible hacerlo, pero existen casos en los que motivos fundamentales que intervinieron en el diseño de la topología del modelo, harían que el mismo se desvirtuará si se modificara, en estos casos la reducción no será posible.
- *Interpolar con otro conjunto de parámetros de un modelo “reducido”, para el cual la secuencia de entrenamiento si resulta adecuada.* Se pueden diseñar ambos al mismo tiempo, el modelo normal $\lambda = (A, B, \Pi)$, y un modelo reducido $\lambda' = (A', B', \Pi')$, que se elige convenientemente a partir del modelo original para lograr que la secuencia de entrenamiento resulte adecuada para él. Luego el modelo interpolado $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\Pi})$ es obtenido con la siguiente fórmula:

$$\bar{\lambda} = e\lambda + (1-e)\lambda' \quad (4.38)$$

Donde e es un valor entre 0 y 1 y representa el peso de los parámetros del modelo original y $(1 - e)$ representa el peso del modelo reducido, de esta forma, si la secuencia de observaciones es bastante completa, e tenderá a 1.0, y si la secuencia de entrenamiento es pequeña e insuficiente, e tenderá a 0.0.

- *Añadir restricciones adicionales a los parámetros del modelo.* Lo cual aseguraría que ningún parámetro sea estimado por debajo de un cierto *nivel* especificado. Como ejemplo, considérese la observación del símbolo v_k y su correspondiente probabilidad de observación para el estado j , $b_j(k)$, entonces δ_{b_j} podría ser un nivel mínimo estipulado y la estimación de $b_j(k)$ se haría teniendo en cuenta la siguiente regla:

$$b_j(k) = \begin{cases} b_j(k), & \text{if } b_j(k) \geq \delta_{b_j} \\ \delta_{b_j}, & \text{caso contrario.} \end{cases} \quad (4.39)$$

4.5.5 Elección del modelo

Uno de los aspectos críticos, es una adecuada elección del modelo. Donde por elección de modelo, se refiere a la elección de la *topología de modelo* (que hace referencia a la forma de las transiciones propias de la cadena de Markov oculta), un ejemplo sería de izquierda a derecha (*left-right*), que se mostró en la figura 3.5 del capítulo anterior, también se refiere a la elección del *tamaño del modelo* (que refiere a la cantidad de estados del mismo) y a la elección de los *símbolos de observación*, que serán posibles de observar con el modelo.

El problema, es que de acuerdo a [Rab89, RJ93], no existe una manera teóricamente correcta de realizar éstas elecciones. En general dependerá del tipo de señal (referirse a la introducción del capítulo 3 - sección 3.1 -) que se esté modelando.

En el capítulo 7 se verá como se eligió la topología, el tamaño y los símbolos de observación del modelo para un caso de estudio en particular, y se podrá apreciar la absoluta dependencia de éstas elecciones de la señal específica que se esté modelando, y del patrón específico que se quiera detectar.

Capítulo 5

ANÁLISIS TÉCNICO BURSÁTIL

“Puedo calcular el movimiento de los cuerpos celestes, pero no la locura de las personas”.

Isaac Newton.

5.1 Introducción

Los analistas de Mercados Financieros utilizan variadas herramientas en la tarea de comprender el comportamiento del mercado y *preveer su futuro*, se puede agrupar dichas herramientas en dos grupos: las que pertenecen al *análisis técnico*²⁶ y las que se corresponden con el *análisis fundamental*. [AP97].

En este capítulo se analizará en detalle la filosofía y fundamento del análisis técnico, introduciendo todos los conceptos necesarios (gráficos, tendencia, medias móviles, etc.) para la total comprensión de los patrones del análisis técnico, ya que su entendimiento es central para este trabajo.

5.1.1 Análisis técnico vs. análisis fundamental

Como una breve descripción del análisis fundamental (en el ámbito específico del análisis de acciones cotizantes de una empresa) se dirá que un analista fundamental recurre a los estados financieros de la empresa, coteja los datos del sector económico a la que pertenece la misma, realiza estudios de mercado, de la economía en general, socio-políticos, etc. Pero el objetivo principal es conocer con detalle el funcionamiento de la empresa, el rumbo que sus directivos le darán y sus

²⁶ En este capítulo se referirá al *análisis técnico bursátil* sencillamente como *análisis técnico*.

posibilidades de crecimiento en su sector y en el mercado en general, con el propósito de determinar el *valor intrínseco*²⁷ de la misma, que puede diferir sustancialmente del *valor de mercado*²⁸.

Es justamente en base a esa diferencia que un analista fundamental toma sus decisiones de compra o venta. Si el valor intrínseco es superior al de mercado será una señal de compra y si por el contrario, el valor intrínseco es inferior al de mercado será una señal de venta.

Es claro que el análisis fundamental requiere de muchos datos (a veces no tan fáciles de conseguir), de un tiempo considerable de análisis y de un conocimiento específico de muchos aspectos (financieros, económicos), pero tal vez lo más difícil, es que se debe poseer cierto conocimiento del funcionamiento de la empresa, pues un inversor que utilice este enfoque para sus inversiones persigue la idea de que “al poseer acciones posee negocios, no pedazos de papel” [Hag98].

La idea de adquirir acciones de una empresa sin conocer las funciones operativas de la misma (lo que incluye productos y servicios, las relaciones laborales, los costos de materia prima, los bienes muebles e inmuebles, las necesidades de inversión y reinversión, las cuentas por cobrar, las necesidades de caja de una empresa, etc.) no tiene sentido para un inversor con un enfoque fundamental.

El enfoque fundamental y el enfoque técnico (que se verá en detalle en este capítulo) tienen acérrimos cultores, muchas veces enfrentados entre sí. El autor de esta tesis al igual que muchos analistas, piensa que se pueden aunar los dos enfoques para obtener mejores resultados, esto es justamente la forma en que proceden las grandes compañías de inversión.

Este trabajo se encuadra en el enfoque técnico y no se harán mayores referencias al análisis fundamental más allá de esta sección, el lector que quiera ampliar sobre la materia, puede encontrar en [Hag98] un excelente libro que trata en detalle las inversiones de *Warren Buffet*, el inversor bursátil más exitoso de la historia y uno de los hombres más ricos del mundo, que usó y usa un enfoque absolutamente fundamental. En [Sat05] puede encontrarse una guía detallada este análisis.

“El análisis técnico es el estudio de los movimientos de mercado²⁹, principalmente mediante el uso de *gráficos*, con el propósito de pronosticar las futuras *tendencias*³⁰”

²⁷ Valor que el analista fundamental considera que vale la empresa luego de su análisis (concepto desarrollado por Benjamín Graham y David Dodd en su libro *Security analysis* (1934)) [Hug98].

²⁸ El valor de mercado de una empresa es el producto entre el valor de cotización de sus acciones y la cantidad de acciones de la sociedad.

de precios” [Mur00], se basa en que el mismo mercado proporciona la mejor información para predecir su futura evolución. [AP97]

Mientras que el análisis fundamental estudia la *causa* del movimiento, el análisis técnico estudia el *efecto*.

A diferencia del análisis fundamental que sólo tiene sentido para analizar el comportamiento en el largo plazo, el análisis técnico se adapta a cualquier dimensión de tiempo (corto, mediano y largo plazo); algunos consideran que el análisis técnico debe usarse para el mediano y corto plazo, dejando el largo plazo para el análisis fundamental, pero en [Mur00] se expresa lo contrario. En la práctica, el análisis técnico ha dado en el largo plazo resultados similares que en el mediano y corto plazo.

Asimismo, debido a la cantidad de datos necesarios para el análisis fundamental, sus cultores deben especializarse; en cambio el análisis técnico puede aplicarse a casi cualquier área de operaciones en múltiples instrumentos financieros y participar en varios mercados simultáneamente. [Mur00]. Los principios del análisis técnico pueden aplicarse a opciones³¹, índices, futuros, divisas, mercaderías, tipos de interés, etc., aunque cada instrumento tiene sus peculiaridades, por ejemplo en el caso de las opciones habría que tener en cuenta además, el precio del activo subyacente, la fecha de vencimiento, el precio de ejercicio³² y el tipo (compra o venta).

Los analistas técnicos buscan extraer de los gráficos creados a partir de series temporales de precios y volúmenes, *patrones* (ver sección 5.7) que se repiten a través del tiempo. Es precisamente en base a la aparición de dichos patrones que el analista descubre indicios de cambio o continuación de *tendencias*, lo que deriva en señales de compra o venta.

En esta tesis se trabajará con el análisis de acciones cotizantes de empresas, pero como los patrones se pueden verificar en las series temporales de cualquier instrumento financiero, se puede hacer fácilmente extensible. El lector que desee ampliar conocimientos sobre opciones puede encontrar una muy buena guía en [For94] y en [GF00].

²⁹ “Movimientos de mercado” incluye precio, volumen e interés abierto (este último sólo se utiliza en futuros y opciones, y representa la cantidad de opciones y/o contratos de futuros pendientes en un determinado periodo)

³⁰ Ver sección 5.4

³¹ “Una opción es un acuerdo entre un comprador (propietario o tenedor) y un vendedor (emisor) que, tras el pago de una retribución, da al comprador el derecho, pero no la obligación de comprar (call) o vender (put) un activo en una fecha determinada o antes de ella”. [For94]

³² Precio fijado en la opción y al cual ésta debe ser ejercida, cabe recordar que una opción puede ejercerse o no (“derecho, pero no obligación”).

La presencia de patrones en las series temporales financieras que se repiten en el tiempo no infiere que sean estacionarias, sólo que eventos positivos o negativos influyen de manera *parecida* sobre la misma. (Es decir: cuando cierto patrón se está desarrollando en una serie, puede inferirse que algo positivo o negativo está sucediendo). Esto último va en contra de la *hipótesis del mercado eficiente* (ver sección 5.2).

John Murphy en [Mur00] establece una subdivisión entre los analistas técnicos, el *chartista tradicional* y el *analista técnico estadístico*. Para el primero los gráficos (ver sección 5.3) son su herramienta principal y todo lo demás es secundario; por el contrario, el segundo toma los principios del análisis técnico, los cuantifica, prueba y optimiza, con el propósito de desarrollar *sistemas mecánicos* de análisis. A su vez, los *técnicos informáticos* se subdividen en aquellos que desarrollan sistemas totalmente mecánicos (con el enfoque de *caja negra*), y aquellos que utilizan la tecnología informática para mejorar o desarrollar indicadores técnicos, manteniendo el control de la interpretación de dichos indicadores, así como el proceso de toma de decisiones.

Este trabajo encuadraría en este último rubro, ya que no se pretende (ni se desea) que automáticamente se ejecuten decisiones de compra o de venta, sino que se contribuya a la detección de patrones que ayuden al analista, junto con otros análisis paralelos, en el proceso de toma de decisiones. La idea es *agilizar* el proceso de análisis mediante la automatización de la detección de patrones, no automatizar la toma de decisiones.

5.1.2 Historia del análisis técnico

Hay consenso entre la mayoría de los técnicos y estudiosos de los mercados de occidente en considerar que lo que hoy se denomina análisis técnico tiene sus orígenes en las teorías propuestas por primera vez por Charles Dow³³ a principios del siglo XX. [Mur00] La primera publicación de un índice bursátil elaborado por Dow apareció el 3 de julio de 1884, dicho índice consistía de una media de 11

³³ Charles Dow y su socio Edward Jones fundaron la Dow Jones & Co. en 1882, el índice industrial que Dow y Jones introdujeron en 1884, y que desde 1928 incluye 30 compañías, es hasta hoy el índice bursátil más importante del mundo.

valores [AP97]. Sin embargo se pueden encontrar referencias de la existencia de análisis técnico mucho antes, guiándonos por [Nis96] se descubre que los japoneses usaron análisis técnico para operar en uno de los primeros mercados de futuros del mundo, que fue el del arroz a comienzos del siglo XVIII. El mercado de futuros del arroz trajo aparejado una gran especulación, y a partir de ésta nació el análisis técnico japonés.

Uno de los comerciantes japoneses más importantes de aquella época descubrió que aunque existía una relación entre el valor de mercado y las fuerzas de la oferta y la demanda, los mercados estaban fuertemente influenciados por las *emociones* de los comerciantes, por lo que llegó a la conclusión que el estudio de las emociones del mercado podía ayudar a predecir los precios, aún más, se dio cuenta que el valor real y el precio de mercado del arroz, podían ser diferentes. Lo que aquel comerciante descubrió, es que el *aspecto psicológico del mercado* era esencial para su éxito comercial. Dicho descubrimiento es válido y aplicable hoy en día a cualquier instrumento financiero (acciones, bonos, divisas, etc.).

Volviendo a los comienzos del análisis técnico en occidente, es importante señalar que dicho análisis jugó un rol fundamental en esos primeros años, ya que las empresas no tenían la obligación de hoy en día de dar precisiones sobre sus estados financieros, por lo tanto era el único análisis posible en aquellos tiempos [AP97].

5.2 Filosofía y fundamento

5.2.1 Hipótesis del análisis técnico

El análisis técnico presupone tres hipótesis fundamentales [Mur00, AP97]:

- Los movimientos de mercado descuentan toda la información (disponible o no).
- Los precios se mueven siguiendo tendencias.
- Lo que ha ocurrido en el mercado antes, ocurrirá en el futuro (la historia se repite).

5.2.1.1 Los movimientos del mercado descuentan toda información

Esta hipótesis se puede considerar como la piedra angular del análisis técnico. El análisis técnico presupone que *cualquier cosa* que pueda afectar al precio (razones fundamentales, políticas, psicológicas, etc.) se refleja realmente en el precio de mercado, esto lleva consecuentemente a la conclusión que basta con realizar un estudio sobre los movimientos de precio.

Lo que esta hipótesis significa es que los movimientos de precio reflejan los *cambios entre la oferta y la demanda*. Esto es, si la oferta supera a la demanda, los precios deberían bajar y si por el contrario la demanda supera a la oferta, los precios deberían subir. Lo último es la base de cualquier pronóstico económico (incluso el fundamental), el análisis técnico invierte la carga de la prueba, es decir, si los precios están subiendo, entonces la demanda debe superar a la oferta y los fundamentos (conocidos o no) deben ser positivos. Por el contrario, si los precios bajan, entonces debe ser que la oferta supera a la demanda y los fundamentos deben ser negativos; pero al analista técnico no le interesan dichos fundamentos (al menos de forma directa), sabe que la causa de los mercados alcistas o bajistas está en la fuerza subyacente de la oferta y la demanda, pero con la hipótesis de que el mercado lo descuenta todo, intenta descubrir los cambios a través del mismo mercado. Como ya se dijo, se interesa por el efecto y no por la causa.

Considerando como válida esta hipótesis, se encuentra aquí un punto a favor del análisis técnico sobre el fundamental, ya que en los comienzos de un movimiento alcista o bajista, en general, los fundamentos son poco o nada conocidos, y cuando llegan a la luz, el movimiento ya ha sido significativo [Mur00].

5.2.1.2 Los precios se mueven siguiendo tendencias.

El concepto de *tendencia* (se verá en detalle en la sección 5.4) es primordial para el análisis técnico. Como se vio al comienzo de este capítulo, el propósito del análisis técnico es identificar a través de la representación gráfica, tendencias en las series temporales para obrar en consecuencia y obtener beneficios de la captación temprana de la gestación de dichas tendencias.

Se presupone que los precios se mueven por tendencias, de esta forma pueden seguir una tendencia alcista, bajista o estable, lo que lleva a la necesidad de estudiar las posibles señales o patrones que determinen la continuación o el cambio de una tendencia.

5.2.1.3 La historia se repite.

El análisis técnico y en general el estudio de los movimientos del mercado está en estrecha relación con el estudio de la psicología humana aplicada al mercado. Los patrones en los gráficos que se han descubierto y clasificado en mas de cien años, muestran que estos patrones que aparecen en los gráficos de precios, son reveladores de la psicología alcista o bajista del mercado, y teniendo como base que dichos patrones funcionaron bien en el pasado, se asume que lo seguirán haciendo en el futuro. [Mur00]

Esto es así, pues la psicología humana de hecho, tiene tendencia a no cambiar; por lo tanto si el comportamiento de las personas en el pasado dio como resultado un movimiento alcista o bajista, con un especial patrón registrado en los gráficos, se asume que si dicho patrón está volviendo a generarse en un gráfico, entonces el movimiento que en el pasado generó dicho patrón debería volver a suceder. Es decir, para discernir el futuro es necesario estudiar el pasado.

5.2.2 Hipótesis del Mercado Eficiente

De las críticas realizadas hacia el análisis técnico, la más seria es la que se tratará en esta sección, denominada la *hipótesis del mercado eficiente*.

La hipótesis del mercado eficiente fue desarrollada por Fama³⁴ en 1965 [Pfi00], en ella se asume que el precio actual de mercado contiene *toda la información disponible* o dicho de otro modo, ningún inversor puede beneficiarse de ganancias *anormales* en base al análisis de patrones de precio pasados o cualquier otra estrategia.

³⁴ Eugene F. Fama desarrolló la hipótesis del mercado eficiente (EMH por sus siglas en inglés) en 1965 como su tesis de doctorado en economía en la Universidad de Chicago, también denominada la teoría del paseo aleatorio o “random walk theory”

Como una breve descripción técnica, esta tesis especifica que los precios fluctúan de forma aleatoria alrededor de su valor intrínseco, siguiendo lo que se conoce como caminata o paseo aleatorio (Random Walk), que no es otra cosa que una cadena de Markov (ver capítulo 3).

Esta tesis infiere que la mejor estrategia es simplemente comprar y mantener, en lugar de cualquier intento de predicción.

Si la hipótesis fuera cierta, ningún análisis, ni técnico, ni fundamental, tendrían sentido. La opinión actual de la comunidad financiera y científica (para desasosiego de Fama) es que existe mucha evidencia de la *ineficiencia de los mercados*. [Pfi00]. Se cita a continuación varias pruebas de esta última afirmación.

En la sección 5.1.2 se vio que el aspecto psicológico del mercado juega un rol fundamental, es precisamente esto que se descubrió hace tantos años una de las razones que pone en aprietos a la hipótesis del mercado eficiente, pero permite vislumbrar que efectivamente el mercado pasa por diferentes períodos o regímenes de comportamiento dependiendo del *estado emotivo* de los agentes que participan en el mismo, lo que indica que los Modelos Ocultos de Markov son una herramienta adecuada para su análisis (ver capítulo 3, sección 3.1).

En la sección 5.1.1 se cito a Warren Buffet [Hug98], la sola existencia de una persona que haya podido obtener ganancias muy superiores al promedio en forma consistente durante tanto tiempo es un ejemplo que contradice la hipótesis, recuérdese que solo es necesario un ejemplo para contradecirla.

Y para el lector que quiera analizar una prueba matemática rigurosa que refuta la tesis puede referirse a [LM99].

Finalmente, es interesante señalar que existe una nueva rama de la economía financiera que se denomina Finanzas Conductistas o (“Behavioral Finance” en inglés), que sostiene lo citado anteriormente, que la psicología humana y los precios están interrelacionados.

El autor de este trabajo concuerda con la no eficiencia del mercado, pero es importante señalar, que aunque Fama tuviera razón, no se afecta al objetivo perseguido por esta tesis, que se limita a detectar patrones provenientes del análisis técnico. No es el objetivo de la misma un juicio de valor sobre la utilidad de dichos patrones.

5.3 Gráficos

Se incluye esta introducción a los gráficos del análisis técnico, pues son la herramienta fundamental del analista técnico y es donde el mismo encuentra los patrones que le revelan los cambios de tendencia. Además en adelante cuando se ejemplifiquen variadas situaciones referidas a tendencias, patrones, indicadores técnicos y otros, se hará uso de los gráficos y es necesario que el lector no habituado al análisis técnico, se familiarice con ellos.

Es importante señalar que los gráficos no provocan subas o bajas en un mercado, sencillamente reflejan la psicología alcista o bajista del mismo. [Mur00]

Existen varios tipos de gráficos, los de *punto y figura*, los *lineales*, los de *barra*, los gráficos de *velas*, etc., sólo se detallarán los lineales y los de barra, que son los que se usarán en este trabajo, en especial los de barra. [Mur00, AP97]

5.3.1 Gráficos lineales

Es el más simple, se elabora utilizando los precios de *cierre* de las cotizaciones (que son los precios más indicativos o *críticos*), dicho precio se grafica con un punto sobre el eje vertical a lo largo del tiempo que se representa sobre el eje horizontal. El tiempo puede corresponder a diferentes períodos (horas, días, semanas, meses), siendo el precio que se grafica, el de cierre del período que se trate, luego se unen los respectivos precios de cierre con una línea conformando el gráfico propiamente dicho. En la figura 5.1 se puede observar un gráfico lineal, con el agregado del volumen (que se verá en todos los gráficos y que se detallará en la sección 5.5).

5.3.2 Gráficos de barra

Es el gráfico usado por excelencia en el análisis técnico, en él se utilizan los tres precios más significativos (*máximo, mínimo, cierre*) del periodo que se trate (hora, día, semana, mes), se utiliza el eje vertical para los precios, trazando una barra vertical sobre el periodo correspondiente, al igual que en el gráfico lineal, el tiempo se representa en el eje horizontal.

El límite superior de la barra corresponde al precio máximo, el inferior al mínimo, y luego se traza un pequeño guión en forma horizontal, partiendo desde la barra y hacia la derecha a la altura del precio de cierre. En la figura 5.2 se puede apreciar un gráfico de barras con su respectivo volumen.

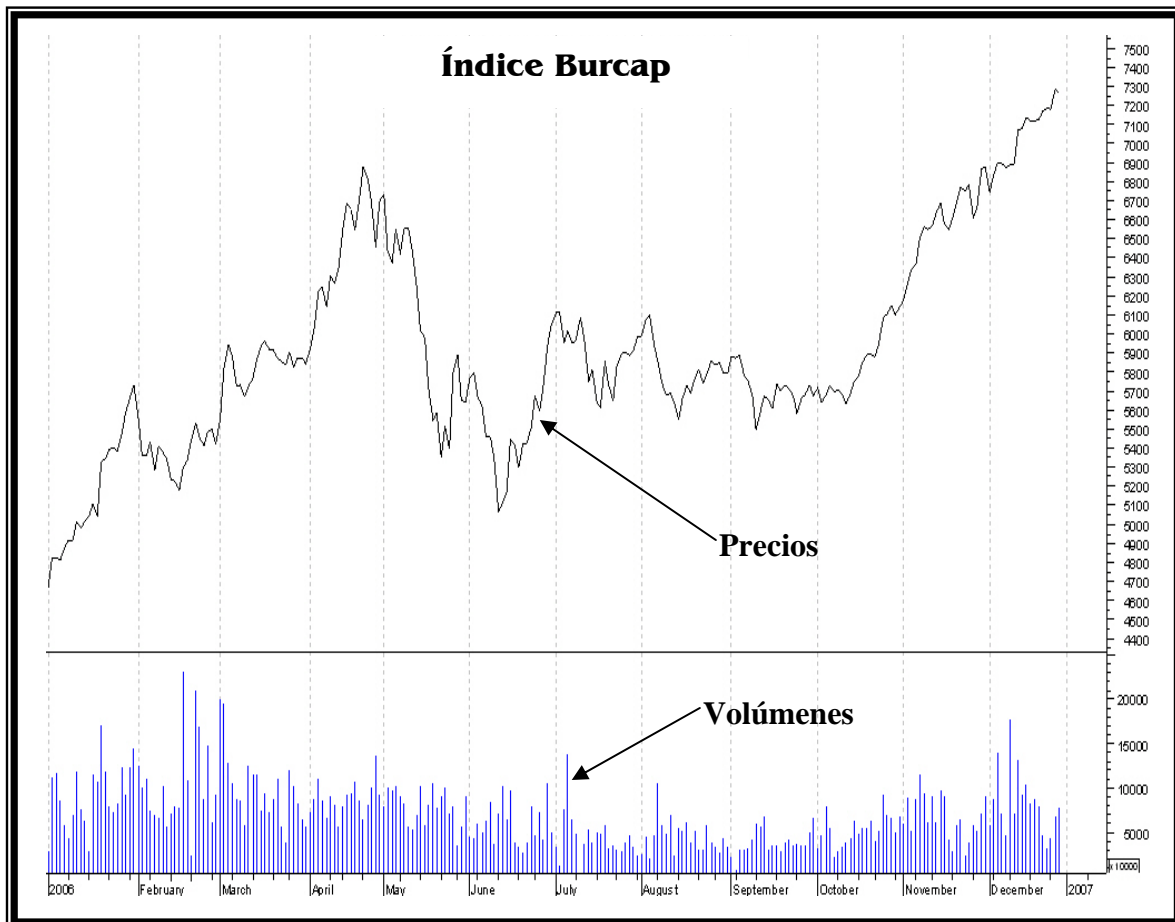


Figura 5.1: Grafico lineal del índice Burcap con volúmenes y periodo diario desde el 02/01/2006 al 28/12/2006.

5.3.3 Escala Simétrica vs. escala logarítmica

Los gráficos de las figura 5.1 y 5.2 son simétricos, puesto que sus escalas son simétricas, es decir a igual diferencia de precio (eje vertical) o de tiempo (eje

horizontal), igual distancia en el gráfico. Como ejemplo, supóngase que la acción ABC tiene un incremento de \$10 a \$20 y luego otro incremento de \$20 a \$30, en el gráfico simétrico ambos incrementos tendrán la misma representación (distancia), lo cual desde el punto de vista del beneficio neto es un error, pues el primero representa un incremento del 100% y el segundo del 50%.

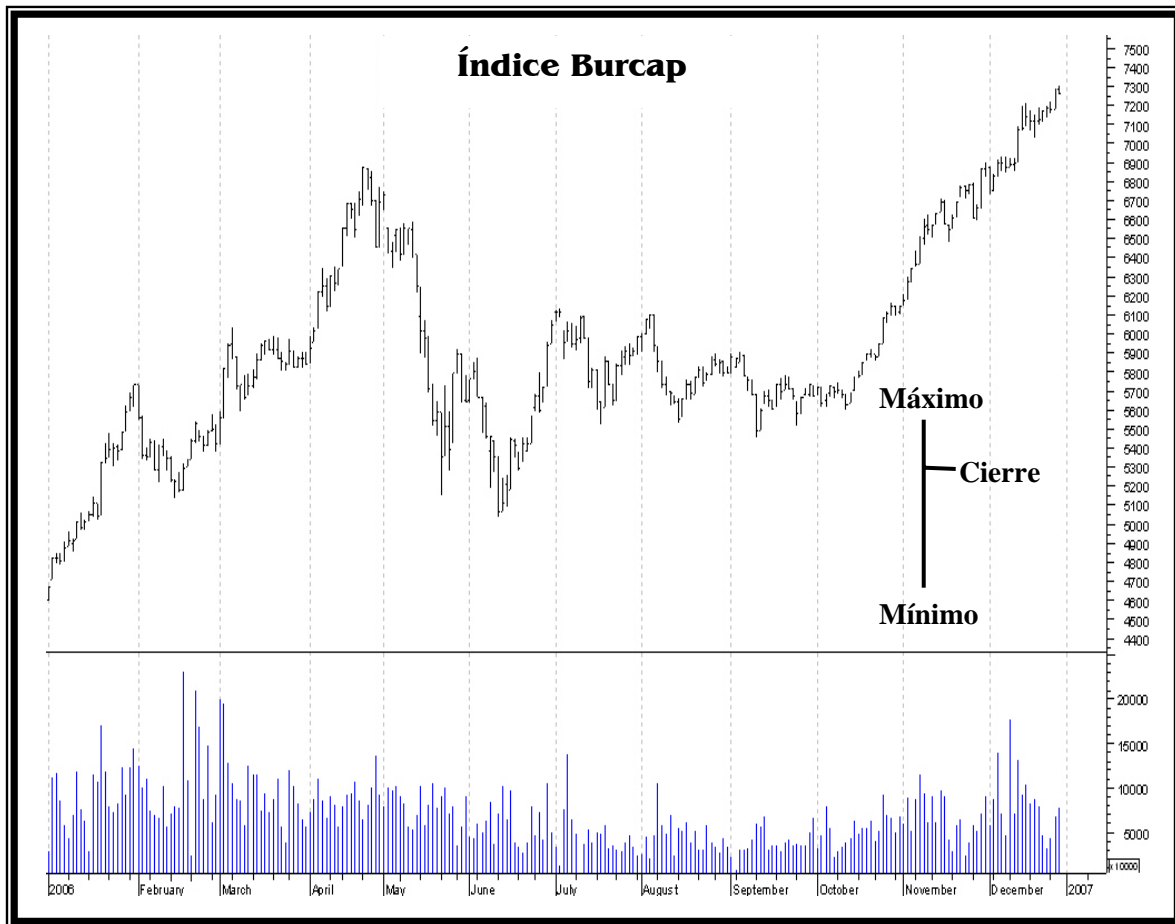


Figura 5.2: Gráfico de barras del índice Burcap con volúmenes y periodo diario desde el 02/01/2006 al 28/12/2006.

Con el fin de subsanar este problema, nacen los gráficos semilogarítmicos, semilogarítmicos porque sólo uno de los ejes tiene escala logarítmica (el vertical usualmente). Así, en un gráfico de esta naturaleza para el ejemplo anterior, se tendría que la representación en el eje vertical para el segundo incremento de \$20 a

\$30, será la mitad del primero de \$10 a \$20, es decir se representan los incrementos porcentuales y no los incrementos nominales.

Aunque los gráficos semilogarítmicos son mejores para apreciar el movimiento en términos del retorno real, los patrones del análisis técnico se verifican mejor en los gráficos simétricos, por esta razón en este trabajo se utilizarán éstos últimos. Sin embargo en gráficos de períodos muy largos, con movimientos significativos de precio, se deben usar los gráficos semilogarítmicos. En la figura 5.3 se puede apreciar un ejemplo de gráfico semilogarítmico.

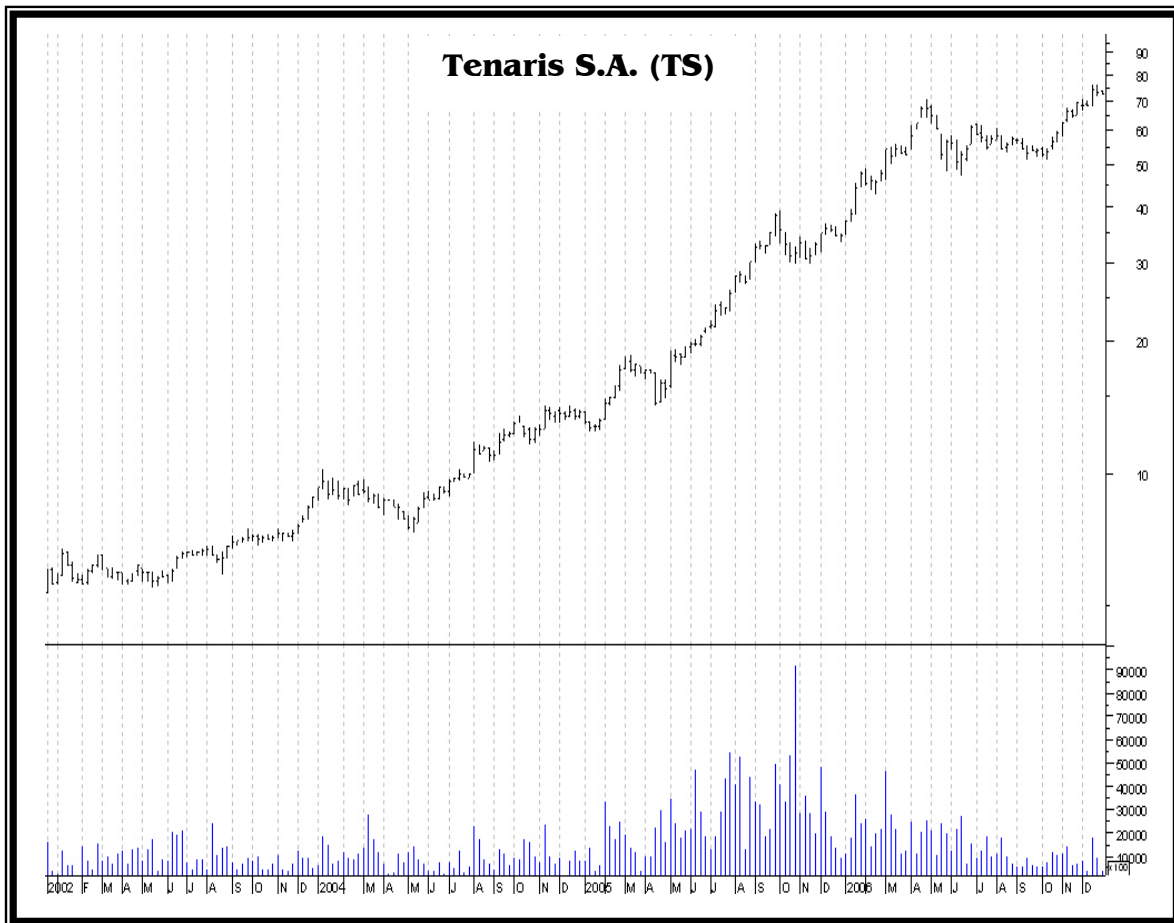


Figura 5.3: Gráfico semilogarítmico de las acciones de la empresa Tenaris S.A. con volúmenes y periodo semanal desde el 20/12/2002 al 28/12/2006.

5.4 Tendencia

Como se vio en la sección 5.2.1, una de las hipótesis del análisis técnico refiere que los precios se mueven siguiendo tendencias, el concepto de tendencia pues, es fundamental para el análisis técnico. Todas las herramientas que utiliza un analista técnico están orientadas a descubrir tendencias en las series temporales, con el fin de aprovecharlas.

En sentido amplio, la tendencia es la dirección en la que el mercado se está moviendo. [Mur00] Pero como los movimientos de mercado son generalmente en *zigzag*, con *crestas* y *valles*, entonces se puede señalar con más precisión que la dirección de esas crestas y valles conforman la tendencia del mercado. Una tendencia ascendente sería representada por una sucesión de crestas y valles cada vez más altos; una tendencia descendente por el contrario la indicaría una sucesión de crestas y valles cada vez más bajos; y si los picos y valles se suceden en forma horizontal, entonces se tendría una tendencia lateral. Si se traza una línea que pase por las crestas o los valles, se obtiene lo que se denomina *líneas de tendencia* en el gráfico.

El mercado puede tener tres tendencias, *ascendente*, *descendente* o *lateral*, aunque esta última puede también definirse como la *ausencia* de una tendencia (bajista o alcista). Rápidamente uno puede intuir que el momento de compra sería al comienzo de una tendencia ascendente, el momento de venta lo sería por el contrario al comienzo de una tendencia descendente y en una tendencia lateral sería prudente mantenerse fuera del mercado. Esta es una visión simplista, pero sirve como una primera aproximación de la aplicación del análisis técnico.

A su vez las tendencias se clasifican por el tiempo que permanecen, puede haber muchas, pero siguiendo la teoría de Dow [AP97], se tiene la tendencia de largo plazo o *primaria* (más de un año), de mediano plazo o *secundaria* (de uno a cuatro meses) y a corto plazo o *terciaria* (de unas horas a varias sesiones); por lo tanto se puede tener una tendencia primaria alcista, conjuntamente con tendencias secundaria y terciaria bajistas por ejemplo. La figura 5.4 muestra un gráfico de largo plazo con algunas de sus tendencias primarias y la proyección de una parte del mismo (mediano y corto plazo) con algunas líneas de tendencia correspondientes a tendencias secundarias y terciarias.



Figura 5.4: Tendencias primaria, secundaria y terciaria. El gráfico madre es semanal y cubre desde el 01/06/1998 al 28/12/2006, y la proyección cubre desde el 02/01/2004 al 31/12/2004 sobre el índice M.AR.

5.5 Volumen

El volumen es la cantidad de activos financieros operados en un periodo de tiempo. Por ejemplo, si se trata de acciones y el periodo es diario, el volumen operado es la cantidad de acciones negociadas en el mercado en el transcurso del día, también se utiliza como volumen al valor monetario transado en un período en particular, es decir, la cantidad de activos operados por el valor al que se operaron [AP97]. Como se vio, el volumen se representa en el gráfico como una línea vertical debajo de las cotizaciones de precios.

El volumen es indicativo de la *fuerza* con la que se produce un movimiento de mercado y puede usarse como confirmación de un movimiento alcista o bajista, de esta forma, si se produce un aumento de precios con un aumento considerable de volumen, entonces se *confirma* que el movimiento tiene fuerza y se podría emprender un fuerte camino alcista; si por el contrario el volumen fuera reducido, habría razones para sospechar que el movimiento alcista no durará demasiado.

Algo parecido sucede en una baja de precios, si el volumen es fuerte, anuncia que la baja puede ser pronunciada y si es débil, puede no durar demasiado. Pero en el caso de las bajas de precio hay que tener especial cuidado, como lo marca [Wei98] y saber en que etapa se encuentra el gráfico (Las *etapas del mercado* se tratan en la sección 5.7.3), pues un valor puede bajar considerablemente aún con volumen reducido.

Por lo tanto en una tendencia alcista debería haber volumen fuerte en las subidas y volumen débil en las correcciones y en una tendencia bajista debería haber volumen fuerte en las bajadas y débil en las correcciones al alza de las bajadas. El volumen actúa como confirmación de los patrones de precios. [Mur00]

En la tabla de la figura 5.5 se puede ver esta relación precio-volumen.

Cotización	crece	baja	baja	crece
Volumen	crece	baja	crece	baja
Señal	alcista	alcista	bajista	bajista

Figura 5.5: Tabla que indica la tendencia subyacente esperada de la relación precio-volumen respectiva.

5.6 Medias móviles

La media móvil dentro de los indicadores técnicos estadísticos es el más usado y al mismo tiempo uno de los más sencillos y efectivos para el seguimiento de las tendencias en series temporales financieras. [Mur00]

La media móvil es un promedio (calculado de diversas maneras como se verá en las siguientes secciones) de una cantidad de datos de períodos específicos. Si por ejemplo, el período es un día, se podría tener una media móvil de 30 días. Se denomina móvil puesto que con el avance de los períodos se avanza también el rango de cálculo para la misma. Siguiendo con el ejemplo, la media de 30 días se calcula sobre los 30 días inmediatos anteriores; al día siguiente el cálculo descartará el día 30 del cálculo anterior e incluirá el día de ayer, y así sucesivamente.

Con frecuencia, se producen distorsiones en las series temporales de precios por motivos estacionales, específicos o extraordinarios. Como consecuencia puede ser dificultoso ver en un gráfico la tendencia real y su posible cambio o continuación. La media móvil ayuda a eliminar estas distorsiones, atenuando las fluctuaciones exageradas de precio. [AP97]

Es básicamente una manera de seguir la tendencia y podría considerarse como una línea de tendencia curva [Mur00]. La media móvil no anticipa el movimiento de la tendencia, sino que lo confirma, es decir, la media móvil sigue al mercado e indica la tendencia del mismo; por lo tanto al igual que en el caso del volumen, puede utilizarse como una herramienta de confirmación.

La media móvil suaviza el movimiento de mercado y permite ver la tendencia subyacente más fácilmente.

La cantidad de períodos utilizados para calcular una media móvil es importante. A mayor cantidad de períodos, más tardará en mostrar la tendencia del mercado, pero menos señales falsas dará; por el contrario una cantidad reducida de períodos en el cálculo mostrará más rápido la tendencia subyacente, pero también dará más señales falsas.

Para el cálculo se utiliza en general el precio de cierre del período, en este trabajo se utilizará dicho precio para su cálculo también, pero algunos analistas prefieren usar otros, como un promedio entre mínimo, máximo y cierre o el *valor de medio punto*, dividiendo el rango de precio del período por dos. [Mur00]

En la figura 5.6 puede observarse un gráfico con medias de distinta cantidad de períodos, eligiendo adecuadamente la cantidad de períodos pueden tenerse medias

móviles que visualicen la tendencias primaria, secundaria y terciaria respectivamente, que es el objetivo del gráfico de la figura 5.6 precisamente.

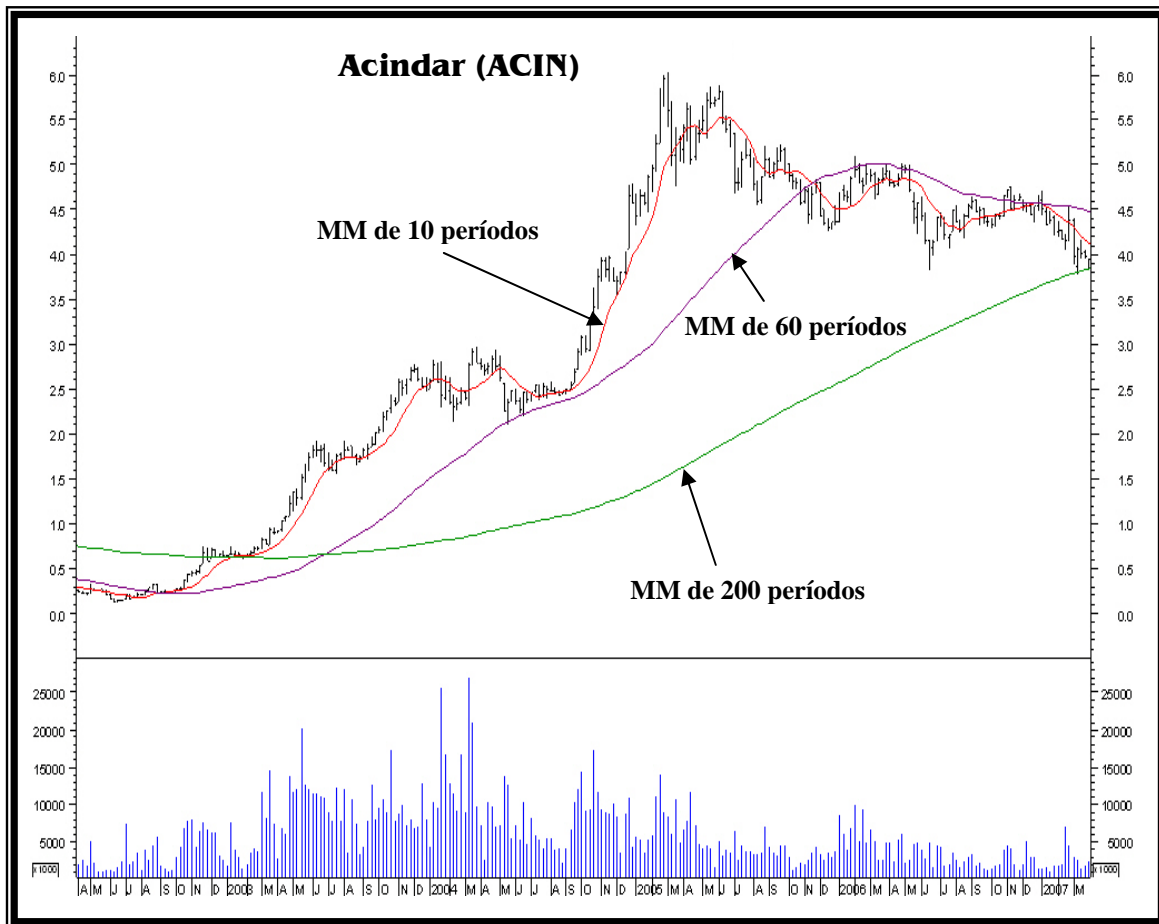


Figura 5.6: Gráfico de Acindar S.A. desde abril del 2002 hasta marzo del 2007, con medias móviles simples de 10, 60 y 200 períodos respectivamente (el periodo es diario en este caso).

5.6.1 Simple

Es la más común y para su cálculo, sencillamente se obtiene el *promedio* o media de los datos de los períodos. Por ejemplo si la media es de 30 períodos, y el período es un día, para calcular el valor de la media móvil se toman los 30 valores

correspondientes de los días inmediatos anteriores, que generalmente, como se vio, será el precio de cierre, y luego se obtiene el promedio de ellos.

A este tipo de medias se le hacen dos críticas básicas:

- Sólo toma en consideración el lapso de tiempo cubierto por la media (si es de 30 periodos, solo considera ese periodo de tiempo)
- Todos los datos del período tienen el mismo peso. (Por ejemplo, en una media de 30 días, el primer día considerado tiene el mismo peso que el último)

El último punto produce señales tardías sobre la tendencia, pues los precios más recientes tienen mayor preponderancia en la misma, por eso algunos analistas piensan que debe darse más peso a los precios recientes.

5.6.2 Ponderadas

En este tipo de media móvil tienen preponderancia los precios más recientes, de esta forma que intenta solucionar una de las críticas a las medias móviles simples.

Existen dos formas de calcularla [AP97], una forma es calcular más de una vez los últimos valores. Otro método es multiplicar por un factor mayor a los precios más recientes y dividir por la suma de los multiplicadores. Por ejemplo si la media es de 5 días, el valor más reciente se multiplica por 5, el siguiente por 4 y así sucesivamente, luego se divide por 15. Más formalmente, si la media es de j días, y sean a_i ($i=1..j$) los precios de los j días anteriores (siendo a_1 el precio del día inmediatamente anterior), entonces la media ponderada (de acuerdo a este último método) se calcula con la siguiente fórmula:

$$\frac{a_1 j + a_2 (j-1) + a_3 (j-2) + \dots + a_j}{\sum_{i=1}^j i} \quad (5.1)$$

Pero esta media mantiene el problema de considerar sólo los precios incluidos dentro del rango de cobertura de la media.

5.6.3 Exponenciales

Esta media soluciona los dos problemas citados anteriormente, los precios más recientes tienen mayor ponderación, por lo que es una media ponderada, pero además toda la serie de precios anterior tiene incidencia en su cálculo. [Mur00]

Para su cálculo [AP97], se utiliza un factor de corrección, calculado dividiendo 2 por el número de períodos de la media (este número se utiliza para obtener el factor de corrección, pero la media tiene en cuenta toda la vida de la serie). Por ejemplo si la media es de 50 períodos, el factor de corrección sería $2/50 = 0,04$.

El cálculo se obtiene multiplicando el factor de corrección por la diferencia entre el precio del periodo y el último valor de la media, y el resultado es sumado al último valor de la media. El valor inicial de esta media se puede obtener por promedio. En la tabla de la figura 5.7 se puede ver un ejemplo del cálculo de esta media.

Sesión	Precio	Última media	Diferencia x factor	Nueva media
0				10
1	12	10	$2 \times 0,04 = 0,08$	10,08
2	11	10,08	$0,92 \times 0,04 = 0,0368$	10,1168
3	10,5	10,1168	$0,3832 \times 0,04 = 0,0153$	10,1321
4	9	10,1321	$-1,1321 \times 0,04 = -0,0453$	10,0868

Figura 5.7: Tabla que muestra el cálculo de una media exponencial de 50 sesiones o periodos (factor de corrección = $2/50 = 0,04$).

La media exponencial tiene una menor probabilidad de mostrar señales falsas, además es más eficiente su cálculo, pues a diferencia de las otras medias vistas, solo es necesario el valor de la media anterior junto con la última cotización. Esta es la media que se usará en este trabajo a la hora de la aplicación de la teoría.

5.7 Patrones del análisis técnico

Como se pudo apreciar a lo largo de este capítulo, la búsqueda de patrones en las series temporales financieras es el objetivo clave de un analista técnico. Estos patrones son indicativos de *cambios de tendencia* o *continuación de tendencia*, además como se verá en la sección 5.7.3 permitirán descubrir en que *etapa del mercado* se encuentra un determinado instrumento financiero.

Formalmente los patrones en series temporales financieras desde el punto de vista del análisis técnico son “imágenes o formaciones que aparecen en los gráficos de precios” [Mur00]. Éstos pueden clasificar en diferentes categorías y poseen valor de predicción. En el capítulo 6 se abordará el tema de los patrones en general.

5.7.1 Patrones de cambio de tendencia

Estos patrones anuncian que se está originando un cambio importante en la tendencia. Los patrones de cambio de tendencia tienen características comunes que se citan a continuación [Mur00]. Como es norma en el análisis técnico, no significa que siempre se van a registrar todas estas características, sólo que existe una importante probabilidad de que así sea, el lector ya habrá descubierto a esta altura que un modelo estocástico es esencial para cualquier estudio sobre el mismo.

- Debe existir una precisa tendencia previa, es decir, no puede haber un cambio de tendencia si no está clara la tendencia anterior.
- En general la primera señal de cambio es la ruptura de una línea de tendencia (ver sección 5.4). Es importante señalar que lo opuesto no es cierto, que se viole una línea de tendencia no necesariamente significa un cambio en la misma, puede significar sólo un cambio en la *pendiente* de la tendencia.
- Cuanto *más grande* sea un patrón, mayor será el movimiento que conlleva su aparición. Donde “más” grande se refiere a la *altura* y *ancho* del patrón, la altura mide la *volatilidad* del mismo y la anchura el *tiempo en que se forma*.
- Los patrones superiores (de precios altos) son más volátiles y más cortos que los inferiores (de precios bajos) que tardan más en formarse y tienen menos volatilidad.

- El volumen es más importante en los movimientos ascendentes que en los descendentes. El volumen como ya se vio (sección 5.5) es importante para la confirmación de los patrones; pero en el lado ascendente es fundamental, mientras que en el descendente pueden caer los precios sin un aumento considerable de volumen.

A continuación se verán algunos de los más significativos [Mur00, AP97].

5.7.1.1 Cabeza y hombros

Es el patrón de cambio más conocido, y probablemente el más exitoso para prever cambios de tendencia [AP97]. En la figura 5.8 se puede ver una representación del mismo, en ella se puede apreciar como una serie de crestas y valles ascendentes empiezan a perder fuerza. Subyacentemente lo que está sucediendo es que las fuerzas de la oferta y la demanda, están en un equilibrio momentáneo.

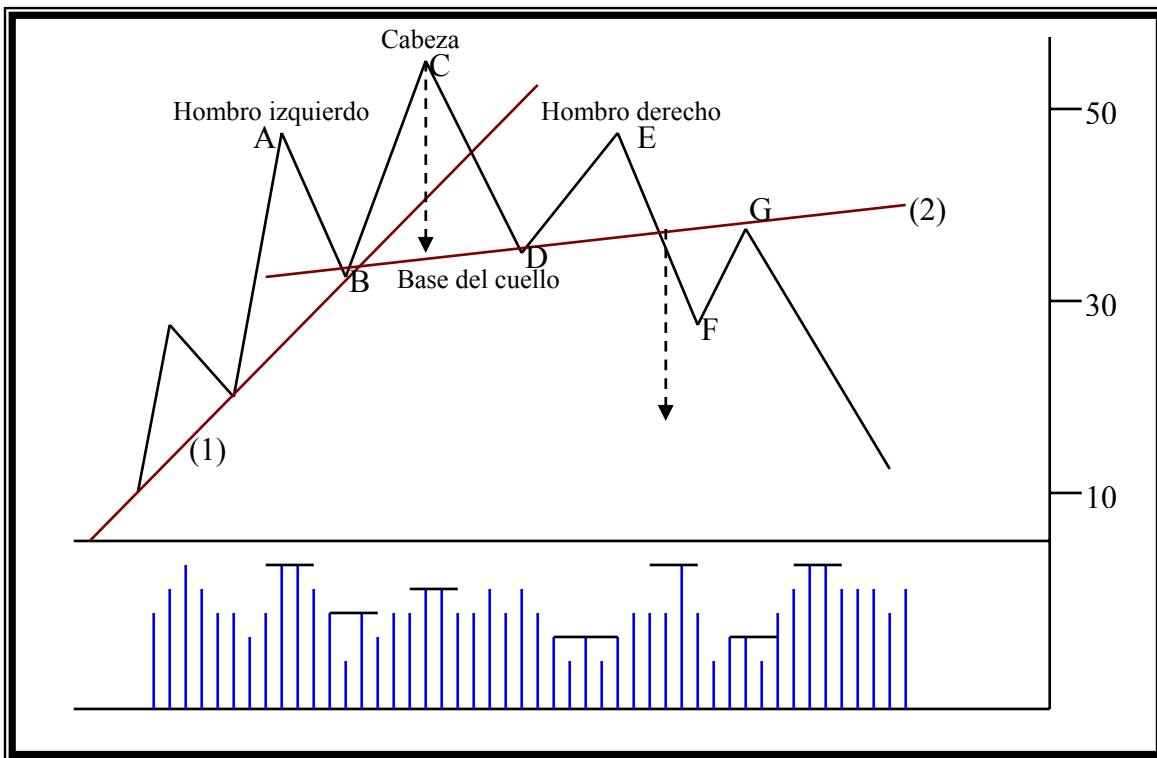


Figura 5.8: Patrón de cabeza y hombros, los puntos A y E son los hombros, izquierdo y derecho respectivamente, el punto C es la cabeza, la flecha punteada es el objetivo de mínima proyectado debajo de la ruptura de la base del cuello.

Un patrón de cabeza y hombros (como puede verse en la figura 5.8) tiene en general las siguientes particularidades [Mur00] (las seis primeras son necesarias para la confirmación del patrón):

1. Una previa tendencia al alza.
2. El hombro izquierdo (punto A) tiene un volumen importante, seguido por una baja al punto B con menor volumen.
3. Una subida hasta un nuevo máximo (punto C), pero con menos volumen que el punto A.
4. Un descenso que lleva los precios debajo de la cresta del punto A, rompe la línea de tendencia (1) y se acerca al mínimo anterior (punto D).
5. Una tercera suba de precios (punto E) con bastante menor volumen y que no puede superar la cabeza (punto C).
6. Una baja que confirma el patrón por debajo de la base del cuello (2).
7. Finalmente con el patrón ya confirmado, una recuperación hasta el (punto G).

Este patrón según [Wei98] y [Mur00] permite establecer un objetivo de precios luego de la confirmación, dicho objetivo se establece como la distancia entre el punto de la cabeza (punto C) y la línea del cuello (2) proyectada hacia abajo desde la misma línea del cuello en el punto donde es vulnerada en sentido descendente. Según [Mur00] este es un objetivo mínimo, es decir, los precios como mínimo llegarán a esa zona de precios.

5.7.1.2 Cabeza y hombros invertidos

Es la imagen inversa del anterior y por ende es una señal de cambio de tendencia de bajista a alcista, en la figura 5.9 se puede ver una representación. Es importante señalar que en este caso es de vital importancia el volumen en la ruptura de la línea del cuello luego del hombro derecho como una confirmación del patrón y de la consecuente tendencia alcista que se avecina.

Además el movimiento de retorno o *estirón* según [Wei98] es más común en el patrón de cabeza y hombros invertido y debe suceder con menor volumen que la ruptura o *fuga* por encima de la base del cuello luego del hombro invertido derecho.

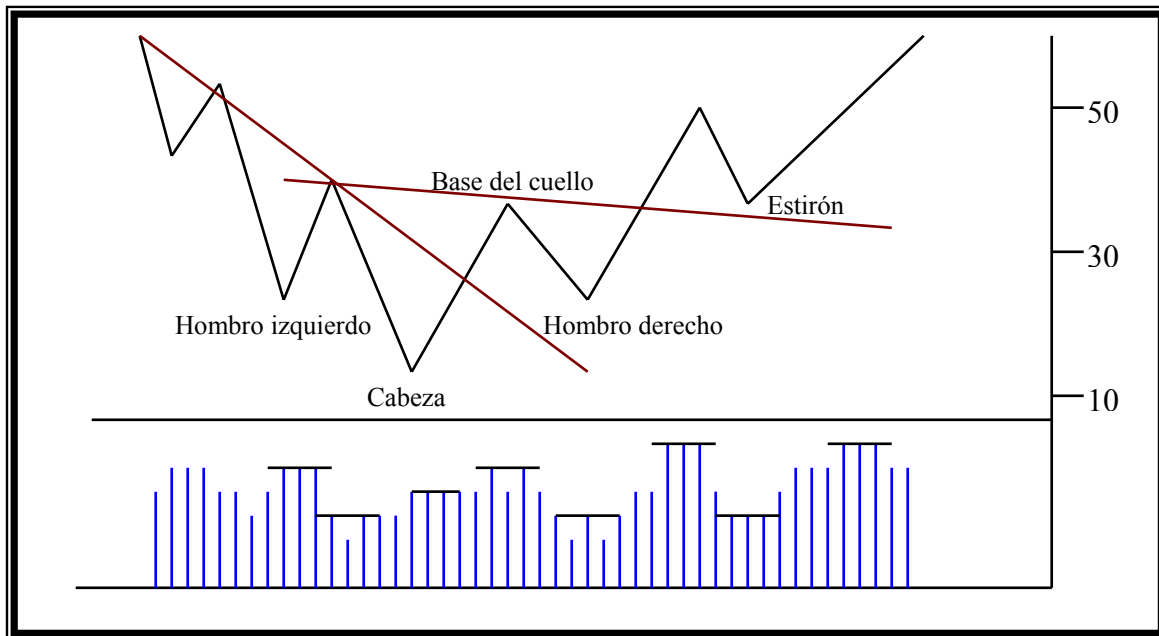


Figura 5.9: Patrón de cabeza y hombros invertidos, notar el aumento significativo de volumen en la ruptura de la línea del cuello.

5.7.1.3 Doble cresta y doble valle

El patrón de doble cresta y doble valle, después del patrón de cabeza y hombros, es el que se ve e identifica con mayor facilidad. En las figuras 5.10a y 5.10b se pueden ver representaciones del patrón de doble cresta y del patrón de doble valle respectivamente.

Como se observa en la figura 5.10a la fuerza alcista disminuye y el máximo del punto C no puede superar al máximo del punto A, se ha quebrado la línea de tendencia respectiva, y finalmente el patrón se confirma cuando los precios descienden por debajo del nivel de precios del punto B.

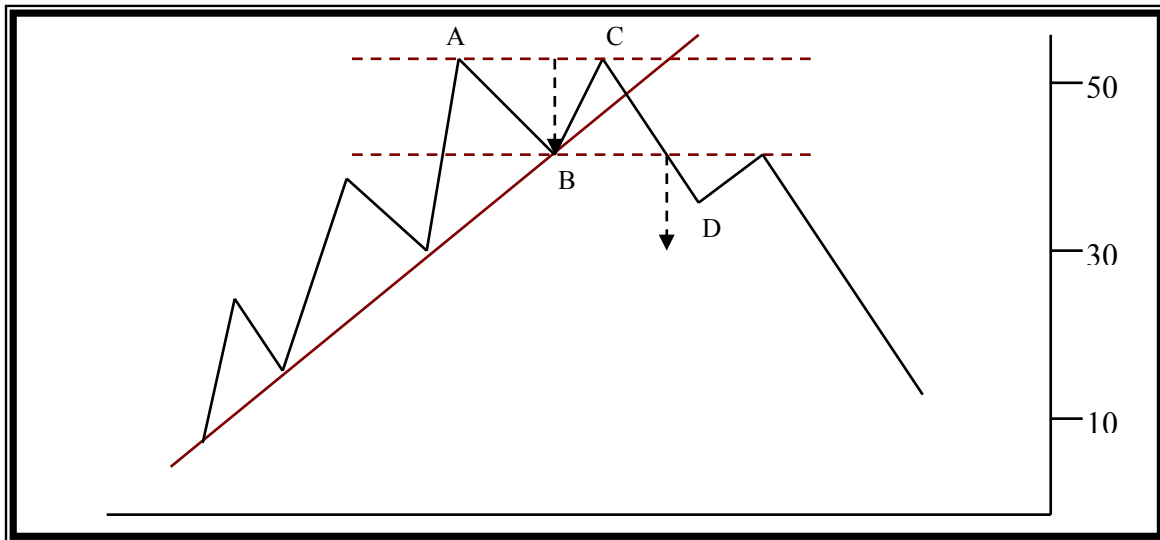


Figura 5.10a: Patrón de doble cresta, la flecha punteada representa el objetivo de mínima de los precios al romper por debajo de la línea del punto B.

El caso del doble valle es análogo, sólo hay que recordar nuevamente, que en el caso de bajista a alcista se debe tener un volumen importante en la ruptura alcista luego del segundo valle.

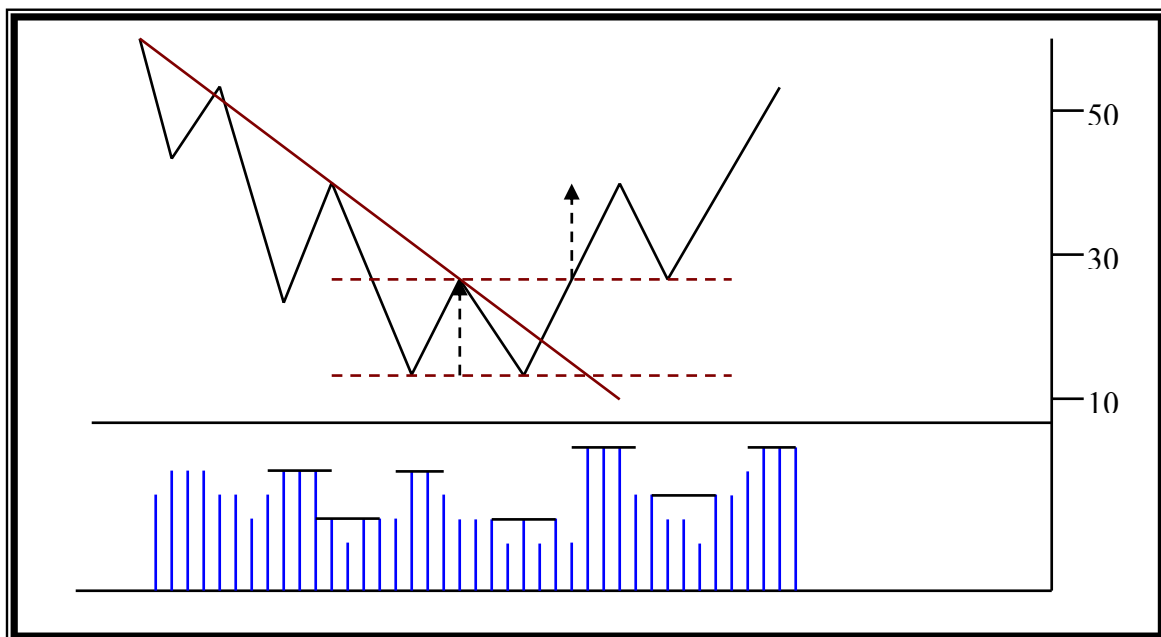


Figura 5.10b: Patrón de doble valle, notar el aumento significativo de volumen en la ruptura de la resistencia, la flecha punteada indica el objetivo de mínima y se proyecta sobre la ruptura de la resistencia.

5.7.1.4 Triple cresta y triple valle

Varios de los conceptos anteriores sirven para analizar este tipo de patrón, sólo que este patrón es mucho más infrecuente que los anteriores. Se caracteriza porque cuando la oferta y demanda se están equilibrando, se producen tres crestas o tres valles. En las figuras 5.11a y 5.11b se puede ver un ejemplo del patrón de doble cresta y doble valle respectivamente.

El tratamiento del patrón de triple cresta es similar al de cabeza y hombros, sólo que todas las crestas están al mismo nivel, y lo visto para dicho patrón es aplicable.

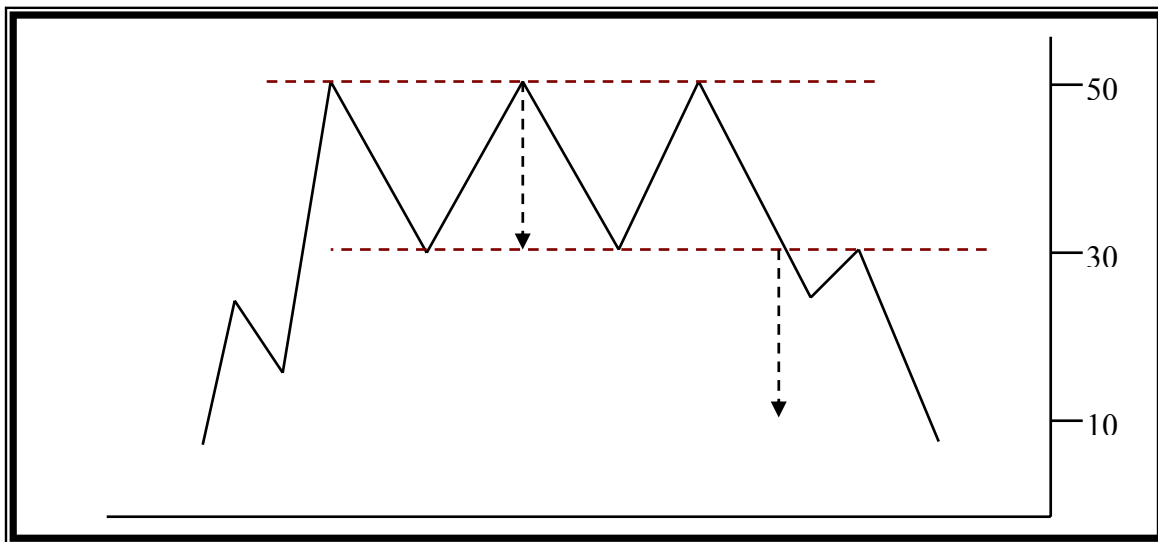


Figura 5.11a: Patrón de triple cresta, la flecha punteada representa el objetivo de mínima de los precios al romper por debajo de la línea de resistencia.

Análogamente, el patrón de triple valle es similar al de cabeza y hombros invertidos y las consideraciones también son afines.

Es importante señalar que estos patrones se confirman solo si se produce una ruptura del soporte o resistencia respectivamente.

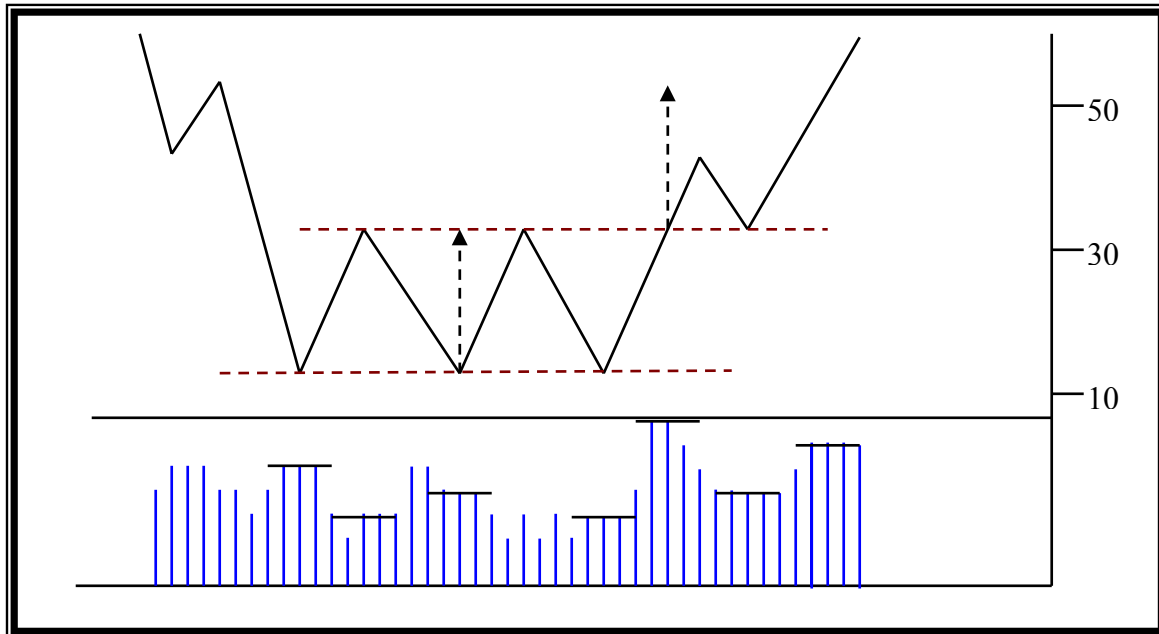


Figura 5.11b: Patrón de triple valle, notar el aumento de volumen en la ruptura de la resistencia, la flecha punteada indica el objetivo de mínima y se proyecta sobre la resistencia.

5.7.1.5 Fondos redondeados y crestas redondeadas

Estos patrones muestran un lento cambio de tendencia bajista a alcista en el caso de los fondos redondeados o *platillos*, y de tendencia alcista a bajista en el caso de las crestas redondeadas. Se puede apreciar una “U” redondeada formada por precios y volúmenes como muestra la figura 5.12a en el caso de los platillos, y una “U” redondeada invertida de los precios, junto con una “U” redondeada normal de volúmenes como muestra la figura 5.12b para el caso de las crestas redondeadas.

Estas formaciones se aprecian en gráficos a largo plazo, semanales o mensuales, ya que estos patrones son lentos, pero luego marcan grandes cambios.

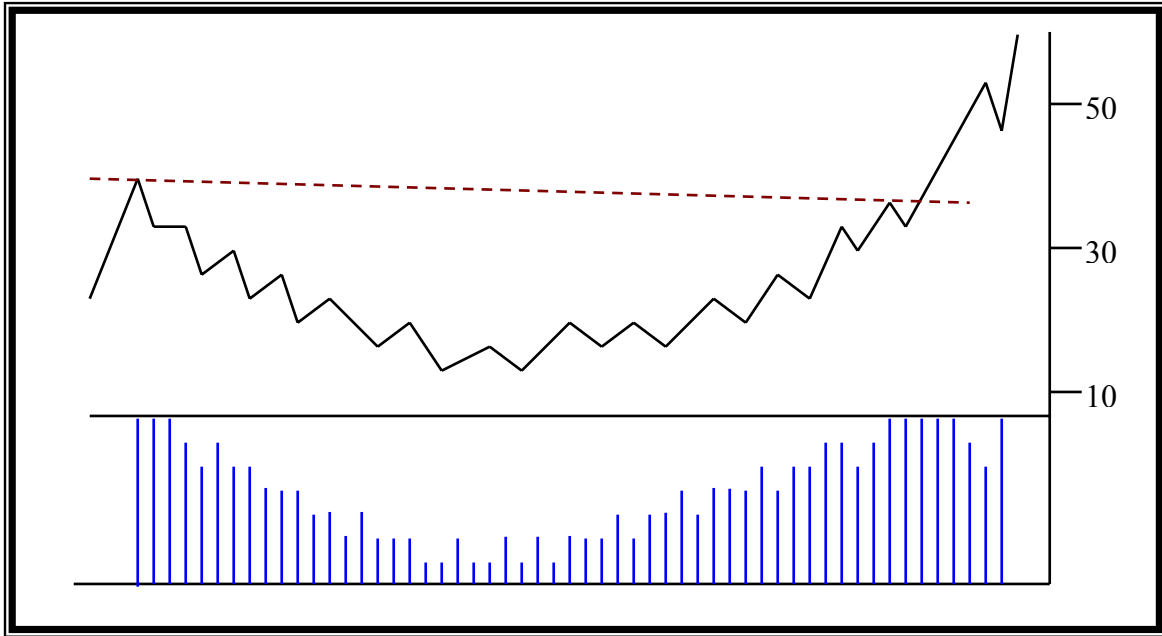


Figura 5.12a: Patrón de platillo o fondo redondeado, notar que el volumen realiza una figura parecida a la que realizan los precios.

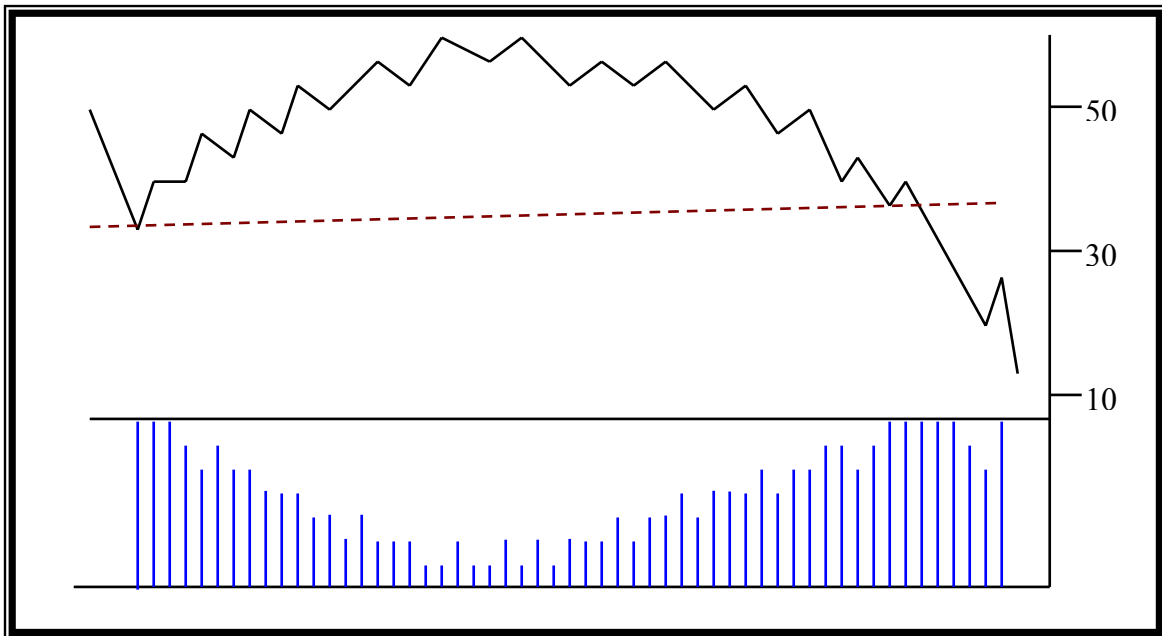


Figura 5.12a: Patrón de cresta redondeada, notar que en este caso el volumen realiza una figura inversa a la que realizan los precios.

5.7.2 Patrones de continuación de tendencia

Estos patrones indican que el movimiento lateral que se está registrando en los precios es una pausa para luego continuar con la tendencia original. Los patrones de continuación tardan menos tiempo en formarse que los de cambio de tendencia.

Es importante recalcar que tanto los patrones de continuidad como los de cambio no son determinísticos, son de naturaleza estocástica, es decir, cuando se forman hay una importante probabilidad de que el movimiento de mercado se comporte como “debería”, pero se debe recordar que también existe la probabilidad “no nula” de que esto no sea así.

5.7.2.1 Triángulos

Existen tres clases de triángulos: *simétricos*, *ascendentes* y *descendentes* [AP97, Mur00]. El triángulo es un patrón intermedio en cuanto al tiempo que tarda en formarse, (de un mes a tres meses generalmente).

El Triángulo simétrico (figura 5.13a) representa una pausa de la tendencia actual, luego de la cual, el movimiento de mercado continuará en la misma dirección. Es decir, si se encuentra en una tendencia ascendente como la de la figura 5.13a, entonces la resolución será alcista.

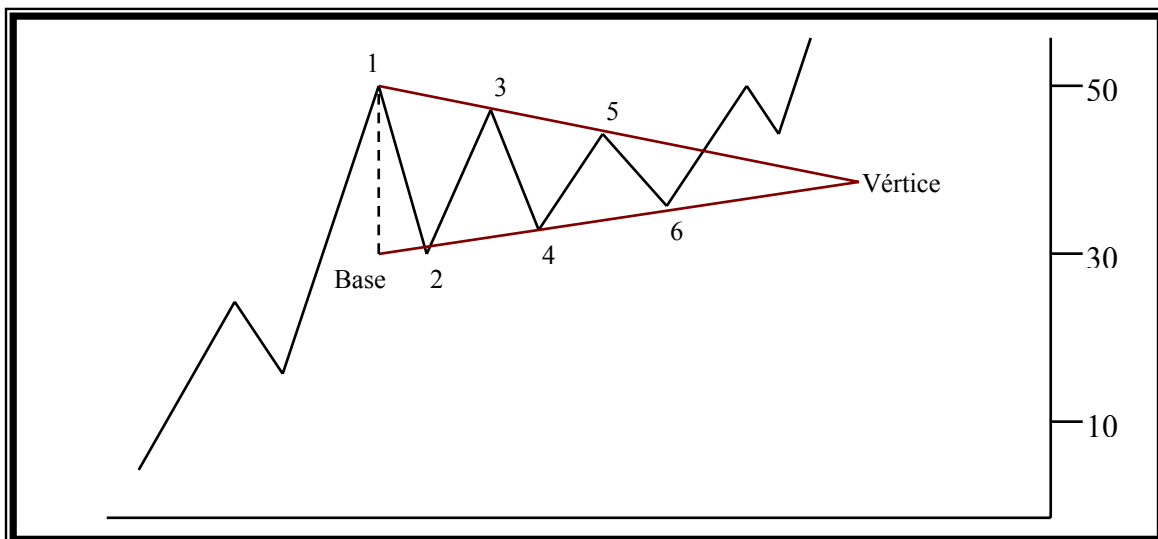


Figura 5.13a: Patrón de triángulo simétrico con seis puntos, el vértice es el tope para la resolución, la parte más ancha es la base.

Para que exista un triángulo (de cualquier tipo) deben haber *cuatro puntos*³⁵ de cambio, en la figura 5.13a se aprecia la existencia de seis.

El triángulo ascendente (figura 5.13b) es una variación del simétrico, pero su resolución es alcista, se considera sin embargo dentro de los patrones de continuidad pues aparece mayoritariamente en recorridos alcistas de los movimientos de mercado.

Nótese que la línea de tendencia superior es plana y la línea de tendencia inferior es alcista, esto demuestra que la fuerza de los compradores es superior a la de los vendedores.

Cuando este patrón aparece en tendencias descendentes, será considerado un patrón de cambio de tendencia de bajista a alcista.

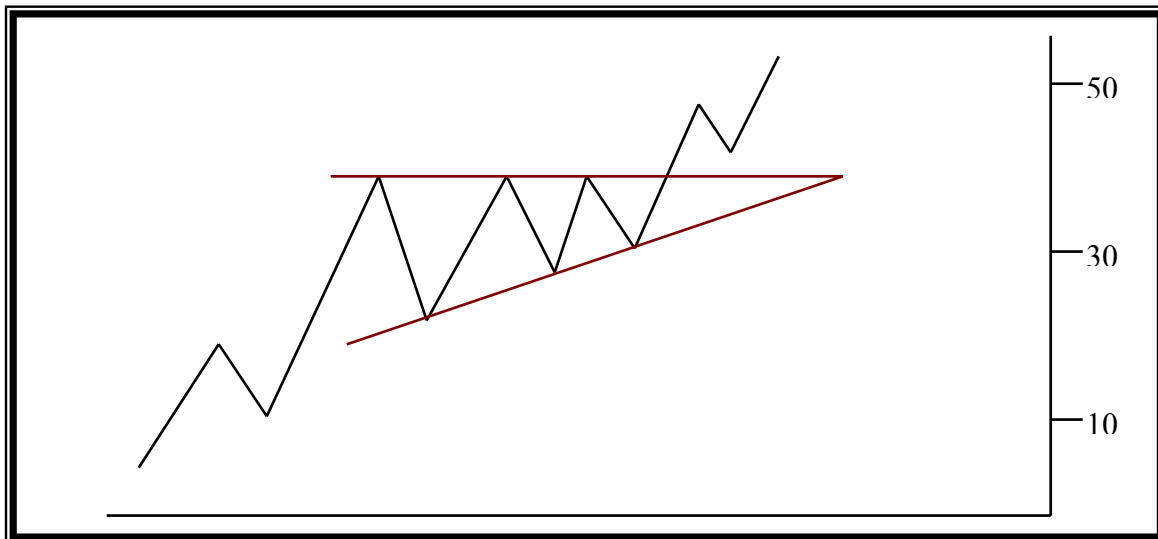


Figura 5.13b: Patrón de triángulo ascendente, notar que la línea de tendencia superior es plana y la inferior es ascendente.

El triángulo descendente (figura 5.13c) tiene (al contrario del ascendente) una resolución a la baja, al igual que el anterior se lo considera de continuidad pues aparece mayoritariamente en movimientos de mercado bajistas.

Nótese que la línea de tendencia inferior es plana y la superior es bajista, denotando en contraposición con la anterior una mayor fuerza por parte de los vendedores.

³⁵ Recuérdese que para una línea de tendencia se necesitan al menos dos puntos, por lo tanto como un triángulo se forma con dos líneas de tendencia convergentes, se necesitan cuatro puntos al menos.

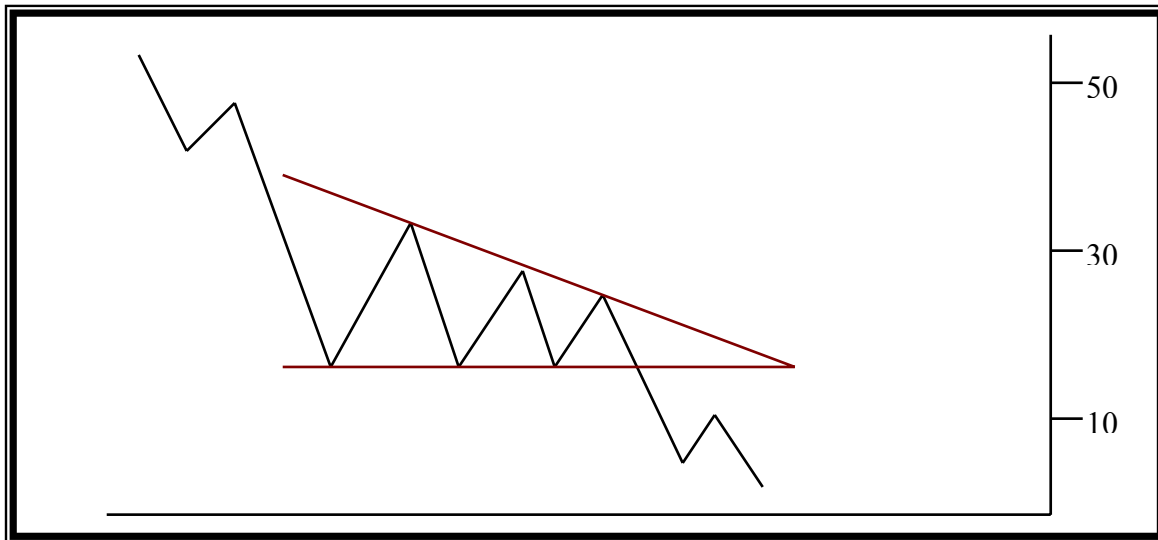


Figura 5.13c: Patrón de triángulo descendente, notar que la línea de tendencia inferior es plana y la superior es descendente.

Cuando este patrón aparece en un recorrido alcista, se lo considera un patrón de cambio de tendencia de alcista a bajista.

Finalmente, se verán algunas características de todos los triángulos:

- El volumen en los triángulos tiende a disminuir a medida que se recorre el canal de convergencia del mismo, pero debe crecer notablemente en la rotura, teniendo en cuenta como en todos los casos que el volumen es vital en las roturas al alza.
- Los triángulos tienen un tiempo tope de resolución, que es el momento de convergencia de las líneas de tendencia.
- Midiendo la altura vertical en la parte más ancha del triángulo, se obtiene el objetivo de mínima que los precios recorrerán en la dirección correspondiente al momento de la resolución.

5.7.2.2 Banderas y banderines

Estos patrones son bastante comunes y representan una breve pausa en un movimiento brusco de precios para luego continuar en la misma dirección. Una de las condiciones para la existencia de este patrón es que este precedido de un movimiento brusco, (casi perpendicular de precio).

Son patrones muy confiables y se desarrollan en periodos muy cortos de tiempo (de una a tres semanas)

Durante su formación se sucede un descenso brusco de volumen, que se incrementa también bruscamente cuando se resuelve.

Ambos patrones se desarrollan aproximadamente a la mitad de un movimiento (alcista o bajista) de mercado, es decir, que se espera que exista una continuación del movimiento inicial del mismo tamaño que desde la ruptura primitiva hasta el patrón.

En la figura 5.14a se puede ver una bandera, nótese que se parece a un pequeño paralelogramo que se inclina en dirección opuesta a la dirección de la tendencia original (que luego se mantendrá).

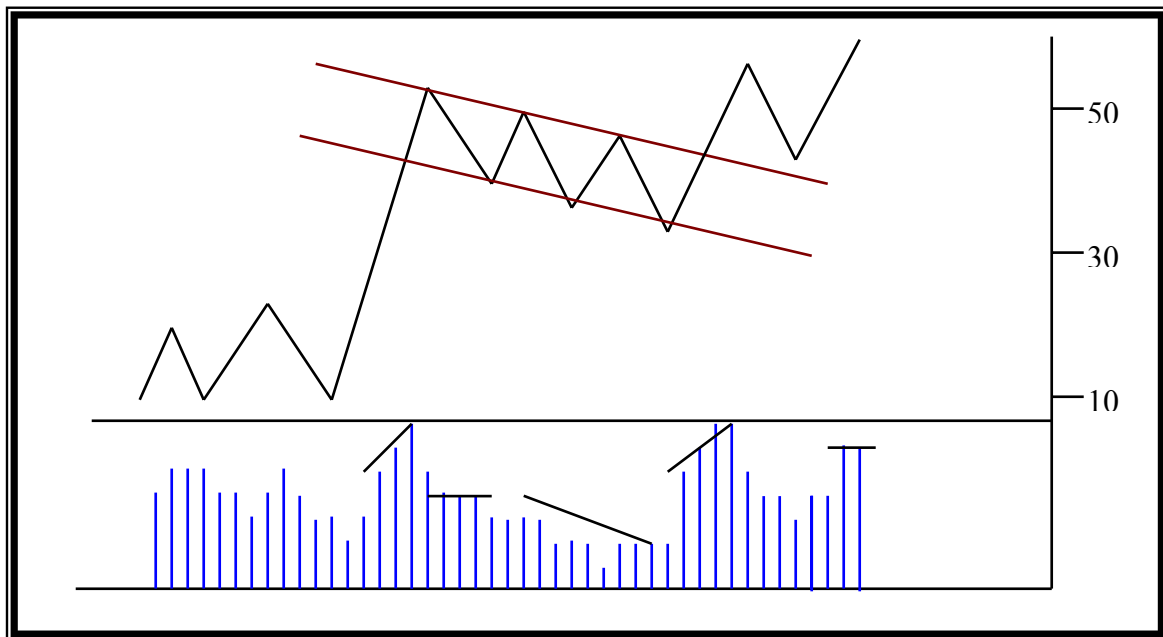


Figura 5.14a: Patrón de bandera alcista, notar que aparece después de un movimiento pronunciado al alza y se inclina en dirección contraria a la tendencia, el volumen se reduce durante la formación e incrementa en la salida.

En la figura 5.14b se puede ver un banderín, nótese que se parece a un triángulo simétrico, pero la diferencia es que no dura más de tres semanas.

Como siempre tanto en las banderas como en los banderines, la confirmación del volumen en el lado alcista es fundamental.

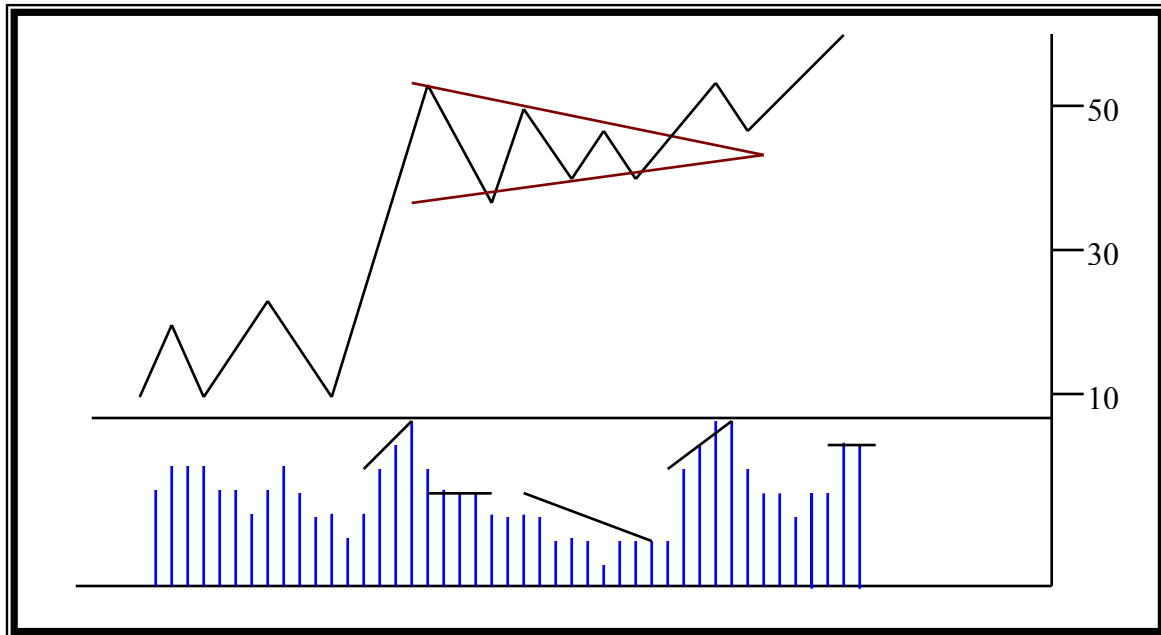


Figura 5.14b: Patrón de banderín alcista, se parece a un triángulo simétrico, la diferencia está en que no dura más de tres semanas.

5.7.2.3 Cuñas

Este patrón es muy parecido a un triángulo simétrico, ya sea por su forma como por su duración, pero la diferencia está en la inclinación de las líneas de tendencia. Mientras que en el triángulo simétrico, una es ascendente y la otra descendente, en la cuña ambas tienen el mismo sentido.

La resolución posterior irá en el sentido opuesto a la inclinación de las líneas de tendencia, es decir, una cuña alcista tendrá resolución bajista y al contrario una cuña bajista tendrá resolución alcista como puede observarse en las figuras 5.15a y 5.15b.

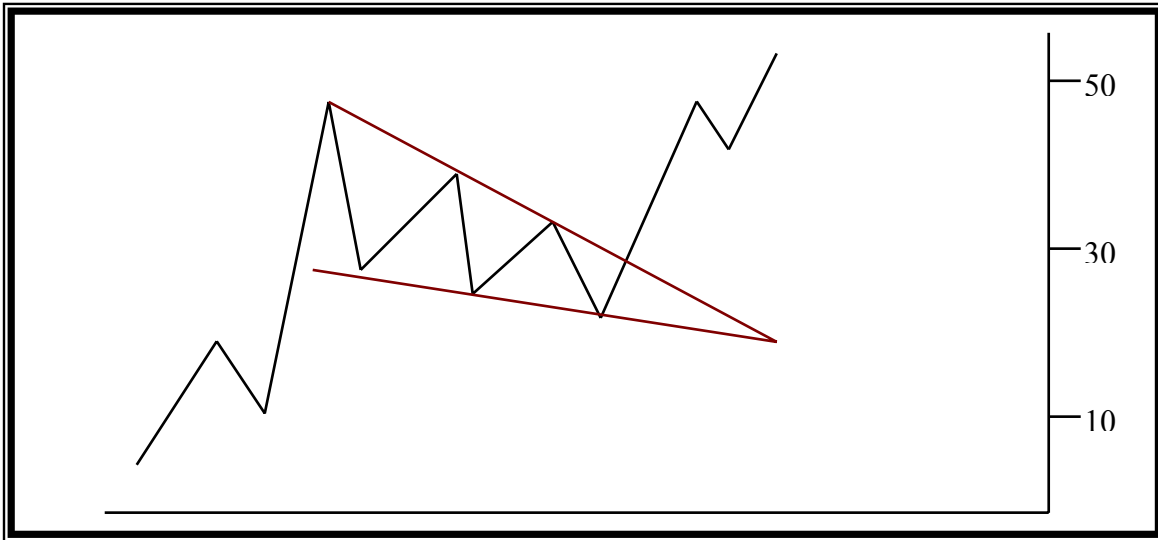


Figura 5.15a: Patrón de cuña alcista, notar que las líneas de tendencia se inclinan en dirección contraria a la tendencia principal.

Las cuñas al igual que los triángulos (ascendentes y descendentes) se consideran de continuación pues la bajista aparece generalmente en un movimiento ascendente de precios y la alcista en un movimiento descendente de precios. Cuando no es así, como en el caso de los triángulos, es indicativo de un cambio de tendencia.

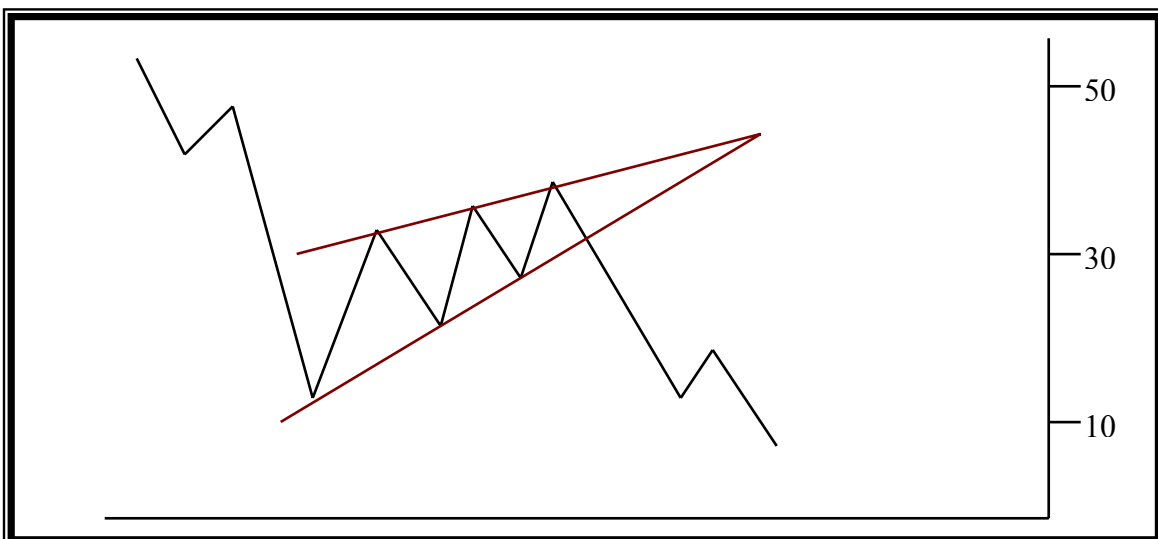


Figura 5.15b: Patrón de cuña bajista, en este caso la inclinación de las líneas de tendencia es hacia arriba.

5.7.3 Etapas del mercado

El siguiente patrón no es un patrón de cambio ni de continuación, pero sin dudas es el más importante de todos los vistos hasta aquí, puesto que permite determinar en todo momento en que *etapa del mercado* se encuentra un determinado instrumento financiero.

La teoría se desarrolla en [Wei98], y como se señala en dicha fuente, siguiendo adecuadamente la misma, se podría prescindir de cualquier otro patrón. En si mismo la teoría captura la esencia del análisis técnico ya que si se identifica con precisión la etapa del mercado en la que se encuentra un instrumento financiero, se conoce la tendencia subyacente en todo momento.

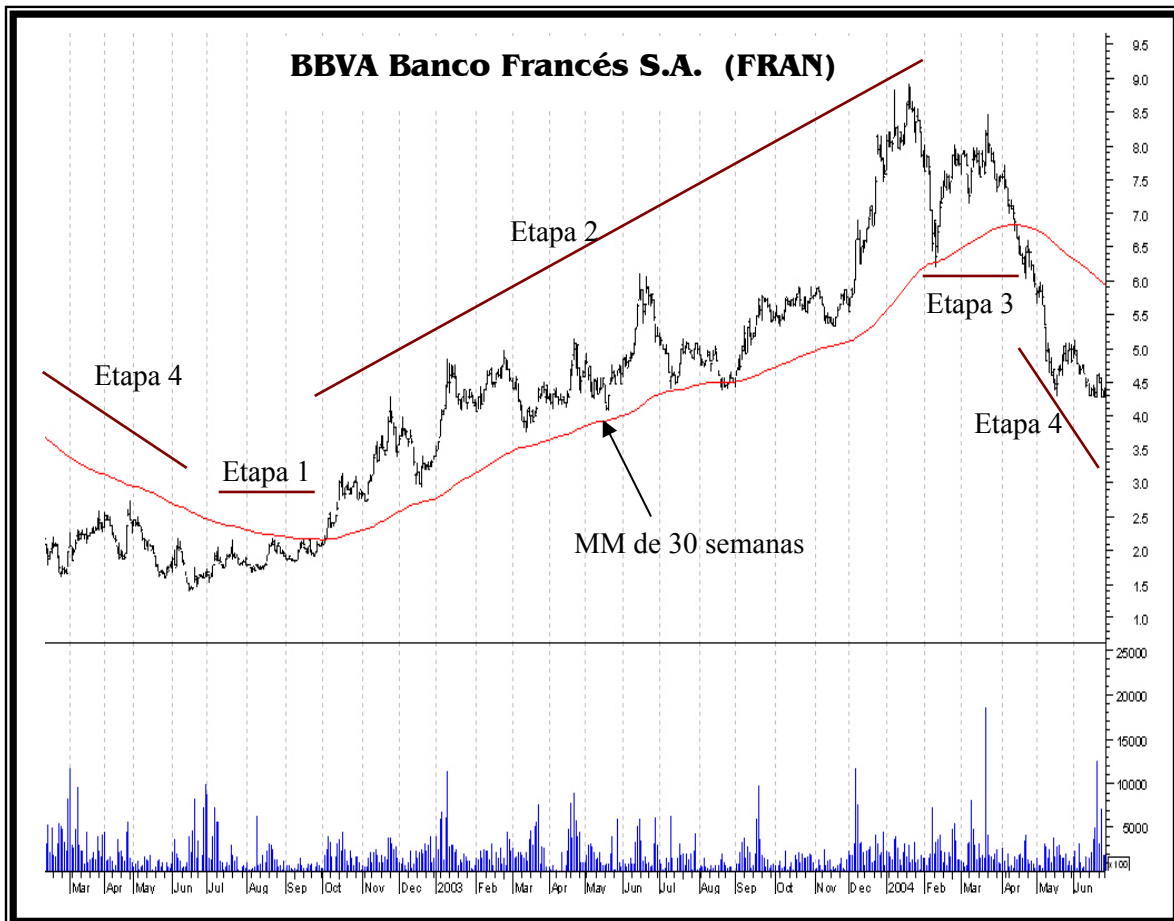


Figura 5.16: Etapas del Mercado, sobre las acciones del Banco Francés S.A. en un gráfico desde febrero de 2002 hasta junio de 2004, con una media móvil de 30 semanas ó 150 días

La teoría dice que todo valor debe encontrarse en una de las *cuatro etapas* del mercado. Las cuatro etapas de un ciclo principal (tendencia primaria, ver sección 5.4) como se muestra en la figura 5.16 son: (1) área de base, (2) etapa alcista, (3) área de techo, y (4) etapa descendente. El basamento de esta teoría es el simple sentido común, existe un viejo dicho que versa: “las ramas de los árboles nunca llegan al cielo, y sus raíces nunca llegan al infierno”, en clara alusión a que irrefutablemente todo valor vivirá ciclos o etapas, el problema está en determinar justamente en cual etapa se encuentra.

5.7.3.1 Etapa 1: la base

Cuando un valor lleve cierto tiempo descendiendo, comenzará a perder fuerza hacia abajo, es decir las fuerzas de la oferta y la demanda empezarán a equilibrarse. En ese momento empezará a sufrir un movimiento *lateral*. Además a medida que la base se forme se verificará un descenso de volumen y a medida que se acerque el final de esta etapa generalmente dicho volumen comenzará a aumentar, aunque los precios se mantengan, lo que indica que la venta ya no hace descender el valor, pues la fuerza de la demanda está empezando a ser significativa.

La media móvil de 30 períodos³⁶ al comienzo de la etapa empieza a perder su pendiente negativa y se convierte en una *pendiente nula* mientras persista esta etapa.

Durante esta etapa se producen movimientos de precios por debajo y por encima de la media móvil, moviéndose lateralmente entre la línea de tendencia que oficia de resistencia superior y la línea de tendencia que hace de soporte inferior de la *gama operativa*³⁷.

Este movimiento puede durar meses o años, por eso no es recomendable comprar en este momento, aún cuando se alcanzara el menor precio, y *siempre existe la posibilidad de que el valor rompa el soporte de la gama operativa y continúe la etapa descendente*. En la figura 5.17 se puede ver un ejemplo de Etapa 1.

³⁶ En el desarrollo de la teoría [Wei98] especifica que es conveniente el uso de esa dimensión temporal para la media. Allí se utiliza una media móvil de 30 semanas, pues usa gráficos semanales.

³⁷ Zona donde los precios se mueven en forma lateral y se pueden establecer con las crestas y valles una línea de tendencia superior (resistencia) y una línea de tendencia inferior (soporte).

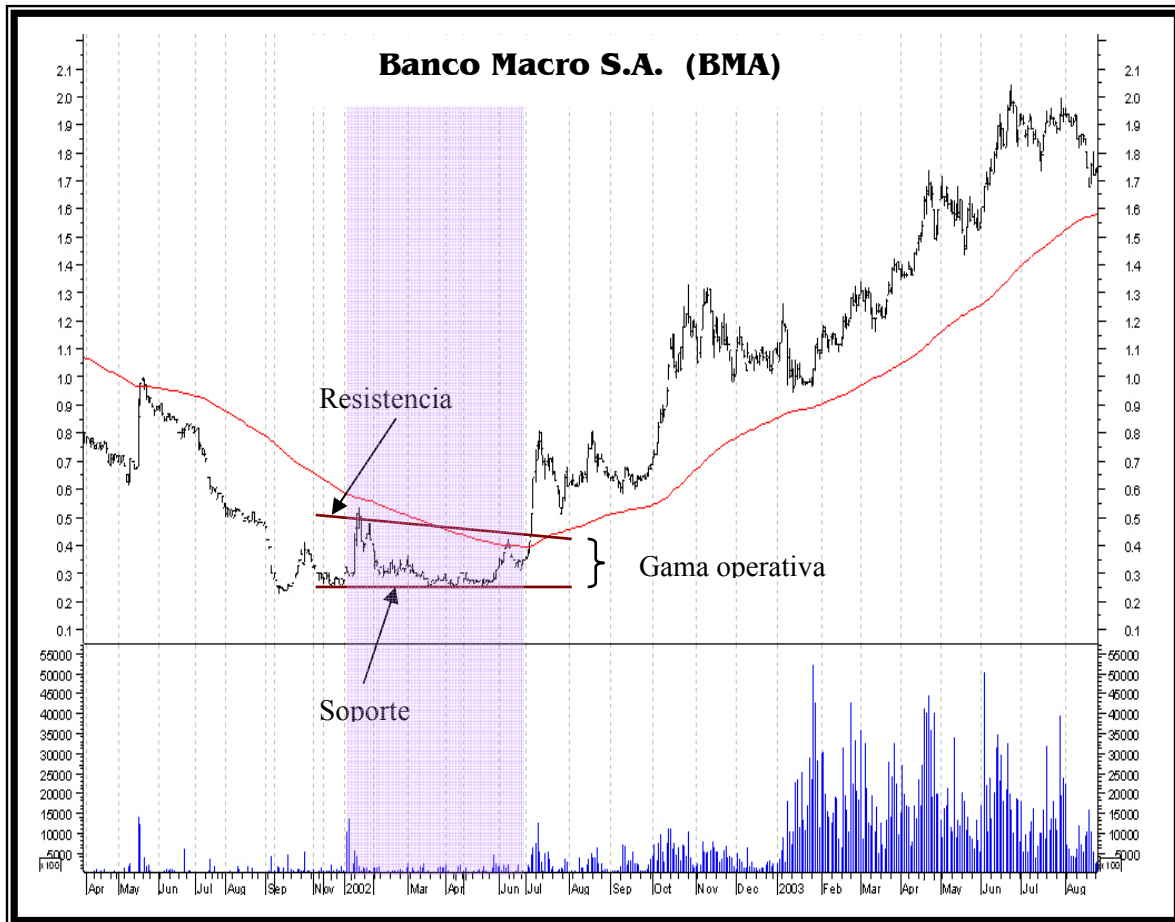


Figura 5.17: Etapa 1: la base, (zona sombreada). Gráfico de Banco Macro S.A. desde abril de 2001 hasta agosto de 2003

5.7.3.2 Etapa 2: alcista

El momento ideal de compra según [Wei98] es cuando el valor ha salido de etapa 1 y empieza a recorrer el camino alcista de etapa 2.

El valor sube encima de su media móvil y produce la ruptura de la línea de tendencia superior de la gama operativa (resistencia). Este movimiento, al igual que en todos los patrones vistos, debe estar acompañado de un importante aumento de volumen como confirmación.

Generalmente luego de esta ruptura el precio baja hacia la zona de la resistencia rota en la subida (movimiento denominado *estirón*), y esta sería la segunda y última oportunidad de compra desde el punto de vista de un inversionista³⁸.

Al comienzo de esta etapa la media móvil empieza a tener *pendiente positiva* y dicha pendiente se mantiene durante toda la etapa. Además el precio se mueve siempre por encima de la media, se producen crestas y valles ascendentes, en donde cada cresta es superior a la anterior y lo propio para los valles.

Hacia el final de la etapa el valor se acerca cada vez más a su media y la pendiente de la misma empieza a reducirse, en ese momento aunque el valor esta todavía en etapa 2, la compra del valor es riesgosa. En la figura 5.18 se puede ver un ejemplo de etapa 2.

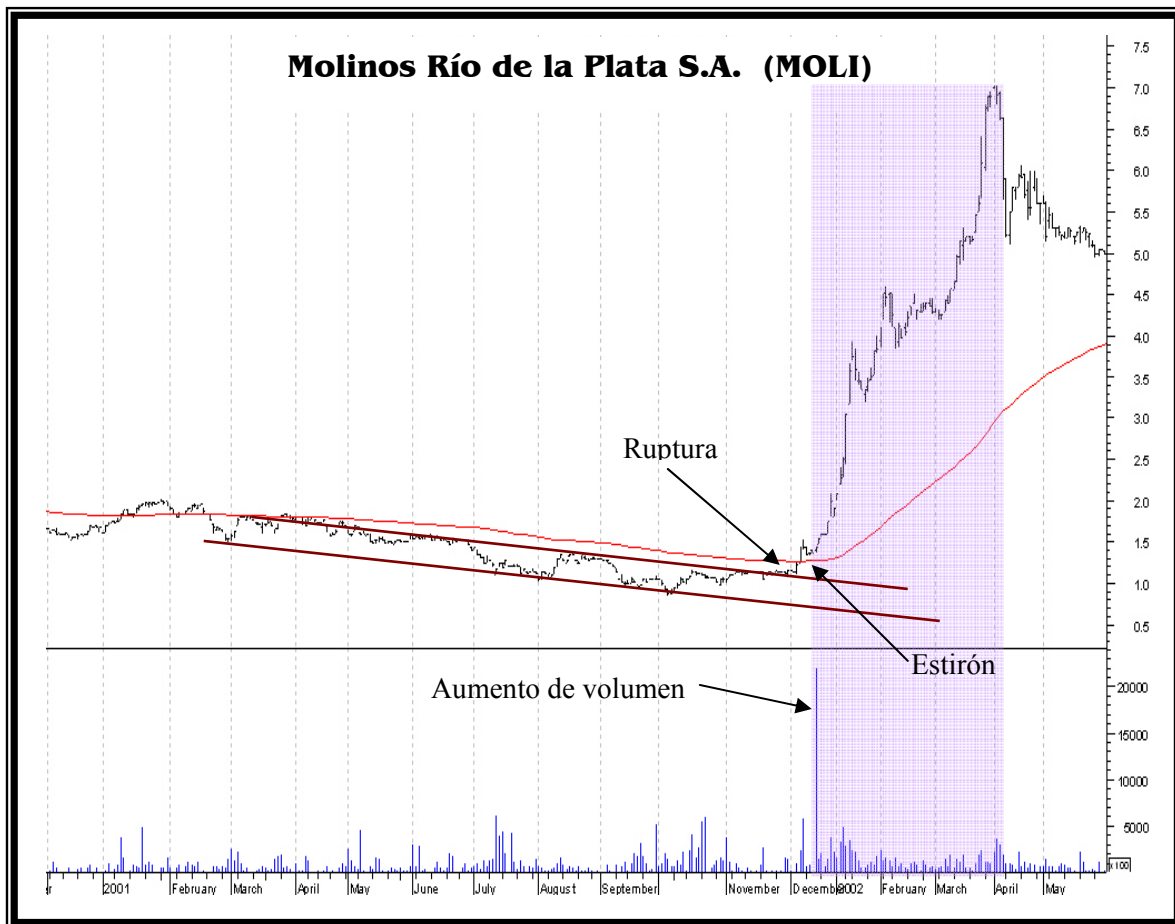


Figura 5.18: Etapa 2: alcista, (zona sombreada). Gráfico de Molinos Río de la Plata S.A. desde diciembre de 2000 hasta junio de 2002, notar el incremento de volumen en la ruptura de la línea de tendencia

³⁸ [Wei98] hace una diferencia entre inversionista (que invierte a mediano y largo plazo) y operador (que invierte en el corto plazo - uno a cuatro meses -)

5.7.3.3 Etapa 3: el techo

Cuando las fuerzas de la oferta y la demanda se equilibran nuevamente, se empieza a revelar un nuevo movimiento lateral de los precios. Esta etapa es el espejo superior de la etapa de base, la etapa 1.

En este momento la media móvil pierde su pendiente positiva y pasa a tener *pendiente nula*, que se mantiene durante esta etapa.

El valor empieza a operar por encima y por debajo de la media móvil, conformando nuevamente una gama operativa con soporte y resistencia.

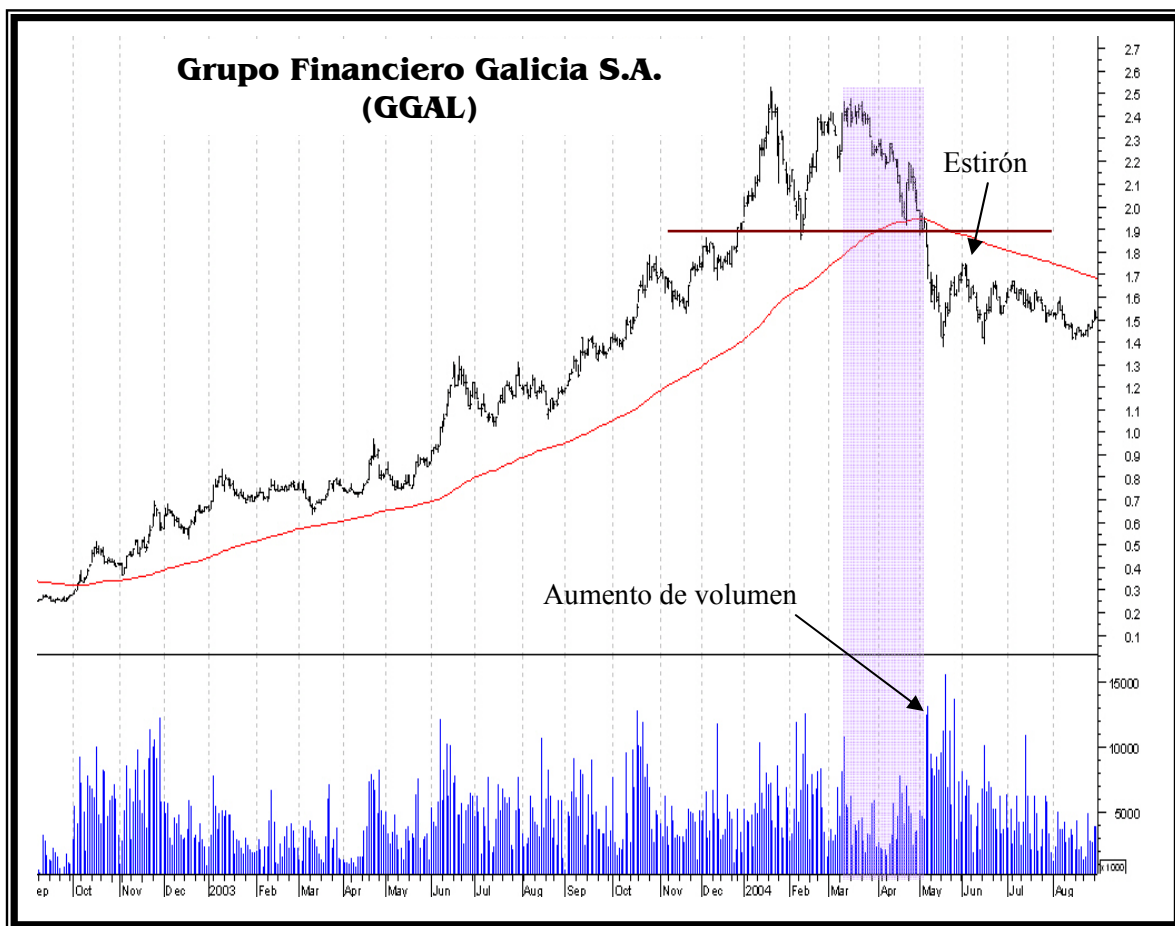


Figura 5.19: Etapa 3: el techo, (zona sombreada). Gráfico de Grupo Financiero Galicia S.A. desde septiembre de 2002 hasta agosto de 2004, notar el incremento de volumen en la ruptura de la línea de tendencia y que la etapa, como se dijo, es corta y volátil.

El volumen operado es generalmente fuerte durante el transcurso de la etapa 3 y los movimientos son bruscos y violentos (recuérdese que los movimientos superiores son más *volátiles* y violentos que los inferiores - sección 5.7.1 -), por lo que en general la etapa 3 es más violenta y *corta* que la etapa 1.

En esta etapa hay que ser muy cauto, se puede vender cuando la media se ha aplanado, pero es bueno recordar que *existe la posibilidad de que el valor rompa la resistencia de la gama operativa y continúe su ascenso*. En la figura 5.19 puede verse un ejemplo de etapa 3.

5.7.3.4 Etapa 4: bajista

Cuando la fuerza de la oferta empieza a superar a la fuerza de la demanda comienza a gestarse la etapa bajista. El valor baja debajo de su media móvil y quiebra la resistencia de su gama operativa de etapa 3. Hay que recordar que a diferencia del caso alcista, no es necesario que en la quiebra ocurra un incremento de volumen, ya que un valor puede “caer por su propio peso”, mientras que en la suba siempre es necesaria la apuesta fuerte de los compradores. Si se produce aumento de volumen, la señal será aún más clara y al igual que en la suba, puede producirse un *estirón* hacia el punto de ruptura. Si todavía no se ha vendido debe venderse en la ruptura, ya que el estirón no siempre ocurre y hay que ser muy precavido en este lado como el lector puede inferir.

La media móvil comienza a tener *pendiente negativa* y se mantiene con dicha pendiente durante toda la etapa, al mismo tiempo el precio del valor se mantiene bajando debajo de la media.

Los precios efectúan valles y crestas descendentes, en donde cada valle produce un nuevo mínimo y cada cresta es inferior que la anterior. En la figura 5.20 puede verse un ejemplo de etapa 4.

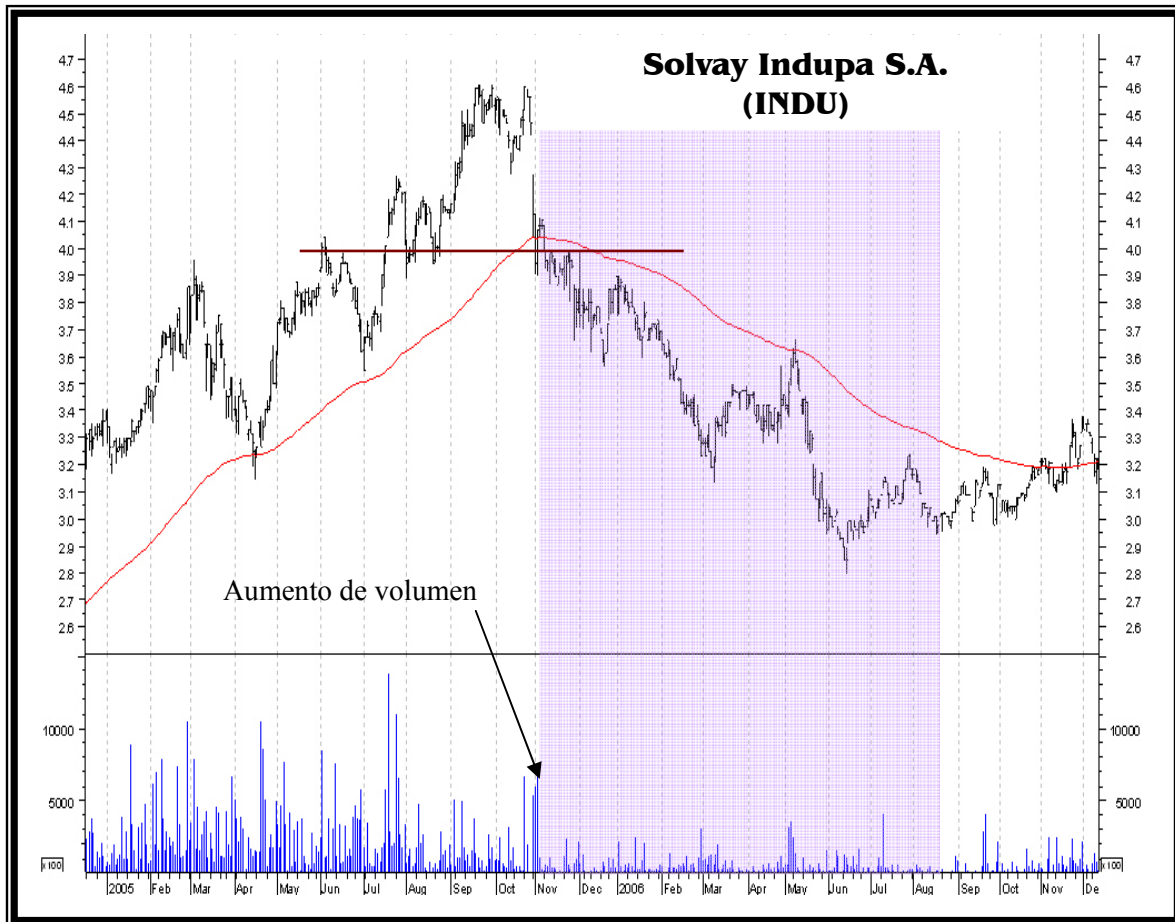


Figura 5.20: Etapa 4: bajista, (zona sombreada). Gráfico de Solvay Indupa S.A. desde diciembre de 2004 hasta diciembre de 2006, notar el incremento de volumen en la ruptura de la línea de tendencia (aunque en las rupturas a la baja su importancia es menor).

Capítulo 6

RECONOCIMIENTO DE PATRONES

“Lo que no se parece a nada no existe”.

Paul Ambroise Valéry.

6.1 Introducción

El reconocimiento de patrones es una característica de todos los seres vivos, pero la forma del reconocimiento difiere. Por ejemplo, un ser humano puede *reconocer* a otro por su imagen en un retrato, por su voz, por un manuscrito, etc. Un perro puede *identificar* a una persona o a otro perro por su olor distintivo a muchos metros de distancia; pero es claro que un ser humano no puede identificar a otro ser humano por el olor, ni un perro puede identificar algo en una fotografía.

Además es importante recalcar que un reconocimiento no siempre se efectúa con algún sentido, por ejemplo en el transcurso de una conversación, se podría identificar un viejo argumento, que fue usado en el pasado. Todos los ejemplos anteriores, son muestras de *reconocimiento*.

En definitiva, el reconocimiento de objetos (físicos o abstractos) es una tarea cotidiana que todos los seres vivos realizan en su interacción con el ambiente en el que se desenvuelven [Mar01, FK99].

El desarrollo de métodos que sean capaces de imitar el proceso de reconocimiento de un objeto ha evolucionado a lo largo del tiempo, encontrándose la necesidad de desarrollar sistemas automatizados *inteligentes* con ese fin. En estos sistemas, los objetos se representan de una manera conveniente para el tipo de procesamiento a la que estarán sujetos. A tales representaciones se las denomina *patrones*.

Por ejemplo, en el tema que compete a este trabajo, el analista técnico bursátil, desea identificar un cambio de la tendencia o una continuación de la tendencia de precios, utilizando para ello el reconocimiento de patrones (que fueron descriptos en

detalle en el capítulo 5) en las series temporales, puesto que éstos patrones, en el pasado, representaron un cambio o una continuación de tendencia.

Se pueden señalar dos tipos de reconocimientos de patrones, el que intenta descubrir la existencia de un patrón u otro, en este caso no se sabe a priori cual patrón se presentará; otra forma de reconocimiento se da cuando se sabe que patrón se presentará, y lo que se desea descubrir es la forma que tendrá el patrón, en este caso, se tiene seguridad de la ocurrencia del patrón, pero éste puede presentarse de distintas maneras.

En este capítulo se mostrará el proceso completo del reconocimiento de patrones, analizando cada una de sus etapas. Esta introducción es necesaria para comprender el proceso de desarrollo utilizado en el caso de estudio del capítulo 7, como así también para ubicar a los HMM en el proceso de reconocimiento de patrones, y en general para entender el proceso por el cual los patrones del análisis técnico bursátil son encontrados y clasificados.

Es importante señalar, que dicha introducción se realizará de manera genérica, abordando sólo los temas básicos que importan para la consecución del presente trabajo; el tema de reconocimiento de patrones es muy amplio y el lector que desee extender sus conocimientos del tema, en [Mur01, FK99] encontrará una completa guía con varios enfoques sobre el tema, en [Fuk90] se trata el tema desde el enfoque estadístico, en [BKPP06] desde el enfoque de la lógica difusa (*Fuzzy logic*) y en [WH03] se puede encontrar una completa aplicación al específico tema del reconocimiento del habla y del lenguaje natural.

6.2 Conceptos y definiciones

En el reconocimiento de patrones existe una relación entre patrones y clases (se busca determinar a que clase pertenece un patrón), lo que intuitivamente lleva a la consideración de la necesidad de una *clasificación*, que se concibe como el acto de decidir si cierta entidad pertenece o no a una determinada categoría, en función de sus características. Por lo tanto, la clasificación implica un proceso, que puede considerarse, de *toma de decisión* por parte del clasificador, por ende el concepto de sistema para reconocimiento de patrones comprende a todos aquellos sistemas, que basándose en patrones, toman una decisión.

Para realizar con éxito el proceso de clasificación se necesita que el sistema cuente con cierta información. Por un lado, se encuentra la noción del problema planteado, en mayor o menor medida, por quién efectúa la clasificación. Por otro lado, está la habilidad de descubrir en la entidad u objeto a clasificar, determinadas características sobresalientes que en conjunción con el conocimiento previo, son útiles para tomar una *decisión* con mayor o menor acierto; de cualquier modo, en el proceso de toma de decisiones, no se requiere conocer todos los detalles del objeto, sino solamente un *conjunto de características* relevantes que lo identifican.

Resumiendo, en un problema de reconocimiento de patrones se tendrá [Mar01]:

- Un conjunto de *clases*, entre las cuales se encuentra la entidad u objeto a reconocer. Como puede inferirse los objetos o entidades a clasificar representan categorías de clases para el clasificador.
- Un cierto conjunto de representaciones (correspondientes a cada clase), denominados *patrones*.
- De cada patrón se puede extraer información que lo caracteriza, es decir, un conjunto de *características*, dichas características son extraídas directamente de los datos.

Ejemplo 6.1 (basado en un ejemplo de [FK99]): Considérese el problema de reconocer los patrones que identifican a una persona en alguna de estas cuatro clases:

- Clase 1: Personas de alta estatura con sobrepeso.
- Clase 2: Personas de alta estatura sin sobrepeso.
- Clase 3: Personas de baja estatura con sobrepeso.
- Clase 4: Personas de baja estatura sin sobrepeso.

Supóngase también que se cuenta con una amplia base de datos con datos sobre las personas a clasificar, como edad, peso, trabajo, estudios, altura, ingresos, etc.

Lo primero que se debe hacer es encontrar las características que determinarán los patrones. Como el problema es simple, sin un amplio conocimiento previo, se puede aseverar que la edad no sería una característica que incluya un patrón para este caso. Una elección razonable sería el par (peso, altura).

Cada par (peso, altura), conformará el patrón de características de una persona en particular, lo que resta es clasificar los patrones de acuerdo a las clases citadas para el problema, es decir, de manera intuitiva, dos patrones distintos deberían asignarse a la misma clase, si sus características son *similares*, y asignarse a clases distintas sin no lo son. En la figura 6.1 puede apreciarse la representación de los espacios de clases, patrones y características para este ejemplo

□

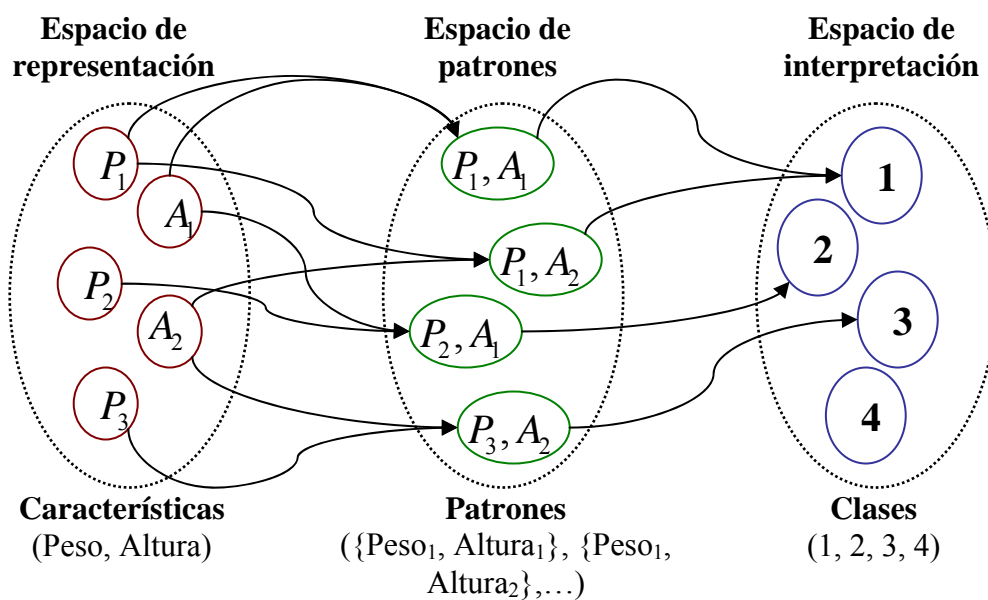


Figura 6.1: Espacios de características, patrones y clases para el ejemplo 6.1.

6.2.1 Clases

Las clases son *categorías* de objetos o entidades asociadas con conceptos o prototipos. Si se piensa en M clases distintas para un problema en particular, se denota el conjunto de clases como [Mar01]:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_M\} \quad (6.1)$$

Donde Ω es el conjunto de todas las clases del problema, conocido como el *espacio de interpretación*. Generalmente se extiende el conjunto de clases, para incorporar una *clase de rechazo* ω_0 , que representa a los patrones que no se corresponden a ninguna de las M clases determinadas para el problema, por lo que se suele definir el siguiente conjunto extendido Ω^* como:

$$\Omega^* = \{\omega_1, \omega_2, \dots, \omega_M, \omega_0\} \quad (6.2)$$

6.2.2 Patrones

En sentido amplio, un patrón es un *conjunto de características que describen el objeto que se quiere identificar*. Así, un patrón p conformado por N características (que pueden representarse en forma de N variables X_i con $i = 1, \dots, N$), se define como [Mur01, Fuk90]:

$$p = \{x_1, x_2, x_3, \dots, x_N\} \quad (6.3)$$

Los distintos tipos de patrones se distinguirán por el significado, la naturaleza y la cantidad de estas características.

6.2.3 Características

Las características son *atributos* o *primitivas* que son útiles en la caracterización de un patrón. Así como los patrones dependen de las clases de objetos que se quieren identificar, las características dependen de los patrones que se quieren reconocer. La

elección adecuada de las características en un problema de reconocimiento de patrones es de vital importancia (como se ejemplifica en el ejemplo 6.1).

En la resolución de problemas se desea contar con un *reducido* número de características, un previo conocimiento del problema puede ayudar en la tarea de encontrar dicho conjunto reducido. Como ejemplo se podría citar el problema de reconocer patrones en un electrocardiograma, (ver ejemplo 6.2), un muestreo de 10 segundos a 500 Hz resultaría en 5000 muestras, usar las 5000 muestras como características para los patrones del problema sería desastroso, afortunadamente como se verá en el ejemplo citado, el conocimiento previo del problema, permite la elección de un reducido conjunto de características.

Si se piensa en N características para un patrón p en particular, entonces el patrón se caracteriza como se dijo en la sección anterior por el conjunto $\{x_1, x_2, \dots, x_N\}$, y dicho conjunto [Mur01, Fuk90]:

$$\{x_1, x_2, \dots, x_N\} \in P \quad (6.4)$$

Donde:

$$P = \{X_1\} \times \{X_2\} \times \{X_3\} \times \dots \times \{X_N\} \quad (6.5)$$

Como puede verse, P es un espacio cuya dimensión esta definida por el número N de características consideradas y es el producto cartesiano de los conjuntos de valores de dichas características. Este espacio, es llamado también, el *espacio de representación*.

Ejemplo 6.2 (basado en un ejemplo de [Mur01]): Supóngase que se desean reconocer patrones provenientes de un electrocardiograma (ECG), para verificar si pertenecen a alguna de estas cuatro clases (ver figura 6.2) que se corresponden con distintas condiciones fisiológicas del corazón:

- N – normal.
- HVI – hipertrofia ventricular izquierda.
- HVD – hipertrofia ventricular derecha.
- IM – infarto de miocardio.

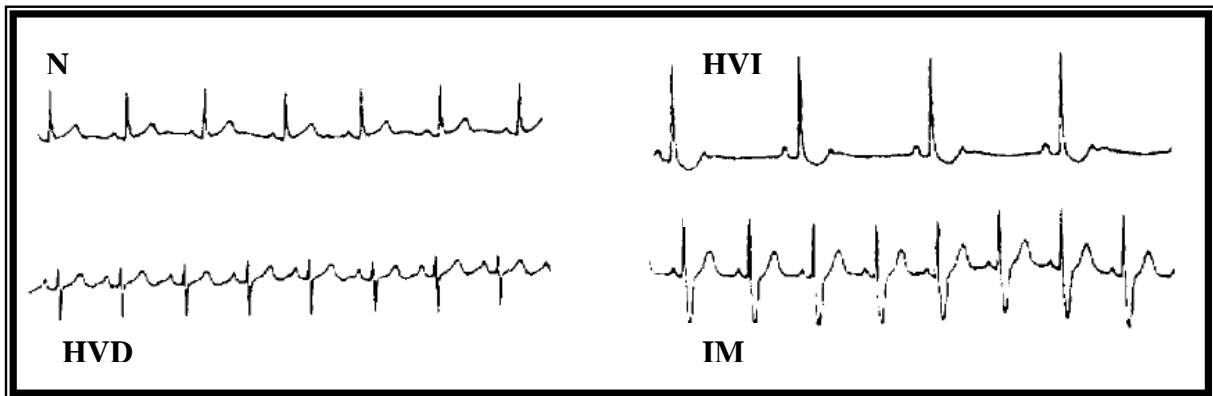


Figura 6.2: Electrocardiogramas de 4 clases de diagnóstico: (N) Normal; (HVI) Hipertrofia ventricular izquierda; (HVD) Hipertrofia ventricular derecha; (IM) Infarto de miocardio.

Como se puede apreciar en la figura 6.2, el ECG, muestra “ondas encapsuladas” que se repiten a tiempos más o menos regulares. En la figura 6.3 se muestran las componentes, que secuencialmente son P, Q, R, S y T. Estas ondas reflejan la actividad eléctrica de las distintas partes del corazón. La onda P, refleja la actividad auricular del corazón, y el resto (Q, R, S y T) reflejan la consiguiente actividad ventricular.

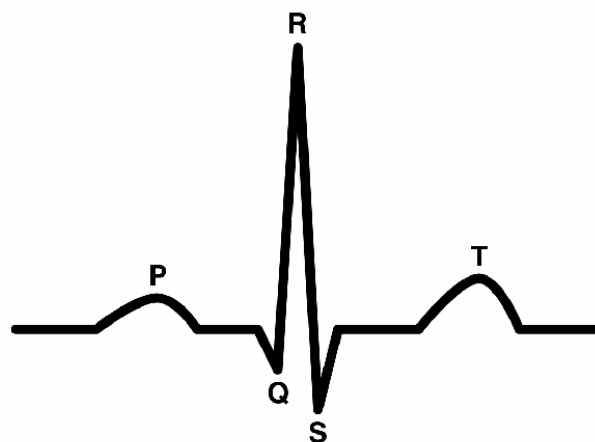


Figura 6.3: Ondas encapsuladas de un ECG, compuestas secuencialmente por las ondas P, Q, R, S y T.

En este caso las clases para el reconocimiento de patrones son las cuatro ya nombradas, los patrones serían la representación de las ondas encapsuladas, es decir, los ECG's y entre las características que se pueden elegir, se tiene la relación entre las amplitudes de las ondas Q y R, se puede utilizar el cociente Q/R.

Si se tomara esa única característica, como puede observarse en la figura 6.2, un cociente muy similar se obtendría para los ECG's que se corresponden a las clases N y HVI, y lo mismo sucedería con los ECG's de las clases IM y HVD. En el primer caso se podría solucionar utilizando características extras, pero el segundo caso es un problema inherente al método, es decir, usando P, Q, R, S y T no se puede tener certeza del reconocimiento y aún un experto en el tema tendría un alto porcentaje de error usando este método estándar. Este ejemplo demuestra que a menudo en el reconocimiento de patrones, los sistemas no se encuentran libres de error y lo que se persigue es limitar dicho error. En la práctica, se busca que los sistemas tengan una performance igual o mejor que la que un experto humano tendría, aunque el sistema no esté cien por ciento libre de error.

□

6.3 Sistemas de reconocimiento de patrones

El objetivo de los sistemas de reconocimiento de patrones es el de realizar un mapeo entre el espacio de representación y el espacio de interpretación. Por ese motivo, la principal tarea de un sistema de reconocimiento de patrones es la asignación de cada patrón de entrada a una clase. Para lo cual, una vez establecido el conjunto de clases, una de las tareas básicas a realizar, es la construcción del *clasificador*, o *regla de clasificación*.

Si un patrón p está caracterizado por N características, entonces la regla de clasificación puede verse como una función d definida de la siguiente manera [FK99]:

$$d : P \rightarrow \Omega^* \quad \forall p \in P, \exists d(p) \in \Omega^* \quad (6.6)$$

Es decir, una función que se define sobre los patrones y posee su imagen en el conjunto extendido de clases para el problema.

Un sistema de reconocimiento de patrones consta generalmente de 3 etapas [Mur01, FK99]:

1. Adquisición de datos
2. Extracción de características
3. Clasificación de patrones

En la figura 6.4 se puede observar una representación del proceso completo.

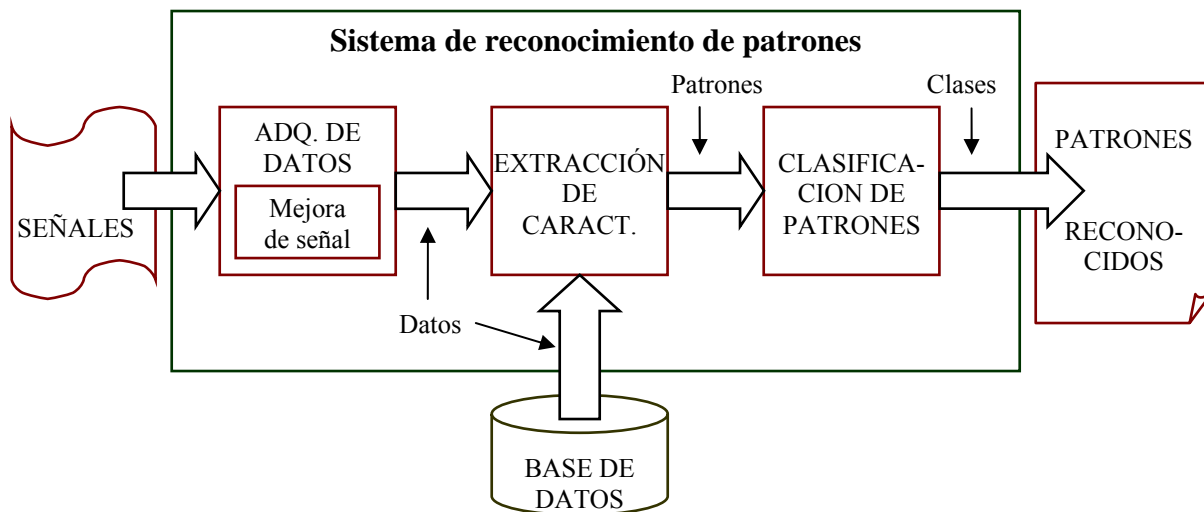


Figura 6.4: Representación de un sistema de reconocimiento de patrones en forma general.

6.3.1 Adquisición de datos

La etapa de adquisición de datos concierne a la obtención de datos directamente desde el objeto a estudiar a través de algún tipo de mecanismo detector de señales, como un sensor, un micrófono, una cámara de video, etc. En numerosos casos consta de una etapa intermedia de acondicionamiento o preparación de la señal (reducción de ruido, mejora de la señal, etc.) con el objetivo de ayudar y hacer que el trabajo de la etapa siguiente sea más efectivo.

A veces, los datos no provienen directamente de la etapa de adquisición, sino que se encuentran almacenados en una base de datos con el propósito de disponer inmediatamente de la cantidad de datos necesarios para el entrenamiento del clasificador (estimación de parámetros en el caso de que en la etapa de clasificación se utilicen HMM), o con el propósito de conseguir medidas comparativas de la capacidad de reconocimiento del sistema con respecto a otros sistemas (o a otras versiones del mismo sistema).

6.3.2 Extracción de características

La etapa de extracción de características³⁹ se refiere al proceso por el cual el sistema adquiere, a través de la aplicación de algoritmos que dependerán del caso, un conjunto limitado de *magnitudes* que constituyen los aspectos prominentes de la entidad u objeto a clasificar a partir de los datos de entrada obtenidos en la etapa de adquisición para ese objeto. Es decir, esta etapa toma como entrada los *datos en bruto* recogidos en la etapa anterior y extrae los *valores* de las características fundamentales, previamente elegidas como tales, para conformar un patrón de características. Dicho patrón es destinado a la etapa siguiente para su clasificación.

Recuerde el lector, que es función del diseñador del sistema de reconocimiento, elegir apropiadamente las características relevantes para el problema, para lo cual como se expresó en la sección 6.2.3 y se pudo constatar en el ejemplo 6.1, es clave el conocimiento previo del problema. En esta sección se presupone que las características adecuadas ya han sido elegidas.

En la etapa de extracción de características hacen su aporte un conjunto de técnicas relativamente contemporáneas a los HMM, englobadas en lo que se conoce como *minería de datos*. Las técnicas de la minería de datos se refieren a la extracción o “minería” de conocimiento a partir de grandes cantidades de datos [HK06].

La minería de datos interviene en las *maneras* en la que los patrones son extraídos [FK99], es decir, interviene en la etapa de extracción de los posibles valores de las características elegidas, aportando algunos lineamientos útiles para su consecución.

³⁹ Genéricamente se denomina extracción de características a la etapa, pero se debe tener claro que lo que se extrae a partir de los datos en bruto, son los *valores* de dichas características, previamente elegidas por el diseñador del sistema.

Haciendo una abstracción hacia el objetivo principal de esta tesis, se podría señalar que la minería de datos ayudará en la tarea de *encontrar los patrones* y los HMM en la tarea de *clasificarlos* (reconocerlos).

A continuación se verán algunas de las técnicas de la minería de datos, que serán útiles para el propósito de este trabajo, el lector que desee ampliar conocimientos sobre este tema, en [HK06] y [Kan03] encontrará dos excelentes opciones.

Es importante señalar, que algunas de las técnicas que serán presentadas, son utilizadas también en la etapa de clasificación; sin embargo en este trabajo, todas las técnicas de minería de datos utilizadas, se usarán en la etapa de extracción de características, y en general serán usadas en el proceso de obtener un conjunto reducido de observaciones a partir de los valores posibles de las características elegidas⁴⁰, para luego usar dichas observaciones en los HMM de la etapa de clasificación.

Existe polémica en cuanto a las fronteras existentes entre la minería de datos y otras disciplinas como estadística, inteligencia artificial, etc. No es objetivo de este trabajo entrar en esa discusión; lo que si es cierto es que la mayoría de las técnicas utilizadas en la minería de datos ya tenían una larga tradición en otros campos como los nombrados, aunque esto no invalida a la minería de datos como un disciplina que toma elementos de otras para enfocarse en un problema particular.

La razón del párrafo anterior, es porque el lector puede encontrar que las técnicas que serán expuestas a continuación, poseen cierta pertenencia también a otras disciplinas.

6.3.2.1 Discretización de las características

Las técnicas de discretización de datos pueden utilizarse para reducir el número de valores de una característica que posea valores continuos, esto se consigue de forma genérica, dividiendo el rango de dichas características en intervalos. Luego las etiquetas de los intervalos serán los nuevos datos *discretos* que reemplacen a los datos originales. El uso de éstas técnicas es obligatorio si en la etapa de clasificación

⁴⁰ Es importante dejar en claro que las técnicas de esta sección se aplicarán sobre los *conjuntos de valores posibles de las características elegidas* gracias al conocimiento previo del problema y no sobre los datos en bruto.

se utilizan clasificadores que trabajen sólo con datos discretos, como es el caso de este trabajo. Por supuesto, si las características toman valores de un conjunto discreto, no serán obligatorias estas técnicas, pero pueden ser útiles para *reducir* los datos aunque sean discretos, pues muchas veces es conveniente un conjunto reducido de valores para las características.

Si la discretización es realizada correctamente, se produce en forma efectiva una simplificación del problema original, con los beneficios colaterales que esto significa.

Las técnicas de discretización pueden dividirse en dos grupos [HK06]:

- *Técnicas supervisadas*: Se refiere a las técnicas que utilizan información de las clases en las que serán englobados los datos; es decir, es conocido de antemano la forma de segmentación en clases de los datos, como así también a que clase debería pertenecer un dato dado en particular.
- *Técnicas no supervisadas*: Por el contrario, en las técnicas no supervisadas no se posee información de clase, o sea que a priori, no se conoce la forma final de la segmentación, ni a que clase debería pertenecer un dato dado.

En cualquier caso, un objetivo comúnmente perseguido por las técnicas de discretización, es el obtener particiones jerárquicas. Para entender el concepto de jerarquía, considérese la característica “edad”, se puede considerar a los valores en bruto de la característica como de “bajo nivel”, y los valores (niño, joven, adulto, anciano) serían de “alto nivel” dentro de una clasificación jerárquica, produciendo la discretización buscada.

Ejemplos de técnicas supervisadas de discretización son:

- la segmentación en intervalos proporcionales previamente elegidos, para lo que puede usarse el algoritmo 4.6 del capítulo 4 (segmental k -means).
- la segmentación manual, producto del conocimiento previo del problema.
- la segmentación basada en *entropía*, que utiliza intervalos que proporcionen mayor *ganancia*, está basado en la teoría de la información.

Ejemplos de técnicas no supervisadas de discretización son:

- Técnicas de *histograma*, ya sea con intervalos iguales en dimensión o con intervalos de frecuencias iguales.
- *Clustering*, que es básicamente la agrupación de los datos por conceptos de similaridad. Un *cluster* es una colección de datos similares entre sí, pero disimilares con los datos de otro cluster.

Los métodos de discretización supervisada pueden incorporar dos características que son relevantes:

- discretización *dinámica*: La clase asignada a un dato en particular puede cambiar con el tiempo, un ejemplo de utilidad de estos sistemas se da cuando el tamaño de las clases varía en relación a la variación de algún dato en bruto de donde se extraen las características.
- discretización *cualitativa*: En estos sistemas dos valores pertenecen a la misma clase mientras no exista un cambio cualitativo entre ellos, es decir se mide la relación entre dos datos consecutivos de acuerdo a algún parámetro elegido, y el segundo dato sólo se asigna a una clase distinta cuando se produce un cambio que sobrepasa un determinado umbral.

Una buena descripción de cada uno de las técnicas mencionadas en esta sección puede encontrarse en [HK06] y en [Kan03].

6.3.2.2 Árboles de decisión

Un eficiente método para la clasificación de datos⁴¹, es la utilización de árboles de decisión (en inglés, *decision trees*) [HK06, Kan03]. Los árboles de decisión conforman uno de los métodos de representación más usados de la lógica.

Un árbol de decisión es un diagrama de flujo, similar a la estructura de árbol, donde cada nodo interno representa el test de un atributo o característica, y cada rama representa un resultado del test de su nodo superior, y cada nodo hoja (o nodo terminal) tiene una etiqueta de clase. En la figura 6.5 puede verse un ejemplo de

⁴¹ Es importante no confundir “clasificación de datos” con “clasificación de patrones”, tema que se tratará en la sección 6.3.3

árbol de decisión, donde A es una característica que puede tomar valores reales, B por el contrario, es una característica que toma valores booleanos, C y D pueden tomar los valores E , F y G . En este caso, el árbol produce siete clases distintas.

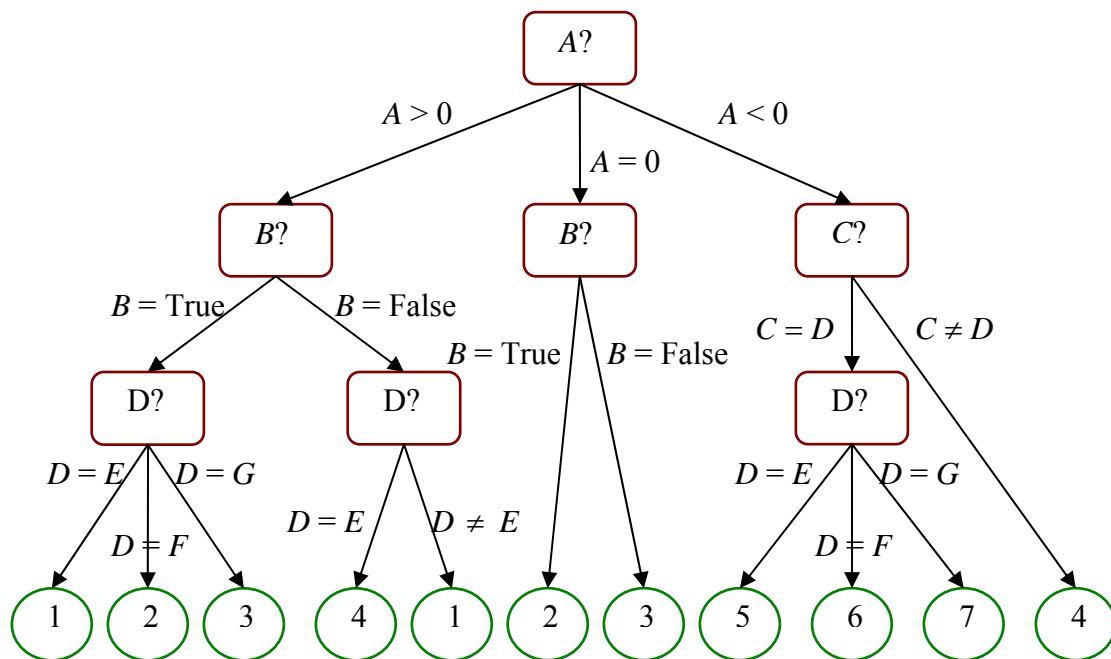


Figura 6.5: Un ejemplo de árbol de decisión donde se produce el test de las características A , B , C y D , el resultado es la clasificación en 7 clases distintas (números 1 al 7 en la figura).

En el análisis de decisión, un árbol de decisión puede utilizarse para representar visualmente y explícitamente las decisiones y la toma de decisiones. En minería de datos, un árbol de decisiones describe datos, pero no decisiones; sino más bien, la resultante de clasificación de los árboles pueden ser un insumo para la toma de decisiones, y esto es compatible con el uso que se le dará en este trabajo.

Existen algoritmos para la generación de árboles de decisión a partir de conjuntos de datos de entrenamiento, un ejemplo son el ID3 y su evolución el C4.5 [HK06, Kan03], ambos diseñados por J. Ross Quinlan. Estos algoritmos son un ejemplo de algoritmos voraces (greedy en inglés), en los que el árbol se construye de la raíz hacia las hojas en forma recursiva utilizando la estrategia de divide y vencerás (divide and conquer). Estos procedimientos se describen en detalle en [HK06] y [Kan03], dichos algoritmos intentan encontrar además el *menor* árbol de decisión

posible. De cualquier manera, ninguno de ellos garantiza encontrar el menor árbol de decisión que sea compatible con los datos de entrenamiento, sólo consiguen una buena aproximación utilizando técnicas basadas en la teoría de la información. De hecho, según [Kan03] este problema es NP-completo.

En el caso de la aplicación de los árboles de decisión a la etapa de extracción de características, la idea es tomar una secuencia de datos brutos de las características seleccionadas, (datos provenientes de la etapa de adquisición, relativos a las características), y generar con el auxilio de un árbol de decisión un conjunto reducido de secuencias de clases de datos, que en el caso específico de este trabajo, podrán ser usadas como secuencias de observaciones que luego servirán para los HMM usados en la etapa de clasificación.

6.3.2.3 Reglas IF-THEN

Una manera de clasificación de datos, es aquella que utiliza *reglas* para conseguir dicho propósito. Un clasificador basado en reglas, utiliza un conjunto de reglas IF-THEN para la clasificación. Una regla IF-THEN es una expresión de la forma [HK06]:

$$\text{IF } \textit{condición} \text{ THEN } \textit{conclusión} \quad (6.7)$$

Considérese el siguiente ejemplo:

Regla 1: IF $A = 1$ AND $B = 2$ THEN $Clase = 3$

Donde la primera parte de la regla (la del IF) es conocida como la *precondición de la regla* y la segunda parte (la del THEN) es conocida como la *consecuencia de la regla*. La precondición de la regla consiste en uno o más test de características y el conectivo lógico AND si hubiera más de un test, y la consecuencia de la regla consiste en la clase que la regla predice para una tupla de características en particular, si la precondición de la regla es cierta para dicha tupla. La Regla 1 del ejemplo anterior puede escribirse de la siguiente manera:

Regla 1: $(A = 1) \wedge (B = 2) \Rightarrow Clase = 3$

Si la precondición de una regla es verdadera para una tupla, entonces se dice que la tupla *satisface* la regla.

Es obvio que no puede haber dos reglas con la misma precondición, pero distintas consecuencias, además se debe tener en cuenta que existen dos problemas que pueden ocurrir si no se tiene cierto cuidado cuando las reglas IF-THEN son elaboradas:

- Un problema es que para una determinada tupla, exista la predicción de más de una clase, debido a que satisfaga a más de una regla IF-THEN
- Otro problema, es que una tupla en particular no satisfaga ninguna regla y por consiguiente no le sea asignada ninguna clase.

Para atacar el primer problema se puede utilizar una estrategia de elegir la precondición más grande, es decir la que más test de características posea; o bien tener un orden en las reglas, de modo que si una tupla satisface dos reglas, sólo se considera la que posea orden mayor.

El segundo problema puede resolverse utilizando una clase por defecto, de modo que si una tupla no satisface ninguna regla, le será asignada dicha clase.

Proposición 6.1

Si los conjuntos de tuplas que satisfacen las precondiciones de las reglas IF-THEN son mutuamente exclusivos y exhaustivos, entonces:

- a) No puede haber tuplas a las que se le asignen más de una clase.*
- b) No puede haber tuplas a las que no pueda asignárseles una clase.*

Demostración: a) Supóngase por el absurdo, que una tupla X satisface las siguientes reglas:

$$\begin{aligned} &\text{IF } A \text{ THEN Clase} = 1, \\ &\text{IF } B \text{ THEN Clase} = 2. \end{aligned}$$

Entonces por definición de regla, X satisface A y X satisface B , lo que implica que X pertenece al conjunto de tuplas que satisfacen la precondición A y también pertenece al conjunto de tuplas que satisfacen la precondición B , pero por hipótesis esto

indicaría que $A = B$ lo que no está permitido; por lo tanto se llegó a un absurdo y el absurdo provino de suponer que X satisface ambas reglas.

b) Como por hipótesis se tiene que los conjuntos de tuplas que satisfacen las precondiciones son mutuamente exclusivos y exhaustivos, la prueba es trivial.

□

Los árboles de decisión (vistos en la sección anterior), pueden ser fácilmente convertidos a la clasificación por reglas IF-THEN [HK06]. La forma consiste en utilizar un camino por regla (desde la raíz hasta una hoja), utilizando las condiciones que figuran en las ramas del árbol en la precondición de la regla (conectándolas con conectivos AND si hubiera más de una rama) y usando la clase que figura en la hoja en la consecuencia de la regla.

Como ejemplo, las reglas IF-THEN que se corresponden al árbol de decisión del ejemplo 6.5 son:

- Regla 1: IF $A > 0$ AND $B = \text{True}$ AND $D = E$ THEN Clase = 1
- Regla 2: IF $A > 0$ AND $B = \text{True}$ AND $D = F$ THEN Clase = 2
- Regla 3: IF $A > 0$ AND $B = \text{True}$ AND $D = G$ THEN Clase = 3
- Regla 4: IF $A > 0$ AND $B = \text{False}$ AND $D = E$ THEN Clase = 4
- Regla 5: IF $A > 0$ AND $B = \text{False}$ AND $D \neq E$ THEN Clase = 1
- Regla 6: IF $A = 0$ AND $B = \text{True}$ THEN Clase = 2
- Regla 7: IF $A = 0$ AND $B = \text{False}$ THEN Clase = 3
- Regla 8: IF $A < 0$ AND $C = D$ AND $D = E$ THEN Clase = 5
- Regla 9: IF $A < 0$ AND $C = D$ AND $D = F$ THEN Clase = 6
- Regla 10: IF $A < 0$ AND $C = D$ AND $D = G$ THEN Clase = 7
- Regla 11: IF $A < 0$ AND $C \neq D$ THEN Clase = 4

Entre cada regla extraída del árbol se ubica una *disyunción* (OR lógico). El conjunto de reglas extraídas de un árbol de decisión cumple la siguiente proposición:

Proposición 6.2

Las reglas IF-THEN extraídas correctamente de un árbol de decisión garantizan que en la clasificación de las tuplas de valores de características mediante la disyunción de dichas reglas, a cada tupla se le asignará “exactamente” una clase.

Demostración: Como se tiene una regla por cada hoja del árbol, y por construcción del árbol, cada tupla puede mapear con una sola hoja. Se tiene que los conjuntos de tuplas que satisfacen las precondiciones de cada regla son mutuamente exclusivos. Por otro lado, como en un árbol de decisión existe una regla para cada posible combinación de valores de características en las tuplas, se tiene que los conjuntos de tuplas que satisfacen las precondiciones de cada regla son exhaustivos.

Lo anterior permite la aplicación de la proposición 6.1, con lo que concluye la prueba.

□

6.3.2.4 Regresión lineal simple

Cuando los datos que se corresponden a una cierta variable *independiente* X están distribuidos de manera lineal o *aproximadamente* lineal con respecto a los valores de otra variable *dependiente* Y , que es función de X (ver figura 6.6), entonces es posible aplicar un modelo de regresión lineal simple.

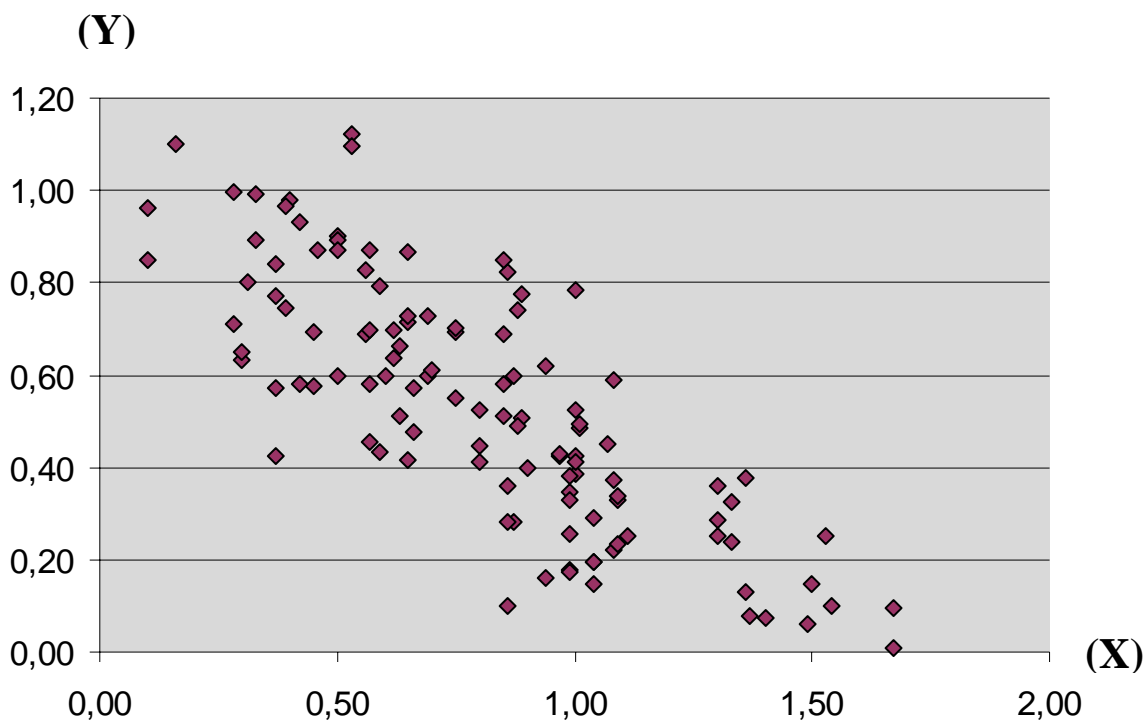


Figura 6.6: Gráfico de dispersión de los valores de Y en función de X

Más formalmente [Dev01], en un modelo de regresión lineal simple existen parámetros β_0 , β_1 y σ^2 tales que para cualquier valor fijo de la variable independiente X , la variable dependiente Y está relacionada con X , a través del modelo dado por la ecuación:

$$Y = \beta_0 + \beta_1 X + e \quad (6.8)$$

Donde la cantidad e es una variable aleatoria de la que se supone que posee una distribución normal con $E(e) = 0$ y $V(e) = \sigma^2$.

Como consecuencia los valores estarán distribuidos de manera aleatoria alrededor de lo que se conoce como *recta de regresión estimada* (ver figura 6.7), pues los parámetros β_0 y β_1 deben ser estimados. El método más comúnmente utilizado para estimar dichos parámetros se conoce como el *método de los mínimos cuadrados*.

Tomando b_0 y b_1 como la ordenada al origen y la pendiente respectivamente de la recta antes citada, entonces la desviación vertical desde cada punto (X_i, Y_i) a la recta $Y = b_0 + b_1 X$ es:

$$Y_i - (b_0 + b_1 X_i) \quad (6.9)$$

La suma de estas desviaciones al cuadrado para los X_i y los Y_i (con $i = 1, 2, \dots, N$), produce la siguiente función:

$$f(b_0, b_1) = \sum_{i=1}^n [Y_i - (b_0 + b_1 X_i)]^2 \quad (6.10)$$

Ahora bien, la idea del método de los mínimos cuadrados, es conseguir b_0 y b_1 de tal modo que se minimice dicha función, para lo cual se utilizan las *estimaciones de mínimos cuadrados* representadas por $\hat{\beta}_0$ y $\hat{\beta}_1$, y son tales que $f(\hat{\beta}_0, \hat{\beta}_1) \leq f(b_0, b_1)$ para cualquier b_0 y b_1 . Finalmente la *recta de mínimos cuadrados* es aquella tal que su ecuación es:

$$Y = \widehat{\beta}_0 + \widehat{\beta}_1 X \quad (6.11)$$

Los valores óptimos para $\widehat{\beta}_0$ y $\widehat{\beta}_1$ se obtienen al tomar las derivadas parciales de $f(b_0, b_1)$ con respecto a b_0 y a b_1 e igualar el resultado de ambas derivadas a cero; con lo que se obtiene (álgebra mediante), un sistema de ecuaciones lineales con valores desconocidos b_0 y b_1 . La única solución a este sistema son justamente las estimaciones de mínimos cuadrados.

La estimación de mínimos cuadrados para la pendiente β_1 de la recta de regresión real es:

$$\widehat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - (\sum X_i)(\sum Y_i) / N}{\sum X_i^2 - (\sum X_i)^2 / N} \quad (6.12)$$

La estimación de mínimos cuadrados para la ordenada al origen β_0 de la misma recta es:

$$\widehat{\beta}_0 = \frac{\sum Y_i - \widehat{\beta}_1 \sum X_i}{N} = \bar{Y} - \widehat{\beta}_1 \bar{X} \quad (6.13)$$

Donde N es la cantidad de datos, $\bar{X} = \frac{\sum X_i}{N}$ y $\bar{Y} = \frac{\sum Y_i}{N}$.

La regresión lineal simple se utiliza como una manera de predicción de datos [HK06], pero también interviene en la etapa de extracción de características como una herramienta para determinar o *aproximar* el valor de una característica.

Como ejemplo se puede citar el que justifica la inclusión de la regresión simple en este trabajo. Considérese como característica la “pendiente de la recta tangente a la media móvil exponencial de 30 períodos” (ver sección 5.6), si la media móvil conformara una función de la que se pudiera tomar una derivada, la solución sería inmediata, pero esto no es posible. Entonces el método de regresión lineal simple ofrece una buena aproximación, utilizando la estimación de β_1 y en el capítulo 7 se verá su aplicación.

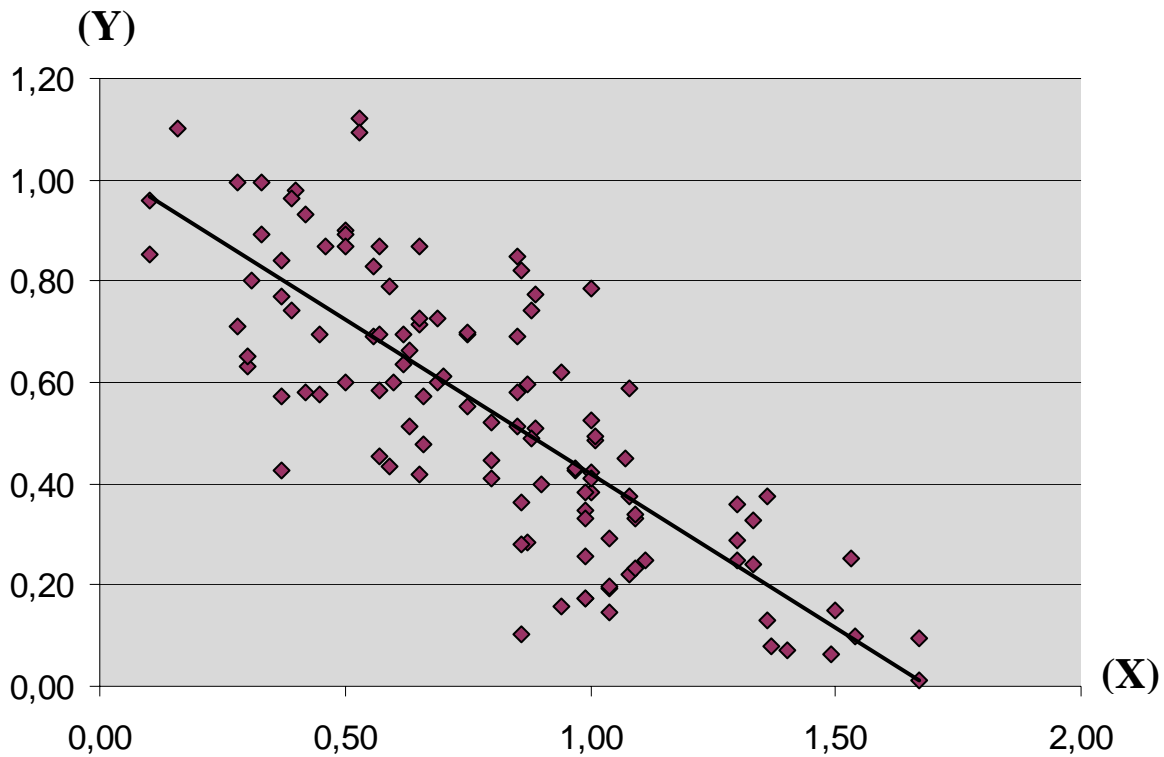


Figura 6.7: Gráfico de dispersión de los valores de Y en función de X con la recta de mínimos cuadrados correspondiente.

Es importante señalar que la aplicación del modelo de regresión lineal al ejemplo anterior, sólo resulta adecuado para una secuencia de valores de la media móvil no demasiado extensa, pues en ella será evidente un patrón *considerablemente* lineal, en concordancia con los requerimientos indicados en [Dev01], aún mas teniendo en cuenta que la media móvil es en esencia una suavización de los precios, lo que hace que los datos estén bastante más cerca de la recta que si se tomarán los precios del activo subyacente; por otro lado si se tomara una secuencia demasiado extensa, podrían tenerse más de una tendencia en la serie, lo que derivaría en no linealidad. Pero para el caso de ejemplo se necesita una muy reducida secuencia como se verá en su aplicación, por lo que el método es viable.

6.3.3 Clasificación de patrones

La etapa de clasificación de patrones es el núcleo, y la fase final de un sistema de reconocimiento de patrones, es la encargada de recibir los valores de las características, previamente extraídos y analizados en la etapa anterior, y proceder a la clasificación de los patrones que puedan encontrarse en las secuencias de dichos valores mediante la aplicación de un *clasificador*.

En general los clasificadores requieren ser *entrenados* previo a su utilización efectiva. Dicho entrenamiento consiste en ofrecer al clasificador, versiones de los patrones a clasificar, para estimar los parámetros que sean necesarios.

Finalmente, en el proceso de reconocimiento, el clasificador asigna al conjunto de valores de entrada una clase correspondiente al patrón de referencia que corresponda a través de la comparación con los patrones de referencia cuyos parámetros fueron estimados, en el caso de que se quiera reconocer un patrón entre varios; o una clase correspondiente a una forma de referencia para un patrón único, a través de la comparación con las posibles formas que el mismo pueda adoptar, en el caso de que lo que se quiera reconocer sea la forma de un patrón. La salida del sistema será el resultado de dicha clasificación y este resultado es en definitiva, el producto final del sistema completo de reconocimiento de patrones.

Existen diferentes enfoques para el diseño del clasificador que pueden ser catalogados como *supervisados* o *no supervisados* de acuerdo a si incluyen información de clase a priori o no. A continuación se mostrará una breve introducción a algunos de estos enfoques.

6.3.3.1 Clasificación por mínima distancia

También denominada clasificación por agrupamiento (*clustering* en inglés). El objetivo de estos clasificadores es el de organizar los patrones en grupos útiles utilizando algún tipo de medida de similitud. Es un método no supervisado de clasificación, pues las clases finales no son definidas de antemano.

Si cada clase está representada por un prototipo llamado centro de grupo (*cluster center*), se puede usar un clasificador basado en la mínima distancia para clasificar un nuevo patrón.

Este tipo de clasificación es útil cuando se quiere extraer algún significado de una gran cantidad de información sin clasificar, o en una fase exploratoria de investigación sobre reconocimiento de patrones, para evaluar similitudes en los datos.

Una buena guía sobre este tipo de clasificación puede encontrarse en [Mur01] y en [FK99], y un tratamiento integral puede consultarse en [HK06] y en [Kan03].

6.3.3.2 Enfoque estadístico

La clasificación estadística es un enfoque clásico en reconocimiento de patrones. Se basa esencialmente en el uso de modelos probabilísticos para la distribución de los valores de las características entre las clases, con el propósito de obtener las funciones de clasificación. La estimación de estas distribuciones se basa en la utilización de un conjunto de *patrones de entrenamiento*, cuya clasificación es conocida a priori. Es un método supervisado de clasificación pues las clases se conocen con antelación.

Los HMM se engloban dentro de esta tipo de clasificación, pero existen muchas variantes de la clasificación estadística, que dependen de si el modelo de distribución usado es *conocido*, *parametrizable*, o no. Recuerde el lector que en un HMM, es oculto.

El enfoque de la clasificación estadística es adecuado cuando los patrones de varias clases se solapan, esto sucede cuando son originadas por alguna distribución estadística.

Una buena guía de esta clasificación puede ser encontrada en [FK99] y también en [Mur01], un tratamiento integral en [Fuk90]. Para la clasificación por HMM referirse a la sección 6.4 de este trabajo, los HMM fueron extensamente introducidos en los capítulos 3 y 4.

6.3.3.3 Clasificación usando lógica difusa

La clasificación que utiliza lógica difusa o (*fuzzy logic* en inglés) se basa en que es muy común que la clasificación se realice con cierto grado de incertidumbre. O bien el resultado mismo de clasificación puede estar en duda; o bien algún patrón puede pertenecer en algún grado a más de una clase. Por ejemplo, una persona 1,75 metros de altura no pertenece plenamente a la clase “alta estatura”, pero al mismo tiempo no puede ser aceptado plenamente en la clase “baja estatura” (pensando en que sólo existen estas dos clases). Por lo tanto, es conveniente el uso de una clasificación difusa, en donde un patrón será miembro de cada clase con algún grado de pertenencia entre 0 y 1. Existen varias alternativas de clasificación difusa, como son la equivalencia en las relaciones y la agrupación difusa o (*fuzzy clustering*).

En ese sentido existen algoritmos que luego de que los centros de grupo (*cluster centers*) son determinados, asignan a cada patrón un conjunto definitivo de grados de pertenencia que establecen el grado de su clasificación en los distintos grupos (clusters).

Una buena guía que trata esta forma de clasificación puede encontrarse en [FK99] ó [Kan03] y un tratamiento integral en [BKKP06]

6.3.3.4 Clasificación estructural

El enfoque estructural en el reconocimiento de patrones a diferencia del resto de los enfoques utiliza la estructura de los patrones (de ahí su nombre). Es decir, en lugar de hacer un análisis basado exclusivamente en las características cuantitativas del patrón, este enfoque enfatiza el análisis de las interrelaciones entre las *primitivas* que permiten la comparación del patrón. De esta forma el reconocimiento puede asumir la forma de *matching* estructural, que permite evaluar si un objeto desconocido o parte de él se relaciona con un prototipo de clase.

Un tipo particular de clasificación estructural es la clasificación sintáctica. Ejemplos de patrones que son típicos del reconocimiento de patrones sintácticos son las huellas dactilares, cromosomas, etc. La analogía entre la estructura de algunos

patrones y la sintaxis de un *lenguaje* (que tiene una sólida base teórica) es la fundamentación para este tipo de clasificación.

Al introducir el concepto de *gramática formal* y lenguaje, se necesita diseñar clasificadores sintácticos que puedan clasificar un determinado patrón que se presenta ahora como una cadena de símbolos (un *string*). En general, dada una determinada clase, se tendrá una gramática cuyo lenguaje consistirá de los patrones designados para dicha clase. Para cada patrón desconocido (*string*), el clasificador sintáctico efectúa un proceso llamado análisis sintáctico (*parsing*) y determina si pertenece al lenguaje (clase). Una excelente guía sobre este enfoque puede encontrarse en [FK99], [Kan03] o también en [Mar01]

6.3.3.5 Clasificación usando redes neuronales

La clasificación por redes neuronales asume que se posee un conjunto de patrones de entrenamiento con sus clasificaciones correctas para entrenar el sistema. La arquitectura de una red, que incluye capa de entrada (*input layer*), capa de salida (*output layer*) y capas ocultas (*hidden layers*) puede ser muy compleja.

La clasificación está caracterizada por un conjunto de pesos y funciones de activación que determinan como cualquier información (señal de entrada – *input signal*) está siendo transmitida a la capa de salida.

Cuando la red neuronal es entrenada con los conjuntos de patrones de entrenamiento, se ajustan los pesos hasta que se obtienen clasificaciones correctas para esos datos de entrenamiento. Luego del entrenamiento, la red podrá ser usada para clasificar patrones desconocidos arbitrarios. Existen muchos clasificadores basados en redes neuronales, desde un simple *perceptrón* hasta clasificadores muy complejos y avanzados.

Una completa guía sobre la clasificación basada en redes neuronales puede ser encontrada en [FK99], [Mar01] y también en [Kan03].

6.4 Sistemas de reconocimiento de patrones basados en HMM

En un sistema de reconocimiento de patrones basado en HMM la característica distintiva es que los patrones no están almacenados físicamente en el sistema, conformando una librería de patrones para su utilización en la etapa de clasificación, sino que están *representados por un modelo*. Un modelo oculto (correspondiente a la cadena de Markov del HMM), es decir, *para cada patrón que se quiera reconocer se debe construir un modelo diferente* (un HMM diferente).

El procedimiento completo se explica a continuación: Luego de la etapa de adquisición de datos, se extrae de los datos en bruto, las características fundamentales, previamente elegidas como tales por el diseñador, cuyos valores luego del proceso que se realiza en la etapa de extracción de características, conformarán los símbolos de observación. Una secuencia de símbolos de observación servirá para:

- *entrenar el modelo* en caso que se trate de una secuencia de observaciones de entrenamiento.
- *reconocer patrones* en el caso contrario.

Los sistemas de reconocimientos de patrones basados en HMM siempre deben ser entrenados, es decir, se deben estimar los parámetros del modelo inicialmente. Un método para este propósito, como se vio en el capítulo 4 sección 4.4, es utilizar el algoritmo de Baum-Welch.

Una vez estimados los parámetros, se puede proceder a utilizar el modelo para el reconocimiento de patrones propiamente dicho.

Considérese una secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, como cada patrón i con $(i = 1, \dots, V)$, está representado por el modelo λ^i , lo que el sistema realiza es calcular $P(O | \lambda^i) \forall i, i = 1, \dots, V$, que como se vio en el capítulo 4 sección 4.2, se puede calcular con el procedimiento de avance o con el de retroceso.

Finalmente el sistema devolverá como clasificado (*reconocido*) el patrón j que se corresponde con el modelo λ^j , para el cual $P(O | \lambda^j)$ sea máxima, es decir, devuelve:

$$j = \arg \max_{1 \leq j \leq V} [P(O | \lambda^j)] \tag{6.14}$$

En la figura 6.8 se puede apreciar una representación del proceso completo.

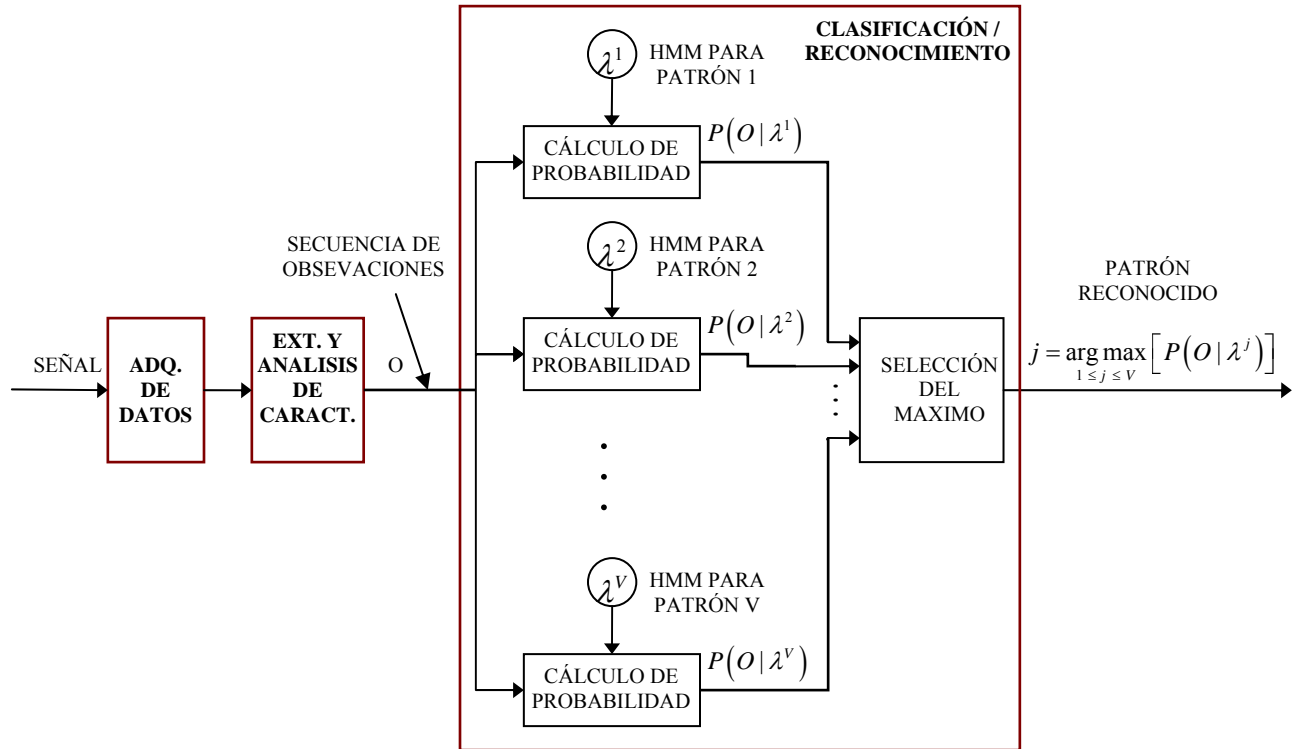


Figura 6.8: Representación de un sistema de reconocimiento de patrones basado en HMM (que reconoce un patrón entre varios).

Se puede conocer la secuencia de estados más probable $Q = (q_1, q_2, \dots, q_T)$, dada la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$ y el modelo λ^j , o sea $P(Q | O, \lambda^j)$, con el algoritmo de Viterbi, visto en el capítulo 4, sección 4.3. Esto es importante, pues permite entre otras cosas, analizar la forma (secuencia de estados ocultos) y el tiempo en el que se formó un patrón en particular (tiempo de permanencia en cada estado), que será fundamental en el caso de que lo que se quiera reconocer sea esto último y no un patrón en sí, la figura 6.9 muestra este tipo de reconocimiento.

Como el cálculo de $P(O | \lambda)$ tiene una complejidad del $O(N^2T)$ de acuerdo a la proposición 4.1 del capítulo 4, se tiene que la etapa de clasificación de un sistema de reconocimiento de patrones basado en HMM (que reconoce un patrón entre varios)

tiene un coste computacional del $O(VN^2T)$, cuya demostración, observando la figura 6.8, es trivial.

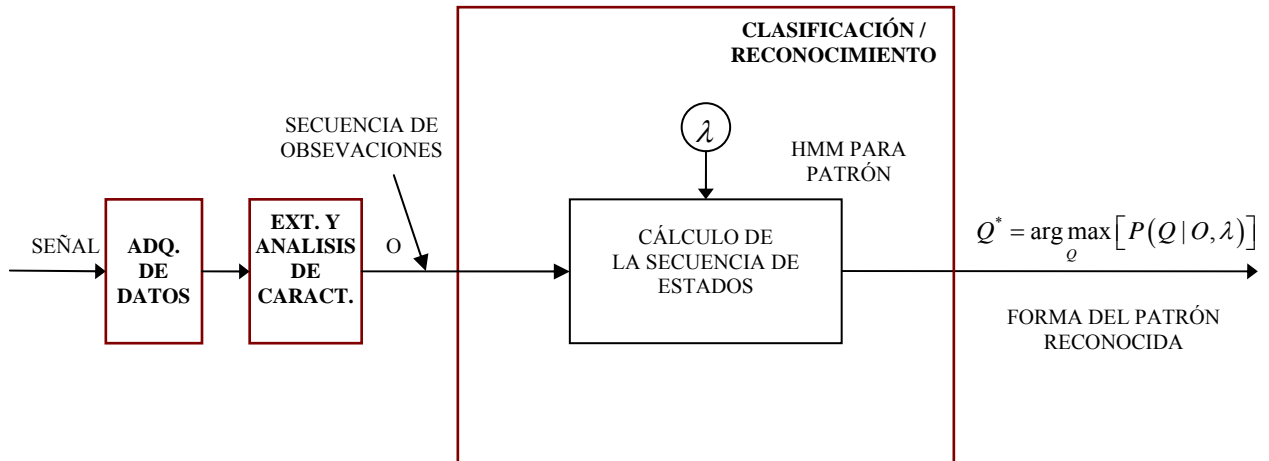


Figura 6.9: Representación de un sistema de reconocimiento de patrones basado en HMM (que reconoce la forma de un patrón).

El sistema de la figura 6.9 tiene un coste computacional del mismo orden que el algoritmo de Viterbi $O(N^2T)$, pues como puede verse en la figura es básicamente el algoritmo que realiza todo el trabajo, luego que los parámetros del modelo sean estimados. En ambos sistemas habría que sumar lo que insumen las etapas previas, pero no son representativas en el cálculo total, en ninguno de los mismos.

Capítulo 7

CASO DE ESTUDIO: IDENTIFICACIÓN DE LAS ETAPAS DEL MERCADO

*“Si buscas resultados distintos,
no hagas siempre lo mismo”.*

Albert Einstein.

7.1 Introducción

En este caso de estudio se describirá el desarrollo de un sistema de reconocimiento de patrones con el uso de HMM en la etapa de clasificación. El mismo se aplicará al reconocimiento de la forma del patrón descrito en la sección 5.7.3 del capítulo 5.

El motivo de la elección de dicho patrón es, como se comentó en la sección citada, que este patrón es el más importante de todos vistos según [Wei98], y de acuerdo a la teoría de [Wei98], toda serie temporal financiera cumple dicho patrón. El mismo captura la esencia del análisis técnico, y podría ser usado prescindiendo de cualquier otro patrón.

Por lo tanto *para el desarrollo del sistema se parte de la hipótesis de que toda serie analizada cumple con el patrón de las etapas del mercado*, del cual se querrá reconocer la forma y características de aparición que presente, las cuales variarán según la serie temporal que se analice y según la dimensión temporal que se utilice.

A lo largo de este capítulo se detallarán los pormenores de cada etapa del sistema, etapa de adquisición de datos, etapa de extracción de características y etapa de clasificación, y se mostrará las decisiones de diseño tomadas en cada una de las etapas. También se mostrarán los pasos seguidos en cuanto a la iniciación de los parámetros del modelo, un tema sensible que requerirá de una atención especial.

Finalmente se muestran los resultados obtenidos con el sistema desarrollado y los detalles de la implementación del mismo.

7.2 Objetivos

El objetivo será, en línea con lo comentado en la introducción, el desarrollo de un sistema de reconocimiento que permita *determinar en que etapa de mercado se encuentra el valor del activo o índice financiero analizado, en cada periodo de tiempo elegido*.

Lo anterior deriva en que se deberá desarrollar un modelo que sea capaz de capturar la idea de Stan Weinstein en su libro [Wei98], es decir, un modelo capaz de modelar las etapas del mercado y por ende capaz de modelar, siguiendo la tesis de Weinstein, toda serie temporal financiera, desde la óptica de la etapa del mercado en la que se encuentra.

7.3 Adquisición de datos

En realidad, a diferencia de otros sistemas de reconocimiento de patrones, como sistemas de reconocimiento de voz, imagen, etc. En el análisis de series temporales financieras, siempre será posible conformar una base de datos previo a la etapa de extracción de características. Aún obteniendo los datos en tiempo real, éstos siempre se pueden adicionar a la base de datos, previo a la extracción de características. Esta será la pauta usada en el desarrollo del sistema.

Por lo tanto, lo primero que debe realizarse es diseñar la base de datos que se utilizará para cada serie financiera correspondiente a un activo o índice en particular. Para el caso, se diseñó una base de datos donde para cada momento de tiempo t (con $t = 150, \dots, T$)⁴², se tiene la siguiente información:

- *Precio de cierre*: Será el último precio registrado para el activo en el periodo elegido, en este trabajo se usará el periodo diario y el precio de cierre del día consecuentemente.

⁴² La razón de empezar en 150 tiene que ver con el cálculo de las medias móviles y se explica más adelante.

- *Volumen*: Será el producto entre la cantidad de *papeles*⁴³ operados del activo por el precio al cual se operaron. (Podría ser también la cantidad de papeles operados, pero la opción elegida representa mejor el volumen real *monetizado*) (ver sección 5.5, capítulo 5)
- *Media Móvil Exponencial de 150 periodos del precio*: Se calculará sobre los valores de los precios, (ver sección 5.6.3, capítulo 5). Se elige esta dimensión temporal para la media móvil, pues se condice con la media móvil de 30 semanas⁴⁴ utilizada en [Wei98] para el desarrollo de la teoría de las etapas del mercado.
- *Media Móvil Exponencial de 150 periodos del volumen*: Se calculará sobre los valores del volumen, y se utilizará para determinar si el volumen es fuerte o no para un determinado periodo, determinar si el volumen es fuerte o no, en [Wei98] es una observación manual, pero en este trabajo se debió encontrar una manera de capturar la idea y automatizar su aplicación.

Por simplicidad, en este desarrollo las bases de datos estarán representadas por un archivo *txt* en formato ASCII con cuatro columnas, una para cada dato descripto anteriormente y ordenados en forma creciente, es decir, el primero corresponde al momento $t = 150$ y el último al momento $t = T$.

Las bases de datos se pueden elaborar fácilmente, se consiguen los datos de las series temporales financieras a través de la fuente deseada. En el caso de este trabajo, como se especificó en el capítulo 1, se usaron datos provenientes de la *Bolsa de Comercio de Buenos Aires*, a través de su sitio en Internet www.bolsar.com. De este sitio se pueden bajar en un archivo con extensión *xls* (con formato para su uso en programas como excel® de Microsoft o StarOffice® de Sun Microsystems por ejemplo), o también en un archivo con extensión *txt* (con formato para uso exclusivo en el programa MetaStock® de Equis Internacional).

Como el archivo *txt* es exclusivo del programa citado y no se puede editar directamente, en este trabajo se bajaron los datos en el archivo *xls*.

Estrictamente hablando, el cálculo de las medias móviles correspondería a la etapa de extracción de características, pero por los siguientes motivos, resulta conveniente en la práctica incluirlas en las bases de datos.

⁴³ En la jerga, un papel de un activo es la unidad indivisible que representa una porción del activo en el mercado, por ejemplo, un papel será una acción, en el caso de que el activo se refiera a una empresa.

⁴⁴ 150 días son 30 semanas bursátiles.

Cuando el cálculo de la media móvil comienza, se descartan los primeros 149 datos pues no habrá valor de la misma para esos días. Y como los datos de las medias móviles son cruciales para las características elegidas como se verá en la próxima sección, en las bases de datos no deberán figurar esos primeros 149 datos iniciales de precio y volumen, sino los datos a partir del 150 en adelante. Por eso es importante la creación de la base de datos con la inclusión de las medias móviles, y una vez que este proceso se inicia, solo es necesario ir agregando los datos diarios. A partir de ese punto, se podrán elegir las secuencias deseadas de 4-uplas, para la posterior extracción de las características.

La media móvil exponencial correspondiente a los precios de cierre del día t , se abreviará $MMEP_t$ y la media móvil respectiva a los volúmenes $MMEV_t$ y en general MME_t . De acuerdo a lo expuesto en la sección 5.6.3 de capítulo 5, el factor de corrección fc para el cálculo de las medias móviles será $1,33 \times 10^{-2}$, y los sucesivos valores de las mismas se calcularán recursivamente de acuerdo a las siguientes fórmulas:

$$MME_{150} = \frac{1}{150} \sum_{t=1}^{150} P_t \quad (7.1)$$

$$MME_{t>150} = (P_t - MME_{t-1}) fc + MME_{t-1} \quad (7.2)$$

Donde P_t se corresponde a los precios de cierre en el caso de $MMEP_t$, y al volumen operado en el caso de $MMEV_t$.

7.4 Extracción de características

A partir de los datos en bruto, adquiridos en la etapa anterior, se deben extraer los valores de las características que se utilizarán en la conformación del patrón. La elección de dichas características como se expuso en el capítulo 6, se encuentra a cargo del diseñador del sistema.

Gracias al conocimiento adquirido del problema específico a resolver, se eligieron las siguientes características, que luego de un procesamiento que se explicará en esta sección, conformaran el patrón a reconocer en la etapa de clasificación. Dichas características son:

- *Pendiente de la recta tangente a la MMEP*
- *Precio de cierre con respecto a la MMEP*
- *Volumen con respecto a la MMEV*

A continuación se describe el procedimiento para la extracción de los valores de cada una de estas características.

7.4.1 Pendiente de la recta tangente a la MMEP

Si la *MMEP* fuera una función perfectamente definida, el cálculo sería sencillo, pues bastaría con obtener la derivada de la función para el momento deseado, como está muy lejos de serlo, para obtener los valores de esta característica se recurrirá a la regresión lineal simple (más precisamente a la utilización del método de los cuadrados mínimos con el objetivo de estimar β_1) (ver capítulo 6, sección 6.3.2.4). Con dicho método se obtendrá una buena aproximación del valor real de la pendiente.

Se tomarán los últimos n valores de la *MMEP* para el cálculo de la fórmula (6.12). El conjunto de valores posibles de esta característica como podrá advertirse, es continuo, y como se utilizan HMM discretos en la etapa de clasificación, es decir, HMM que sólo trabajan con observaciones discretas, se debe discretizar el conjunto de valores.

Sea R_t el resultado del cálculo de $\hat{\beta}_1$ para el periodo t , entonces se aplicará una función de discretización $f: \mathbb{R} \rightarrow \{-1, 0, 1\}$ que para todo $t > 4$ se define a continuación:

$$f(R_{t>4}) = \begin{cases} 0 & \text{si } |R_t| \leq \overbrace{\left(\frac{1}{n} \sum_{i=t-4}^t MMEP_i \right)}^{(A)} cf \\ -1 & \text{si no se cumple (A) y } R_t < 0 \\ 1 & \text{caso contrario} \end{cases} \quad (7.3)$$

Como podrá inferirse con la función anterior se busca obtener simplemente los valores que interesan, esto es, (0) si la pendiente es nula, (-1) si es negativa y (1) si es positiva. Nótese que se usan los últimos n valores de la $MMEP$ para calcular $\hat{\beta}_1$ y se usan los mismos valores para el cálculo de la función. Además se utiliza un coeficiente cf para el cálculo de la misma, con el que se busca definir las fronteras de lo que será una pendiente nula. La razón de utilizar un promedio de las medias móviles que se utilizan para el cálculo de la pendiente y normalizarlas mediante un coeficiente de porcentaje, es la de hacer que la frontera de lo que será la pendiente nula se adapte al cambio de magnitud en los precios del activo subyacente. En otras palabras, hacer que la frontera sea dinámica y se adapte al cambio en los precios (discretización dinámica). Por otro lado los únicos parámetros disponibles para definir una frontera que sea dinámica con respecto a los precios, son el mismo precio y la $MMEP$. Se optó por esta última pues se utiliza para el cálculo mismo de la pendiente y es más estable.

En este trabajo se usó un valor para n de 5 y un valor para cf de $1,25 \times 10^{-3}$; ambos valores fueron obtenidos empíricamente, el primero puesto que se desean señales rápidas de cambio de pendiente, a mayor número, será más fiel la señal, pero también más tardía, debe tenerse presente que no se puede utilizar un número *grande* de n por las restricciones de linealidad del caso. En el caso del coeficiente sucede algo parecido, un coeficiente pequeño producirá cambios de pendiente más rápidamente, pero la posibilidad de señales falsas en el contexto de las series financieras aumentará también. En resumen se debe buscar un equilibrio, en la tabla de la figura 7.1 se puede apreciar la relación existente entre las dimensiones de n y cf , y la rapidez y/o fiabilidad de la señal.

n	grande	pequeño
cf	grande	pequeño
rapidez de la señal	baja	alta
fiabilidad de la señal	alta	baja

Figura 7.1: Relación existente entre el tamaño de n y cf y la rapidez y/o fiabilidad de la señal.

Finalmente, nótese que para los primeros $(n - 1)$ días (los primeros 4 días en el caso de este trabajo), no se tendrá el valor de la pendiente. Para estos días se utilizará el valor de la pendiente del día n . Es una buena aproximación puesto que n es pequeño.

7.4.2 Precio de cierre con respecto a la *MMEP*

Siendo P_t el precio de cierre del día t , para obtener los valores de esta característica, que al igual que la anterior, tomará valores en un conjunto de tres posibles, se utiliza la función $f : \mathbb{R}^{\geq 0} \times \mathbb{R}^{\geq 0} \rightarrow \{-1, 0, 1\}$ definida con sigue:

$$f(P_t, MMEP_t) = \begin{cases} 1 & \text{si } P_t > MMEP_t(1 + cf) \\ -1 & \text{si } P_t < MMEP_t(1 - cf) \\ 0 & \text{caso contrario} \end{cases} \quad (7.4)$$

Donde el coeficiente debe ser elegido de manera que represente adecuadamente la relación para la tendencia que desee ser capturada. En la sección 5.4 del capítulo 5 se vieron las tres tendencias según la teoría de Dow: primaria, secundaria y terciaria. Para una tendencia de corto plazo o terciaria debería elegirse un valor pequeño y para una tendencia primaria o de largo plazo un valor más grande. En este trabajo se consideró adecuado, teniendo en cuenta que para las etapas del mercado (sección

5.7.3 del capítulo 5) se utiliza la tendencia principal o primaria, el uso de un coeficiente de 0,10 que resultó adecuado para el caso.

7.4.3 Volumen con respecto a la *MMEV*

Siendo V_t el volumen operado el día t , para obtener los valores de esta característica se utiliza una función idéntica a la de la sección anterior, de precios con respecto a su media móvil, sólo que en lugar de P_t será V_t , y en lugar de $MMEP_t$, será $MMEV_t$.

Se deben tener en cuenta las mismas consideraciones que en el caso anterior y también en este caso se utilizó un coeficiente de 0,10.

7.4.4 Árbol de decisión

Una vez que se tienen los valores de todas las características para un día en particular, se utilizará el árbol de decisión⁴⁵ de la figura 7.2 para generar los símbolos de observación que serán los utilizados en el HMM en la etapa de clasificación. En la raíz del árbol, el test se corresponde con los valores posibles de la pendiente de la recta tangente a la *MMEP*, en el segundo nivel el test es con respecto a los posibles valores de la relación precio-*MMEP* y en el tercer nivel el test se refiere a los posibles valores de la relación volumen-*MMEV*.

Como se puede apreciar en la figura, se dispondrán de 27 símbolos de observación, por lo tanto el HMM usado en la etapa de clasificación, contará con un alfabeto que se corresponde al conjunto $V = \{0, 1, 2, \dots, 26\}$ de símbolos observables.

Las secuencias de observaciones, tanto para la fase de entrenamiento del modelo, como para la fase de reconocimiento, serán sucesiones de símbolos pertenecientes al conjunto V .

⁴⁵ Referirse al capítulo 6, sección 6.3.2.2

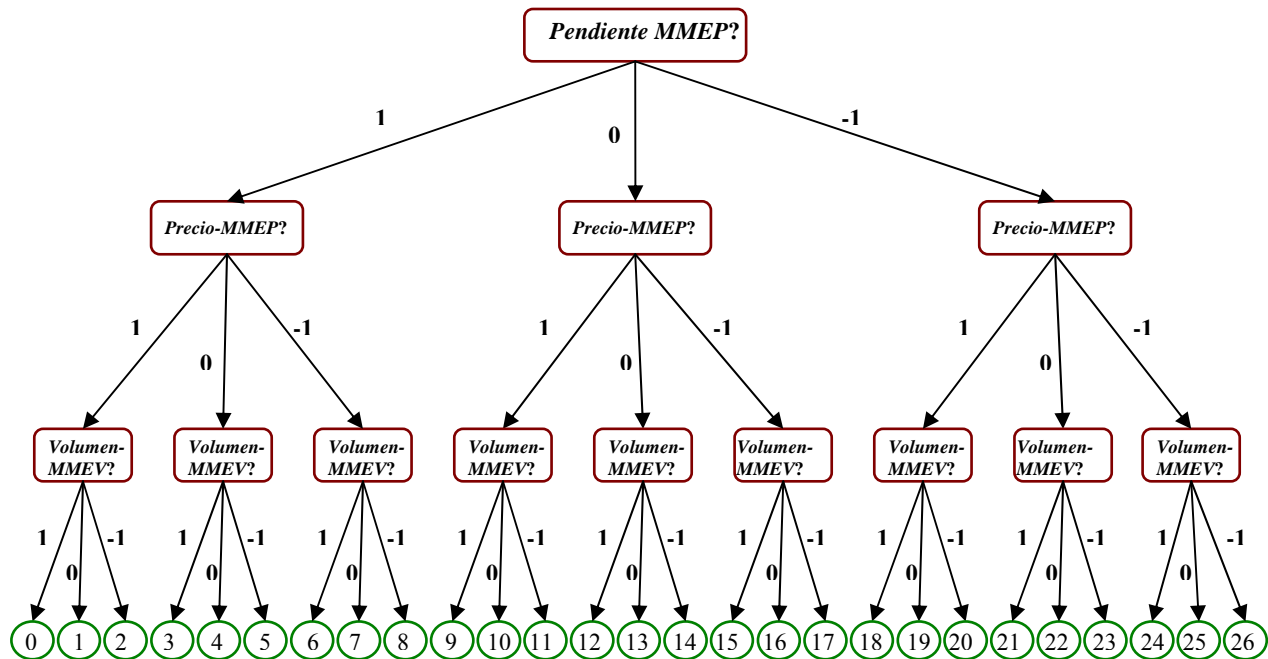


Figura 7.2: Árbol de decisión para la generación de las observaciones a partir de los valores de las características.

Para hacer efectiva la generación de las observaciones, se convierte el árbol de la figura 7.2 a las reglas IF-THEN usando una regla por cada camino desde la raíz a cada hoja (referirse al capítulo 6, sección 6.3.2.3), y la proposición 6.2 garantiza que a cada tupla (3-upla en este caso) se le asigna *exactamente* una clase, es decir, un símbolo de observación. Por lo tanto, se tendrán 27 reglas IF-THEN que se desprenden del árbol de decisión.

7.5 Clasificación / Reconocimiento

Como se expresó en la introducción, se parte de la hipótesis de que toda serie temporal financiera cumple con el patrón de las etapas del mercado⁴⁶, en línea con la teoría de [Wei98], pero el problema fundamental que concierne a la etapa de clasificación, será *diseñar la topología adecuada* para el HMM que permita cumplir

⁴⁶ Referirse al capítulo 5, sección 5.7.3

con semejante hipótesis. En este trabajo se aporta la siguiente topología de HMM representada en la figura 7.3 con dicho fin.

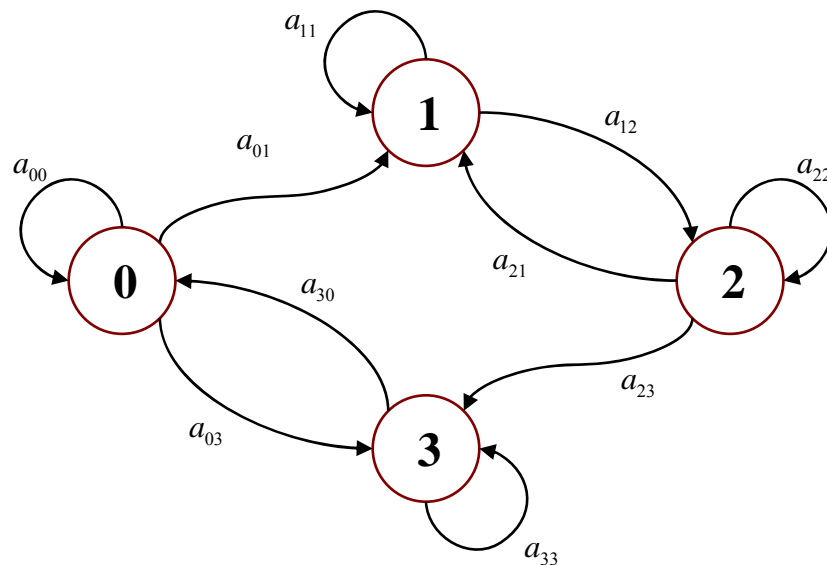


Figura 7.3: Topología de HMM para modelar las *etapas del mercado* de una serie temporal financiera.

Donde el estado (0) representa a la etapa 1 (de base), el estado (1) representa a la etapa 2 (alcista), el estado (2) representa a la etapa 3 (el techo), y por último, el estado (3) representa a la etapa 4 (bajista). Además cada estado tendrá la misma probabilidad de ser el estado inicial y contará con un función de distribución de probabilidad para con los 27 símbolos de observación que provienen de la etapa de extracción de características.

Supóngase que en un determinado momento la serie temporal se encuentra en la etapa 1 (estado 0), entonces los precios, que están moviéndose *lateralmente*, pueden tener uno de los tres siguientes comportamientos:

- *seguir lateralizando.*
- *empezar un camino alcista.*
- *comenzar un camino bajista.*

Las tres posibilidades están contempladas en el modelo propuesto, algo similar sucede si los precios están en la etapa 3 (estado 2) ya que tanto en la base, como en

el techo, se producen movimientos laterales de precio, que se diferencian por el contexto en el que aparecen (el techo precedido de una fase alcista y la base precedida de una fase bajista).

Pero nótese que si los precios están en la etapa alcista/bajista, antes de cambiar de dirección, necesariamente deben lateralizar, al menos la *MME* lo hará y luego, o bien continúan con la tendencia previa (transiciones del estado 2 al estado 1, y del estado 0 al estado 3), o bien cambian la tendencia anterior (transiciones del estado 2 al estado 3, y del estado 0 al estado 1), es por eso que desde los estados 1 y 3 no se puede retroceder. Esto refleja perfectamente la idea de [Wei98] en cuanto a que tanto en las bajas, como en las subas, puede haber pausas para luego continuar con la dirección original.

De esta forma el modelo se adapta a los movimientos por etapas de cualquier serie temporal financiera.

7.6 Iniciación de los parámetros del modelo

La iniciación de los parámetros del modelo (de los parámetros del HMM) es un primer paso obligado en sistemas de reconocimiento de patrones que utilicen HMM en la etapa de clasificación. Este proceso se realiza una vez, y luego el modelo se puede utilizar para el reconocimiento. Sin embargo, puede efectuarse un reentrenamiento cada vez que se lo juzgue conveniente, por lo tanto se debe proveer la iniciación de los parámetros del modelo presentado en la sección anterior.

La teoría, como se vio en el capítulo 4, sección 4.4, provee un algoritmo, el algoritmo de Baum-Welch, con el fin de obtener la estimación de los parámetros del modelo. Pero como se expresó en la sección 4.5.1 del mismo capítulo, debido a que el algoritmo sólo encuentra un máximo local de la función de probabilidad, se debe tratar de inicializar los parámetros de la mejor manera posible.

En el caso de la distribución de probabilidad inicial de los estados, como se estableció que cada estado tendrá igual probabilidad de ser el inicial, una iniciación uniforme es adecuada, por lo tanto, este parámetro se inicia de la siguiente forma:

$$\pi_i = \frac{1}{4}, \quad i = 0, 1, 2, 3 \quad (7.5)$$

Como no es posible una mejor aproximación, y en línea con lo expresado en el capítulo 4 (sección 4.5.1), también se iniciará la distribución de probabilidad de transiciones entre estados de manera uniforme, respetando las restricciones estocásticas, entonces para el caso se tendrá la siguiente inicialización:

$$\begin{aligned}
 a_{0j} &= \frac{1}{3}, & j &= 0,1,3 \\
 a_{1j} &= \frac{1}{2}, & j &= 1,2 \\
 a_{2j} &= \frac{1}{3}, & j &= 1,2,3 \\
 a_{3j} &= \frac{1}{2}, & j &= 0,3
 \end{aligned} \tag{7.6}$$

En cuanto a las distribuciones de probabilidad más sensibles y complicadas de estimar, que son las distribuciones de probabilidad de cada estado para con los símbolos observables $b_j(k)$, $j=0,1,2,3$, $k=0,1,2,\dots,26$, se usará el algoritmo *segmental k-means*, (algoritmo 4.6), introducido con ese fin en la sección 4.5.1 del capítulo 4.

Para el paso 1 del algoritmo se utilizarán las estimaciones iniciales citadas anteriormente para π_i , $i=0,1,2,3$ y para a_{ij} , $i, j=0,1,2,3$, pero para el caso de $b_j(k)$ se usará una inicialización manual producto del conocimiento del problema, pues se puede tener una mejor aproximación que si fuera uniforme o aleatoria.

Esto es así, pues es claro analizando el árbol de decisión de la figura 7.2, que las observaciones que se corresponden a una pendiente de la recta tangente a la *MMEP* negativa no pueden pertenecer al estado 1, del mismo modo las observaciones que se corresponden con una pendiente positiva no pueden pertenecer al estado 3. Nótese también que si la pendiente es positiva o negativa, tendrá una baja probabilidad de estar en el estado 0 o en el estado 2 (aunque no nula en este caso, pues se pueden deber al comienzo o al final de una etapa alcista/bajista).

El análisis del párrafo anterior conllevó a proponer la siguiente inicialización para $b_j(k)$ y completar el paso 1 del algoritmo *segmental k-means*:

Para $j = 0, 2$ y $k = 0, 1, 2, \dots, 26$ se tiene:

$$\begin{aligned}
 b_j(k) &= 1 \times 10^{-3}, & k &= 0, 26 \\
 b_j(k) &= 2 \times 10^{-3}, & k &= 1, 2, 24, 25 \\
 b_j(k) &= 6 \times 10^{-3}, & k &= 3, 23 \\
 b_j(k) &= 7 \times 10^{-3}, & k &= 4, 5, 21, 22 \\
 b_j(k) &= 8 \times 10^{-3}, & k &= 6, 7, 8, 18, 19, 20 \\
 b_j(k) &= 8,33 \times 10^{-2}, & k &= 9, 10, 11, 15, 16, 17 \\
 b_j(k) &= 1,33 \times 10^{-1}, & k &= 12, 13, 14
 \end{aligned} \tag{7.7}$$

Para $j = 1$ y $k = 0, 1, 2, \dots, 26$ se tiene:

$$\begin{aligned}
 b_1(k) &= 1,67 \times 10^{-1}, & k &= 0, 1, 2 \\
 b_1(k) &= 1 \times 10^{-1}, & k &= 3, 4, 5 \\
 b_1(k) &= 3,33 \times 10^{-2}, & k &= 6, 7, 8 \\
 b_1(k) &= 1,33 \times 10^{-2}, & k &= 9, 10, 11, 12, 13, 14 \\
 b_1(k) &= 6,67 \times 10^{-3}, & k &= 15, 16, 17 \\
 b_1(k) &= 0, & k &\geq 18
 \end{aligned} \tag{7.8}$$

Para $j = 3$ y $k = 0, 1, 2, \dots, 26$ se tiene:

$$\begin{aligned}
 b_3(k) &= 1,67 \times 10^{-1}, & k &= 24, 25, 26 \\
 b_3(k) &= 1 \times 10^{-1}, & k &= 21, 22, 23 \\
 b_3(k) &= 3,33 \times 10^{-2}, & k &= 18, 19, 20 \\
 b_3(k) &= 1,33 \times 10^{-2}, & k &= 12, 13, 14, 15, 16, 17 \\
 b_3(k) &= 6,67 \times 10^{-3}, & k &= 9, 10, 11 \\
 b_3(k) &= 0, & k &\leq 8
 \end{aligned} \tag{7.9}$$

En el paso 2 del algoritmo *segmental k-means* se obtiene la secuencia de estados más probable para la secuencia de observaciones utilizada y para el modelo que se inicializó antes. Esta secuencia se obtiene a través del algoritmo de Viterbi (capítulo 4, sección 4.3.1).

Finalmente, con dicha secuencia, el paso 3 del algoritmo provee la *estimación inicial* de las distribuciones de probabilidad de cada estado con respecto a los símbolos observables. Recordando el algoritmo 4.6, se obtiene:

$$\hat{b}_j(k) = \frac{\text{cantidad de símbolos } v_k \text{ observados en el estado } j}{\text{cantidad de símbolos observados en el estado } j} \quad (7.10)$$

para $j = 0, 1, 2, 3, k = 0, 1, 2, \dots, 26$

Nótese que el denominador es igual a la cantidad de veces que aparece el estado j en la secuencia de estados devuelta por el algoritmo de Viterbi.

Ya con todos los parámetros iniciados, se puede entrenar el modelo a través del algoritmo 4.5 (de Baum-Welch) (capítulo 4, sección 4.4.1)

El HMM resultante del entrenamiento, es decir, el HMM con los parámetros estimados por el algoritmo de Baum-Welch (previa estimación inicial de parámetros antes vista), se *almacena*, pues será utilizado de ahora en más para encontrar la secuencia de estados más probable que corresponda, para cada una de las futuras secuencias de observaciones que se analicen con el sistema.

7.7 Resultados

En este caso de estudio se confeccionó una base de datos con los precios de cierre y volúmenes operados diarios de las acciones de las empresas cotizantes en la bolsa de comercio de Buenos Aires: Banco Francés S.A. y Grupo Financiero Galicia S.A.

Los datos de Banco Francés S.A. (FRAN) abarcan desde el 02/01/1997 hasta el 28/12/2006, en total 2471 datos, de los cuales luego de los cálculos de las *MMEs*, se descartaron los primeros 149 datos (ver sección 7.3), por lo que la base de FRAN quedo conformada por 2322 datos múltiples.

Los datos de Grupo Financiero Galicia S.A. (GGAL) abarcan desde el 26/07/2000 hasta el 28/12/2006, en total 1586 datos, que luego del descarte al calcular las *MMEs*, resultan en una base de datos para GGAL de 1437 datos múltiples.

Para la etapa de entrenamiento se utilizaron los primeros 1000 datos de la base de FRAN. Luego de la extracción de características para los datos de entrenamiento se obtuvo una secuencia de símbolos observables de largo $T = 1000$, con dicha secuencia y la iniciación de parámetros citado en la sección anterior, se procedió a aplicar el algoritmo segmental k-means para obtener una estimación inicial de las distribuciones de probabilidad de cada estado para con los 27 símbolos observables. En las tablas a, b, c y d de la figura 7.4 se muestran los resultados:

$b_0(9) = 4/303$	$b_0(10) = 2/303$	$b_0(11) = 1/303$
$b_0(12) = 90/303$	$b_0(13) = 34/303$	$b_0(14) = 163/303$
$b_0(15) = 3/303$	$b_0(16) = 1/303$	$b_0(17) = 5/303$

Figura 7.4 a: Distribución de probabilidad del estado 0 para con los símbolos de observación, $b_0(k) = 0$ para $k < 9$ y $k > 17$.

$b_1(0) = 67/189$	$b_1(1) = 20/189$	$b_1(2) = 70/189$
$b_1(3) = 8/189$	$b_1(4) = 2/189$	$b_1(5) = 22/189$

Figura 7.4 b: Distribución de probabilidad del estado 1 para con los símbolos de observación, $b_1(k) = 0$ para $k > 5$.

$b_2(9) = 3/251$	$b_2(10) = 2/251$	$b_2(11) = 8/251$
$b_2(12) = 47/251$	$b_2(13) = 18/251$	$b_2(14) = 168/251$
$b_2(15) = 2/251$	$b_2(16) = 1/251$	$b_2(17) = 2/251$

Figura 7.4 c: Distribución de probabilidad del estado 2 para con los símbolos de observación, $b_2(k) = 0$ para $k < 9$ y $k > 17$.

$b_3(21) = 6/257$	$b_3(22) = 1/257$	$b_3(23) = 16/257$
$b_3(24) = 72/257$	$b_3(25) = 16/257$	$b_3(26) = 146/257$

Figura 7.4 d: Distribución de probabilidad del estado 3 para con los símbolos de observación, $b_3(k) = 0$ para $k < 21$.

Con los datos obtenidos se pueden hacer las siguientes reflexiones. En las distribuciones resultantes, existe probabilidad nula de observar los símbolos 6, 7, 8, 18, 19 y 20. Se esperaba que fuera muy baja y por eso se iniciaron con una probabilidad muy pequeña, el resultado confirmó la presunción y es un resultado razonable, pues los símbolos 6, 7 y 8 se corresponden con una pendiente de la recta tangente a la *MMEP* positiva, con precios por debajo de la *MMEP*, lo cual es bastante improbable. Del mismo modo los símbolos 18, 19 y 20 se corresponden con una pendiente negativa y precios por encima de dicha *MMEP* y también resulta improbable, pues debe recordarse que la MME (ver capítulo 5, sección 5.6.3) reacciona rápidamente a los precios recientes. Este resultado llevaría a la necesidad de hacer más pruebas con el objeto de eliminar dichas clases de observaciones, o bien agruparlas en torno a clases vecinas.

También es visible, como estaba previsto, que las etapas 2 y 4 (estados 1 y 3) son como espejos y que tienen probabilidad nula de observar símbolos que se corresponden con la pendiente de la recta tangente a la *MMEP* nula o contraria al movimiento. Esto es, para la etapa 2, *alcista*, no puede ocurrir que la pendiente sea negativa y viceversa para el caso de la etapa 4, *bajista*.

Es de destacar la importancia del símbolo 2 en la etapa alcista y del símbolo 24 en la etapa bajista, esto significa que en las subas, a pesar del fuerte volumen inicial esperado, una gran cantidad de tiempo se tiene volúmenes bajos (lo que es un dato curioso). En el caso de la etapa bajista y el símbolo 24, es normal, pues en las bajas, inicialmente, como se comentó en el capítulo 5, el volumen puede aumentar al principio.

Finalmente es de destacar los parecidos entre las distribuciones de los estados 0 y 2 (etapa de base y de techo respectivamente) como era de esperar, aunque es curioso que en ambos casos el símbolo 14 fuera el más probable, es decir, pendiente de la recta tangente a la *MMEP* nula, con precios entorno a la *MMEP* y volúmenes bajos; y que el volumen fuerte (símbolo 12) fuera más probable en el estado 0 (base), cuando generalmente se piensa lo contrario apoyándose en la literatura (que en la

etapa de techo, correspondiente al estado 2, se registran los mayores volúmenes, junto con una mayor volatilidad).

Con el HMM corregido en sus distribuciones $b_j(k)$, $j = 0, 1, 2, 3$, $k = 0, 1, 2, \dots, 26$ se procedió a estimar los parámetros con el algoritmo de Baum-Welch con la misma secuencia de 1000 datos de FRAN. Un problema que puede ocurrir al reestimar los parámetros con dicho algoritmo con una sola secuencia de observaciones, es que la distribución de probabilidad inicial de los estados va a quedar signada por el estado que azarosamente resulte el primero en dicha secuencia, es decir dicho estado tendrá probabilidad 1 de ser el estado inicial. Una solución es usar múltiples secuencias de entrenamiento, como se vio en el capítulo 4, sección 4.5.2. Pero en este caso, como se estableció al diseñar el modelo que los cuatro estados poseen la misma probabilidad de ser el estado inicial, no se reestimarán con Baum-Welch las distribuciones de probabilidad inicial, es decir, se consideran como finales las citadas en la fórmula (7.5).

En la figura 7.5 a, b, c y d se muestran las distribuciones a_{ij} , $i, j = 0, 1, 2, 3$ y $b_j(k)$, $j = 0, 1, 2, 3$, $k = 0, 1, 2, \dots, 26$ respectivamente para cada estado, y como resultado de aplicar el algoritmo de Baum-Welch.

$a_{00} = 9,62 \times 10^{-1}$	$a_{01} = 1,74 \times 10^{-2}$	$a_{03} = 2,08 \times 10^{-2}$
$b_0(9) = 1,6 \times 10^{-2}$	$b_0(10) = 2 \times 10^{-3}$	$b_0(11) = 1 \times 10^{-3}$
$b_0(12) = 3,04 \times 10^{-1}$	$b_0(13) = 1,15 \times 10^{-1}$	$b_0(14) = 5,33 \times 10^{-1}$
$b_0(15) = 9 \times 10^{-3}$	$b_0(16) = 3 \times 10^{-3}$	$b_0(17) = 1,9 \times 10^{-2}$

Figura 7.5 a: Distribución de probabilidad de las transiciones desde el estado 0 y de dicho estado para con los símbolos de observación, $b_0(k) = 0$ para $k < 9$ y $k > 17$.

$a_{11} = 9,7 \times 10^{-1}$		$a_{12} = 2,97 \times 10^{-2}$
$b_1(0) = 4,94 \times 10^{-1}$	$b_1(1) = 1,06 \times 10^{-1}$	$b_1(2) = 3,27 \times 10^{-1}$
$b_1(3) = 2,9 \times 10^{-2}$	$b_1(4) = 6 \times 10^{-3}$	$b_1(5) = 3,8 \times 10^{-2}$

Figura 7.5 b: Distribución de probabilidad de las transiciones desde el estado 1 y de dicho estado para con los símbolos de observación, $b_1(k) = 0$ para $k > 5$.

$a_{21} = 3,01 \times 10^{-2}$	$a_{22} = 9,59 \times 10^{-1}$	$a_{03} = 1,11 \times 10^{-2}$
$b_2(9) = 2,5 \times 10^{-2}$	$b_2(10) = 4 \times 10^{-3}$	$b_2(11) = 1,3 \times 10^{-2}$
$b_2(12) = 1,87 \times 10^{-1}$	$b_2(13) = 8,7 \times 10^{-2}$	$b_2(14) = 6,66 \times 10^{-1}$
$b_2(15) = 8 \times 10^{-3}$	$b_2(16) = 4 \times 10^{-3}$	$b_2(17) = 6 \times 10^{-3}$

Figura 7.5 c: Distribución de probabilidad de las transiciones desde el estado 2 y de dicho estado para con los símbolos de observación, $b_2(k) = 0$ para $k < 9$ y $k > 17$.

$a_{30} = 2,13 \times 10^{-2}$	$a_{33} = 9,79 \times 10^{-1}$	
$b_3(21) = 1,9 \times 10^{-2}$	$b_3(22) = 7 \times 10^{-3}$	$b_3(23) = 3,4 \times 10^{-2}$
$b_3(24) = 1,87 \times 10^{-1}$	$b_3(25) = 7,9 \times 10^{-2}$	$b_3(26) = 6,74 \times 10^{-1}$

Figura 7.5 d: Distribución de probabilidad de las transiciones desde el estado 3 y de dicho estado para con los símbolos de observación, $b_3(k) = 0$ para $k < 21$.

De los datos anteriores es de destacar, que en contraposición a la inicialización uniforme de la distribución de probabilidad de transición para cada estado, el algoritmo de Baum-Welch, propinó en todos los caso una alta probabilidad de permanecer en el mismo estado, como era de esperarse.

Finalmente con el HMM *entrenado y almacenado*, se procedió a obtener la secuencia de estados más probable con el algoritmo de Viterbi, dada una secuencia de observaciones de test y dado el modelo entrenado.

Como test se utilizó la secuencia de observaciones producida en la etapa de extracción de características del sistema a partir de los datos de la base de datos de FRAN desde el 01/02/2002 hasta el 30/06/2004 (un total de 584 observaciones). En la figura 7.6 puede observarse la *secuencia de estados* que se obtuvo.

En conclusión, la *forma del patrón* buscada para esta secuencia de observaciones en particular, *indica en que etapa del mercado se encuentra el precio de FRAN para todo tiempo t*.

En la figura 7.7 se superponen los datos obtenidos en un gráfico con los datos de precio y volumen operado para el período citado, permitiendo visualizar la efectividad del reconocimiento, en comparación con la aproximación manual que se efectuó en la figura 5.16 del capítulo 5; el lector podrá apreciar que la secuencia no

fue elegida al azar, ya que es el mismo período de dicha figura y permite comparar el resultado obtenido. En ese sentido el resultado, como puede verse, es más que satisfactorio, pues incluso permite corregir la aproximación manual que se había realizado en la figura 5.16. En dicha figura la etapa 2 se había marcado desde aproximadamente comienzos de octubre de 2002 hasta fines de enero de 2004, pero evidentemente, en agosto de 2003 existía un ingreso a etapa 3 que no había sido señalado en la figura original.

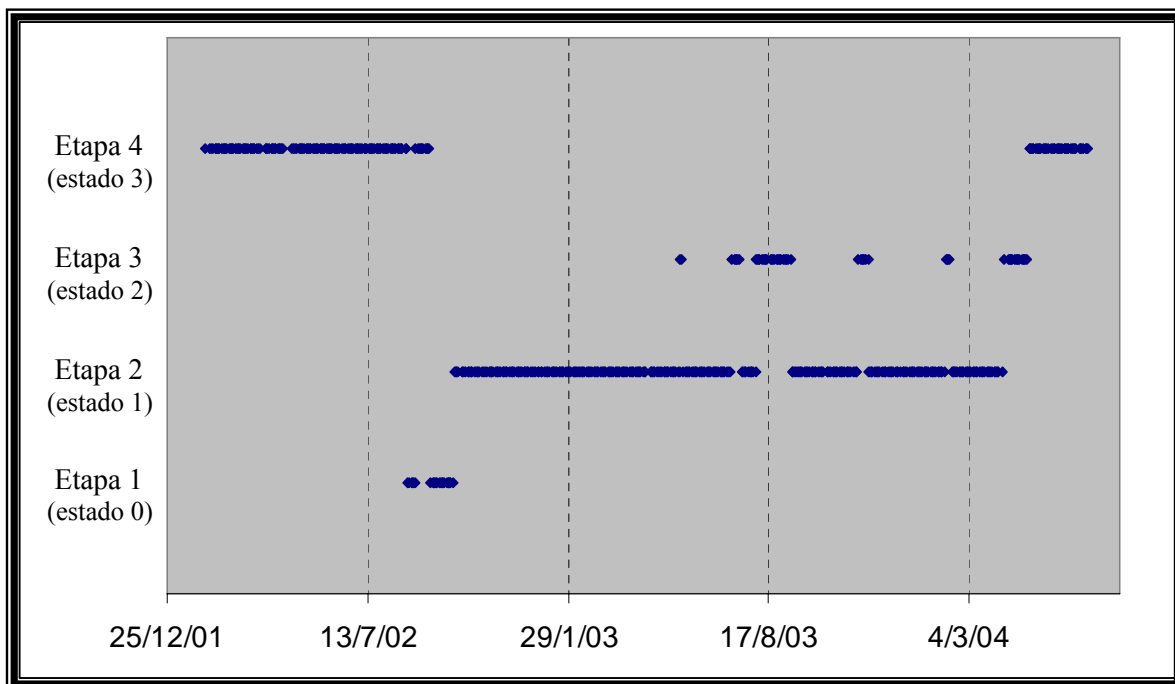


Figura 7.6: Secuencia de estados obtenida como producto del sistema de reconocimiento de patrones a partir de los datos de FRAN desde el 01/02/2002 hasta el 30/06/2004.

Para verificar que el HMM entrenado es de utilidad para identificar la forma del patrón de etapas del mercado en cualquier serie temporal financiera se efectuó un análisis similar al confeccionado para FRAN, para la secuencia de observaciones generada luego de la etapa de extracción de características, a partir de los datos obtenidos de la base de datos correspondiente a las acciones del Grupo Financiero Galicia S.A. (GGAL), para el periodo 03/10/2002 hasta el 20/08/2004, (un total de 469 datos). En la figura 7.8 puede verse la secuencia de estados obtenida, superpuesta con el gráfico correspondiente, y puede constatar que los resultados

fueron satisfactorios. Estos datos tampoco fueron elegidos al azar, se anima al lector a comparar la figura 7.8 con la figura 5.19 del capítulo 5.

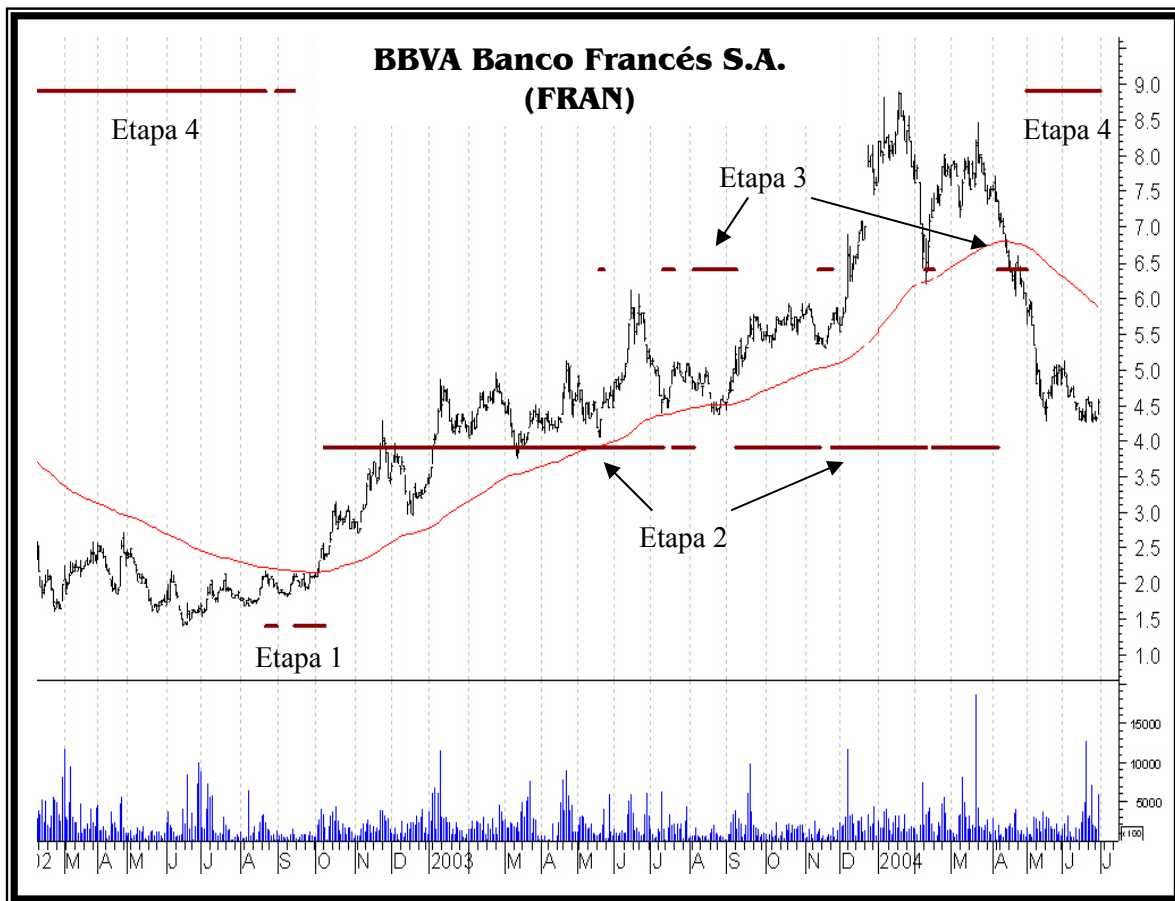


Figura 7.7: La secuencia de estados obtenida de la figura 7.6 superpuesta con el gráfico de FRAN desde el 01/02/2002 hasta el 30/06/2004.

Como corolario de los resultados, es interesante comentar que el conocimiento en todo momento de la etapa en que se encuentran los precios de un determinado activo financiero, permite desarrollar variadas estrategias con el objetivo de sacar provecho de la teoría de las etapas de mercado.

De acuerdo a [Wei98] una decisión de compra estaría enfocada en los comienzos de la etapa 2 (estado 1), por el contrario, una decisión de venta lo estaría en los comienzos de la etapa 4 (estado 3). Pero en el lado bajista se debe ser más cuidadoso, por lo tanto el sólo hecho de producirse una pausa de etapa 3 (estado 2) debería alertar sobre un posible cambio en la tendencia.

En la secuencia de datos obtenida, la indicación de compra para el analista estaría representada por el momento en el que el sistema informa que se encuentra en etapa 2, previo a encontrarse en etapa 1 (el momento en el cual se produce una transición desde el estado 0 hacia el estado 1 en el HMM), y la indicación de venta estaría representada por el instante en que el sistema informa que se encuentra en etapa 4, previo a encontrarse en etapa 3 (cuando se produce una transición desde el estado 2 al estado 3 en el HMM), aunque se deberían tomar precauciones cuando se produce una transición desde el estado 1 al estado 2 del HMM (cuando el sistema informa que se encuentra en la etapa 3, previo a estar en la etapa 2).

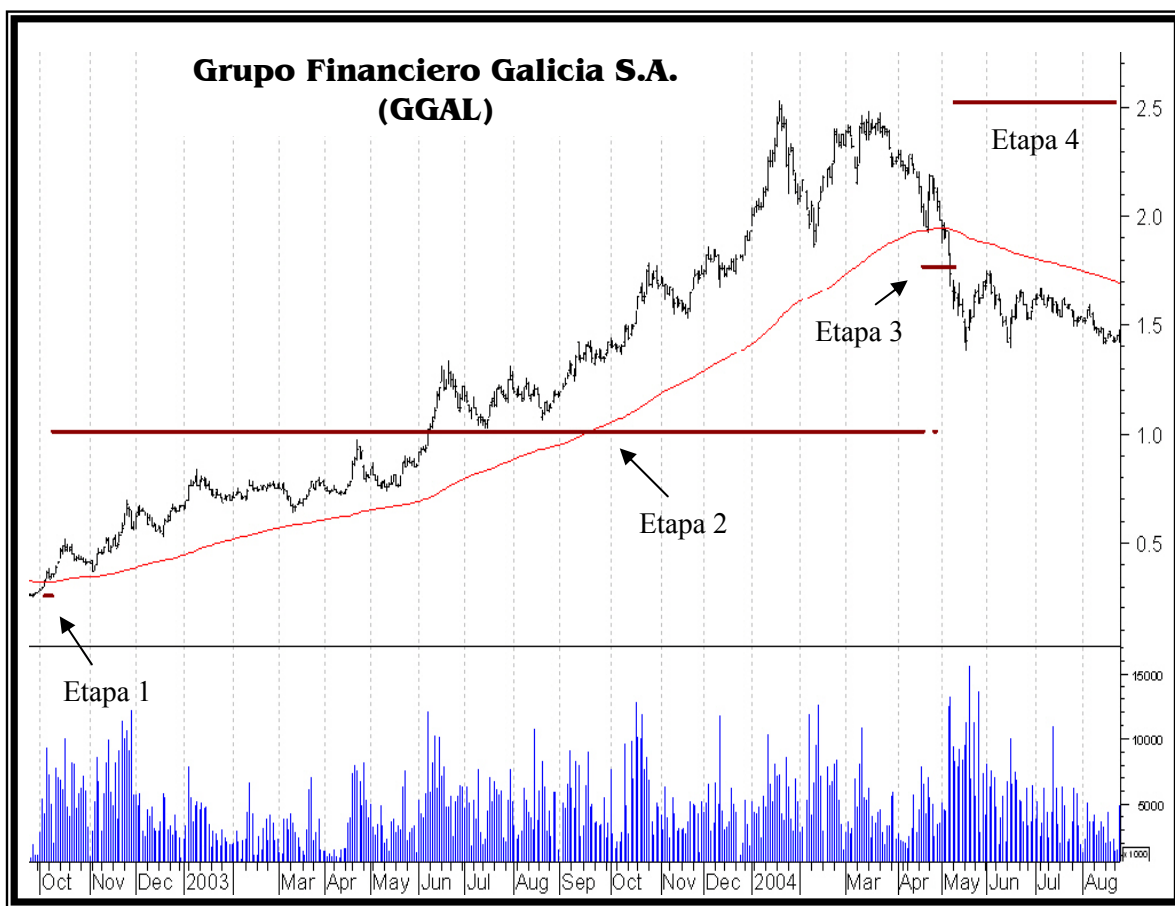


Figura 7.8: Secuencia de estados obtenida por el sistema para GGAL a partir de la secuencia de observaciones generada en la etapa de extracción de características desde la base de datos respectiva, desde el 03/10/2002 hasta el 20/08/2004, superpuesta sobre el gráfico correspondiente para dicho periodo de GGAL.

7.8 Detalles de implementación

Se utilizaron preceptos de ingeniería del software para el desarrollo del sistema de reconocimiento, fundamentalmente diseño modular. El principio de diseño por módulos se adapta perfectamente al problema y permite descomponer el sistema de forma tal que se obtienen los siguientes beneficios:

- Cada módulo se puede diseñar independientemente del resto.
- Cada módulo puede ser comprendido sin necesidad de comprender los restantes.
- Es posible cambiar la implementación de un módulo, sin necesidad de cambiar o afectar la implementación del resto de los módulos.
- Se puede encarar una gran modificación al sistema como un conjunto de pequeñas modificaciones.

Nótese que es fundamental para el sistema de reconocimiento de patrones desarrollado en este caso de estudio y para cualquier sistema de reconocimiento de patrones contar con dichas ventajas, pues por ejemplo, en este caso de estudio el sistema reconoce las distintas formas que puede adoptar el patrón de las etapas del mercado, pero puede requerirse que el sistema reconozca un patrón entre varios, y si el sistema no estuviera diseñado con un principio de diseño para el cambio (como lo es el diseño modular), debería modificarse todo el sistema. Los sistemas de reconocimiento de patrones son susceptibles de cambios, y deben diseñarse de manera que resulte sencillo modificarlos. En la figura 7.9 se puede ver el diseño modular elegido para el sistema, que por otra parte, al lector le resultará natural tanto por la teoría desarrollada hasta este punto, como por el desarrollo de este caso de estudio.

Para la implementación del sistema de reconocimiento se utilizó el lenguaje JAVA por dos motivos:

- Se adapta naturalmente al diseño modular (diseño orientado a objetos), donde cada módulo se corresponderá a una *clase* de JAVA.
- Posee una gran portabilidad (es independiente de la plataforma), esto permitirá que funcione en casi cualquier sistema operativo.

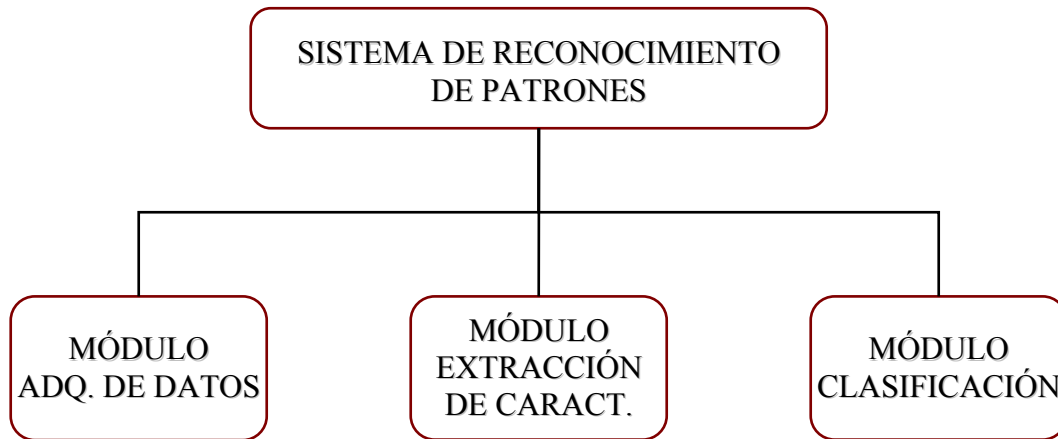


Figura 7.9: Estructura de módulos del sistema de reconocimiento de patrones.

El punto en contra de JAVA es su velocidad, y posiblemente en un sistema comercial, específicamente diseñado para un sistema operativo en particular, convendría implementarlo en C++, pero desde la óptica académica de este trabajo y por los motivos antes expuestos, resulta conveniente el uso de JAVA para este caso de estudio.

JAVA es un lenguaje de programación, en el que se utilizan conjuntos de clases, denominados *librerías* de JAVA, para la solución de diferentes problemas. En ese aspecto se utilizaron librerías básicas del lenguaje para el desarrollo, con la excepción de la librería *Jahmm*⁴⁷ utilizada en el módulo Clasificación para la implementación de los HMM, específicamente se utilizó la versión 0.3.3 de dicha librería. Existen muchas librerías para la implementación de los HMM en diversos lenguajes de programación, aparte de la nombrada, se pueden citar por ejemplo:

- HTK (*Hidden Markov Models Toolkit*)⁴⁸ que está orientada al reconocimiento del habla.
- HMMER⁴⁹ que está orientada al análisis de secuencias biológicas.

⁴⁷ Librería JAVA para la implementación de los HMM, desarrollada por Jean-Marc François y que puede conseguirse en el sitio web del autor <http://www.run.montefiore.ulg.ac.be/~francois/software/jahmm/>

⁴⁸ Desarrollada en ANSI C por el departamento de ingeniería de la universidad de Cambridge, se puede conseguir en el sitio web: <http://htk.eng.cam.ac.uk/>

Las razones para la elección de *Jahmm* (aparte de que está desarrollada en JAVA) son las siguientes:

- Es una librería con la suficiente optimización y utilización en diferentes claustros académicos, lo que garantiza su buen funcionamiento.
- No tiene una orientación específica como las nombradas anteriormente, por lo que puede ser fácilmente adaptada para cualquier uso en particular.
- Está totalmente basada en [RJ86, Rab89, RJ93], que representan la base teórica referente a los HMM de este trabajo también.
- Es ideal para su aplicación a HMM discretos, ya que éstos eran para los cuales se diseñó la librería, aunque después se añadió la posibilidad de manejar HMM continuos.

Se utilizó para este trabajo la versión 0.3.3 de la librería, pues es una versión totalmente estable y probada, que incorpora todos los requerimientos necesarios en cuanto a HMM discretos. Las versiones posteriores, mejoraron aspectos relacionados a HMM continuos, agregaron manejo por línea de comando y cosas de menor importancia. Lo único relevante para este trabajo que se añadió después, fue la posibilidad de que el usuario generara sus propios HMM a través de un archivo de texto en formato ASCII, pero en este trabajo se diseñó un formato propio, mejorado con respecto a el utilizado por el autor de la librería, que se detallará más adelante, por lo que se hace innecesaria la utilización de una versión posterior a la 0.3.3.

Además de los módulos fundamentales de la figura 7.9, se añadió el módulo *Herramientas*, cuya finalidad será entre otras cosas, proveer métodos para lectura y escritura de HMM y/o secuencias de estados. Cada módulo se corresponde con una clase del programa Java elaborado. A continuación se detallan los métodos principales de cada clase.

Clase **ADQ_DATOS**:

- Método *adquirir_datos*: Este método recibe como parámetro el nombre de un archivo y devuelve los datos (en bruto) recogidos del mismo. De acuerdo a lo expresado en la sección 7.3, este archivo representará la base de datos

⁴⁹ Desarrollada en ANSI/POSIX C por Sean Eddy [Edd97] del departamento de genética del instituto médico Howard Hughes, se puede conseguir en el sitio web: <http://hmmer.janelia.org/>

correspondiente a algún activo en particular y tendrá formato ASCII, con los datos multidimensionales en columnas según la necesidad del caso.

Este será el módulo que se modifique si el origen y/o formato de los archivos de entrada fueran cambiados.

Clase **EXT_CHARACTERISTICAS**:

- Método *obtener_sec_obs*: Recibe como parámetro datos multidimensionales, (que pueden ser los recogidos en el módulo ADQ_DATOS), y devuelve luego de todo el proceso descrito en la sección 7.4 la secuencia de observaciones que utilizará el HMM.

Este será el módulo modificado si las características buscadas en los datos cambiasen o si el procedimiento para extraer dichas características cambiase.

Clase **CLASIFICACION**:

- Método *entrenar_modelo*: Recibe como parámetros el nombre del archivo donde se encuentra almacenado un HMM⁵⁰ y una secuencia de observaciones, (que puede ser la generada por el módulo EXT_CHARACTERISTICAS), y devuelve un HMM entrenado. El HMM almacenado (que será inicial) se corresponde al descrito en la sección 7.6, y este método deberá usar el algoritmo segmental k-means tal cual se describió en dicha sección, antes de reestimar los parámetros del modelo mediante el algoritmo de Baum-Welch.
- Método *obtener_sec_estados*: Recibe como parámetros una secuencia de observaciones y el nombre del archivo donde se encuentra almacenado un HMM entrenado y devuelve la secuencia de estados más probable dados los parámetros. Este método es el que reconoce la forma del patrón.
- Método *segmentalKmeans*: Recibe como parámetros un HMM y la secuencia de observaciones y devuelve un HMM con las distribuciones de probabilidad de los estados con respecto a las observaciones estimadas según el algoritmo *segmental k-means*. Este método es llamado por el método *entrenar_modelo*.

⁵⁰ En todos los casos donde el nombre de archivo recibido no se corresponda con un archivo que posee almacenado un HMM, no será aceptado dicho archivo y se devolverá el error respectivo.

Clase **HERRAMIENTAS**:

- Método *almacenar_HMM*: Recibe como parámetros un HMM y un nombre de archivo y almacena el HMM en un archivo con el nombre recibido.
- Método *almacenar_sec_est*: Recibe como parámetro una secuencia de estados, (que puede ser la generada con el módulo CLASIFICACION), y almacena la misma en un archivo de nombre “sec_est.txt”
- Método *leer_HMM*: Recibe como parámetro el nombre de archivo donde se encuentra almacenado un HMM y devuelve dicho HMM para su utilización posterior.

Con respecto a los métodos anteriores, es necesario hacer algunas aclaraciones. En primer lugar la librería *Jahmm* otorga la posibilidad de llamar al algoritmo de Baum-Welch de forma simple o con la estrategia de escalado vista en el capítulo 4, sección 4.5.3. En este trabajo se uso por los motivos descritos en dicha sección, la versión que utiliza coeficientes de escalado. En segundo lugar, el archivo “sec_est.txt” se almacena con un formato adecuado para su posterior utilización en análisis, gráficos, etc., resultando el formato más conveniente, un formato ASCII con una columna a un dato por línea.

En tercer lugar y lo más importante, la librería *Jahmm* en su versión 0.3.3 no permite ni la edición, ni la creación de HMM en forma manual como ya se comentó, pero en este trabajo se agregó la posibilidad de que el usuario pueda crear manualmente un HMM a través de un archivo, esto será de vital importancia a la hora de introducir el HMM inicial para su posterior entrenamiento. Con ese fin el método *leer_HMM* no sólo lee archivos guardados en formato exclusivo de la librería *Jahmm*, sino también los archivos en formato ASCII con un formato propio aportado por este trabajo y que será mostrado a continuación.

Nótese que debe mantenerse la posibilidad de leer los HMM en formato exclusivo de la librería, pues en dicho formato se guardarán, por ejemplo, los HMM que resulten del entrenamiento.

En la figura 7.10 se puede apreciar el formato del archivo ASCII para la creación de HMM y que luego podrá ser reconocido por el programa. Como ejemplo del formato mostrado en la figura 7.10, en el Apéndice B (B.1) puede observarse el archivo correspondiente al HMM inicial que sirvió para este trabajo y que se elaboró

de acuerdo a lo comentado en la sección 7.6 para el paso 1 del algoritmo *segmental k-means*.

```
HMM

ESTADOS [nro_estados]
ALFABETO [tam_alfabeto]

ESTADO
PI [prob_inicial]
A [prob_transición_1] [prob_transición_2] ... [prob_transición_j]
B [prob_obs_1] [prob_obs_2] ... [prob_obs_M]
.
.
.
.
```

} nro_estados

Figura 7.10: Formato del archivo conteniendo un HMM para ser leído por el sistema desarrollado.

El sistema resultado de este desarrollo puede ser manejado por línea de comandos. A continuación se da un ejemplo de una secuencia típica de reconocimiento, con el pseudo-código correspondiente a la ejecución del programa para cada una de estas opciones.

Primero se debería entrenar el modelo con una secuencia de entrenamiento:

```
java -jar SRP_HMM.jar sec_ent.txt /e hmm_inicial.hmm hmm_entrenado.hmm
```

Utilizando la descripción de métodos y módulos antes descripta, la figura 7.11 muestra el pseudo-código que representa la ejecución de la sentencia de la línea anterior. Nótese que el método *entrenar_modelo*, deberá llamar, entre otros, a los métodos *leer_HMM* de la clase HERRAMIENTAS y *segmental k-means* de la clase CLASIFICACION.

```
datos = ADQ_DATOS.adquirir_datos("sec_ent.txt")
sec_observaciones = EXT_CARACTERISTICAS.obtener_sec_obs(datos)
hmm = CLASIFICACION.entrenar_modelo(sec_observaciones, "hmm_inicial.hmm")
HERRAMIENTAS.almacenar_HMM(hmm, "hmm_entrenado.hmm")
```

Figura 7.11: Pseudo-código correspondiente una instrucción por línea de comandos que muestra un ejemplo de entrenamiento del modelo.

Luego obtener la secuencia de estados más probable para una secuencia de observaciones:

```
java -jar SRP_HMM.jar sec_test.txt hmm_entrenado.hmm
```

La figura 7.12 muestra el pseudo-código que se corresponde a la ejecución de la sentencia de la línea de comando anterior. Nótese que el método *obtener_sec_est* deberá llamar, entre otros, al método *leer_HMM* de la clase HERRAMIENTAS.

```
datos = ADQ_DATOS.adquirir_datos("sec_test.txt")
sec_observaciones = EXT_CARACTERISTICAS.obtener_sec_obs(datos)
sec_est = CLASIFICACION.obtener_sec_est(sec_observaciones, "hmm_entrenado.hmm")
HERRAMIENTAS.almacenar_sec_est(sec_est)
```

Figura 7.12: Pseudo-código correspondiente una instrucción por línea de comandos que muestra un ejemplo de obtención de secuencia de estados más probable.

Capítulo 8

CONCLUSIONES Y FUTUROS DESARROLLOS

“Al principio vienen necesariamente a la mente la fantasía y la fábula. Desfilan después los cálculos matemáticos, y sólo al final la realización corona el pensamiento”.

Konstantín Eduardovich Tsiolkovski.

8.1 Conclusiones

En línea con los objetivos propuestos para este trabajo, a continuación se muestran las conclusiones de mayor relevancia que se recogieron durante la elaboración del mismo.

El primer objetivo propuesto era el desarrollo de una teoría que sirviera de base para la elaboración de sistemas automáticos de reconocimiento de patrones del análisis técnico bursátil. En relación al mismo las conclusiones que se obtuvieron fueron las siguientes:

- El desarrollo de la teoría a través del estudio del análisis de series temporales, los modelos ocultos de Markov, el análisis técnico bursátil y los sistemas de reconocimiento de patrones, resultó completa y fructífera, permitiendo aplicar ideas desarrolladas en principio para otras disciplinas como el reconocimiento de voz, reconocimiento de imágenes digitalizadas, análisis de series financieras, etc., para el estudio de los patrones del análisis técnico bursátil.

- Los HMM resultaron una pieza fundamental en el desarrollo de la teoría, permitiendo por un lado modelar las series temporales financieras (no estacionarias y altamente volátiles por naturaleza) a través de las características elegidas para las mismas, y por otro lado identificar un conjunto de estados finitos a través de los que se desarrollan los patrones del análisis técnico bursátil.

El segundo objetivo se refería a la aplicación de la teoría desarrollada, y las conclusiones obtenidas fueron las siguientes:

- La aplicación al caso de estudio “Las etapas del mercado” mostró que la teoría desarrollada fue efectiva para su consecución y resultado ampliamente satisfactoria para el caso. Como puede apreciarse en los resultados de la aplicación se logra capturar la idea de Stan Weinstein en su libro [Wei98], donde el autor expresa: “*una mirada al gráfico vale más que mil predicciones*”, aunque en este trabajo se *automatizó* el proceso.
- Los resultados empíricos de la aplicación en las series temporales financieras de test fueron satisfactorias, produciendo la correcta determinación de la etapa de mercado en la que se encuentra el precio del activo en cada momento de tiempo (según la dimensión temporal elegida).
- La implementación del sistema de reconocimiento de patrones del análisis técnico bursátil, gracias al desarrollo previo de la teoría, resultó accesible y sin mayores complicaciones.

El objetivo secundario, pero no menos importante, era dotar a la teoría de un carácter general para permitir su aplicación a otros ámbitos. Dicho objetivo ha sido cumplido con la forma en que se elaboró cada capítulo y con la presentación de la metodología de los sistemas de reconocimiento de patrones basados en HMM de manera general.

Finalmente es importante puntualizar los aspectos más difíciles concernientes a la aplicación de la teoría que tuvieron que ser encarados, las herramientas que se utilizaron para lograr superarlos y las conclusiones que se obtuvieron en el proceso.

En cuanto a la etapa de extracción de características de los sistemas de reconocimiento de patrones, fue fundamental a la hora de la aplicación de la teoría el desarrollo de los métodos de minería de datos incluidos en el capítulo 6. Asimismo la mayor dificultad en dicho momento estuvo representada por la *elección* de las características de las que luego se extraerían los valores, que previo proceso, permitirían obtener la secuencia de observaciones del modelo.

En cuanto a los HMM, pertenecientes a la etapa de clasificación del sistema de reconocimiento de patrones, la inclusión de los problemas de implementación y sus posibles soluciones en el capítulo 4, fue fundamental para la resolución de dichos problemas en el momento de la aplicación de la teoría. La mayor dificultad relativa a los HMM a la hora de su aplicación fue la elección de la topología adecuada para el caso de estudio.

Es interesante señalar que en ambas etapas del sistema de reconocimiento de patrones, la mayor dificultad estuvo representada por cuestiones que son privativas del conocimiento del problema y son responsabilidad del diseñador del sistema, y sobre lo que no existe una solución *teóricamente correcta* como se expresó oportunamente, pues estas elecciones dependen del tipo de señal que se esté analizando y del patrón que se quiera reconocer en ella.

Por supuesto, una vez elegidas las características y la topología del modelo, se podría comparar el rendimiento y la eficacia del sistema con respecto a otros modelos y/o otros conjuntos de características, pero no se tendrá la certeza de que el modelo y las características que resulten con mejor rendimiento, sean los mejores posibles, solamente que son los mejores del conjunto analizado. Nótese que la definición de *mejor* depende también del problema específico.

8.2 Aportes de esta tesis

Los aportes principales de esta tesis se detallan a continuación:

- *Taxonomía de series temporales financieras*: La investigación del estado del arte de los diferentes modelos utilizados para modelar series temporales financieras es el primer aporte de este trabajo. La clasificación aportada permite identificar las particularidades que son comunes en los modelos mostrados por distintos autores, así como los aspectos que las diferencian.
- *Desarrollo de un sistema automático de reconocimiento de patrones del análisis técnico bursátil*: Teniendo en cuenta que el único trabajo encontrado al respecto es el de John Murphy, y que presenta las limitaciones citadas en los capítulos 1 (sección 1.2) y capítulo 2 (sección 2.4.3), el desarrollo de la teoría que permite la elaboración del sistema citado es uno de los aportes fundamentales de este trabajo.
- *Aplicación de los HMM al reconocimiento de patrones del análisis técnico bursátil*: Dado que no se han encontrado trabajos al respecto, la aplicación de los HMM a dichos sistemas constituye otro de los aportes de esta tesis.
- *Demostraciones de las proposiciones del capítulo 4*: Todas las proposiciones enunciadas en el capítulo 4 (cuya demostración se incluye en el apéndice A), figuraban sin prueba en todos los textos consultados. A muchos lectores, entre los que se incluye el autor de esta tesis, les puede resultar complejo entender las igualdades que las proposiciones enuncian con respecto a un tema novel en la literatura como lo son los HMM, por lo tanto las demostraciones (que no son triviales), son un aporte al correcto entendimiento del funcionamiento de los algoritmos de los HMM.
- *Topología del HMM para las etapas del mercado*: Dado que como fue expresado en la sección anterior, uno de los procedimientos más complejos es elegir adecuadamente la topología del HMM para un problema en particular. La topología propuesta para el caso de estudio de las etapas del mercado constituye otra contribución de esta tesis.

8.3 Líneas de continuación

Las líneas de continuación y/o futuros desarrollos posibles a partir de esta tesis se detallan a continuación:

Con respecto a la teoría:

- Investigar el uso de IOHMM (modelo oculto de Markov de entrada y salida) como reemplazo de los HMM en la etapa de clasificación de los sistemas de reconocimiento de patrones, esta variación de los HMM poseen además de un conjunto de símbolos de observación (*outputs*), un conjunto de símbolos de entrada (*inputs*), y tienen otras peculiaridades como por ejemplo la cadena de Markov oculta puede ser *no homogénea*. La presentación de estos modelos puede verse en [BF95], y una aplicación a series de retornos financieros puede consultarse en [BLD02].
- Investigar y desarrollar la teoría que permita utilizar más de un modelo en la etapa de clasificación del sistema, para luego tomar la decisión de la mayoría. Como puede ser el uso de los tres siguientes: DTW (algoritmo dinámico de tiempo), HMM y ANN (Algoritmo basado en redes neuronales), la utilización de los tres citados, en el contexto del reconocimiento de voz se muestra en [LMM07].

Con respecto a la aplicación de la teoría:

- Aplicar la teoría a un caso de estudio que se corresponda con un sistema de reconocimiento que reconozca un patrón entre varios patrones bursátiles, como pueden ser: sistemas de reconocimiento de patrones de cambio de tendencia, sistemas de reconocimiento de patrones de continuación de tendencia o sistemas que reconozcan ambos tipos de patrones.
- Aplicar la teoría al reconocimiento de patrones en series temporales de otros ámbitos, como pueden ser: series temporales climatológicas, series temporales de la investigación médica, etc.

APÉNDICE A

Demostraciones del Capítulo 4

A.1 Demostración de la proposición 4.2

$$\gamma_t(i) = \frac{P(q_t = i, O | \lambda)}{\sum_{j=1}^N P(q_t = j, O | \lambda)}$$

Demostración: Por definición se tiene que $\gamma_t(i) = P(q_t = i | O, \lambda)$, luego:

$$\begin{aligned} \gamma_t(i) &= P(q_t = i | O, \lambda) \stackrel{(1)}{=} \frac{P(q_t = i, O, \lambda)}{P(O, \lambda)} \\ &\stackrel{(2)}{=} \frac{P(q_t = i, O | \lambda) P(\lambda)}{P(O, \lambda)} \\ &\stackrel{(3)}{=} \frac{P(q_t = i, O | \lambda)}{P(O | \lambda)} \end{aligned}$$

Donde en (1) y (3) se aplica la definición de probabilidad condicional [Dev01] y en (2) la regla de multiplicación. Lo anterior implica que sólo se debe demostrar que

$$P(O | \lambda) = \sum_{j=1}^N P(q_t = j, O | \lambda).$$

Recuérdese que $P(O | \lambda)$ es la probabilidad de obtener la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, dado el modelo; asimismo, $P(q_t = j, O | \lambda)$ es la probabilidad conjunta de que en el tiempo t , el estado sea j y se obtenga la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, dado el modelo; por lo tanto $\sum_{j=1}^N P(q_t = j, O | \lambda)$ es

la probabilidad de que el estado en el momento t sea *cualquiera*⁵¹, y se obtenga la secuencia de observaciones $O = (o_1, o_2, \dots, o_T)$, dado el modelo, lo que es claramente igual a $P(O | \lambda)$. Más formalmente, se incluye la siguiente prueba:

$$\begin{aligned}
 P(O | \lambda) &\stackrel{(1)}{=} \frac{P(O, \lambda)}{P(\lambda)} \stackrel{(2)}{=} \frac{1}{P(\lambda)} \sum_{j=1}^N P(O, \lambda | q_t = j) P(q_t = j) = \\
 &\stackrel{(3)}{=} \frac{1}{P(\lambda)} \sum_{j=1}^N P(O, q_t = j, \lambda) \stackrel{(4)}{=} \frac{1}{P(\lambda)} \sum_{j=1}^N P(O, q_t = j | \lambda) P(\lambda) = \\
 &= \sum_{j=1}^N P(O, q_t = j | \lambda)
 \end{aligned}$$

Donde la primera igualdad (1), es por definición de probabilidad condicional [Dev01], para la segunda igualdad (2), nótese que los N eventos *distintos* que se corresponden a las N ocurrencias posibles del estado $j \in \{1, 2, \dots, N\}$ en el momento t , son *mutuamente excluyentes* y *exhaustivos*, por lo tanto, aplicando la ley de probabilidad total [Dev01], se tiene la equivalencia citada; en la tercera y cuarta igualdad, (3) y (4), se aplica nuevamente la definición de probabilidad condicional junto con la regla de multiplicación.

□

A.2 Demostración de la proposición 4.3

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}$$

Demostración: De la proposición anterior se sabe que:

⁵¹ Recuerde el lector que son N estados y que en el momento t el proceso *debe estar en uno y solo uno* de los N estados, además, aquí se asume que cada estado pertenece al conjunto $\{1, 2, \dots, N\}$.

$$\gamma_t(i) = \frac{P(q_t = i, O | \lambda)}{\sum_{j=1}^N P(q_t = j, O | \lambda)}$$

Recuérdese que:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \lambda) \text{ y } \beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda)$$

Pero $\alpha_t(i)$ se puede expresar como:

$$\begin{aligned} \alpha_t(i) &= P(o_1, o_2, \dots, o_t, q_t = i | \lambda) \stackrel{(1)}{=} \frac{P(o_1, o_2, \dots, o_t, q_t = i, \lambda)}{P(\lambda)} = \\ (2) &= \frac{1}{P(\lambda)} P(o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i, \lambda) = \\ (3) &= P(q_t = i | \lambda) P(o_1, o_2, \dots, o_t | q_t = i, \lambda) \end{aligned}$$

Donde en (1) y en (3) se hace uso de la definición de probabilidad condicional, y en (2) se utiliza la regla de la multiplicación.

Luego:

$$\begin{aligned} \alpha_t(i) \beta_t(i) &= \\ &= P(q_t = i | \lambda) P(o_1, o_2, \dots, o_t | q_t = i, \lambda) P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda) = \\ (1) &= P(q_t = i | \lambda) P(o_1, o_2, \dots, o_t, o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda) = \\ (2) &= P(q_t = i | \lambda) \frac{P(o_1, o_2, \dots, o_t, o_{t+1}, o_{t+2}, \dots, o_T, q_t = i, \lambda)}{P(q_t = i, \lambda)} = \\ (3) &= P(q_t = i | \lambda) \frac{P(o_1, o_2, \dots, o_t, o_{t+1}, o_{t+2}, \dots, o_T, q_t = i | \lambda)}{P(q_t = i, \lambda)} P(\lambda) = \\ (4) &= P(o_1, o_2, \dots, o_t, o_{t+1}, o_{t+2}, \dots, o_T, q_t = i | \lambda) = P(O, q_t = i | \lambda) \end{aligned}$$

Donde en (1) se utiliza la *independencia* entre observaciones, en (2) y (4) la definición de probabilidad condicional y en (3) la regla de multiplicación.

□

A.3 Demostración de la proposición 4.4

$$\delta_{t+1}(i) = \left[\max_{1 \leq i \leq N} \delta_t(i) a_{ij} \right] b_j(o_{t+1})$$

Demostración: Primero se debe analizar el cálculo de $\delta_t(i) a_{ij}$ que se muestra a continuación.

$$\delta_t(i) a_{ij} = \left(\max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t | \lambda) \right) P(q_{t+1} = j | q_t = i, \lambda)$$

Pero como la maximización no incluye los estados de los momentos t y $t + 1$, se puede tomar:

$$\delta_t(i) a_{ij} = \max_{q_1, q_2, \dots, q_{t-1}} \left[P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t | \lambda) P(q_{t+1} = j | q_t = i, \lambda) \right]$$

Como:

$$\begin{aligned} P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t | \lambda) &= \frac{P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t, \lambda)}{P(\lambda)} = \\ &= \frac{P(q_1, q_2, \dots, q_{t-1}, o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i, \lambda)}{P(\lambda)} = \\ &= P(q_1, q_2, \dots, q_{t-1}, o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i | \lambda) \end{aligned}$$

Donde la primera y última igualdad es por definición de probabilidad condicional y la segunda por regla de la multiplicación.

Entonces:

$$\begin{aligned}
 & P(q_1, q_2, \dots, q_t = i, o_1, o_2, \dots, o_t | \lambda) P(q_{t+1} = j | q_t = i, \lambda) = \\
 & = P(q_1, q_2, \dots, q_{t-1}, o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i | \lambda) P(q_{t+1} = j | q_t = i, \lambda) = \\
 & \stackrel{(1)}{=} P(q_1, q_2, \dots, q_{t-1}, q_{t+1} = j, o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i | \lambda) = \\
 & \stackrel{(2)}{=} \frac{P(q_1, q_2, \dots, q_{t-1}, q_{t+1} = j, o_1, o_2, \dots, o_t, q_t = i, \lambda) P(q_t = i | \lambda)}{P(q_t = i, \lambda)} = \\
 & \stackrel{(3)}{=} \frac{P(q_1, q_2, \dots, q_{t-1}, q_{t+1} = j, o_1, o_2, \dots, o_t, q_t = i | \lambda) P(\lambda) P(q_t = i | \lambda)}{P(q_t = i, \lambda)} = \\
 & \stackrel{(4)}{=} P(q_1, q_2, \dots, q_{t-1}, q_t = i, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda)
 \end{aligned}$$

Donde en (1) se utiliza el hecho de que q_{t+1} depende solamente de q_t y del modelo, pero no de los estados anteriores y menos de las observaciones, en (2) y (4) se aplicó la definición de probabilidad condicional y en (3) la regla de la multiplicación.

Por lo tanto:

$$\delta_t(i) a_{ij} = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_{t-1}, q_t = i, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda)$$

Luego como se debe tomar $\max_{1 \leq i \leq N} \delta_t(i) a_{ij}$ esto indica que el estado en el momento t , también debe maximizarse, es decir, forma parte de la mejor secuencia de estados que termina, en este caso, en el estado j , y por lo tanto:

$$\max_{1 \leq i \leq N} \delta_t(i) a_{ij} = \max_{q_1, q_2, \dots, q_{t-1}, q_t} P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda)$$

Finalmente como:

$$b_j(o_{t+1}) = P(o_{t+1} | q_{t+1} = j, \lambda)$$

Y utilizando el hecho de que la maximización no incluye el estado del momento $t + 1$, se tiene que:

$$\begin{aligned} & \left[\max_{1 \leq i \leq N} \delta_t(i) a_{ij} \right] b_j(o_{t+1}) = \\ & = \max_{q_1, q_2, \dots, q_{t-1}, q_t} \left[P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda) P(o_{t+1} | q_{t+1} = j, \lambda) \right] \end{aligned}$$

Pero como:

$$\begin{aligned} & P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda) = \\ & = \frac{P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t, \lambda)}{P(\lambda)} = \\ & = \frac{P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t | q_{t+1} = j, \lambda) P(q_{t+1} = j, \lambda)}{P(\lambda)} = \\ & = P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) \end{aligned}$$

Donde la primera y última igualdad es por definición de probabilidad condicional y la segunda por regla de la multiplicación.

Se tiene que:

$$\begin{aligned} & P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t | \lambda) P(o_{t+1} | q_{t+1} = j, \lambda) = \\ & = P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) P(o_{t+1} | q_{t+1} = j, \lambda) = \\ & \stackrel{(1)}{=} P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t, o_{t+1} | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) = \\ & \stackrel{(2)}{=} \frac{P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t, o_{t+1}, q_{t+1} = j, \lambda)}{P(q_{t+1} = j, \lambda)} P(q_{t+1} = j | \lambda) = \\ & \stackrel{(3)}{=} \frac{P(q_1, q_2, \dots, q_{t-1}, q_t, o_1, o_2, \dots, o_t, o_{t+1}, q_{t+1} = j | \lambda) P(\lambda)}{P(q_{t+1} = j, \lambda)} P(q_{t+1} = j | \lambda) = \\ & \stackrel{(4)}{=} P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t, o_{t+1} | \lambda) \end{aligned}$$

En (1) se utiliza el hecho de que o_{t+1} depende solo de q_{t+1} y del modelo, en (2) y (4) se usó la definición de probabilidad condicional y en (3) la regla de la multiplicación.

Lo que implica que:

$$\begin{aligned} \left[\max_{1 \leq i \leq N} \delta_t(i) a_{ij} \right] b_j(o_{t+1}) &= \\ &= \max_{q_1, q_2, \dots, q_{t-1}, q_t} P(q_1, q_2, \dots, q_{t-1}, q_t, q_{t+1} = j, o_1, o_2, \dots, o_t, o_{t+1} | \lambda) \end{aligned}$$

Con lo que concluye la prueba. □

A.4 Demostración de la proposición 4.6

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^N \sum_{l=1}^N \alpha_t(k) a_{kl} b_l(o_{t+1}) \beta_{t+1}(l)}$$

Demostración: Por definición se tiene que $\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda)$, luego:

$$\begin{aligned} P(q_t = i, q_{t+1} = j | O, \lambda) &\stackrel{(1)}{=} \frac{P(q_t = i, q_{t+1} = j, O, \lambda)}{P(O, \lambda)} = \\ &\stackrel{(2)}{=} \frac{P(q_t = i, q_{t+1} = j, O | \lambda) P(\lambda)}{P(O, \lambda)} \stackrel{(3)}{=} \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)} \end{aligned}$$

Donde en (1) y (3) se utilizó la definición de probabilidad condicional, y en (2) la regla de multiplicación.

A continuación se demostrará que:

$$P(O, \lambda) = \sum_{k=1}^N \sum_{l=1}^N P(q_t = k, q_{t+1} = l, O | \lambda)$$

De la demostración para la proposición 4.2 se tiene que:

$$P(O | \lambda) = \sum_{k=1}^N P(q_t = k, O | \lambda)$$

Luego se tiene que:

$$\begin{aligned} P(O | \lambda) &= \sum_{k=1}^N P(q_t = k, O | \lambda) \stackrel{(1)}{=} \frac{1}{P(\lambda)} \sum_{k=1}^N P(q_t = k, O, \lambda) \\ &\stackrel{(2)}{=} \frac{1}{P(\lambda)} \sum_{k=1}^N \sum_{l=1}^N P(q_t = k, O, \lambda | q_{t+1} = l) P(q_{t+1} = l) = \\ &\stackrel{(3)}{=} \frac{1}{P(\lambda)} \sum_{k=1}^N \sum_{l=1}^N P(q_t = k, q_{t+1} = l, O, \lambda) = \\ &\stackrel{(4)}{=} \frac{1}{P(\lambda)} \sum_{k=1}^N \sum_{l=1}^N P(q_t = k, q_{t+1} = l, O | \lambda) P(\lambda) = \\ &= \sum_{k=1}^N \sum_{l=1}^N P(q_t = k, q_{t+1} = l, O | \lambda) \end{aligned}$$

Donde la primera igualdad (1), es por definición de probabilidad condicional, para la segunda igualdad (2), notar que los N eventos *distintos* que se corresponden a las N ocurrencias posibles del estado $l \in \{1, 2, \dots, N\}$ en el momento $t + 1$, son *mutuamente excluyentes y exhaustivos*, por lo tanto, aplicando la ley de probabilidad total, se tiene la equivalencia citada; en la tercera y cuarta igualdad, (3) y (4), se aplica nuevamente la definición de probabilidad condicional junto con la regla de multiplicación.

Finalmente se dará una demostración formal de:

$$P(q_t = i, q_{t+1} = j, O | \lambda) = \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$$

Primero se verá el cálculo de $\alpha_t(i) a_{ij}$:

$$\begin{aligned}
 \alpha_t(i)a_{ij} &= P(o_1, o_2, \dots, o_t, q_t = i | \lambda) P(q_{t+1} = j | q_t = i, \lambda) = \\
 (1) &= \frac{P(o_1, o_2, \dots, o_t, q_t = i, \lambda)}{P(\lambda)} P(q_{t+1} = j | q_t = i, \lambda) = \\
 (2) &= \frac{P(o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i, \lambda)}{P(\lambda)} P(q_{t+1} = j | q_t = i, \lambda) = \\
 (3) &= P(o_1, o_2, \dots, o_t | q_t = i, \lambda) P(q_t = i | \lambda) P(q_{t+1} = j | q_t = i, \lambda) = \\
 (4) &= P(o_1, o_2, \dots, o_t, q_{t+1} = j | q_t = i, \lambda) P(q_t = i | \lambda) = \\
 (5) &= \frac{P(o_1, o_2, \dots, o_t, q_t = i, q_{t+1} = j, \lambda)}{P(q_t = i, \lambda)} P(q_t = i | \lambda) = \\
 (6) &= \frac{P(o_1, o_2, \dots, o_t, q_t = i, q_{t+1} = j | \lambda) P(\lambda)}{P(q_t = i, \lambda)} P(q_t = i | \lambda) = \\
 (7) &= P(o_1, o_2, \dots, o_t, q_t = i, q_{t+1} = j | \lambda)
 \end{aligned}$$

Donde en (1), (3), (5) y (7) se utilizó la definición de probabilidad condicional, en (2) y (6) la regla de multiplicación y en (4) la independencia de q_{t+1} con respecto a las observaciones o_1, o_2, \dots, o_t .

A continuación se mostrará el cálculo de $\alpha_t(i)a_{ij}b_j(o_{t+1})$:

$$\begin{aligned}
 \alpha_t(i)a_{ij}b_j(o_{t+1}) &= P(o_1, o_2, \dots, o_t, q_t = i, q_{t+1} = j | \lambda) P(o_{t+1} | q_{t+1} = j, \lambda) = \\
 (1) &= \frac{P(o_1, o_2, \dots, o_t, q_t = i, q_{t+1} = j, \lambda)}{P(\lambda)} P(o_{t+1} | q_{t+1} = j, \lambda) = \\
 (2) &= P(o_1, o_2, \dots, o_t, q_t = i | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) P(o_{t+1} | q_{t+1} = j, \lambda) = \\
 (3) &= P(o_1, o_2, \dots, o_t, o_{t+1}, q_t = i | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda)
 \end{aligned}$$

Donde en (1) se recurrió a la definición de probabilidad condicional, en (2) a la regla de multiplicación, y en (3) se usó el hecho de que o_{t+1} depende sólo de q_{t+1} y del modelo.

Por último se verá el cálculo de $\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$:

Recuérdese que $\beta_{t+1}(j) = P(o_{t+2}, o_{t+3}, \dots, o_T | q_{t+1} = j, \lambda)$, luego:

$$\begin{aligned}
 \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) &= \\
 &= P(o_1, o_2, \dots, o_t, o_{t+1}, q_t = i | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) \cdot \\
 &\quad \cdot P(o_{t+2}, o_{t+3}, \dots, o_T | q_{t+1} = j, \lambda) = \\
 (1) &= P(o_1, o_2, \dots, o_t, o_{t+1}, o_{t+2}, \dots, o_T, q_t = i | q_{t+1} = j, \lambda) P(q_{t+1} = j | \lambda) = \\
 (2) &= \frac{P(O, q_t = i, q_{t+1} = j, \lambda)}{P(q_{t+1} = j, \lambda)} P(q_{t+1} = j | \lambda) = \\
 (3) &= \frac{P(O, q_t = i, q_{t+1} = j | \lambda) P(\lambda)}{P(q_{t+1} = j, \lambda)} P(q_{t+1} = j | \lambda) = \\
 (4) &= P(q_t = i, q_{t+1} = j, O | \lambda)
 \end{aligned}$$

Donde en (1) se usó el hecho de que las observaciones $o_{t+2}, o_{t+3}, \dots, o_T$ sólo dependen de $q_{t+2}, q_{t+3}, \dots, q_T$ respectivamente y del modelo, en (2) y (4) se recurrió a la definición de probabilidad condicional, y en (3) a la regla de multiplicación.

Con lo que concluye la demostración. □

A.5 Demostración de la proposición 4.7

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \tag{4.27}$$

Demostración: De la proposición 4.2 se tiene que:

$$\gamma_t(i) = \frac{P(q_t = i, O | \lambda)}{P(O | \lambda)}$$

Y en la primera parte de la demostración de la proposición 4.6, se probó que:

$$\xi_t(i, j) = \frac{P(q_t = i, q_{t+1} = j, O | \lambda)}{P(O | \lambda)}$$

Por lo tanto, sólo se debe probar la siguiente igualdad:

$$P(q_t = i, O | \lambda) = \sum_{j=1}^N P(q_t = i, q_{t+1} = j, O | \lambda)$$

Dicha prueba se exhibe a continuación:

$$\begin{aligned} P(q_t = i, O | \lambda) &= \frac{P(q_t = i, O, \lambda)}{P(\lambda)} = \\ (1) &= \frac{1}{P(\lambda)} \sum_{j=1}^N P(q_t = i, O, \lambda | q_{t+1} = j) P(q_{t+1} = j) = \\ (2) &= \frac{1}{P(\lambda)} \sum_{j=1}^N P(q_t = i, q_{t+1} = j, O, \lambda) = \\ (3) &= \frac{1}{P(\lambda)} \sum_{j=1}^N P(q_t = i, q_{t+1} = j, O | \lambda) P(\lambda) = \\ &= \sum_{j=1}^N P(q_t = i, q_{t+1} = j, O | \lambda) \end{aligned}$$

Donde la primera igualdad (1), se debe a que los N eventos *distintos* que se corresponden a las N ocurrencias posibles del estado $j \in \{1, 2, \dots, N\}$ en el momento $t + 1$, son *mutuamente excluyentes* y *exhaustivos*, por lo tanto se aplica la ley de probabilidad total, la segunda igualdad (2) es debida a la definición de probabilidad condicional, y en la tercera (3) se aplicó la regla de multiplicación.

Con lo que concluye la demostración.

□

APÉNDICE B

B.1 Archivo de ejemplo para el ingreso manual de un HMM

Ejemplo de archivo con el formato especificado en la sección 7.8 del capítulo 7 (que se corresponde a la figura 7.10 de dicho capítulo), donde se almacena un HMM para su posterior uso en el sistema de reconocimiento de patrones. Este archivo se utilizó para ingresar al sistema el HMM inicial en la fase de entrenamiento de la implementación del caso de estudio de dicho capítulo.

HMM

ESTADOS 4

ALFABETO 27

ESTADO

PI 0.25

A 0.333 0.333 0.0 0.333

B 0.001 0.002 0.002 0.006 0.007 0.007 0.008 0.008 0.008
0.0833 0.0833 0.0833 0.133 0.133 0.133 0.0833 0.0833 0.0833
0.008 0.008 0.008 0.007 0.007 0.006 0.002 0.002 0.001

ESTADO

PI 0.25

A 0.0 0.5 0.5 0.0

B 0.166 0.166 0.166 0.1 0.1 0.1 0.033 0.033 0.033 0.0133
0.0133 0.0133 0.0133 0.0133 0.0133 0.0066 0.0066 0.0066 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0

ESTADO

PI 0.25

A 0.0 0.333 0.333 0.333

B 0.001 0.002 0.002 0.006 0.007 0.007 0.008 0.008 0.008
0.0833 0.0833 0.0833 0.133 0.133 0.133 0.0833 0.0833 0.0833
0.008 0.008 0.008 0.007 0.007 0.006 0.002 0.002 0.001

ESTADO

PI 0.25

A 0.5 0.0 0.0 0.5

B 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0066 0.0066 0.0066

0.0133 0.0133 0.0133 0.0133 0.0133 0.0133 0.033 0.033 0.033

0.1 0.1 0.1 0.166 0.166 0.166

BIBLIOGRAFÍA

[AP97] AMAT, ORIOL; PUIG, XAVIER “Análisis Técnico Bursátil” (Sexta Edición) (1997) Gestión 2000.

[Ben99] BENGIO, YOSHUA “Markovian Models for Sequential” Data (1999) Dept. Informatique et Recherche Opérationnelle – Université de Montreal, Montreal, Qc H3C-3J7, Canada

[BF95] BENGIO, YOSHUA; FRASCONI, PAOLO “An Input Output HMM Architecture” (1995) Université de Montréal, Università di Firenze.

[BH04] BHAR, RAMAPRASAD; HAMORI, SHIGEYUKI “Hidden Markov Models - Applications to Financial Economics” (2004) Springer Science + Business Media, Inc.

[BKKP06] BEZDEK, JAMES C.; KELLER, JAMES; KRISNAPURAM, RAGHU; PAL, NIKHIL R. “Fuzzy Models and Algorithms for Pattern Recognition and Image Processing” (softcover printing) (2005) Springer.

[BLD02] BENGIO, YOSHUA; LAUZON, VINCENT-PHILIPPE; DUCHARME, RÉJEAN “Experiments on the Application of IOHMMs to Model Financial Returns Series” (2002) Scientific Series, Cirano.

[BLN03] BASHARIN, GELY P.; LANGVILLE, AMY N.; NAUMOV, VALERIY A. “The life and work of A.A. Markov” (2003) Science Direct

[Bul06] BULLA, JAN “Application of Hidden Markov Models and Hidden Semi-Markov Models to Financial Time Series” (2006) PhD Dissertation Faculty of Economics and Business Administration of the Georg-August-University of Göttingen.

[Cha96] CHATFIELD, CHRIS “The Analysis of Time Series: An introduction” (Fifth Edition) (2002) John Wiley & Sons.

[Chu67] CHUN, K.L., “Markov chains with stationary transition probabilities”, (1967) Springer – Verlag.

[CMR07] CAPPE, OLIVIER; MOULINES, ERIC; RYDEN, TOBIAS “Inference in Hidden Markov Models” (2005) Springer

[Dev01] DEVORE, JAY L. "Probabilidad y estadística para ingeniería y ciencias" (Quinta Edición) (2001) Internacional Thomson Editores S.A.

[Din65] DINKIN E.B. "Markov Processes" Vol. I (1965) Springer – Verlag

[Doo90] DOOB, JOSEPH L. "Stochastic Processes" (1990) (Reedición del clásico de 1953) John Wiley & Sons.

[Edd97] EDDY, SEAN R. "Hidden Markov Model and Large-Scale Genome Analysis" (1997) Dept. of Genetics, Washington University School of Medicine.

[FD03] FRANSES, PHILIP HANS; VAN DIJK, DICK "Nonlinear Time Series Models in Empirical Finance" (2003) Cambridge University Press.

[FK99] FRIEDMAN, MENAHEM; KANDEL, ABRAHAM "Introduction to pattern recognition: statistical, structural, neural, and fuzzy logic approaches" (1999) Imperial College Press, World Scientific Publishing Co. Pte. Ltd.

[For94] FORD, DAVID "The investors Guide to trade options" Financial Times, London (Versión en español) (1994) Folio S.A.

[Fuk90] FUKUNAGA, KEINOSUKE "Introduction to Statistical Pattern Recognition" (Second Edition) (1990) Academic Press.

[GF00] GONZALEZ, JULIAN; FERRUZ, LUIS "Aspectos Contables y Financieros de las Opciones" (1994) Gestión 2000.

[Hag98] HAGSTROM, ROBERT G. Jr. "The Warren Buffet Way" (1995) Jhon Wiley & Sons, Inc. (Versión en español) (Primera edición) (1998) Gestión 2000.

[Ham94] HAMILTON, JAMES D. "Time Series Analysis" (1994) Princeton University Press.

[HK06] HAN, JIAWEI; KAMBER, MICHELINE "Data Mining: Concepts and Techniques" (Second Edition) (2006) Elsevier Inc.

[HOZ03] HAYASHI, MIWA; OMAN, CHARLES M.; ZUSCHLAG, MICHAEL "Hidden Markov Models as a Tool to Measure Pilot Attention Switching" (2003) Massachusetts Institute of Technology & US. Department of Transportation Volpe Center.

[HPS72] HOEL, P.G., PORT, S.C. Y STONE, C.J., "Introduction to stochastic processes" (1972) Prospect Heights, Wareland.

[Kan03] KANTARDZIC, MEHMED “Data Mining: Concepts, Models, Methods, and Algorithms” (2003) John Wiley & Sons.

[KBM93] KROGH, ANDERS; BROWN, MICHAEL; MIAN, SAIRA “Hidden Markov Models in Computational Biology: Applications to Protein Modeling” (1993) Computer and Information Sciences. Sinsheimer Laboratories. University of California.

[KS04] KANTZ, HOLGER; SCHREIBER, TOMAS “Nonlinear Time Series Analysis” (2004) Cambridge University Press.

[Lan99] LANE, TERRAN “Hidden Markov Model for Human/Computer Interface Modeling” (1999) Purdue University.

[Law95] LAWLER, GREGORY F. “Introduction to stochastic processes” (1995) Chapman & Hall.

[LM99] LO, ANDREW W.; MACKINLAY, CRAIG A “A Non-Random Walk Down Wall Street” (1999) Princeton University Press

[LMM07] DE LUNA ORTEGA, CARLOS ALEJANDRO; MARTÍNEZ ROMO, JULIO CÉSAR; MORA GONZÁLEZ, MIGUEL “Reconocimiento de Voz con Redes Neuronales, DTW y Modelos Ocultos de Markov” (2007) Universidad de Guadalajara.

[Mar01] MARQUES DE SA, J. P “Pattern Recognition: Concepts, Methods and Applications” (2001) Springer.

[ME07] MAMON, ROGEMAR S.; ELLIOTT, ROBERT J. “Hidden Markov Models in Finance” (2007) Springer.

[Mur00] MURPHY, JOHN J. “Technical análisis of the Financial Markets” (1999) New York Institute of Finance (Versión en español) (Primera edición) (2000) Gestión 2000.

[Nel95] RANDOLPH NELSON “Probability, Stochastic Processes & Queuing Theory” (1995) Springer – Verlag.

[NFS03] NWE, TIN LAY; FOO, SAY WEI; DE SILVA, LIYANAGE C. “Speech emotion recognition using hidden Markov models” (2003) Elsevier Computer Science.

[Nis96] NISON, STEVE “beyond candlesticks” (1994) Jhon Wiley & Sons, Inc. (Versión en español) (Primera edición) (1996) Gesmovasa.

[OFC04] OATES, TIM; FIROIU, LAURA; COHEN, PAUL R. "Clustering Time Series with Hidden Markov Models and Dynamic Time Warping" (2004) Computer Science Department, University of Massachusetts.

[Pfi00] PFISTER ALEXANDER "Heterogeneous Trade Intervals in an Agent Based Financial Market" Working Paper N.99 (2003) Working Paper Series.

[PL99] PENTLAND, ALEX; LIU, ANDREW "Modeling and Prediction of Human Behavior" (1999) Massachusetts Institute of Technology.

[Rab89] RABINER, LAWRENCE R. "A tutorial on hidden Markov models and selected applications in speech recognition" IEEE, VOL. 77, NO. 2, (1989)

[RJ86] RABINER, L.R.; JUANG, B.H. "A introduction to Hidden Markov Models" IEEE, ASSP MAGAZINE, (1986)

[RJ93] RABINER, L.R.; JUANG, B.H. "Fundamentals of Speech Recognition" (1993) Prentice-Hall Internacional, Inc.

[Ros96] ROSS, SHELDON M. "Stochastic Processes (Second Edition)" (1996) Jhon Wiley & Sons.

[Ros97] ROSS, SHELDON M. "Introduction to probability models (Sixth Edition)" (1997) Academia Press.

[Sat05] SATCHWELL, CHRIS "Pattern Recognition and Trading Decisions" (2005) McGraw-Hill.

[SS06] SHUMWAY, ROBERT H.; STOFFER, DAVID S. "Time Series Analysis and Its Applications" (Second Edition) (2006) Springer.

[Ste05] STEEB, WILLI-HANS "The Nonlinear Workbook" (Third Edition) (2005) World Scientific Publishing Co. Pte. Ltd.

[TCPN05] TAGHVA, KAZEM; COOMBS, JEFFREY; PEREDA, RAY; NARTKER, THOMAS "Address Extraction Using Hidden Markov Models" (2005) Information Science Research Institute University of Nevada, Las Vegas.

[Tsa02] TSAY, RUEY S. "Analysis of Financial Time Series" (2002) John Wiley & Sons.

[Wei98] WEINSTEIN, STAN “Stan Weinstein’s Secrets for Profiting in Bull and Bear Markets” Dow Jones, Irwin (Versión en español) (Segunda edición) (1998) Gesmovasa.

[WH03] CHOU, WU; JUANG, BIING HWANG “Pattern Recognition in Speech and Language Processing” (2003) CRC Press LLC.

[WS98] WEIGEND, ANDREAS S.; SHI, SHANMING “Predicting Daily Probability Distributions of S&P500 Returns” (1998) Working Paper IS-98-23 New York University.

[YX94] YANG, TIE; XU, YANGSHENG “Hidden Markov Model for Gesture Recognition” (1994) The Robotics Institute, Carnegie Mellon University.