



UNIVERSIDAD AUTÓNOMA METROPOLITANA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
UNIDAD IZTAPALAPA

**Modelos Ocultos de Markov:
una Aplicación a las Finanzas**

Tesis que presenta

Ana Gabriela López Gutiérrez

Para obtener el grado académico de

Maestro en Ciencias

(Matemáticas Aplicadas e Industriales)

Directora de tesis: Dra. Blanca Rosa Pérez Salvador

Jurado calificador:

Presidente: Dra. Patricia Saavedra Barrera

Secretario: Dr. Juan González Hernández

Vocal: Dra. Blanca Rosa Pérez Salvador



México, D.F. Junio 2016



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No. 00142

Matrícula: 2123802819

MODELOS OCULTOS DE MARKOV:
UNA APLICACION A LAS FINANZAS

En la Ciudad de México, se presentaron a las 15:00 horas del día 29 del mes de junio del año 2016 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DRA. PATRICIA SAAVEDRA BARRERA
DRA. BLANCA ROSA PEREZ SALVADOR
DR. JUAN GONZALEZ HERNANDEZ

Bajo la Presidencia de la primera y con carácter de Secretario el último, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRA EN CIENCIAS (MATEMÁTICAS APLICADAS E INDUSTRIALES)

DE: ANA GABRIELA LOPEZ GUTIERREZ

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

APROBAR

Acto continuo, la presidenta del jurado comunicó a la interesada el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

REVISÓ

LIC. JULIO CESAR DE LARA ISASSI
DIRECTOR DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JOSE GILBERTO CORDOBA HERRERA

PRESIDENTA

DRA. PATRICIA SAAVEDRA BARRERA

VOCAL

DRA. BLANCA ROSA PEREZ SALVADOR

SECRETARIO

DR. JUAN GONZALEZ HERNANDEZ

Agradecimientos

En primer lugar quiero agradecer a Dios por permitirme terminar mi maestría y darme las fuerzas necesarias para lograrlo.

A mi familia por estar siempre conmigo y apoyarme en todo momento, por animarme y recordarme siempre que podía salir adelante, por confiar en mí. A mi esposo por brindarme su apoyo, por toda su paciencia y por animarme siempre a seguir adelante. A mi hijo que se ha convertido en un motivo más para seguir luchando.

A mi asesora, la Dra. Blanca Rosa Pérez Salvador por confiar en mí, por el tiempo dedicado, por su paciencia y por los consejos profesionales mencionados para que esta tesis quedara lo mejor posible. A todos mis maestros de la UAM-I que me formaron académicamente para llegar hasta aquí.

A mis sinodales la Dra. Patricia Saavedra Barrera y el Dr. Juan González Hernández, por su apoyo en la revisión y corrección de esta tesis, gracias por enriquecer este trabajo para que quedara lo mejor posible.

A mis amigos, que no por ser los últimos en mencionar son menos importantes, gracias por brindarme su amistad y estar conmigo en todo momento.

Índice general

1. Conceptos Básicos de procesos estocásticos	3
1.1. Procesos de Markov	3
1.1.1. Ecuaciones de Chapman-Kolmogorov	5
1.1.2. Clasificación de Estados de una Cadena de Markov	6
1.2. Proceso de Movimiento Browniano	10
2. Modelos Ocultos de Markov	15
2.1. Tres Problemas Básicos de los HMMs	23
2.1.1. Problema de Evaluación y el Algoritmo Forward-Backward	23
2.1.2. El Problema de Decodificación y el Algoritmo de Viterbi	26
2.1.3. El Problema de Aprendizaje	29
3. Modelo de Black-Scholes	35
3.1. Ecuación de Black-Scholes	39
3.1.1. Solución de la ecuación de Black-Scholes	41
3.1.2. Extensión de Markov del modelo binomial de múltiples periodos	47
3.2. La extensión de Markov del modelo de Black-Scholes	50
4. Modelo financiero de Markov usando Modelos Ocultos de Markov	57
4.1. Algoritmo	58
4.2. Resultados	60

4.2.1. Resultados con los datos en la primera ventana de tiempo	60
4.2.2. Resultados con los datos en la segunda ventana de tiempo	68
4.2.3. Comparación de precios futuros calculados con los métodos revisados .	75
5. Conclusiones	77
Programas en R	79

Introducción

De acuerdo a Mirás (2000), en los últimos años, los mercados financieros de capitales y derivados han experimentado un enorme auge, hasta el punto de convertirse en una de las industrias de mayor crecimiento y prosperidad. Este período de apogeo ha impulsado un estricto estudio de estos mercados mediante modelos matemáticos. Uno de los grandes problemas en las finanzas modernas es el de valorar (poner precio) a los, cada vez más numerosos y sofisticados, productos financieros: futuros, derivados, opciones, swaps, etc. La moderna teoría de la valuación de activos comienza con la publicación de la famosa fórmula de valuación de Black y Scholes (1973). Esta elegante fórmula cambió para siempre la forma en la que, tanto los teóricos académicos como los profesionales de los mercados, entienden la valuación de derivados. Desde su presentación, ha sido estudiada, analizada y contrastada en los mercados reales de opciones y futuros de todo el mundo. De hecho, pocas teorías han sufrido y resistido una revisión empírica tan rigurosa. La teoría de Black-Scholes ha salido airosa de este escrutinio no sólo por su flexibilidad y grado de aplicación sino porque la mayor parte de las ideas de las modernas teorías de valuación ya se encuentran originalmente en ella. El modelo de Black-Scholes ha servido de base para numerosas generalizaciones y extensiones por parte de académicos y de profesionales de las finanzas. Sin duda, la espectacular expansión de los productos financieros derivados está intrínsecamente unida a este formidable modelo.

Sin embargo, como el ambiente donde se desarrolla el proceso financiero puede cambiar constantemente de “bueno” a “malo”, de “malo” a “normal”, etc. y los valores de las acciones pueden estar afectados por este ambiente cambiante, siguiendo este sentido Jansen et al (2009) proponen modelar este ambiente cambiante mediante una cadena de Markov, con el espacio de estados dado por “bueno”, “normal”, “malo”, etc. Lo más importante del modelo de Markov de Black-Scholes es estimar apropiadamente los parámetros de la matriz de transición. Este problema

no está aun razonablemente resuelto. Por otro lado, Zuchini y McDonald (2009) introducen los modelos ocultos de Markov para analizar series de tiempo usando la programación R. Con base en esto, Tuyen (2010) usa los modelos ocultos de Markov, para encontrar el modelo de Markov que mejor se ajusta a los datos históricos de la bolsa de valores de Vietnam del Norte. Siguiendo esta idea, en esta tesis consideraremos los modelos ocultos de Markov para estimar los precios de las acciones, y a diferencia de Tuyen (2010), aquí se considera que la tasa de interés también depende de los estados del proceso de Markov.

La tesis consta de 4 capítulos, en el capítulo uno se hace una revisión de los procesos estocásticos y del movimiento Browniano. En el Capítulo 2 se estudia los modelos ocultos de Markov, y se presentan sus 3 problemas básicos y los respectivos algoritmos. En el Capítulo 3 se estudia el modelo de Black-Scholes, y se presenta su extensión, el modelo de Markov de Black-Scholes. Por último en el Capítulo 4 se desarrolla una aplicación de los modelos ocultos de Markov, como un modelo financiero de Markov, utilizando las tasas de interés interbancaria de equilibrio (TIIE) a 28 días del banco de México y los precios de las acciones de la bolsa mexicana de valores.

Capítulo 1

Conceptos Básicos de procesos estocásticos

En este capítulo se presentan algunos conceptos importantes necesarios para el desarrollo de esta tesis. La información presentada se fundamenta en los libros de Ross (2009) y Karlin (1975).

Definición 1.1. *Un proceso estocástico $\{X_t : t \in T\}$, el cual se denota por X_t , es una colección de variables aleatorias con respecto al parámetro T . Los valores posibles que tome el proceso X_t le llamaremos **espacio de estados** y es denotado por I . Si I es numerable le llamaremos **proceso con estados discretos**. Cuando T es numerable, diremos que X_t es un **proceso estocástico a tiempo discreto** y si $T = [0, \infty)$ entonces diremos que X_t es un **proceso estocástico a tiempo continuo**.*

Los elementos principales que distinguen los procesos estocásticos son: el espacio de estados I , el espacio de parámetros T y las relaciones de dependencia entre las variables aleatorias, las cuales abreviaremos como v.a.

1.1. Procesos de Markov

Definición 1.2. *Un proceso de Markov es un proceso estocástico X_t , en el cual dado el presente, el futuro del proceso no depende de su pasado. Matemáticamente, un proceso es de Markov si*

$$P(a < X_t \leq b \mid X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n) = P(a < X_t \leq b \mid X_{t_n} = x_n)$$

siempre que $t_1 < t_2 < \dots < t_n < t$.

Sea A un intervalo en \mathbb{R} , la función

$$P(X_t \in A \mid X_s = x), \quad t > s,$$

es llamada la **función de probabilidad de transición**. Cuando el proceso de Markov tiene un espacio de estados finito o numerable es llamado **Cadena de Markov**. Un proceso de Markov para el cual las relaciones $\{X_t, t \in [0, \infty)\}$ son funciones continuas es conocido como **proceso de difusión**. El proceso de Poisson es una cadena de Markov a tiempo continuo y el movimiento Browniano es un proceso de difusión.

Una cadena de Markov a tiempo discreto X_n es un proceso estocástico cuyo espacio de estados es un conjunto finito o numerable, y para el cual $T = \{0, 1, 2, \dots\}$, en donde la probabilidad de que una v.a. X_{n+1} esté en el estado j , dado que X_n está en el estado i es llamada la **probabilidad de transición de un paso** y es denotada por $P_{ij}^{n, n+1}$, es decir:

$$P_{ij}^{n, n+1} = P(X_{n+1} = j \mid X_n = i).$$

Cuando las probabilidades de transición de un paso son independientes del tiempo, es decir del valor de n , decimos que el Proceso de Markov tiene **probabilidades de transición estacionarias u homogéneas**; en este caso tenemos que $P_{ij}^{n, n+1} = P_{ij}$. Nos referimos a $\mathbf{P} = [P_{ij}]$ como la matriz de probabilidades de transición del proceso:

$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & & & & \end{pmatrix}$$

Las propiedades de esta matriz son:

- i) $P_{ij} \geq 0, \quad i, j = 0, 1, 2, \dots$
- ii) $\sum_{j=0}^{\infty} P_{ij} = 1, \quad i = 0, 1, 2, \dots$

Para el resto de este trabajo consideraremos cadenas de Markov con espacio de estados finito, $I = \{1, 2, \dots, m\}$. Para que una cadena de Markov esté completamente definida, es necesario especificar la distribución de probabilidad del estado inicial X_0 , es decir un vector

$$\mathbf{p} = (p_1, p_2, \dots, p_m),$$

tal que

$$p_i \geq 0 \quad , \quad \forall i \in I,$$

$$\sum_{i \in I} p_i = 1.$$

Para toda i , p_i representa la **probabilidad inicial** de empezar en i :

$$p_i = P(X_0 = i).$$

1.1.1. Ecuaciones de Chapman-Kolmogorov

Definimos ahora las probabilidades de transición de n pasos P_{ij}^n como la probabilidad de que un proceso en el estado i esté en el estado j después de n transiciones adicionales. Es decir,

$$P_{ij}^n = P\{X_{n+k} = j \mid X_k = i\}, \quad n \geq 0, i, j \geq 0.$$

Las *ecuaciones de Chapman-Kolmogorov* proporcionan un método para calcular estas probabilidades de transición de n pasos. Estas ecuaciones son

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \quad \forall n, m \geq 0, \forall i, j.$$

Formalmente, tenemos el siguiente teorema.

Teorema 1.1. *Si la matriz de probabilidades de transición de un paso de una cadena de Markov es $\mathbf{P} = [P_{ij}]$, entonces*

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \quad \forall n, m \geq 0, \forall i, j. \quad (1.1)$$

Donde definimos

$$P_{ij}^0 = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Demostración.

$$\begin{aligned}
 P_{ij}^{n+m} &= P\{X_{n+m} = j \mid X_0 = i\} \\
 &= \sum_{k=0}^{\infty} P\{X_{n+m} = j, X_n = k \mid X_0 = i\} \\
 &= \sum_{k=0}^{\infty} P\{X_{n+m} = j \mid X_n = k, X_0 = i\} P\{X_n = k \mid X_0 = i\} \\
 &= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n.
 \end{aligned}$$

□

Si suponemos que $\mathbf{P}^{(n)}$ denota la matriz de probabilidades de transición de n pasos P_{ij}^n , entonces la ecuación (1.1) afirma que

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$$

donde el punto representa la multiplicación de matrices. Por inducción se puede demostrar que

$$\mathbf{P}^{(n)} = \mathbf{P}^n. \quad (1.2)$$

La probabilidad de que el proceso empiece en el estado i al tiempo n es:

$$p_i(n) = P(X_n = i).$$

Si suponemos que $\mathbf{p}(n) = (p_1(n), p_2(n), \dots, p_m(n))$ entonces se puede resolver en notación matricial que $\mathbf{p}(n) = \mathbf{p} \cdot \mathbf{P}^n$.

1.1.2. Clasificación de Estados de una Cadena de Markov

Definición 1.3. Una matriz de transición \mathbf{P} es **regular** si existe un entero positivo k , tal que todos los elementos de la matriz $\mathbf{P}^{(k)}$ son estrictamente positivos.

De la relación (1.2), \mathbf{P} es regular si y sólo si existe un entero $k > 0$ tal que todos los elementos de la k -ésima potencia de \mathbf{P} son estrictamente positivos.

Definición 1.4. El estado j se dice que es **accesible** desde el estado i (escribimos $i \rightarrow j$), si para algún entero $n \geq 0$, $P_{ij}^n > 0$.

Definición 1.5. Dos estados i y j se dice que están **comunicados** si y sólo si $i \rightarrow j$ y $j \rightarrow i$, ó si $j = i$. Escribimos $i \leftrightarrow j$.

El concepto de comunicación es una relación de equivalencia.

Definición 1.6. Decimos que una cadena de Markov es **irreducible** si la relación de equivalencia induce una sola clase. En otras palabras un proceso es irreducible si todos los estados se comunican entre sí.

Definición 1.7. Definimos el periodo del estado i , escribiendo $d(i)$, como el máximo común divisor del conjunto de todos los enteros $n \geq 1$ para el cual

$$P_{ii}^{(n)} > 0.$$

Si $P_{ii}^{(n)} = 0$ para toda $n \geq 1$ definimos $d(i) = 0$.

Definición 1.8. Si $d(i) > 1$, el estado i se dice que es **periódico** con período $d(i)$. Si $d(i) = 1$, entonces el estado i es **aperiódico**.

Observación: Si \mathbf{P} es regular, entonces todos los estados son aperiódicos, ver Janssen y Manca (2006), pp. 109.

Definición 1.9. Una cadena de Markov cuyos estados son todos aperiódicos es conocida como **Cadena de Markov Aperiódica**.

Definición 1.10. Un estado i se dice que es **esencial** si y sólo si se comunica con cada estado que conduce a j ; de lo contrario se dice que es no esencial.

Claramente, si \mathbf{P} es regular, la cadena de Markov es irreducible y aperiódica, ver Janssen y Manca (2006), pp. 109. Tal cadena de Markov se dice que es **ergódica**. En otras palabras tenemos la siguiente definición.

Definición 1.11. Una cadena de Markov se dice que es una cadena de Markov **ergódica** si desde un estado, se puede llegar a cualquier otro estado (no necesariamente en un paso), es decir, $\exists k \in \mathbb{N}$ tal que $P_{ij}^k > 0 \forall i, j \in I$.

Definición 1.12. Considere un estado i arbitrario pero fijo. Definimos para cada número entero $n \geq 1$,

$$f_{ii}^n = P(X_n = i, X_v \neq i, v = 1, 2, \dots, n-1 \mid X_0 = i).$$

En otras palabras f_{ii}^n es la probabilidad que comenzando desde el estado i , el primer retorno al estado i ocurre en la n -ésima transición. Decimos que el estado i es **recurrente** si y sólo si $\sum_{i=1}^{\infty} f_{ii}^n = 1$, en caso contrario diremos que es **transitorio**.

Teorema 1.2. Sea \mathbf{P} la matriz de transición de una cadena de Markov ergódica, entonces el límite $\lim_{n \rightarrow \infty} p_i(n) = \pi_i$ existe independientemente de la distribución inicial $\mathbf{p} = (p_1, p_2, \dots, p_m)$ y $\{\pi_i\}$ es la única solución del sistema

$$\begin{cases} \pi_j = \sum_{i \in I} \pi_i p_{ij} \\ \sum_{i \in I} \pi_i = 1 \end{cases} \quad (1.3)$$

Cualquier conjunto $\{\pi_i\}$ que satisface las ecuaciones anteriores recibe el nombre de **distribución estacionaria**.

Demostración. Primero demostraremos la segunda parte.

Supongamos que el siguiente límite existe

$$\lim_{n \rightarrow \infty} p_j(n) = \pi_j, \quad j \in I, \quad (1.4)$$

empezando con $X_0 = i$.

Ahora, por la Ley de la probabilidad total tenemos

$$\begin{aligned} p_j(n+1) &= P(X_{n+1} = j) \\ &= \sum_{k \in I} P(X_{n+1} = j, X_n = k) \\ &= \sum_{k \in I} P(X_{n+1} = j \mid X_n = k) P(X_n = k) \\ &= \sum_{k \in I} p_k(n) p_{kj}. \end{aligned} \quad (1.5)$$

Suponiendo que podemos cambiar límites con sumas finitas, de la ecuación (1.5) tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} p_j(n+1) &= \lim_{n \rightarrow \infty} \sum_{k \in I} p_k(n) p_{kj} \\ &= \sum_{k \in I} \lim_{n \rightarrow \infty} p_k(n) p_{kj} \\ &= \sum_{k \in I} \pi_k p_{kj}. \end{aligned}$$

Así, por la ecuación (1.4)

$$\pi_j = \sum_{k \in I} \pi_k p_{kj}.$$

Ahora, como

$$\sum_{j \in I} p_j(n) = 1,$$

tenemos

$$\begin{aligned} \sum_{j \in I} \pi_j &= \sum_{j \in I} \lim_{n \rightarrow \infty} p_j(n) \\ &= \lim_{n \rightarrow \infty} \sum_{j \in I} p_j(n) \\ &= \lim_{n \rightarrow \infty} 1 \\ &= 1. \end{aligned}$$

Así,

$$\sum_{j \in I} \pi_j = 1.$$

Por lo tanto, $\{\pi_i\}$ cumple el sistema de ecuaciones (1.3).

El siguiente resultado

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j \quad (1.6)$$

es conocido como resultado *ergódico*, ya que la ecuación (1.6) es independiente del estado inicial i .

Para una distribución inicial \mathbf{p} tenemos

$$\lim_{n \rightarrow \infty} p_i(n) = \lim_{n \rightarrow \infty} \sum_{j \in I} p_j p_{ji}^{(n)},$$

pues

$$\begin{aligned} p_i(n) &= P(X_n = i) \\ &= \sum_{j \in I} P(X_n = i, X_0 = j) \\ &= \sum_{j \in I} P(X_n = i | X_0 = j) P(X_0 = j) \\ &= \sum_{j \in I} p_{ji}^{(n)} p_j. \end{aligned}$$

Luego, por la ecuación (1.6)

$$\begin{aligned}
 \lim_{n \rightarrow \infty} p_i(n) &= \sum_{j \in I} \lim_{n \rightarrow \infty} p_{ji}^{(n)} p_j \\
 &= \sum_{j \in I} \pi_i p_j \\
 &= \pi_i \sum_{j \in I} p_j \\
 &= \pi_i \cdot 1 \\
 &= \pi_i.
 \end{aligned}$$

Por lo tanto $\lim_{n \rightarrow \infty} p_i(n) = \pi_i$, es decir el límite existe independientemente de la distribución inicial \mathbf{p} . □

1.2. Proceso de Movimiento Browniano

El proceso de movimiento Browniano, a veces llamado el proceso de Winer, es uno de los procesos estocásticos más útiles en la teoría de la probabilidad aplicada. Tiene su origen en la física como una forma de modelar el movimiento Browniano. Este proceso se ha utilizado de forma beneficiosa en áreas tales como las pruebas estadísticas de bondad de ajuste, el análisis de los niveles de precios en el mercado de valores, y la mecánica cuántica.

Definición 1.13. *Un proceso estocástico $W = \{W_t, t \geq 0\}$ se dice que es un **proceso de movimiento Browniano o proceso de Wiener** con media μ y varianza σ^2 si:*

- W se adapta a la filtración básica¹
- W tiene incrementos estacionarios e independientes;
- Para todo $s, t, (0 \leq s < t), W_t - W_s$ tiene una distribución normal $N(\mu(t-s), \sigma^2(t-s))$;

¹El proceso estocástico X es adaptado a la filtración f si, para toda t , la v.a. X_t es \mathcal{F}_t -medible. Es decir, $\forall t \in T$:

$$X_t^{-1}(B) = \{w : X_t(w) \in B\} \in \mathcal{F}_t, \quad \forall B \in \mathfrak{B}.$$

Se dice que X es adaptado a la filtración básica cuando se trabaja en un espacio de probabilidad básico completo filtrado $(\Omega, \mathcal{F}, P, (\mathcal{F}_t, t \in [0, \infty)))$.

- $P(W_0 = x) = 1$, para algún $x \in \mathbb{R}$.

Cuando $\mu = 0$, $\sigma = 1$ y $x = 0$, el proceso es llamado **movimiento browniano estándar**.

Vamos ahora a dar las propiedades más importantes del movimiento Browniano estándar.

Propiedad 1. Si W es un movimiento browniano, entonces existe una modificación de W , el proceso W^* , tal que W^* tiene trayectorias continuas casi seguramente.

Propiedad 2. Si W es un movimiento browniano estándar, entonces W es una martingala.²

Propiedad 3. Si W es un movimiento browniano estándar, entonces el proceso Q donde $Q = (W_t^2 - t, t \in [0, \infty))$ es una martingala.

Propiedad 4. Si W es un movimiento browniano estándar, entonces para casi todo w , la trayectoria $w \mapsto W_t(w)$ no es de variación acotada en todo intervalo cerrado $[a, b]$.

Esto explica por qué es necesario para los modelos de finanzas y de seguros definir un nuevo tipo de integral, llamado integral estocástica o de Itô, si queremos integrar con respecto a W .

Debemos, a menos que se indique lo contrario, suponer a lo largo de esta tesis que trabajamos con un movimiento Browniano estándar.

Veamos ahora una aplicación del movimiento Browniano a las finanzas.

Variación del precio de las acciones

Sea S_i el precio de una determinada acción observada al término del i -ésimo día. Es común definir el retorno del precio de esta acción como

$$R_i = \frac{S_i - S_{i-1}}{S_{i-1}}. \quad (1.7)$$

Observemos que este retorno corresponde al incremento del día i -ésimo (positivo o negativo) respecto del precio de la acción del día anterior S_{i-1} . Vamos a presentar un modelo que se

²Sea X un proceso estocástico real definido en un espacio de probabilidad completo (Ω, \mathcal{F}, P) . El proceso X es llamado una martingala si:

1. $\forall t \geq 0, \exists E(X_t)$,
2. $s < t \Rightarrow E(X_t | X_s) = X_s$, casi seguramente.

puede catalogar de ingenuo, pero sin lugar a duda nos ayudará a aclarar conceptos del cálculo estocástico. Supongamos que la sucesión $\{R_i, i = 0, 1, \dots\}$ admite la siguiente estructura

$$R_i = \mu + \varepsilon_i, \quad (1.8)$$

donde μ es el retorno esperado y las variables ε_i son independientes y se distribuyen según una distribución normal $N(0, \sigma^2)$; de manera tal que las R_i son independientes e idénticamente distribuidas según una distribución normal $N(\mu, \sigma^2)$. Considerando las ecuaciones (1.7) y (1.8), se sigue que:

$$\frac{S_i - S_{i-1}}{S_{i-1}} = \mu + \sigma \eta_i,$$

donde η_i son i.i.d. según una distribución normal $N(0, 1)$, de manera que

$$S_i - S_{i-1} = \mu S_{i-1} + \sigma S_{i-1} \eta_i, \quad i = 0, 1, \dots \quad (1.9)$$

Observemos que esta ecuación en diferencias finitas actúa en el espacio de tiempo discreto $\{0, 1, 2, \dots\}$. Extendamos nuestro espacio de tiempo a $[0, \infty)$, y efectuemos una partición, siempre discreta, de manera de proponer un modelo levemente más general, y que en cualquier caso cuando la partición coincida con el conjunto $\{0, 1, 2, \dots\}$ nos resulte la ecuación (1.9). Para $0 = t_0 < t_1 < \dots < t_n < \dots$, hagamos

$$S_{t_i} - S_{t_{i-1}} = \mu S_{t_{i-1}}(t_i - t_{i-1}) + \sigma S_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \quad (1.10)$$

donde $\{W_t; t \geq 0\}$ es un movimiento browniano.

Si la partición $0 = t_0 < t_1 < \dots < t_n < \dots$ es tal que $t_i - t_{i-1} = \Delta t$, entonces si $\Delta t = 1$ la ecuación (1.10) es en esencia la ecuación (1.9). Se puede demostrar, mediante técnicas avanzadas del cálculo estocástico, que toda vez que esta partición sea más fina, es decir cuando $\Delta t \rightarrow 0$, entonces el proceso S_{t_i} definido por la ecuación (1.10) converge (en distribución) a un proceso de difusión S_t , que satisface la ecuación diferencial estocástica

$$dS_t = \mu S_t dt + \sigma S_t dW. \quad (1.11)$$

Obtener la trayectoria de la evolución del precio de la acción S_t , significa resolver la integración estocástica de $\int S_t dW$.

La solución de la ecuación (1.11) es

$$S_t = S_0 \cdot e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}. \quad (1.12)$$

Esto significa que $\frac{S_t}{S_0}$ tiene una distribución lognormal³ con parámetros $((\mu - \frac{\sigma^2}{2})t; \sigma^2 t)$.

³La v.a. no negativa X tiene una *distribución lognormal* con parámetros μ, σ - la cual puede ser escrita como $X \prec LN(\mu, \sigma)$ - si la v.a. $\log X$ tiene una distribución normal con parámetros μ, σ^2 .

Capítulo 2

Modelos Ocultos de Markov

En este capítulo estudiaremos los modelos ocultos de Markov, los cuales son un autómata¹ de estados finitos capaz de producir a su salida una secuencia de símbolos observados.

Definición 2.1. *Un Modelo Oculto de Markov (HMM), ver Figura 2.1, es un autómata estocástico finito. Se puede resumir como una especie de doble proceso estocástico con los siguientes aspectos:*

- *El primer proceso estocástico es un proceso de Markov con un conjunto finito de estados, donde cada uno de estos está asociado con una distribución de probabilidad multidimensional. Las transiciones entre los diferentes estados están estadísticamente organizadas por un conjunto de probabilidades conocidas como probabilidades de transición.*
- *El segundo proceso estocástico, es un conjunto de posibles resultados observables en el tiempo cuya función de distribución depende de cada estado del primer proceso estocástico; esto es, cada estado del primer proceso estocástico nos genera una función de distribución de probabilidades para la variable observable.*
- *El investigador solo tiene información de las variables observables del segundo proceso estocástico porque los estados del primer proceso estocástico son desconocidos; o permanecen “ocultos”, de ahí el nombre de “Modelo Oculto de Markov”.*

¹Un autómata es un modelo matemático para una máquina de estado finito. Una máquina de estado finito es una máquina que, dada una entrada de símbolos, “salta” a través de una serie de estados de acuerdo a una función de transición. Esta función de transición dice al autómata a qué estado cambiar dado determinado estado y símbolo.

Los elementos que constituyen un Modelo Oculto de Markov son cinco:

1. Un conjunto de m estados, definido por

$$I = \{I_1, I_2, \dots, I_m\}. \quad (2.1)$$

Los estados deben estar conectados entre sí, de forma que cualquiera de ellos pueda ser alcanzable desde al menos otro de los estados.

2. Un conjunto de n simbolos observables,

$$V = \{v_1, \dots, v_n\}.$$

Si las observaciones son continuas entonces V es un intervalo.

3. Una matriz de probabilidades de transición, $\mathbf{P} = [p_{ij}]$. Esta matriz es cuadrada de dimensión m y cada elemento p_{ij} corresponde a la probabilidad de transición del estado I_i al estado I_j , por comodidad escribiremos el estado I_i como i y el estado I_j como j

$$p_{ij} = P(X_t = j \mid X_{t-1} = i) \quad 1 \leq i, j \leq m, \quad (2.2)$$

donde X_t indica el estado en el que se encuentra el modelo en el instante t . Las probabilidades de transición deben satisfacer las restricciones estocásticas normales, $p_{ij} \geq 0$, $1 \leq i, j \leq m$ y $\sum_{j=1}^m p_{ij} = 1$, $1 \leq i \leq m$. Por el punto 1, esta matriz de transición \mathbf{P} , es ergódica.

4. Un conjunto de parámetros $B = \{b_j(k)\}$, donde $b_j(k)$ es la probabilidad de que el simbolo v_k sea emitido en el estado j .

$$b_j(k) = P(O_t = v_k \mid X_t = j), \quad 1 \leq j \leq m, \quad 1 \leq k \leq n \quad (2.3)$$

donde v_k denota el k -ésimo simbolo observado, y O_t representa la observación en el instante de tiempo t . Estas probabilidades deben satisfacer:

$$b_j(k) \geq 0, \quad 1 \leq j \leq m, \quad 1 \leq k \leq n \quad \text{y} \quad \sum_{k=1}^n b_j(k) = 1, \quad 1 \leq j \leq m.$$

Si las observaciones son discretas utilizaremos distribuciones de probabilidad, y si son continuas usaremos funciones de densidad de probabilidad.

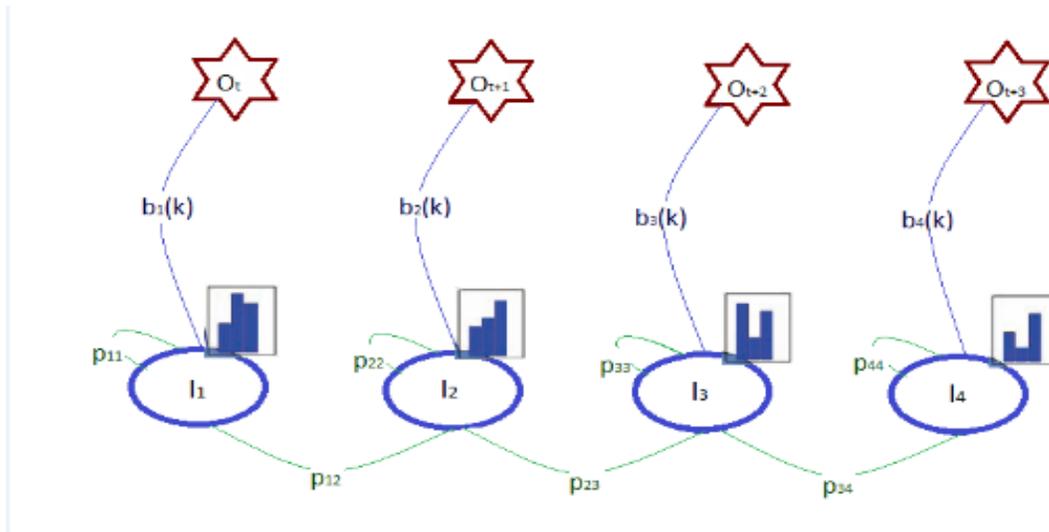


Figura 2.1: Ejemplo de un HMM. Cada estado I_j genera una función de distribución para los observables O_t .

5. Un conjunto de probabilidades para los estados en el tiempo inicial $\mathbf{p} = \{p_i\}$, donde p_i es la probabilidad de que en el tiempo $t = 0$ se esté en el estado i :

$$p_i = P(X_0 = i), \quad 1 \leq i \leq m, \quad (2.4)$$

y satisface las condiciones:

$$0 \leq p_i \leq 1, \quad 1 \leq i \leq m; \quad \sum_{i=1}^m p_i = 1.$$

Un HMM queda definido completamente al especificar los tres elementos: \mathbf{p} , P y B , que identifican al modelo. El modelo se identifica con la letra λ :

$$\lambda = (\mathbf{p}, P, B). \quad (2.5)$$

En la figura (2.1) se esquematiza un modelo oculto de Markov con cuatro estados $\{I_1, I_2, I_3, I_4\}$, los estados se caracterizan por los ovalos, las flechas que salen de estos ovalos indican cuales son los estados que están comunicados a un paso; aquí se ve que las probabilidades de transición que son positivas son $p_{11}, p_{12}, p_{22}, p_{23}, p_{33}, p_{34}$ y p_{44} . Los histogramas representan la función de probabilidad de los observables, cada estado genera una distinta función de probabilidad representada por los histogramas mostrados. Finalmente lo que observamos son las realizaciones de $O_t, O_{t+1}, O_{t+2}, O_{t+3}$.

En el caso de que las observaciones O_t sean variables aleatorias discretas, tenemos:

$$\begin{aligned} P(O_t = v_k) &= \sum_{i \in I} P(X_t = i)P(O_t = v_k | X_t = i) \\ &= \sum_{i \in I} p_i(t)b_i(k), \end{aligned} \quad (2.6)$$

debido a que $p_i(t) = P(X_t = i)$ para $t = 0, 1, 2, \dots, n$ e $i \in I$. Notemos que $\mathbf{p}(t) = \mathbf{p}\mathbf{P}^t$, de ahí que

$$P(O_t = v_k) = \mathbf{p}\mathbf{P}^t\mathbf{B}(k)\mathbf{1}', \quad (2.7)$$

en el cual $\mathbf{p}(t)$ es el vector renglón de elementos $p_i(t)$ y $\mathbf{1}'$ es el vector columna de elementos 1. $\mathbf{B}(k)$ es la matriz diagonal con i -ésimo elemento en la diagonal $b_i(k)$. Así, tenemos la función de verosimilitud como sigue:

$$L_n = \mathbf{p}\mathbf{B}(v_1)\mathbf{P}\mathbf{B}(v_2) \cdots \mathbf{P}\mathbf{B}(v_n)\mathbf{1}'. \quad (2.8)$$

La modelación estocástica de un HMM consiste de dos pasos:

- El primer paso es definir la arquitectura del modelo.
- El segundo paso es definir el algoritmo de aprendizaje y de operación.

La tarea del algoritmo de aprendizaje es encontrar el mejor conjunto de estados de transición y las probabilidades de observación. Por lo tanto, se da una secuencia de salida o un conjunto de estas secuencias.

Vamos a presentar un ejemplo para aclarar el concepto de los modelos ocultos de Markov.

Ejemplo 1. El ejemplo se presenta en dos casos. En el primer caso el proceso de Markov no se encuentra oculto; en el segundo caso se considera oculto el proceso de Markov y solo se tienen los datos observables.

Primer caso: dos procesos estocásticos, el primero es un proceso de Markov.

Considere que se tiene cuatro urnas con bolas de los siguientes colores:

Urna	Color de bola			
	Azul	Amarillo	Verde	Rojo
1	2	4	1	5
2	3	1	4	6
3	5	2	3	2
4	4	3	0	2

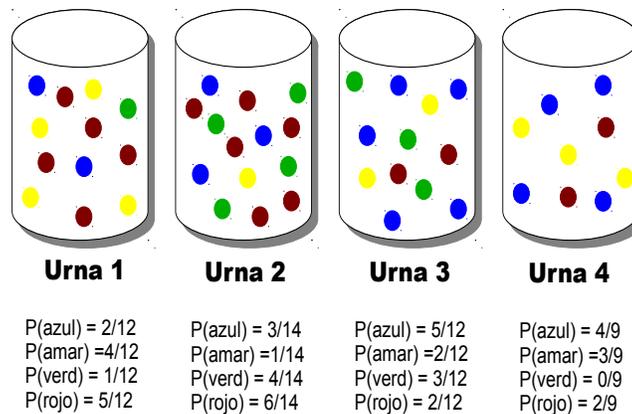


Figura 2.2: 4 urnas con bolas de 4 colores.

Una persona selecciona una urna al azar sucesivamente y de la urna seleccionada se extrae una bola, anota el color y la regresa, y pasa a la siguiente urna. La selección de la urna en el tiempo $t = 1$ es con igual probabilidad, $p_i = \frac{1}{4} = 0.25$, esta es la probabilidad inicial. La selección de las urnas subsecuentes se hace en el tiempo t , $t = 2, 3, 4, \dots$ y depende del color de la bola seleccionada de la urna en el tiempo $t - 1$, con la siguiente regla.

Color de la bola seleccionada de la urna en el paso $t - 1$	Urna seleccionada en el paso t
Azul	Urna 1
Amarillo	Urna 2
Verde	Urna 3
Rojo	Urna 4

De esta manera, la selección de la urna en el paso t , solo depende del color de la bola seleccionada de la urna en el paso $t - 1$, esto es, las urnas son los estados en un proceso de Markov y las probabilidades de transición de la urna i a la urna $j = 1, 2, 3, 4$; son dadas por

De la urna i a la urna 1:

$$P_{i1} = \frac{\# \text{ de bolas azules en la urna } i}{\text{Total de bolas en la urna } i}.$$

De la urna i a la urna 2:

$$P_{i2} = \frac{\# \text{ de bolas amarillas en la urna } i}{\text{Total de bolas en la urna } i}.$$

De la urna i a la urna 3:

$$P_{i3} = \frac{\# \text{ de bolas verdes en la urna } i}{\text{Total de bolas en la urna } i}.$$

De la urna i a la urna 4:

$$P_{i4} = \frac{\# \text{ de bolas rojas en la urna } i}{\text{Total de bolas en la urna } i}.$$

De esta manera, la matriz de transición es

$$P = \begin{pmatrix} \frac{2}{12} & \frac{4}{12} & \frac{1}{12} & \frac{5}{12} \\ \frac{3}{14} & \frac{1}{14} & \frac{4}{14} & \frac{6}{14} \\ \frac{5}{12} & \frac{2}{12} & \frac{3}{12} & \frac{2}{12} \\ \frac{4}{9} & \frac{3}{9} & \frac{0}{9} & \frac{2}{9} \end{pmatrix}$$

El segundo proceso estocástico, O_t , es generado por la urna seleccionada en el tiempo t . Este proceso genera una sucesión de variables aleatorias O_1, O_2, O_3, \dots , normales cuyos parámetros dependen de la urna seleccionada en el paso t .

Con este experimento se tienen dos procesos estocásticos relacionados, uno que nos describe el número de la urna seleccionada en el tiempo t , y que corresponde a un proceso de Markov con probabilidad de transición dado por p_{ij} y que denotaremos como: X_1, X_2, X_3, \dots , y el segundo nos describe las variables O_1, O_2, O_3 , etc. que tienen las siguientes funciones de probabilidad condicionadas al valor de X_t :

$$\begin{aligned} O_t|X_t = 1 &\sim N(3, 1); \\ O_t|X_t = 2 &\sim N(6, 1); \\ O_t|X_t = 3 &\sim N(9, 1); \\ O_t|X_t = 4 &\sim N(12, 1). \end{aligned}$$

Segundo caso:

Considere que hay k urnas con bolas de k colores, los colores los identificaremos como *color 1* = c_1 , *color 2* = c_2, \dots , *color k* = c_k . A estas urnas está conectado un mecanismo autónomo que selecciona las urnas sucesivamente. En el tiempo $t = 1$ la urna seleccionada es con igual probabilidad, esto es, $p_i = \frac{1}{k}$. La selección de las subsecuentes urnas dependen de la urna seleccionada en el paso anterior. Es decir, de la urna seleccionada en el paso t , el mecanismo elige una bola, detecta su color y la regresa a la urna. La selección de la urna en el paso $t + 1$ depende del color de la bola seleccionada en el paso t , mediante la siguiente regla:

Si la bola extraída de la urna en el paso t es del color j , entonces la urna seleccionada en el paso $t + 1$ es la número j , $j = 1, 2, \dots, k$. De esta manera, la probabilidad de pasar de la urna i a la urna j es

$$p_{ij} = \frac{\# \text{ de bolas del color } j \text{ en la urna } i}{\# \text{ de bolas en la urna } i}.$$

Se desconoce el número k , y que urnas son seleccionadas en cada tiempo $t = 1, 2, \dots$. El proceso de selección de las urnas se mantiene oculto. Lo único que observamos es una sucesión de variables aleatorias O_1, O_2, O_3, \dots con distribución condicionado al estado X_t dado por

$$O_t|X_t = i \sim N(\mu_i, \sigma_i^2).$$

El problema es estimar el número k , y la matriz de transición $P = (p_{ij})$ conociendo solo los valores O_t .

Ejemplo 2. Suponga que el proceso de Markov tiene tres estados, $X_{t=1,2,3}$, y la probabilidad de transición entre estos tres estados es $p_{ij} = P(X_{t+1} = j|X_t = i)$, en consecuencia la matriz de

transición es

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

Si el proceso de Markov está en el estado i , esto es $X_t = i$, el segundo proceso estocástico ocurre de acuerdo a la función normal, $O_t | X_t = i \sim N(3i, 1)$.

Esto es,

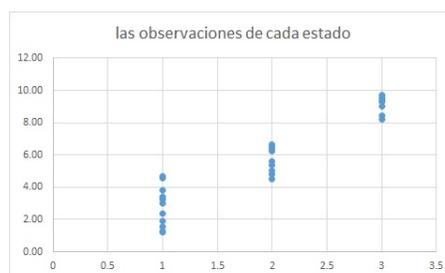
- Si $X_t = 1$ entonces $O_t \sim N(3, 1)$
- Si $X_t = 2$ entonces $O_t \sim N(6, 1)$
- Si $X_t = 3$ entonces $O_t \sim N(9, 1)$

Veamos una posible realización de este proceso.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X_t	2	1	1	3	1	2	3	1	3	2	2	1	1	3
O_t	5.21	2.26	2.64	11.67	2.65	3.72	7.69	3.10	8.59	5.13	7.55	4.50	1.83	8.45

t	15	16	17	18	19	20	21	22	23	24	25	26	27	28
X_t	3	1	2	3	2	1	1	2	1	2	3	1	2	3
O_t	8.75	2.38	6.64	7.82	4.46	1.72	4.86	6.47	1.78	6.39	7.98	4.52	6.50	8.90

Si agrupamos las observaciones de acuerdo a cada estado, se pueden ver las tres distribuciones.



2.1. Tres Problemas Básicos de los HMMs

En esta sección vamos a analizar los tres problemas básicos de los modelos ocultos de Markov:

- Problema de evaluación,
- Problema de decodificación y
- Problema de aprendizaje.

2.1.1. Problema de Evaluación y el Algoritmo Forward-Backward

El problema se describe así: Dado el modelo $\lambda = (\mathbf{p}, P, B)$ y una sucesión de observables $O = O_1, O_2, \dots, O_T$, se debe calcular eficientemente $P(O | \lambda)$. Esta probabilidad puede ser calculada por el uso de argumentos probabilísticos simples. Existe un método de cálculo con complejidad considerablemente baja que utiliza la variable auxiliar

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, X_t = i | \lambda), \quad (2.9)$$

$\alpha_t(i)$ se llama **variable forward**, y O_1, O_2, \dots, O_T es la sucesión parcial de observables.

Proposición 2.1. Dada $\alpha_t(t)$ la variable forward ya definida, se satisface la relación recursiva

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \sum_{i=1}^m \alpha_t(i) p_{ij}, \quad 1 \leq j \leq m, \quad 1 \leq t \leq T-1 \quad (2.10)$$

con $\alpha_1(j) = p_j b_j(O_1)$, $1 \leq j \leq m$.

Demostración. Tenemos que

$$\begin{aligned} \alpha_{t+1}(j) &= P(O_1, O_2, \dots, O_t, O_{t+1}, X_{t+1} = j | \lambda) \\ &= P(O_{t+1} | O_1, O_2, \dots, O_t, X_{t+1} = j, \lambda) P(O_1, O_2, \dots, O_t, X_{t+1} = j, \lambda), \end{aligned}$$

por la propiedad de Markov y por la ley de probabilidad total se sigue que

$$\alpha_{t+1}(j) = P(O_{t+1} | X_{t+1} = j, \lambda) \sum_{i=1}^m P(O_1, O_2, \dots, O_t, X_t = i, X_{t+1} = j, \lambda),$$

luego

$$\begin{aligned}\alpha_{t+1}(j) &= b_j(O_{t+1}) \sum_{i=1}^m P(X_{t+1} = j \mid O_1, O_2, \dots, O_t, X_t = i, \lambda) P(O_1, O_2, \dots, O_t, X_t = i \mid \lambda) \\ &= b_j(O_{t+1}) \sum_{i=1}^m P(X_{t+1} = j \mid X_t = i, \lambda) P(O_1, O_2, \dots, O_t, X_t = i \mid \lambda),\end{aligned}$$

finalmente, por la ecuación (2.9)

$$\alpha_{t+1}(j) = b_j(O_{t+1}) \sum_{i=1}^m \alpha_t(i) p_{ij}.$$

□

La variable forward en el tiempo final T , $\alpha_T(i)$, $1 \leq i \leq m$ se puede calcular utilizando esta fórmula de recursión.

Proposición 2.2. *Sea $\alpha_T(i)$ la variable forward en el tiempo final T , entonces $P(O \mid \lambda)$ es igual a:*

$$P(O \mid \lambda) = \sum_{i=1}^m \alpha_T(i). \quad (2.11)$$

Demostración.

$$\begin{aligned}P(O \mid \lambda) &= P(O_1, O_2, \dots, O_T \mid \lambda) \\ &= \sum_{i=1}^m P(O_1, O_2, \dots, O_T, X_T = i \mid \lambda) \\ &= \sum_{i=1}^m \alpha_T(i).\end{aligned}$$

□

Este método se conoce comúnmente como algoritmo forward. De igual manera, se puede calcular la probabilidad $P(O \mid \lambda)$ usando una variable backward $\beta_t(i)$ definida de manera similar.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T \mid X_t = i, \lambda). \quad (2.12)$$

$\beta_t(i)$ es la probabilidad de la sucesión parcial de observables $O_{t+1}, O_{t+2}, \dots, O_T$ dado que el estado actual es i . $\beta_t(i)$ también se puede calcular eficientemente mediante el uso de la recursividad.

Proposición 2.3. Sea $\beta_t(i)$ la variable backward, antes definida, entonces se tiene que

$$\beta_t(i) = \sum_{j=1}^m \beta_{t+1}(j) p_{ij} b_j(O_{t+1}), \quad 1 \leq i \leq m, \quad 1 \leq t \leq T-1 \quad (2.13)$$

donde $\beta_T(i) = 1, 1 \leq i \leq m$.

Demostración. Por definición (ecuación (2.12)), tenemos

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T | X_t = i, \lambda),$$

y por la ley de probabilidad total se tiene

$$\begin{aligned} \beta_t(i) &= \sum_{j=1}^m P(O_{t+1}, O_{t+2}, \dots, O_T, X_{t+1} = j | X_t = i, \lambda) \\ &= \sum_{j=1}^m P(O_{t+1} | O_{t+2}, \dots, O_T, X_{t+1} = j, X_t = i, \lambda) \\ &\quad P(O_{t+2}, \dots, O_T, X_{t+1} = j | X_t = i, \lambda), \end{aligned}$$

la observación O_{t+1} depende solo del estado X_{t+1} y es independiente de los estados anteriores a X_{t+1} y de los observables; y nuevamente por la probabilidad condicional se sigue que

$$\begin{aligned} \beta_t(i) &= \sum_{j=1}^m P(O_{t+1} | X_{t+1} = j, \lambda) P(O_{t+2}, \dots, O_T | X_{t+1} = j, X_t = i, \lambda) \\ &\quad P(X_{t+1} = j | X_t = i, \lambda) \\ &= \sum_{j=1}^m b_j(O_{t+1}) \beta_{t+1}(j) p_{ij}. \end{aligned}$$

□

Proposición 2.4. Dada $\alpha_t(i)$ y $\beta_t(i)$ antes definidas, entonces se tiene que

$$\alpha_t(i) \beta_t(i) = P(O, X_t = i | \lambda), \quad 1 \leq i \leq m, \quad 1 \leq t \leq T. \quad (2.14)$$

Demostración. Haciendo uso de la probabilidad condicional tenemos

$$\begin{aligned} P(O, X_t = i | \lambda) &= P(O_1, O_2, \dots, O_t, O_{t+1}, \dots, O_T, X_t = i | \lambda) \\ &= P(O_{t+1}, \dots, O_T | O_1, \dots, O_t, X_t = i, \lambda) P(O_1, \dots, O_t, X_t = i | \lambda) \\ &= P(O_{t+1}, \dots, O_T | X_t = i, \lambda) \alpha_t(i) \\ &= \alpha_t(i) \beta_t(i). \end{aligned}$$

□

Así que hay dos maneras de calcular $P(O | \lambda)$, ya sea utilizando la variable forward o la variable backward,

$$P(O | \lambda) = \sum_{i=1}^m P(O, X_t = i | \lambda) = \sum_{i=1}^m \alpha_t(i) \beta_t(i). \quad (2.15)$$

La ecuación (2.15) puede ser muy útil, especialmente en derivar las fórmulas para la formación basada en el gradiente.

2.1.2. El Problema de Decodificación y el Algoritmo de Viterbi

Este problema se describe así: se debe encontrar la sucesión de estados X_1, \dots, X_T más probable dada la sucesión de observables $O = O_1, O_2, \dots, O_T$ y el modelo $\lambda = (\mathbf{p}, P, B)$.

Primero hay que definir el significado de “sucesión de estados más probable”. En un primer enfoque, se quiere encontrar el estado más probable X_t y luego se concatena con las subsecuentes X_{t+1} más probable, pero, debido a que este enfoque a veces no da como resultado una sucesión de estados significativa, se utiliza otro método, conocido comúnmente como el algoritmo de Viterbi. Usando el algoritmo de Viterbi, toda la sucesión de estados con máxima probabilidad es encontrada.

Este algoritmo utiliza el término $\delta_t(i)$ que se define como una variable auxiliar tal que proporciona la más alta probabilidad de la sucesión parcial de observables y de estados hasta el tiempo t , dado que el estado X_t es igual a i ; esto es

$$\delta_t(i) = \max_{X_1, X_2, \dots, X_{t-1}} P(X_1, X_2, \dots, X_{t-1}, X_t = i, O_1, O_2, \dots, O_t | \lambda). \quad (2.16)$$

Proposición 2.5. Sea $\delta_t(i)$, definido anteriormente, entonces

$$\delta_{t+1}(j) = b_j(O_{t+1}) \left[\max_{1 \leq i \leq m} \delta_t(i) p_{ij} \right], \quad 1 \leq i \leq m, 1 \leq t \leq T - 1 \quad (2.17)$$

con $\delta_1(j) = p_j b_j(O_1)$, $1 \leq j \leq m$.

Demostración. Primero analizemos el cálculo de $\delta_t(i) p_{ij}$.

$$\delta_t(i) p_{ij} = \left[\max_{X_1, \dots, X_{t-1}} P(X_1, \dots, X_{t-1}, X_t = i, O_1, \dots, O_t | \lambda) \right] P(X_{t+1} = j | X_t = i, \lambda),$$

como la maximización no incluye los estados de los momentos t y $t + 1$, tenemos

$$\delta_t(i)p_{ij} = \max_{X_1, \dots, X_{t-1}} [P(X_1, \dots, X_{t-1}, X_t = i, O_1, \dots, O_t | \lambda)P(X_{t+1} = j | X_t = i, \lambda)].$$

Por otro lado, por la definición de probabilidad condicional, se tiene para

$$\begin{aligned} P(X_1, \dots, X_{t-1}, X_t = i, O_1, \dots, O_t | \lambda) &= \frac{P(X_1, \dots, X_{t-1}, X_t = i, O_1, \dots, O_t, \lambda)}{P(\lambda)} \\ &= \frac{P(X_1, \dots, X_{t-1}, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i, \lambda)}{P(\lambda)} \\ &= P(X_1, \dots, X_{t-1}, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i | \lambda). \end{aligned}$$

Entonces

$$\begin{aligned} &P(X_1, \dots, X_{t-1}, X_t = i, O_1, \dots, O_t | \lambda)P(X_{t+1} = j | X_t = i, \lambda) \\ &= P(X_1, \dots, X_{t-1}, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i | \lambda)P(X_{t+1} = j | X_t = i, \lambda), \end{aligned}$$

como X_{t+1} solamente depende de X_t y del modelo, pero no de los estados anteriores y menos de las observaciones, se tiene que

$$\begin{aligned} &P(X_1, \dots, X_{t-1}, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i | \lambda)P(X_{t+1} = j | X_t = i, \lambda) \\ &= P(X_1, \dots, X_{t-1}, X_{t+1} = j, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i | \lambda). \end{aligned}$$

Por la definición de probabilidad condicional, tenemos

$$\begin{aligned} &P(X_1, \dots, X_{t-1}, X_{t+1} = j, O_1, \dots, O_t | X_t = i, \lambda)P(X_t = i | \lambda) \\ &= \frac{P(X_1, \dots, X_{t-1}, X_{t+1} = j, O_1, \dots, O_t, X_t = i, \lambda)P(X_t = i | \lambda)}{P(X_t = i, \lambda)} \\ &= \frac{P(X_1, \dots, X_{t-1}, X_{t+1} = j, O_1, \dots, O_t, X_t = i | \lambda)P(\lambda)P(X_t = i | \lambda)}{P(X_t = i, \lambda)} \\ &= P(X_1, \dots, X_{t-1}, X_t = i, X_{t+1} = j, O_1, \dots, O_t | \lambda). \end{aligned}$$

Por lo tanto,

$$\delta_t(i)p_{ij} = \max_{X_1, \dots, X_{t-1}} P(X_1, \dots, X_{t-1}, X_t = i, X_{t+1} = j, O_1, \dots, O_t | \lambda).$$

Ahora, como se debe tomar $\max_{1 \leq i \leq m} \delta_t(i)p_{ij}$ 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:

$$\max_{1 \leq i \leq m} \delta_t(i)p_{ij} = \max_{X_1, \dots, X_{t-1}, X_t} P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t | \lambda).$$

Finalmente como

$$b_j(O_{t+1}) = P(O_{t+1} | X_{t+1} = j, \lambda),$$

Y usando el hecho de que la maximización no incluye el estado del momento $t + 1$, se tiene que

$$\begin{aligned} b_j(O_{t+1}) & \left[\max_{1 \leq i \leq m} \delta_t(i) p_{ij} \right] \\ & = \max_{X_1, \dots, X_{t-1}, X_t} [P(O_{t+1} | X_{t+1} = j, \lambda) P(X_1, \dots, X_t, X_{t+1} = j, O_1, \dots, O_t | \lambda)], \end{aligned}$$

pero por la definición de probabilidad condicional, se tiene que para

$$\begin{aligned} & P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t | \lambda) \\ & = \frac{P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t, \lambda)}{P(\lambda)} \\ & = \frac{P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t | X_{t+1} = j, \lambda) P(X_{t+1} = j, \lambda)}{P(\lambda)} \\ & = P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t | X_{t+1} = j, \lambda) P(X_{t+1} = j | \lambda). \end{aligned}$$

Luego, ya que O_{t+1} depende solo de X_{t+1} y del modelo

$$\begin{aligned} & P(O_{t+1} | X_{t+1} = j, \lambda) P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t | \lambda) \\ & = P(O_{t+1} | X_{t+1} = j, \lambda) P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t | X_{t+1} = j, \lambda) P(X_{t+1} = j | \lambda) \end{aligned}$$

nuevamente por la definición de probabilidad condicional

$$\begin{aligned} & P(O_{t+1} | X_{t+1} = j, \lambda) P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t | X_{t+1} = j, \lambda) P(X_{t+1} = j | \lambda) \\ & = P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t, O_{t+1} | X_{t+1} = j, \lambda) P(X_{t+1} = j | \lambda) \\ & = \frac{P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t, O_{t+1}, X_{t+1} = j, \lambda)}{P(X_{t+1} = j, \lambda)} P(X_{t+1} = j | \lambda) \\ & = \frac{P(X_1, \dots, X_{t-1}, X_t, O_1, \dots, O_t, O_{t+1}, X_{t+1} = j | \lambda) P(\lambda)}{P(X_{t+1} = j, \lambda)} P(X_{t+1} = j | \lambda) \\ & = P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t, O_{t+1} | \lambda). \end{aligned}$$

Por lo tanto

$$\begin{aligned} b_j(O_{t+1}) \left[\max_{1 \leq i \leq m} \delta_t(i) p_{ij} \right] & = \max_{X_1, \dots, X_{t-1}, X_t} P(X_1, \dots, X_{t-1}, X_t, X_{t+1} = j, O_1, \dots, O_t, O_{t+1} | \lambda) \\ & = \delta_{t+1}(j). \end{aligned}$$

□

Así que, la sucesión de estados que maximiza la probabilidad $P(X_1, \dots, X_T, O_1, \dots, O_T \mid \lambda)$ es aquella que maximiza $\delta_T(j)$, $1 \leq j \leq m$.

Por lo que, se satisface la siguiente relación

$$\max_{X_1, X_2, \dots, X_T} P(X_1, X_2, \dots, X_T, O_1, \dots, O_T \mid \lambda) = \max_{1 \leq j \leq m} \delta_T(j).$$

Puesto que el objetivo es identificar la sucesión de estados que maximiza la probabilidad, se debe almacenar para cada tiempo t y estado i , los argumentos de la sucesión X_1, X_2, \dots, X_T que dan el valor $\delta_t(i)$, con este objetivo se define la variable $\varphi_{t+1}(j) = \arg\{\max_{1 \leq i \leq m} \delta_t(i) p_{ij}\}$.

Siempre se tiene un apuntador al estado ganador en el proceso de obtención del máximo. Este resulta en el estado j^* , donde $j^* = \arg\{\max_{1 \leq j \leq m} \delta_T(j)\}$. Partimos de este estado y rastreamos la sucesión de estados como el puntero que indica cada estado.

Este algoritmo es muy usado para reconocimiento de voz, biología molecular, fonemas, palabras, codificadores, entre otros. A cada secuencia de estados le corresponde una secuencia de etiquetas de clasificación, es decir, palabras, caracteres, fonemas, sufijos.

2.1.3. El Problema de Aprendizaje

En este problema nos hacemos la siguiente pregunta, ¿Cómo podemos ajustar los parámetros del HMM de manera que un determinado conjunto de observaciones esté representado por el modelo de la mejor manera para el uso previsto?

Dependiendo de la aplicación, la cantidad que debe optimizarse durante el proceso de aprendizaje difiere. Por lo que hay varios criterios de optimización para el aprendizaje. En la literatura, podemos encontrar dos principales criterios de optimización: Máxima Verosimilitud (ML por sus siglas en inglés) y Máxima Información Mutua (MMI por sus siglas en inglés).

Describiremos el criterio de máxima verosimilitud ya que es el que se ocupará en esta tesis.

Criterio de Máxima Verosimilitud

Dado el HMM λ_w de clase w , tratamos de maximizar la probabilidad de una determinada sucesión de observaciones O^w , la pertenencia a una determinada clase w , que corresponden a los

parámetros del modelo λ_w . Matemáticamente, esta probabilidad se puede expresar como

$$L_{tot} = P(O^w | \lambda_w). \quad (2.18)$$

Quitaremos el subíndice y superíndice w porque consideramos una sola clase w a la vez, la ML puede darse como

$$L_{tot} = P(O | \lambda). \quad (2.19)$$

El modelo $\lambda = (\mathbf{p}, P, B)$ que maximiza L_{tot} no puede ser resuelto analíticamente. Usando un procedimiento iterativo, como el de Baum-Welch o un método basado en el gradiente, podemos maximizarlo localmente eligiendo los parámetros apropiados del modelo.

Algoritmo de Baum-Welch

El algoritmo de Baum-Welch es también conocido como algoritmo Forward-Backward. Este método se puede derivar como se conoce en la literatura mediante el uso simple de argumentos de conteo de ocurrencia o utilizando el cálculo para maximizar la cantidad auxiliar

$$Q(\lambda, \bar{\lambda}) = \sum_X P(X | O, \lambda) \log[P(O, X, \bar{\lambda})] \quad (2.20)$$

sobre $\bar{\lambda}$.

Adicionalmente a las variables forward y backward, tenemos que definir dos variables auxiliares más. La primera de estas variables es

$$\xi_t(i, j) = P(X_t = i, X_{t+1} = j | O, \lambda), \quad (2.21)$$

la cual puede escribirse como

$$\xi_t(i, j) = \frac{P(X_t = i, X_{t+1} = j, O | \lambda)}{P(O | \lambda)}. \quad (2.22)$$

Proposición 2.6. Sea $\xi_t(i, j)$ definida anteriormente, entonces se satisface la ecuación

$$\xi_t(i, j) = \frac{\alpha_t(i) p_{ij} \beta_{t+1}(j) b_j(O_{t+1})}{\sum_{i=1}^m \sum_{j=1}^m \alpha_t(i) p_{ij} \beta_{t+1}(j) b_j(O_{t+1})}. \quad (2.23)$$

Demostración. Tenemos de la ecuación (2.22)

$$\xi_t(i, j) = \frac{P(X_t = i, X_{t+1} = j, O_1, \dots, O_T | \lambda)}{P(O_1, \dots, O_T | \lambda)}. \quad (2.24)$$

Trabajemos primero con el numerador de la ecuación (2.24), y apliquemos la propiedad de Markov

$$\begin{aligned} & P(X_t = i, X_{t+1} = j, O_1, \dots, O_T | \lambda) \\ &= P(O_{t+1} | X_t = i, X_{t+1} = j, O_1, \dots, O_t, O_{t+2}, \dots, O_T, \lambda) \\ &\quad P(X_t = i, X_{t+1} = j, O_1, \dots, O_t, O_{t+2}, \dots, O_T | \lambda) \\ &= P(O_{t+1} | X_{t+1} = j)P(X_t = i, X_{t+1} = j, O_1, \dots, O_t, O_{t+2}, \dots, O_T | \lambda) \\ &= b_j(O_{t+1})P(O_{t+2}, \dots, O_T | X_t = i, X_{t+1} = j, O_1, \dots, O_t, \lambda) \\ &\quad P(X_t = i, X_{t+1} = j, O_1, \dots, O_t | \lambda) \\ &= b_j(O_{t+1})P(O_{t+2}, \dots, O_T | X_{t+1} = j, \lambda)P(X_{t+1} = j | X_t = i, O_1, \dots, O_t, \lambda) \\ &\quad P(X_t = i, O_1, \dots, O_t | \lambda) \\ &= b_j(O_{t+1})\beta_{t+1}(j)P(X_{t+1} = j | X_t = i)P(O_1, \dots, O_t, X_t = i | \lambda) \\ &= b_j(O_{t+1})\beta_{t+1}(j)p_{ij}\alpha_t(i). \end{aligned} \quad (2.25)$$

Ahora trabajemos con el denominador de la ecuación (2.24), por la ley de la probabilidad total tenemos

$$P(O_1, \dots, O_T | \lambda) = \sum_{i=1}^m \sum_{j=1}^m P(O_1, \dots, O_T, X_t = i, X_{t+1} = j | \lambda),$$

por la ecuación (2.25) se sigue

$$P(O_1, \dots, O_T | \lambda) = \sum_{i=1}^m \sum_{j=1}^m \alpha_t(i)p_{ij}\beta_{t+1}(j)b_j(O_{t+1}). \quad (2.26)$$

Sustituyendo la ecuación (2.25) y (2.26) en la ecuación (2.24) tenemos

$$\xi_t(i, j) = \frac{\alpha_t(i)p_{ij}\beta_{t+1}(j)b_j(O_{t+1})}{\sum_{i=1}^m \sum_{j=1}^m \alpha_t(i)p_{ij}\beta_{t+1}(j)b_j(O_{t+1})}.$$

□

La segunda variable es la probabilidad a posteriori

$$\gamma_t(i) = P(X_t = i | O, \lambda). \quad (2.27)$$

Proposición 2.7. Sea $\gamma_t(i)$ definida anteriormente, entonces

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^m \alpha_t(i)\beta_t(i)}. \quad (2.28)$$

Demostración. De la ecuación (2.27) tenemos

$$\gamma_t(i) = \frac{P(X_t = i, O | \lambda)}{P(O | \lambda)}. \quad (2.29)$$

Observe el numerador de la ecuación (2.29)

$$\begin{aligned} P(X_t = i, O | \lambda) &= P(X_t = i, O_1, \dots, O_t, O_{t+1}, \dots, O_T | \lambda) \\ &= P(O_{t+1}, \dots, O_T | X_t = i, O_1, \dots, O_t, \lambda)P(X_t = i, O_1, \dots, O_t | \lambda) \\ &= P(O_{t+1}, \dots, O_T | X_t = i, \lambda)P(O_1, \dots, O_t, X_t = i | \lambda) \\ &= \beta_t(i)\alpha_t(i). \end{aligned} \quad (2.30)$$

Ahora, trabajando con el denominador de la ecuación (2.29), y aplicando la ley de probabilidad total y la probabilidad condicional tenemos

$$\begin{aligned} P(O | \lambda) &= P(O_1, \dots, O_T | \lambda) \\ &= \sum_{i=1}^m P(O_1, \dots, O_T, X_t = i | \lambda) \\ &= \sum_{i=1}^m P(O_{t+1}, \dots, O_T | X_t = i, O_1, \dots, O_t, \lambda)P(X_t = i, O_1, \dots, O_t | \lambda) \\ &= \sum_{i=1}^m P(O_{t+1}, \dots, O_T | X_t = i, \lambda)P(O_1, \dots, O_t, X_t = i | \lambda) \\ &= \sum_{i=1}^m \alpha_t(i)\beta_t(i). \end{aligned} \quad (2.31)$$

Sustituyendo la ecuación (2.30) y (2.31) en la ecuación (2.29) se tiene

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^m \alpha_t(i)\beta_t(i)}.$$

□

Proposición 2.8. Dadas las variables $\gamma_t(i)$ y $\xi_t(i, j)$ antes definidas, se satisface la relación

$$\gamma_t(i) = \sum_{j=1}^m \xi_t(i, j), \quad 1 \leq i \leq m, 1 \leq t \leq T. \quad (2.32)$$

Demostración. En efecto,

$$\xi_t(i, j) = P(X_t = i, X_{t+1} = j \mid O, \lambda)$$

y

$$\begin{aligned} \sum_{j=1}^m \xi_t(i, j) &= \sum_{j=1}^m P(X_t = i, X_{t+1} = j \mid O, \lambda) \\ &= P(X_t = i \mid O, \lambda) \\ &= \gamma_t(i). \end{aligned}$$

□

Para maximizar $P(O \mid \lambda)$, ahora podemos describir el proceso de aprendizaje de Baum-Welch. El proceso de aprendizaje de Baum-Welch supone que existe un modelo inicial $\lambda = (\mathbf{p}, P, B)$ y para este modelo se calculan los términos $\alpha_t(i)$ y $\beta_t(i)$. Luego se calculan los términos $\xi_t(i, j)$ y $\gamma_t(i)$. Enseguida se calculan las siguientes ecuaciones que se conocen como fórmulas de re-estimación y se utilizan para actualizar los parámetros del HMM:

$$\bar{\mathbf{p}}_i = \gamma_1(i), \quad 1 \leq i \leq m, \quad (2.33)$$

$$\bar{p}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad 1 \leq i \leq m, 1 \leq j \leq m, \quad (2.34)$$

$$b_j(\bar{k}) = \frac{\sum_{t=1, o_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}, \quad 1 \leq j \leq m, 1 \leq k \leq n. \quad (2.35)$$

Como existen parámetros desconocidos en el HMM, para estimarlos se usa el algoritmo EM, que se estudia en la siguiente sección.

Algoritmo EM

El algoritmo EM es un método iterativo para realizar la estimación de máxima verosimilitud cuando faltan algunos de los datos, y aprovecha el hecho de que los datos completos de la log-verosimilitud puede ser sencilla para maximizar aún cuando la probabilidad de que los datos observados no lo es. Por “datos completos de la log-verosimilitud”(CDLL) nos referimos a la

log-verosimilitud de los parámetros de interés θ , basado tanto en los datos observados y los datos faltantes.

El algoritmo puede ser descrito informalmente como sigue.

Elegir valores iniciales para los parámetros θ que desea calcular. A continuación, se repiten los siguientes pasos:

- **Paso E.** Calcular las esperanzas condicionales de los datos que faltan, dadas las observaciones y dadas las estimaciones actuales de θ . Más precisamente, calcular las esperanzas condicionales de las funciones de los datos que faltan que aparecen en los datos completos de log-verosimilitud.
- **Paso M.** Maximizar, con respecto a θ , los datos completos de log-verosimilitud con las funciones de los datos que faltan reemplazados en el mismo por parte de sus esperanzas condicionales.

Estos dos pasos son repetidos hasta que se satisface algún criterio de convergencia, por ejemplo, hasta que el cambio resultante en θ es menor que algún umbral. El valor resultante de θ es entonces un punto estacionario de la probabilidad de los datos observados. En algunos casos, el punto estacionario alcanzado puede ser un máximo local (a diferencia del global) o un punto silla.

Capítulo 3

Modelo de Black-Scholes

En 1973 Fisher Black, Myron Scholes y Robert Merton lograron uno de los mayores avances en la valuación de opciones hasta ese momento, conocido como el modelo de Black-Scholes, que ha tenido una gran influencia en la manera en que los agentes valúan y cubren opciones. Ha sido también un punto de referencia para el desarrollo y éxito de la ingeniería financiera desde entonces.

En este capítulo desarrollaremos el modelo de Black-Scholes. Empezaremos dando los conceptos necesarios para el desarrollo del mismo.

Definición 3.1. *Un activo es cualquier posesión que pueda producir beneficios económicos.*

Un portafolio es un conjunto de activos, que pueden ser acciones, derivados, bonos, etc.

En la realidad existen costos para realizar operaciones financieras. Estos **costos de transacción** pueden depender de una transacción de un activo subyacente o un derivado, de una compra o una venta, etc.

Definición 3.2. *La tasa de interés libre de riesgo es aquella generada por una inversión “segura”.*

Los primeros instrumentos derivados básicos se denominan opciones vanilla y se definen como:

Definición 3.3. *Una opción call (respectivamente opción put) es un contrato que da derecho a comprar(respectivamente vender) un activo financiero, llamado activo subyacente, a un precio*

*fijo, llamado **precio de ejercicio**, en el final del tiempo de contrato, llamado **tiempo de madurez** ó **tiempo de ejercicio**, también previsto en el contrato.*

Si una opción se puede ejercer antes del tiempo de madurez se dice que es de tipo **Americano**; en caso que la opción solo se pueda ejercer en el tiempo de madurez es una opción de tipo **Europeo**.

Usaremos la siguiente notación: K para el precio de ejercicio, T para el tiempo de madurez y S para el valor del activo subyacente a madurar.

Nos enfocaremos en una opción de compra Europea y vamos a representar por C la prima de la opción call, la cual es la estimación del costo de la opción.

Vamos a cubrir el problema principal para las opciones vanilla, es decir, la fijación de precios de estos productos opcionales. Dentro de un marco de teoría económica-financiera se tiene que dar el valor de la prima C en función de la madurez T y el valor S del activo en el momento 0.

En terminos más generales, como el titular de una opción puede vender su opción en el mercado de opciones en cualquier momento t , $0 < t < T$, también es necesario darle el valor justo de la opción en el momento t sabiendo que el activo subyacente tiene, en este momento, el valor de $S = S(t)$, y el valor justo de mercado está representada por

$$C(S, \tau) \tag{3.1}$$

donde

$$\tau = T - t \tag{3.2}$$

representa la madurez calculada en el tiempo t .

A veces, también es útil representar el valor call como una función del tiempo $C(S, t)$.

Para discutir este problema de precios, es absolutamente necesario dar hipótesis sobre el proceso estocástico

$$S = (S(t), 0 \leq t \leq T). \tag{3.3}$$

En cuanto al marco de la teoría económico-financiero, se adopta el supuesto de mercado eficiente.

Definición 3.4. *Un mercado eficiente es cuando toda la información disponible en el momento t está al alcance de los inversores y se refleja en los valores de los activos.*

De acuerdo a esta definición, en un mercado eficiente las transacciones que tienen una alta rentabilidad anormal no son posibles y se satisfacen los siguientes supuestos:

1. Ausencia de costos de transacción;
2. posibilidad de ventas en corto;
3. disponibilidad de toda la información a todos los agentes económicos;
4. perfecta divisibilidad de los activos;
5. mercado financiero en tiempo continuo.

Definición 3.5. *Un bono cupón cero es un título emitido por una entidad que no paga intereses durante su vida, sino que lo hace íntegramente en el momento en el que se amortiza el capital. En compensación, su precio, muy sensible al tipo de interés, es inferior a su valor nominal.*

Además, el mercado se ha **completado**, lo que significa que existe bonos cupón cero y sin riesgo para todos los vencimientos posibles.

Tengamos en cuenta que la palabra “información” que se usa en el punto 3 puede tener diferentes interpretaciones: débil, semi-fuerte o fuerte, dependiendo si se basa en los precios del pasado, en toda la información pública, o finalmente, en toda la información posible que el agente pueda encontrar.

Según Fama (1965), las suposiciones eficientes justifican el modelo de “caminata aleatoria” en tiempo discreto, diciendo que si $R_i(s)$ representa el incremento de un activo i entre s y $s + 1$, tenemos

$$\Delta R_i(s) = \mu_i + \varepsilon_i(s), \quad (3.4)$$

siendo μ_i una constante y $\varepsilon_i(s)$ una sucesión de v.a. no correlacionadas de media 0, a veces llamadas **errores**.

Aunque la hipótesis de la eficiencia parece ser natural, algunos estudios empíricos muestran que este no es siempre el caso, en particular, ya que algunos agentes pueden tener acceso a información privilegiada, en principio, prohibido por el control de autoridad legal.

No obstante, estos agentes deben usar la información correspondiente, se notará rápidamente esos mercados y se restaurará el equilibrio entre los agentes.

Esta posibilidad, también llamada caso de la *información asimétrica*, fue estudiada por Spencer, Akerlof y Stiglitz, quienes se les concedió el Premio Nobel de Economía en 2001.

Los modelos de información asimétrica siempre deben ser modelados a corto plazo que rechazan la hipótesis de **ausencia de arbitraje (AOA)**, es decir, ganar dinero sin ningún tipo de inversión.

Para completar, notemos que ahora es posible construir modelos sin el supuesto AOA pero haciendo suposiciones sobre la evolución de los activos de tiempo y una elección de posibles escenarios, por lo que el inversor puede predecir lo que ocurrirá si se producen estos escenarios.

Definición 3.6. *Se llama **función de pago** de un derivado, activo o portafolio al resultado final de la inversión.*

La función de pago de una opción call es $\max\{S_T - K, 0\}$ ya que si $S_T > K$ se ejerce a K y se vende a S_T , lo que da una ganancia de $S_T - K$. En el otro caso la opción no se ejerce y la función de pago es cero.

Otro concepto de suma importancia en finanzas es el concepto de *volatilidad*.

Definición 3.7. *Se llama **volatilidad** a una medida de la frecuencia e intensidad de los cambios del precio de un activo o de un tipo. Se puede medir de muchas maneras pero, a efectos de valuar una opción o instrumento que contenga una opción incorporada, la volatilidad se mide normalmente como desviación típica del porcentaje anual de cambio cuando ese porcentaje cambia continuamente. La volatilidad se puede estimar a partir de datos históricos para conseguir una volatilidad histórica o se puede extraer del precio de una opción con el uso de un modelo apropiado para la determinación del precio de opciones para obtener una volatilidad implícita.*

3.1. Ecuación de Black-Scholes

Black y Scholes utilizaron un modelo de tiempo continuo para el activo subyacente introducido por Samuelson (1965). En esta sección deduciremos la ecuación de Black-Scholes.

Consideremos un espacio completo de probabilidad filtrado, el proceso estocástico

$$S = (S(t), t \geq 0) \quad (3.5)$$

ahora representará la evolución temporal del activo subyacente.

La suposición básica es que la dinámica estocástica del proceso S está dada por

$$\begin{aligned} dS &= \mu S dt + \sigma S dW, \\ S(0) &= S_0, \end{aligned}$$

donde el proceso $W = (W(t), t \in [0, T])$ es un proceso Browniano estándar y dW representa la componente de incertidumbre de la variación instantánea del precio. Este modelo representa el precio de un activo con “riesgo”. Cabe recalcar que el precio tanto de un derivado como del subyacente tiene este término estocástico de riesgo o incertidumbre. Nos proponemos eliminar la componente estocástica determinada por dW .

Consideremos la función $V(S, t)$ que representa el valor de una opción a través del tiempo t con precio inicial $S_0 > 0$, para los efectos de la deducción no importa si esta opción es call o put. En este caso se construye un portafolio de la siguiente manera, lo que se denomina una estrategia de cobertura:

$$\pi = (-\Delta, 1),$$

donde $-\Delta$ son las unidades del activo subyacente y 1 es el derivado.

El valor del portafolio lo determinamos como un producto interno de \mathbb{R}^2 , de esta manera:

$$\Pi = \pi \cdot (S, V) = (-\Delta, 1)(S, V) = -\Delta S + V.$$

Luego, tomando diferencias instantáneas de tiempo en ambos lados de la ecuación anterior tenemos:

$$d\Pi = -\Delta dS + dV. \quad (3.6)$$

Aplicando el Lema de Itô (ecuación (1.12)), tenemos

$$dV(S, t) = \frac{\partial V}{\partial S}(S_0, t_0)dS + \frac{\partial V}{\partial t}(S_0, t_0)dt + \frac{1}{2} \frac{\partial^2 V}{\partial S^2}(S_0, t_0)(dS)^2. \quad (3.7)$$

Considerando que $dS = \mu Sdt + \sigma SdW$ y sustituyendo la ecuación (3.7) en la ecuación (3.6) se tiene

$$d\Pi = -\Delta(\mu Sdt + \sigma SdW) + \frac{\partial V}{\partial S}(\mu Sdt + \sigma SdW) + \frac{\partial V}{\partial t}dt + \frac{1}{2} \frac{\partial^2 V}{\partial S^2}(\mu Sdt + \sigma SdW)^2,$$

de esta ecuación y los resultados del lema de Itô, agrupamos los términos que tienen dW y procuramos que se cancelen (para eso está la Δ , que es la variable de la ecuación cuyo valor cumple dicha cancelación), resulta entonces

$$d\Pi = \sigma S \left(\frac{\partial V}{\partial S} - \Delta \right) dW + \left(\mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \mu \Delta S \right) dt. \quad (3.8)$$

Tomando

$$\Delta = \frac{\partial V}{\partial S}.$$

y sustituyendo en la ecuación (3.8) tenemos

$$\begin{aligned} d\Pi &= \sigma S \left(\frac{\partial V}{\partial S} - \frac{\partial V}{\partial S} \right) dW + \left(\mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} - \mu \frac{\partial V}{\partial S} S \right) dt \\ &= \left(\frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} \right) dt. \end{aligned} \quad (3.9)$$

Por la hipótesis de no arbitraje, como P es un portafolio libre de riesgo tenemos que su retorno es igual al de un bono de tasa r cuya fórmula está dada por:

$$\frac{d\Pi}{\Pi} = rdt \Rightarrow d\Pi = \Pi rdt. \quad (3.10)$$

Igualando la ecuación (3.9) y la ecuación (3.10) tenemos

$$\Pi rdt = \left(\frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t} \right) dt.$$

Cancelando dt y sustituyendo $\Pi = -\Delta S + V = -\frac{\partial V}{\partial S} S + V$, nos queda

$$-\frac{\partial V}{\partial S} S r + V r = \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}.$$

Finalmente, despejando rV , llegamos a la ecuación de Black-Scholes:

$$rV = rS \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}. \quad (3.11)$$

3.1.1. Solución de la ecuación de Black-Scholes

Vamos a obtener la solución de la ecuación de Black-Scholes para calcular el valor de la opción Call Europea sobre un activo de precio S con precio de ejercicio K y tiempo de madurez T . En este caso llamaremos $V = C$, la ecuación (3.11) queda entonces como

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0,$$

con las condiciones de frontera

$$C(0, t) = 0, \quad C(S, t) \sim S \text{ si } S \rightarrow \infty$$

ya que cuando el precio del activo es nulo, también debe serlo el de la opción (es claro que no se va a ejercer). Y cuando el precio tiende a infinito, $S - K$ se va a aproximar a S . También recordemos la condición final, es decir la función de pago de la opción

$$C(S, T) = \max\{S - K, 0\}.$$

Con todo lo anterior, podemos describir el modelo de la siguiente manera:

$$\begin{cases} \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0, & S \in (0, \infty), t \in [0, T) \\ C(S, T) = \max\{S - K, 0\}, & S \in (0, \infty) \\ C(0, t) = 0, & t \in [0, T) \\ C(S, t) \sim S, & t \in [0, T), S \rightarrow \infty \end{cases} \quad (3.12)$$

Nos concentraremos en las dos primeras ecuaciones de (3.12), pues posteriormente veremos que las últimas dos, que describen el comportamiento de C en los bordes, también se van a satisfacer. Entonces nuestro modelo queda como sigue:

$$\begin{cases} \frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS \frac{\partial C}{\partial S} - rC = 0, & S \in (0, \infty), t \in [0, T) \\ C(S, T) = \max\{S - K, 0\}, & S \in (0, \infty) \end{cases} \quad (3.13)$$

Para resolver esta ecuación, hagamos primero los siguientes cambios de variables

$$x = \ln \frac{S}{K}, \quad \tau(t) = \frac{1}{2}\sigma^2(T - t), \quad v(x, \tau) = \frac{C(S, t)}{K},$$

es decir

$$S = Ke^x, \quad t = T - \frac{2\tau(t)}{\sigma^2}, \quad C(S, t) = Kv(x, \tau).$$

Calculando las derivadas parciales concernientes a la primer ecuación de (3.13) tenemos

$$\frac{\partial C}{\partial t} = -\frac{1}{2}\sigma^2 K \frac{\partial v}{\partial \tau}, \quad (3.14)$$

$$\frac{\partial C}{\partial S} = \frac{K}{S} \frac{\partial v}{\partial x}, \quad (3.15)$$

$$\frac{\partial^2 C}{\partial S^2} = \frac{K}{S^2} \left(-\frac{\partial v}{\partial x} + \frac{\partial^2 v}{\partial x^2} \right). \quad (3.16)$$

Como $\sigma(T) = 0$, también tenemos una condición inicial para v a partir de la condición final de C :

$$C(S, T) = Kv(x, 0) = \max\{Ke^x - K, 0\} \Rightarrow v(x, 0) = \max\{e^x - 1, 0\}$$

que resulta una condición inicial. Sustituyendo las ecuaciones (3.14-3.16) en el modelo (3.13) tenemos

$$-\frac{1}{2}\sigma^2 K \frac{\partial v}{\partial \tau} + \frac{1}{2}\sigma^2 S^2 \left(\frac{K}{S^2} \left(-\frac{\partial v}{\partial x} + \frac{\partial^2 v}{\partial x^2} \right) \right) + rS \left(\frac{K}{S} \frac{\partial v}{\partial x} \right) - rKv(x, \tau) = 0,$$

lo cual, simplificando nos da

$$\frac{1}{2}\sigma^2 \frac{\partial v}{\partial \tau} = -\frac{1}{2}\sigma^2 \frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 v}{\partial x^2} + r \frac{\partial v}{\partial x} - rv.$$

Luego el modelo (3.13) queda como

$$\begin{cases} \frac{1}{2}\sigma^2 \frac{\partial v}{\partial \tau} = -\frac{1}{2}\sigma^2 \frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 v}{\partial x^2} + r \frac{\partial v}{\partial x} - rv, & x \in \mathbb{R}, \tau \in [0, T\sigma^2/2) \\ v(x, 0) = \max\{e^x - 1, 0\}, & x \in \mathbb{R}, \end{cases}$$

y si llamamos $K = \frac{2r}{\sigma^2}$ el modelo queda como

$$\begin{cases} \frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial x^2} + (K - 1) \frac{\partial v}{\partial x} - Kv, & x \in \mathbb{R}, \tau \in [0, T\sigma^2/2) \\ v(x, 0) = \max\{e^x - 1, 0\}, & x \in \mathbb{R}, \end{cases} \quad (3.17)$$

Hagamos un último cambio de variables. Proponemos

$$v(x, \tau) = e^{ax+b\tau} u(x, \tau)$$

con a y b parámetros a determinar. Las derivadas parciales de v resultan

$$\begin{aligned} \frac{\partial v}{\partial x} &= ae^{ax+b\tau} u + e^{ax+b\tau} \frac{\partial u}{\partial x}, \\ \frac{\partial v}{\partial \tau} &= be^{ax+b\tau} u + e^{ax+b\tau} \frac{\partial u}{\partial \tau}, \\ \frac{\partial^2 v}{\partial x^2} &= a^2 e^{ax+b\tau} u + 2ae^{ax+b\tau} \frac{\partial u}{\partial x} + e^{ax+b\tau} \frac{\partial^2 u}{\partial x^2}, \end{aligned}$$

sustituyéndolas en la primera ecuación de (3.17) y simplificando el término $e^{ax+b\tau}$ tenemos

$$bu + \frac{\partial u}{\partial \tau} = a^2u + 2a\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2} + (K-1)\left(au + \frac{\partial u}{\partial x}\right) - Ku.$$

Como a y b eran arbitrarios elegimos a y b de tal manera que se anulen los coeficientes que multiplican a u y a $\frac{\partial u}{\partial x}$, es decir

$$b = a^2 + (K-1)a - K, \quad y \quad 0 = 2a + (K-1).$$

Al resolver el sistema de ecuaciones, resultan

$$a = -\frac{K-1}{2} \quad y \quad b = -\frac{(K+1)^2}{4}.$$

Con esta elección de parámetros la primera ecuación de (3.17) queda

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2},$$

cuya condición inicial resulta

$$\begin{aligned} u(x, 0) &= v(x, 0)e^{\frac{1}{2}(K-1)x} \\ &= \max\{e^x - 1, 0\}e^{\frac{1}{2}(K-1)x} \\ &= \max\{e^{\frac{1}{2}(K+1)x} - e^{\frac{1}{2}(K-1)x}, 0\}, \end{aligned}$$

con lo cual, tenemos el modelo

$$\begin{cases} \frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}, & x \in \mathbb{R}, \tau \in [0, T\sigma^2/2) \\ u(x, 0) = u_0(x) = \max\{e^{\frac{1}{2}(K+1)x} - e^{\frac{1}{2}(K-1)x}, 0\}, & x \in \mathbb{R}, \end{cases} \quad (3.18)$$

Esta es la ecuación del calor cuya solución se encuentra determinada por

$$u(x, \tau) = \frac{1}{2\sqrt{\pi\tau}} \int_{-\infty}^{\infty} u_0(s) e^{-\frac{(x-s)^2}{4\tau}} ds$$

que es la convolución entre la condición inicial y la solución fundamental de la ecuación del calor. Evaluemos esta integral haciendo el cambio

$$s = w\sqrt{2\tau} + x,$$

con lo que resulta

$$\begin{aligned} u(x, \tau) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_0(w\sqrt{2\tau} + x) e^{-\frac{1}{2}w^2} dw \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \max\{e^{\frac{1}{2}(K+1)(w\sqrt{2\tau}+x)} - e^{\frac{1}{2}(K-1)(w\sqrt{2\tau}+x)}, 0\} e^{-\frac{1}{2}w^2} dw. \end{aligned}$$

Para eliminar el máximo, haremos uso de lo siguiente:

$$\begin{aligned} e^{\frac{1}{2}(K+1)s} - e^{\frac{1}{2}(K-1)s} &\geq 0 \\ \Leftrightarrow e^{\frac{1}{2}(K+1)s} &\geq e^{\frac{1}{2}(K-1)s} \\ \Leftrightarrow \frac{K+1}{2}s &\geq \frac{K-1}{2}s \\ \Leftrightarrow s &\geq 0, \end{aligned}$$

por lo cual el integrando no será nulo cuando $w\sqrt{2\tau} + x \geq 0$, es decir, si

$$w \geq -\frac{x}{\sqrt{2\tau}},$$

por lo que la integral resulta

$$u(x, \tau) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{1}{2}(K+1)(w\sqrt{2\tau}+x)} e^{-\frac{1}{2}w^2} dw - \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{1}{2}(K-1)(w\sqrt{2\tau}+x)} e^{-\frac{1}{2}w^2} dw.$$

Llamaremos A_1 y A_2 a cada una de las integrales anteriores. Haremos los cálculos para la primera pues serán análogos los de la segunda.

Sacando del integrando el término que no depende de w y juntando los exponentes tenemos

$$\begin{aligned} A_1 &= \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{1}{2}(K+1)(w\sqrt{2\tau}+x)} e^{-\frac{1}{2}w^2} dw \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(K+1)x} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{1}{2}(K+1)w\sqrt{2\tau} - \frac{1}{2}w^2} dw. \end{aligned}$$

Completando los cuadrados en el exponente se tiene

$$A_1 = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(K+1)x} \int_{-\frac{x}{\sqrt{2\tau}}}^{\infty} e^{\frac{1}{4}(K+1)^2\tau - \frac{1}{2}(w - \frac{1}{2}(K+1)\sqrt{2\tau})^2} dw.$$

Sacamos el término que no depende de w y sea

$$\rho = w - \frac{1}{2}(K+1)\sqrt{2\tau}$$

con lo cual, haciendo el cambio de variable nos queda

$$A_1 = e^{\frac{1}{2}(K+1)x + \frac{1}{4}(K+1)^2\tau} \frac{1}{\sqrt{2\pi}} \int_{-\frac{x}{\sqrt{2\tau}} - \frac{1}{2}(K+1)\sqrt{2\tau}}^{\infty} e^{-\frac{1}{2}\rho^2} d\rho.$$

Si definimos

$$d_1 = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(K+1)\sqrt{2\tau}, \quad (3.19)$$

la integral resulta

$$\begin{aligned} A_1 &= e^{\frac{1}{2}(K+1)x + \frac{1}{4}(K+1)^2\tau} \frac{1}{\sqrt{2\pi}} \int_{-d_1}^{\infty} e^{-\frac{1}{2}\rho^2} d\rho \\ &= e^{\frac{1}{2}(K+1)x + \frac{1}{4}(K+1)^2\tau} N(d_1) \end{aligned}$$

donde $N(\cdot)$ es la función de probabilidad de la distribución normal estándar

$$N(d_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d_1} e^{-\frac{1}{2}s^2} ds.$$

El cálculo para A_2 es idéntico al de A_1 , reemplazando $(K-1)$ por $(K+1)$ en todo el análisis.

Es decir, resulta

$$A_2 = e^{\frac{1}{2}(K-1)x + \frac{1}{4}(K-1)^2\tau} N(d_2)$$

con

$$d_2 = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(K-1)\sqrt{2\tau},$$

así que tenemos una fórmula explícita para $u(x, \tau)$ dada por

$$u(x, \tau) = e^{\frac{1}{2}(K+1)x + \frac{1}{4}(K+1)^2\tau} N(d_1) - e^{\frac{1}{2}(K-1)x + \frac{1}{4}(K-1)^2\tau} N(d_2).$$

Ahora volvemos a cambiar las variables para obtener una expresión para $C(S, t)$. En primer lugar tenemos que

$$v(x, \tau) = e^{-\frac{1}{2}(K-1)x - \frac{1}{4}(K+1)^2\tau} u(x, \tau),$$

sustituyendo $u(x, \tau)$ tenemos

$$\begin{aligned} v(x, \tau) &= e^{-\frac{1}{2}(K-1)x - \frac{1}{4}(K+1)^2\tau} (e^{\frac{1}{2}(K+1)x + \frac{1}{4}(K+1)^2\tau} N(d_1) - e^{\frac{1}{2}(K-1)x + \frac{1}{4}(K-1)^2\tau} N(d_2)) \\ &= e^x N(d_1) - e^{K\tau} N(d_2). \end{aligned}$$

Usando los cambios de variable dados en un principio

$$x = \ln \frac{S}{K}, \quad \tau(t) = \frac{1}{2} \sigma^2 (T - t), \quad v(x, \tau) = \frac{C(S, t)}{K}, \quad K = \frac{2r}{\sigma^2},$$

tenemos

$$\frac{C(S, t)}{K} = e^{\ln \frac{S}{K}} N(d_1) - e^{-\frac{2r}{\sigma^2} \frac{1}{2} \sigma^2 (T-t)} N(d_2),$$

luego, simplificando la ecuación anterior, tenemos la fórmula de Black-Scholes:

$$C(S, t) = SN(d_1) - Ke^{-r(T-t)} N(d_2), \quad (3.20)$$

y por otro lado

$$\begin{aligned} d_1 &= \frac{x}{\sqrt{2\tau}} + \frac{1}{2}(K+1)\sqrt{2\tau} \\ &= \frac{\ln \frac{S}{K}}{\sqrt{2\frac{1}{2}\sigma^2(T-t)}} + \frac{1}{2} \left(\frac{2r}{\sigma^2} + 1 \right) \sqrt{2\frac{1}{2}\sigma^2(T-t)} \\ &= \frac{\ln \frac{S}{K}}{\sigma\sqrt{T-t}} + \left(\frac{r}{\sigma^2} + \frac{1}{2} \right) \sigma\sqrt{T-t} \\ &= \frac{\ln \frac{S}{K} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}. \end{aligned} \quad (3.21)$$

Análogamente,

$$d_2 = \frac{\ln \frac{S}{K} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}. \quad (3.22)$$

Por lo tanto el modelo de Black-Scholes queda como sigue:

$$\begin{aligned} C(S, t) &= SN(d_1) - Ke^{-r(T-t)} N(d_2), \\ d_1 &= \frac{\ln \frac{S}{K} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned}$$

donde $N(\cdot)$ es la función de distribución acumulativa de una distribución normal estándar.

Antes de pasar a la extensión de Markov del Modelo de Black-Scholes veremos un tema que será de gran importancia en la demostración del teorema de la siguiente sección. Cabe mencionar que el tema que a continuación estudiaremos es una extensión de Markov del modelo binomial o también conocido como modelo CRR (Cox-Ross-Rubinstein).

3.1.2. Extensión de Markov del modelo binomial de múltiples periodos

Consideraremos un modelo de múltiples periodos en el intervalo $[0, n]$, donde n es un número entero mayor a 1, siempre bajo la suposición de ausencia de arbitraje.

Para obtener resultados útiles, se puede seguir la presentación fundamental del modelo CRR, ver[2], adaptado al entorno de Markov, de tal manera que se puedan encontrar resultados manejables:

1. Conociendo todos los estados X_0, \dots, X_n

Comenzaremos con un modelo a tiempo discreto con n periodos y supondremos que, X_0, \dots, X_n con $X_0 = i, X_n = j$ y $S(0) = S_0$ son conocidos.

Supondremos que el valor del activo $S(n)$ al tiempo n esta dado por:

$$S(n) = V_{X_0 X_1} \cdots V_{X_{n-1} X_n} S_0$$

donde $V_{X_k X_{k+1}}$ define un movimiento ascendente con parámetro $u_{ij} > 1$ o un movimiento descendente con parámetro $d_{ij} < 1$, con probabilidad dada por

$$V_{X_{n-1} X_n} = \begin{cases} u_{ij} & \text{con probabilidad } q_{ij}, \\ d_{ij} & \text{con probabilidad } 1 - q_{ij}. \end{cases}$$

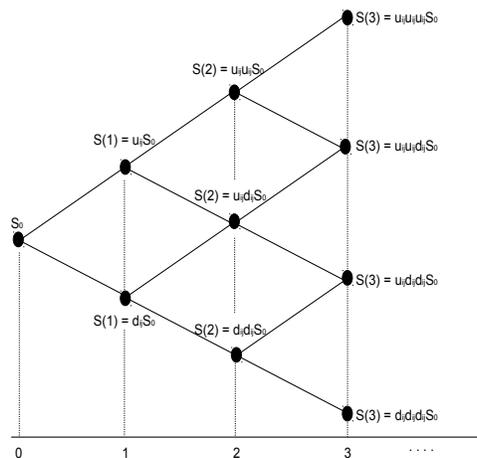


Figura 3.1: Modelo de multiples periodos.

Más aún, supongamos que, para cada n , las v.a. $V_{X_0X_1}, \dots, V_{X_{n-1}X_n}$ son condicionalmente independientes dadas X_0, \dots, X_n .

Además se tiene la hipótesis estándar $d_{ij} < r_{ij} < u_{ij}$, donde r_{ij} es la tasa de interés libre de riesgo.

Si la v.a. M_n representa el número total de movimientos ascendentes sobre $[0, n]$, el valor del activo al tiempo n esta dado por:

$$S(n) = (u_{ij})^{M_n} (d_{ij})^{n-M_n} S_0,$$

luego

$$\ln \frac{S(n)}{S_0} = M_n \ln(u_{ij}) + (n - M_n) \ln(d_{ij}).$$

Dado $X_0 = j_0, \dots, X_n = j_n, S(0) = S_0$, la distribución condicional de M_n es una distribución binomial con parámetros (n, q_{ij}) . Entonces se sigue que

$$E \left(\ln \frac{S(n)}{S_0} \mid X_0 = j_0, \dots, X_n = j_n, S(0) = S_0 \right) = n(q_{ij} \ln(u_{ij}) + (1 - q_{ij}) \ln(d_{ij})) \quad (3.23)$$

$$\text{Var} \left(\ln \frac{S(n)}{S_0} \mid X_0 = j_0, \dots, X_n = j_n, S(0) = S_0 \right) = n \left[q_{ij}(1 - q_{ij}) \left(\ln \frac{u_{ij}}{d_{ij}} \right)^2 \right] \quad (3.24)$$

La probabilidad neutral al riesgo¹ sobre los n períodos, esta dada por

$$\tilde{q}_{ij_1} = \frac{r_{ij_1} - d_{ij_1}}{u_{ij_1} - d_{ij_1}},$$

pues

$$\begin{aligned} d_{ij} &< r_{ij} < u_{ij} \\ \Rightarrow 0 &< r_{ij} - d_{ij} < u_{ij} - d_{ij} \\ \Rightarrow 0 &< \frac{r_{ij} - d_{ij}}{u_{ij} - d_{ij}} < 1, \end{aligned}$$

¹Se define la medida de probabilidad condicional neutral al riesgo como

$$\tilde{q}_{X_0X_1} = \frac{r_{X_0X_1} - d_{X_0X_1}}{u_{X_0X_1} - d_{X_0X_1}}.$$

es claro que, bajo nuestras suposiciones, para cada n , dado $X_0, \dots, X_n, S(0) = S_0$ con $X_0 = i, X_n = j$, tenemos un modelo CRR. Luego, obtenemos el valor del call europeo con precio de ejercicio y madurez n como el valor presente de la esperanza de la “ganancia” en el tiempo n bajo la medida de riesgo neutral, es decir:

$$C(S_0, 0 \mid X_0 = i, X_1, \dots, X_n = j) = \frac{1}{v_{ij}^n} \sum_{k=0}^n \binom{n}{k} \tilde{q}_{ij}^k (1 - \tilde{q}_{ij})^{n-k} \max\{u_{ij}^k d_{ij}^{n-k} S_0 - k, 0\}.$$

De aquí se obtiene la siguiente expresión (ver Cox y Rubinstein):

$$C(S_0, n \mid X_0 = i, X_1, \dots, X_n = j) = \begin{cases} S_0 B(b_{ij}; n, \tilde{q}'_{ij}) - \frac{k}{v_{ij}^n} B(b_{ij}; n, \tilde{q}_{ij}), & \text{si } b_{ij} < n, \\ 0 & \text{si } b_{ij} > n, \end{cases} \quad (3.25)$$

donde $B(x; m, \alpha)$ es el valor de la función de distribución binomial complementaria con parámetros complementarios m, α en el punto x y

$$b_{ij} = \left[\frac{\ln(k/d_{ij}^n S_0)}{\ln(u_{ij}/d_{ij})} + 1 \right],$$

$$\tilde{q}'_{ij} = \frac{u_{ij}}{r_{ij}} q_{ij}.$$

El resultado anterior puede ser visto como la extensión a tiempo discreto de la fórmula de Black-Scholes dado el entorno:

$$X_0 = i, \dots, X_n = j, S(0) = S_0.$$

2. Conociendo $X_0 = i$

Si solo conocemos el estado inicial del entorno $X_0 = i$, es claro que el valor del call está dado por

$$C_i(S_0, n) = \sum_{j=1}^m p_{ij}^{(n)} C_{ij}(S_0, n) \quad (3.26)$$

donde $[p_{ij}^{(n)}] = \mathbf{P}^n$ y $C_{ij}(S_0, n)$ esta dado por la ecuación (3.25).

3. Conociendo $X_n = j$

El uso de la fórmula de Bayes proporciona el siguiente resultado,

$$\begin{aligned} P(X_0 = i \mid X_n = j) &= \frac{P(X_0 = i, X_n = j)}{P(X_n = j)} \\ &= \frac{p_i p_{ij}^{(n)}}{\sum_{k=0}^m p_k p_{kj}^{(n)}}, \end{aligned} \quad (3.27)$$

por lo cual, el valor del call dado $X_n = j$, representado por $C^j(S_0, n)$, esta dado por

$$C^j(S_0, n) = \sum_{i=1}^m \frac{p_i p_{ij}^{(n)}}{\sum_{k=0}^m p_k p_{kj}^{(n)}} C_{ij}(S_0, n). \quad (3.28)$$

4. Sin tener conocimiento del entorno

Si no tenemos conocimiento alguno sobre el estado inicial del entorno pero conocemos su distribución de probabilidad dada por

$$\begin{aligned} P(X_0 = i) &= p_i, \quad \forall i \in I, \\ P(X_1 | X_0 = i) &= p_{ij}, \quad \forall i, j \in I, \end{aligned}$$

el valor del call denotado por $C(S_0, n)$ esta dado por

$$C(S_0, n) = \sum_{i=1}^m p_i C_i(S_0, n), \quad (3.29)$$

o por

$$C(S_0, n) = \sum_{j=1}^m \sum_{k=1}^m p_k p_{kj}^{(n)} C^j(S_0, n). \quad (3.30)$$

3.2. La extensión de Markov del modelo de Black-Scholes

En esta sección, se presenta una nueva extensión de la ecuación de Black-Scholes (1973) en las finanzas estocásticas con la introducción de un entorno aleatorio económico y financiero usando procesos de Markov, que se deben a Janssen y Manca (1999).

La construcción de este modelo se inicia desde el modelo tradicional de Cox, Ross, Rubinstein (CRR) de un periodo para obtener un nuevo modelo a tiempo continuo que satisface la hipótesis de ausencia de arbitraje.

Una de las principales aplicaciones potenciales de este modelo se refiere a la posibilidad de obtener una nueva forma de utilizar la fórmula de Black-Scholes con información relacionada con el entorno económico y financiero, en particular en relación con la volatilidad del activo subyacente.

Considere el modelo de Black-Scholes para la evolución de un activo que tiene valor conocido $S(0) = S_0$ al tiempo 0 y valor aleatorio S_t al tiempo t . El entorno económico y financiero se

define con las v.a. X_0, X_t que representan el estado del entorno al tiempo 0 y al tiempo t , respectivamente. Estas v.a. toman sus valores en el espacio de estados $I = \{1, 2, \dots, m\}$, las cuales están definidas en el espacio de probabilidad por:

$$P(X_0 = i) = p_i, \quad \forall i \in I, \quad (3.31)$$

$$P(X_t = j | X_0 = i) = p_{ij}, \quad \forall i, j \in I. \quad (3.32)$$

Introducimos las siguientes funciones $r_{X_0, X_t}, \sigma_{X_0, X_t}$ de X_0, X_t , las cuales son las v.a. de tasa de interés sin riesgo y volatilidad del activo respectivamente, con estado X_0 al tiempo 0 y X_t al tiempo t . Entonces tendremos las matrices $\mathbf{r} = [r_{ij}]$, $\boldsymbol{\sigma} = [\sigma_{ij}]$. Para utilizar la notación clásica de Black y Scholes, se puede definir la intensidad de la tasa de interés instantánea ρ_{ij} con $\rho_{ij} = \log(1 + r_{ij})$.

Teorema 3.1. [*Janssen y Manca 2009*] *Bajo la suposición de que la cadena de Markov de matriz de transición \mathbf{P} del proceso de entorno es ergódica y dado que el estado inicial del entorno es $i \in I$ y el estado del entorno al tiempo t es $j \in I$, la tasa sin riesgo está dada por ρ_{ij} y la volatilidad anual por σ_{ij} , entonces tenemos los siguientes resultados concernientes al precio del call Europeo al tiempo 0 con precio de ejercicio K y tiempo de madurez t :*

a) *Conociendo los estados $X_0 = i, X_t = j$, el valor del call está dado por*

$$\begin{aligned} C_{ij}(S_0, t) &= S_0 N(d_{ij1}) - Ke^{-\rho_{ij}t} N(d_{ij2}), \\ d_{ij1} &= \frac{1}{\sigma_{ij}\sqrt{t}} \left[\log \frac{S_0}{K} + \left(\rho_{ij} - \frac{\sigma_{ij}^2}{2} \right) t \right], \\ d_{ij2} &= d_{ij1} - \sigma_{ij}\sqrt{t}. \end{aligned}$$

b) *Conociendo el estado $X_0 = i$, el valor del call representado por $C_i(S_0, t)$ está dado por*

$$C_i(S_0, t) = \sum_{j \in I} \pi_j \cdot C_{ij}(S_0, t).$$

c) *Conociendo el estado $X_t = j$, el valor del call representado por $C^j(S_0, t)$ está dado por*

$$C^j(S_0, t) = \sum_{i \in I} p_i \cdot C_{ij}(S_0, t).$$

d) Sin conocer el entorno de ningún estado, el valor del call representado por $C(S_0, t)$ esta dado por

$$C(S_0, t) = \sum_{i \in I} p_i \cdot C_i(S_0, t).$$

o por

$$C(S_0, t) = \sum_{j \in I} \pi_j \cdot C^j(S_0, t).$$

Demostración. a) Para demostrar esta parte consideraremos un modelo discreto de la cadena de Markov límite de múltiples periodos, con n periodos, donde cada periodo tiene longitud h de modo que tenemos observaciones equidistantes al tiempo $0, h, 2h, \dots, nh$ con $n = \lfloor t/h \rfloor$. Además asumiremos que en el modelo de tiempo discreto aproximado, el proceso del entorno es una cadena de Markov ergódica homogénea definida por las relaciones (3.31) y (3.32) y que, para cada n , dados $X_0, \dots, X_n, S(0) = S_0$ con $X_0 = i, X_n = j$, elegimos, en cada subintervalo $[kh, (k+1)h]$, los siguientes parámetros ascendentes y descendentes:

$$u_{j_k, j_{k+1}} = e^{\sigma_{ij} \sqrt{\frac{t}{n}}}, \quad d_{j_k, j_{k+1}} = e^{-\sigma_{ij} \sqrt{\frac{t}{n}}}, \quad (3.33)$$

$$q_{j_k, j_{k+1}} = \frac{1}{2} + \frac{1}{2} \frac{\mu_{ij}}{\sigma_{ij}} \sqrt{\frac{t}{n}}, \quad (3.34)$$

dependiendo por lo tanto de las dos matrices no negativas $m \times m$;

$$[\mu_{ij}], \quad [\sigma_{ij}].$$

De las relaciones (3.23) y (3.24), se sigue que, para toda n :

$$E \left(\ln \frac{S(n)}{S_0} \mid X_0 = j_0, \dots, X_n = j_n, S(0) = S_0 \right) = \mu_{ij} t,$$

$$\text{Var} \left(\ln \frac{S(n)}{S_0} \mid X_0 = j_0, \dots, X_n = j_n, S(0) = S_0 \right) = \sigma_{ij}^2 t - \frac{\mu_{ij}^2 t^2}{n}.$$

Nuestro condicionamiento implica que podemos seguir el razonamiento de Cox y Rubinstein, ver Janssen y Manca (2010), tenemos por el teorema central del límite, cuando $n \rightarrow \infty$:

$$\ln \frac{S(t)}{S_0} \sim N(\mu_{ij} t, \sigma_{ij}^2 t),$$

donde i es el entorno del estado inicial observado a $t = 0$ y j el entorno del estado al tiempo t .

En cuanto a las tasas de interés sin riesgo, podemos suponer que, para toda i y j , existe $v_{ij} > 1$ tal que la nueva tasa de retorno para todos los periodos $(kh, (k+1)h)$, denotada por \hat{r}_{ij} , para $n \rightarrow \infty$, satisface la siguiente condición:

$$(1 + r_{ij})^n \rightarrow (1 + \hat{r}_{ij})^t. \quad (3.35)$$

Ahora supongamos que $C_{ij}(S_0, n)$ representa el valor de una opción call Europea al tiempo 0, con madurez n y precio de ejercicio K cuando el estado inicial es i y en el tiempo n es j .

Usando la misma técnica de la demostración de la fórmula de Black-Scholes dada por Cox y Rubinstein pero con parámetros que dependen de todos los entornos de los estados i y j , obtenemos bajo las condiciones (3.33)-(3.35), para t fijo:

$$C_{ij}(S_0, n) \rightarrow C_{ij}(S_0, t)$$

donde:

$$C_{ij}(S_0, t) = S_0 N(d_{ij,1}) - Kr_{ij}^{-t} N(d_{ij,2}), \quad (3.36)$$

$$d_{ij,1} = \frac{\ln \frac{S_0}{Kr_{ij}^{-t}}}{\sigma_{ij}\sqrt{t}} + \frac{1}{2}\sigma_{ij}\sqrt{t}, \quad (3.37)$$

$$d_{ij,2} = d_{ij,1} - \sigma_{ij}\sqrt{t}. \quad (3.38)$$

Este resultado da el valor del call al tiempo 0 con i como el entorno de estado inicial y j como el entorno del estado observado al tiempo t , representado ahora por X_t .

Si queremos usar la notación tradicional en el marco de Black y Scholes, podemos definir la intensidad de las tasas de interés instantaneas ρ_{ij} tal que:

$$r_{ij} = e^{\rho_{ij}}$$

de modo que, la fórmulas (3.36)-(3.38) se convierten en:

$$C_{ij}(S_0, t) = S_0 N(d_{ij,1}) - Ke^{-\rho_{ij}t} N(d_{ij,2}),$$

$$d_{ij,1} = \frac{1}{\sigma_{ij}\sqrt{t}} \left[\ln \frac{S}{K} + \left(\rho_{ij} - \frac{\sigma_{ij}^2}{2} \right) t \right],$$

$$d_{ij,2} = d_{ij,1} - \sigma_{ij}\sqrt{t}.$$

b) Para esta parte consideremos la ecuación (3.26)

$$C_i(S_0, n) = \sum_{j=1}^m p_{ij}^{(n)} C_{ij}(S_0, n).$$

Ahora supongamos que $n \rightarrow \infty$ y cambiando el límite con la suma finita, se tiene

$$\begin{aligned} \lim_{n \rightarrow \infty} C_i(S_0, n) &= \lim_{n \rightarrow \infty} \sum_{j=1}^m p_{ij}^{(n)} C_{ij}(S_0, n) \\ &= \sum_{j=1}^m \lim_{n \rightarrow \infty} p_{ij}^{(n)} C_{ij}(S_0, n) \end{aligned}$$

Por la suposición de ergodicidad sobre el entorno de la cadena de matriz \mathbf{P} , podemos usar la ecuación (1.6); y por el resultado a), tenemos

$$C_i(S_0, t) = \sum_{j \in I} \pi_j \cdot C_{ij}(S_0, t).$$

c) Consideremos la ecuación (3.28),

$$C^j(S_0, n) = \sum_{i=1}^m \frac{p_i p_{ij}^{(n)}}{\sum_{k=0}^m a_k p_{kj}^{(n)}} C_{ij}(S_0, n).$$

Procediendo como en el resultado b) tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} C^j(S_0, n) &= \lim_{n \rightarrow \infty} \sum_{i=1}^m \frac{p_i p_{ij}^{(n)}}{\sum_{k=0}^m p_k p_{kj}^{(n)}} C_{ij}(S_0, n) \\ &= \sum_{i=1}^m \lim_{n \rightarrow \infty} \frac{p_i p_{ij}^{(n)}}{\sum_{k=0}^m p_k p_{kj}^{(n)}} C_{ij}(S_0, n) \\ &= \sum_{i=1}^m \frac{\lim_{n \rightarrow \infty} p_i p_{ij}^{(n)}}{\lim_{n \rightarrow \infty} \sum_{k=0}^m p_k p_{kj}^{(n)}} \lim_{n \rightarrow \infty} C_{ij}(S_0, n) \\ &= \sum_{i=1}^m \frac{p_i \pi_j}{\sum_{k=0}^m \lim_{n \rightarrow \infty} p_k p_{kj}^{(n)}} C_{ij}(S_0, t) \\ &= \sum_{i=1}^m \frac{p_i \pi_j}{\sum_{k=0}^m p_k \pi_j} C_{ij}(S_0, t) \\ &= \sum_{i=1}^m \frac{p_i}{\sum_{k=0}^m p_k} C_{ij}(S_0, t) \\ &= \sum_{i=1}^m p_i C_{ij}(S_0, t). \end{aligned}$$

Por lo tanto,

$$C^j(S_0, t) = \sum_{i \in I} p_i \cdot C_{ij}(S_0, t).$$

d) Para esta última parte consideraremos las ecuaciones (3.29) y (3.30),

$$C(S_0, n) = \sum_{i=1}^m p_i C_i(S_0, n), \quad (3.39)$$

$$C(S_0, n) = \sum_{j=1}^m \sum_{k=1}^m p_k p_{kj}^{(n)} C^j(S_0, n). \quad (3.40)$$

Utilizando los resultados b) y c), y procediendo de manera análoga tenemos para la ecuación (3.39)

$$\begin{aligned} \lim_{n \rightarrow \infty} C(S_0, n) &= \lim_{n \rightarrow \infty} \sum_{i=1}^m p_i C_i(S_0, n) \\ &= \sum_{i=1}^m \lim_{n \rightarrow \infty} p_i C_i(S_0, n) \\ &= \sum_{i=1}^m p_i C_i(S_0, t). \end{aligned}$$

Por lo tanto

$$C(S_0, t) = \sum_{i=1}^m p_i C_i(S_0, t).$$

Ahora para la ecuación (3.39), tenemos

$$\begin{aligned} \lim_{n \rightarrow \infty} C(S_0, n) &= \lim_{n \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^m p_k p_{kj}^{(n)} C^j(S_0, n) \\ &= \sum_{j=1}^m \sum_{k=1}^m \lim_{n \rightarrow \infty} p_k p_{kj}^{(n)} \lim_{n \rightarrow \infty} C^j(S_0, n) \\ &= \sum_{j=1}^m \sum_{k=1}^m p_k \pi_j C^j(S_0, t) \\ &= \sum_{j=1}^m \pi_j C^j(S_0, t). \end{aligned}$$

Por lo tanto

$$C(S_0, t) = \sum_{j=1}^m \pi_j C^j(S_0, t).$$

□

El problema principal del modelo de Markov de Black-Scholes es estimar los parámetros de las matrices \mathbf{P} , $\sigma = [\sigma_{ij}]$ y $\rho = \rho_{ij}$.

Este modelo proporciona la posibilidad de tomar en cuenta las previsiones realizadas por los inversionistas de manera tal como para incorporarlos en su propia valuación de opciones y también se puede utilizar para los modelos con crisis financiera, así como para la construcción de escenarios, y en particular en el caso de estrés en un enfoque de tipo VaR.

Capítulo 4

Modelo financiero de Markov usando Modelos Ocultos de Markov

Jansen et al (2009) consideraron que el ambiente en donde se desarrolla el proceso financiero se realiza en un ambiente que cambia de manera aleatoria de bueno a malo, de malo a bueno, de bueno a regular, etc.. Entonces, este autor propone modelar este ambiente cambiante por medio de una cadena de Markov. De este modelo lo más complicado es estimar apropiadamente los parámetros del modelo de Markov, en este sentido Tuyen (2010) propone utilizar los modelos ocultos de Markov para identificar los parámetros del modelo de Markov en la extensión de Markov del modelo de Black-Scholes. Tuyen aplicó este método a los datos de la bolsa de valores de VietNam del norte para determinar los precios de las acciones. En este capítulo se aplica la misma metodología propuesta por Tuyen a los datos de la bolsa mexicana de valores. A diferencia de Tuyen, aquí también consideraremos que las tasas de interés son afectadas por los estados de la cadena de Markov.

En el proceso financiero se desconoce todo acerca del modelo de Markov oculto, lo más básico en el proceso financiero es que se desconoce el número de estados existentes, por lo que, determinar el número de estados más probables dadas las variables observables, será el primer paso a seguir. Para esto, se resuelve el problema de evaluación, esto es, se realiza el algoritmo Forward-Backward, con este algoritmo se resuelve el problema de aprendizaje, esto es, se realiza el algoritmo EM, usando los valores de m igual a $m = 2, 3, 4, 5, 6, \dots$, y posteriormente, se elige el mejor valor de m con base en los criterios AIC y BIC, los cuales se definen mediante

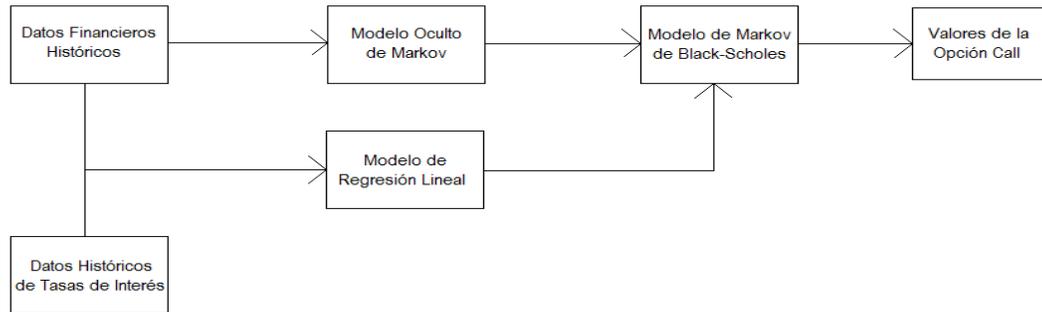


Figura 4.1: Operación del modelo financiero de Markov usando HMM

las fórmulas:

$$AIC = 2k - \ln(L) \quad y \quad BIC = -2\ln(L) + k\ln(n)$$

donde k es el número de parámetros en el modelo estadístico, L es el máximo valor de la función de verosimilitud para el modelo estimado y n es el número de datos para la muestra.

Para cada valor de m , el algoritmo recursivo se alimenta con datos iniciales para los valores del vector \mathbf{p} , de la matriz de transición P , y de los parámetros de las variables μ_i y σ_i^2 .

En particular, la matriz de transición inicial es igual a la matriz cuyos elementos son todos iguales a $\frac{1}{m}$, ya que esta matriz es ergódica y el teorema 3.1 pide que dicha matriz sea así, el programa realizado al proponer una matriz ergódica nos devuelve otra matriz ergódica.

Posteriormente, ya con un valor de m seleccionado se resuelve el problema de decodificación realizando el algoritmo de Viterbi.

4.1. Algoritmo

La operación del modelo se ilustra en la figura 4.1.

El proceso para estimar los parámetros ocultos del proceso de Markov suponiendo que el número de estados es m , sigue los pasos:

- Entren los datos historicos y el valor de m .
 - Entren los valores iniciales propuestos de los parámetros de la distribución de los observables $\{\mu_i, \sigma_i\}_{i=1}^m$,
 - entran los valores iniciales propuestos de la distribución inicial $\mathbf{p} = \{p_1, p_2, \dots, p_m\}$ y
 - entran los valores iniciales propuestos de la matriz de transición $\mathbf{P} = [p_{ij}]$.
- Se realiza el algoritmo Forward-Backward.
- Se realiza el algoritmo EM.
- Una vez que se realiza este algoritmo para $m = 2, 3, \dots$ se elige la m que presenta mejores valores de AIC y BIC, y se obtienen los nuevos valores de $\{\mu_i, \sigma_i\}_{i=1}^m$, \mathbf{p} y \mathbf{P} .
- Con la matriz \mathbf{P} obtenida, se estima la distribución limite π_i , $i = 1, 2, \dots, m$ de la cadena de Markov.
- Usando el algoritmo de Viterbi se encuentra la sucesión de estados $\{j_0, j_1, \dots, j_n\}$ más probable correspondiente a la sucesión de observaciones $\{o_0, o_1, \dots, o_n\}$.

Para obtener los precios de las acciones con la extensión de Markov del modelo de Black-Scholes, además de la matriz de transición P , la distribución limite π_i y la distribución inicial p_i se requiere también la matriz de volatilidad $\sigma = [\sigma_{ij}]$ y la matriz de tasa de interés libre de riesgo $\rho = \rho_{ij}$, que no son obtenidas del proceso HMM. Tuyen (2010), en su artículo propone un nuevo método de obtención de estas matrices. De la sucesión de estados $\{j_0, j_1, \dots, j_n\}$, obtenida del algoritmo se dividen los datos históricos $\{o_0, o_1, \dots, o_n\}$ en m grupos G_1, G_2, \dots, G_m , cada uno correspondiente a los m estados de la cadena de Markov. Para σ_{ij} , se calcula la desviación estandar del conjunto $G_i \cup G_j$. Así se obtiene la matriz $\sigma = [\sigma_{ij}]$.

Para la matriz ρ , suponemos que la tasa de interés r es estable en el tiempo. Para estimar r_{ij} , se propone usar un modelo de regresión lineal. Primero, se usa un modelo de regresión lineal para estimar los parámetros β_0, β_1 en la relación $r_i = \beta_1 O_i + \beta_0 + \varepsilon$. Luego se calcula $r_i = \beta_1 \mu_i + \beta_0$ para toda $i = 1, \dots, m$. Finalmente, se estima $r_{ij} = r_{ji} = \frac{r_i + r_j}{2}$. Así se tiene la matriz $[\rho_{ij}]$.

Para la distribución inicial \mathbf{p} , como el tiempo $t = 0$ se puede tomar en cualquier momento en el futuro, se supone que $p_i = \pi_i$ para toda $i = 1, \dots, m$.

4.2. Resultados

Para aplicar el método se utilizaron datos históricos diarios de acciones de la Bolsa Mexicana de Valores (<http://mx.investing.com/equities/bolsa-mexicana-de-valores-a-historical-data>) y datos de las tasas de interés interbancaria de equilibrio (TIIE) a 28 días, del Banco de México. Para probar si el tamaño de la muestra afecta a los resultados, consideramos dos ventanas de tiempo, la primera va del dos de enero del dos mil trece, al treinta de enero del dos mil quince, con un total de 523 datos, la segunda ventana de tiempo fue del dieciocho de julio al diecisiete de octubre del dos mil catorce, con un total de 65 datos. Los valores iniciales utilizados para la aplicación fueron los siguientes: para los 523 datos el valor de m seleccionado fue 5, y los datos iniciales fueron: $\mu = (25, 26.1, 27, 27.7, 28.4)$; $\sigma = (1, 2, 3, 4, 5)$ y los elementos del vector p y de la matriz \mathbf{P} son igual a 0.2. Mientras que para los 65 datos m fue igual a 3 y los valores iniciales fueron: $\mu = (25.5, 27, 28.9)$; $\sigma = (1, 2, 3)$ y los elementos del vector p y de la matriz \mathbf{P} son igual a 1/3.

Los resultados obtenidos se presentan en las dos secciones siguientes.

4.2.1. Resultados con los datos en la primera ventana de tiempo

En la figura 4.2 se muestra la gráfica del precio de las acciones (observaciones) y la gráfica de las tasas de interés.

En el cuadro 4.1 se presentan los valores de los criterios AIC y BIC asociados a los modelos obtenidos mediante el algoritmo EM para los distintos valores de m .

	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
AIC	2036.827	1809.249	1535.716	1359.201	1333.392
BIC	2058.125	1856.104	1616.648	1482.728	1508.035

Cuadro 4.1: AIC y BIC

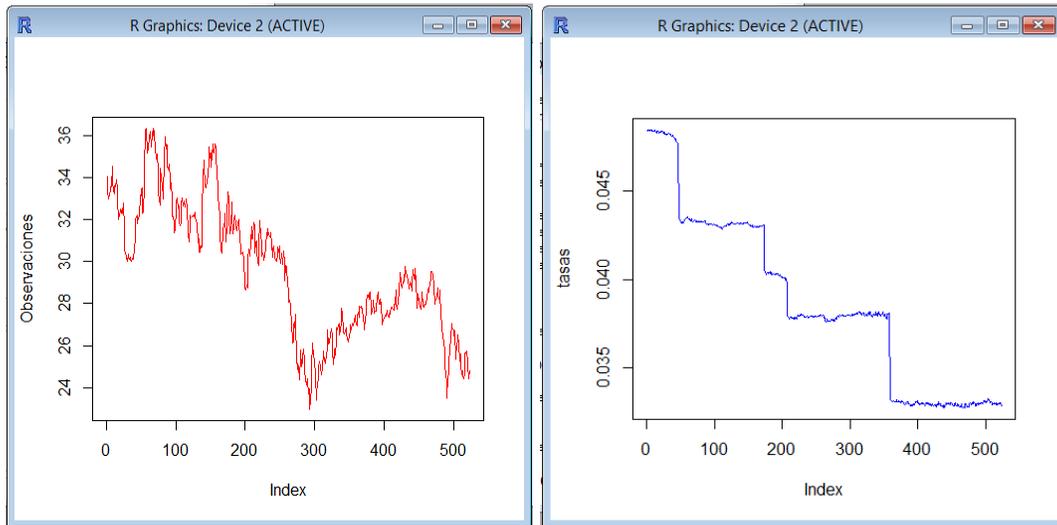


Figura 4.2: Gráfica de datos hitóricos y tasas de interés.

Obsérvese que el AIC menor ocurrió cuando $m = 6$, mientras que el BIC menor ocurrió cuando $m = 5$, en este sentido los dos criterios nos indican un número diferente, sin embargo, al revisar el histograma de los datos, vemos como se manifiestan 5 distintas distribuciones normales, que estarán relacionadas con los estados del proceso de Markov. Por lo que decidimos considerar 5 estados en el proceso de Markov para seguir con el análisis, ver figura 4.3.

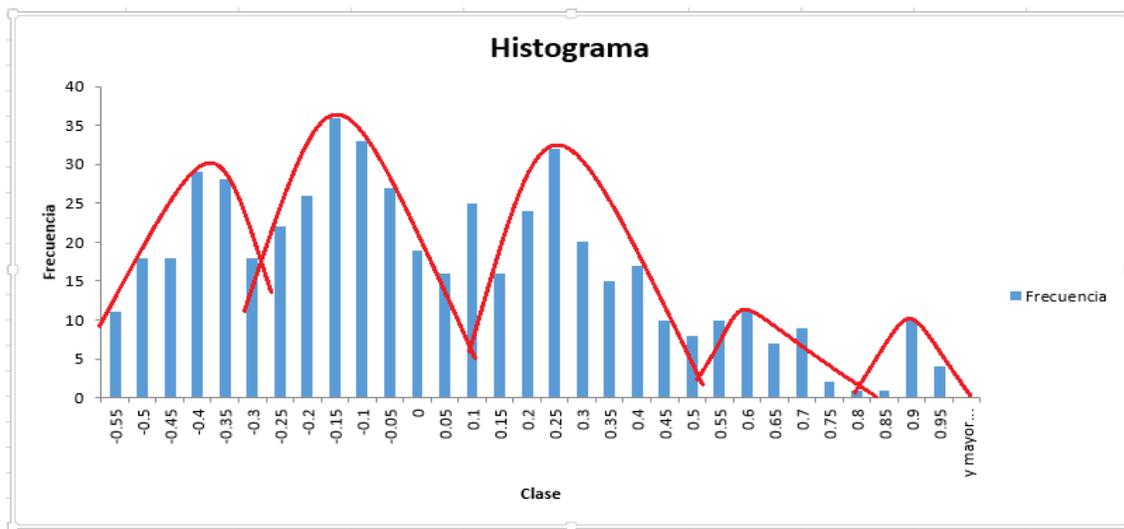


Figura 4.3: Histograma de los datos históricos diarios de la bolsa mexicana de valores. Se puede ver las 5 distribuciones normales. Las tres primeras distribuciones tienen mayor probabilidad de ocurrir, las dos últimas tienen una frecuencia menor.

En el cuadro 4.2 se muestran los parámetros obtenidos para las distribuciones Normales. Estos parámetros se obtuvieron con el algoritmo EM desarrollado en el capítulo 2.

Parametros	Estado 1	Estado 2	Estado 3	Estado 4	Estado 5
μ	25.6814	28.2444	30.3843	32.0240	34.4754
σ	1.0259	0.7238	0.2865	0.5999	1.0313

Cuadro 4.2: Parámetros de las distribuciones Normales.

La matriz de transición P se estimó con el algoritmo forward-backward y el algoritmo EM, ambos presentados en el capítulo 2. La estimación resultante fue la siguiente

$$\mathbf{P} = \begin{pmatrix} 0.9920 & 0.0079 & 0.0000 & 0.0000 & 0.0000 \\ 0.0142 & 0.9787 & 0.0070 & 0.0000 & 0.0000 \\ 0.0000 & 0.0300 & 0.8787 & 0.0911 & 0.0000 \\ 0.0000 & 0.0000 & 0.0080 & 0.9117 & 0.0202 \\ 0.0000 & 0.0000 & 0.0000 & 0.0386 & 0.9613 \end{pmatrix}$$

está matriz puede ayudar a los inversionistas a tomar decisiones al predecir los cambios en el mercado de valores. Está matriz es regular y su distribución límite es:

$$\pi = \left(0.152221 \quad 0.085758 \quad 0.020803 \quad 0.240395 \quad 0.500821 \right)$$

La matriz $\sigma = [\sigma_{ij}]$. Para estimar esta matriz, se utiliza la sucesión de estados más probable que se obtuvo con el algoritmo de Viterbi. Dividimos los datos históricos $\{o_0, o_1, \dots, o_n\}$ en m grupos G_1, G_2, \dots, G_m correspondientes a m estados de la cadena de Markov. Para σ_{ij} , se calculó la desviación estandar del conjunto $G_i \cup G_j$. Luego la matriz estimada es

$$\sigma = \begin{pmatrix} 0.020411 & 0.019124 & 0.023132 & 0.026276 & 0.029353 \\ 0.016977 & 0.013639 & 0.015095 & 0.019145 & 0.021493 \\ 0.020121 & 0.012874 & 0.011282 & 0.016128 & 0.018474 \\ 0.021971 & 0.015513 & 0.014546 & 0.016570 & 0.018067 \\ 0.025584 & 0.017922 & 0.016766 & 0.017035 & 0.018178 \end{pmatrix}$$

Caso r calculado con regresión lineal

Siguiendo el artículo de Tuyen L. (2010) se obtuvo una ecuación lineal de primer orden para las tasas de interés r en función de los precios históricos O , la relación obtenida es

$$r = 0.001058 * O + 0.007407.$$

Con la que se calcula las tasas de interés para cada uno de los 5 estados. Se obtuvo que r_i es igual a:

$$\left(0.034577 \quad 0.037289 \quad 0.039553 \quad 0.041288 \quad 0.043881 \right)$$

La matriz $[r_{ij}]$ se estimó de la siguiente manera $r_{ij} = r_{ji} = \frac{r_i + r_j}{2}$.

$$\mathbf{r} = \begin{pmatrix} 0.034577 & 0.035933 & 0.037065 & 0.037933 & 0.039229 \\ 0.035933 & 0.037289 & 0.038421 & 0.039288 & 0.040585 \\ 0.037065 & 0.038421 & 0.039553 & 0.040420 & 0.041717 \\ 0.037933 & 0.039288 & 0.040420 & 0.041288 & 0.042585 \\ 0.039229 & 0.040585 & 0.041717 & 0.042585 & 0.043881 \end{pmatrix}$$

Al usar la relación $\rho_{ij} = \log(1 + r_{ij})$, se obtiene la matriz $[\rho_{ij}]$.

$$[\rho_{ij}] = \begin{pmatrix} 0.033993 & 0.035303 & 0.036395 & 0.037231 & 0.038480 \\ 0.035303 & 0.036611 & 0.037701 & 0.038536 & 0.039783 \\ 0.036395 & 0.037701 & 0.038791 & 0.039625 & 0.040871 \\ 0.037231 & 0.038536 & 0.039625 & 0.040458 & 0.041703 \\ 0.038480 & 0.039783 & 0.040871 & 0.041703 & 0.042946 \end{pmatrix}$$

Con los datos obtenidos de σ_{ij} y ρ_{ij} , se aplicó el modelo de Markov de Black-Scholes para $S(0) = 24.8$, $K = 29.8$ y $t = 6$ meses. Con estos datos se obtuvieron los siguientes resultados:

Al conocer el estado inicial $X_0 = i$ y el estado final $X_n = j$, los valores de la opción $C_{ij}(S_0, t)$ para los diferentes estados se muestra en el cuadro 4.3.

$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$
1 \rightarrow 1	0.778571	2 \rightarrow 1	0.839512	3 \rightarrow 1	1.016377
1 \rightarrow 2	0.881368	2 \rightarrow 2	0.935326	3 \rightarrow 2	1.064389
1 \rightarrow 3	1.074157	2 \rightarrow 3	1.088693	3 \rightarrow 3	1.198017
1 \rightarrow 4	1.222410	2 \rightarrow 4	1.243888	3 \rightarrow 4	1.343160
1 \rightarrow 5	1.411234	2 \rightarrow 5	1.425028	3 \rightarrow 5	1.524189
4 \rightarrow 1	1.140977	5 \rightarrow 1	1.341597		
4 \rightarrow 2	1.198190	5 \rightarrow 2	1.381227		
4 \rightarrow 3	1.329935	5 \rightarrow 3	1.509239		
4 \rightarrow 4	1.454580	5 \rightarrow 4	1.620390		
4 \rightarrow 5	1.627682	5 \rightarrow 5	1.790584		

Cuadro 4.3: Valores de la opción call conociendo el estado i y el estado j .

Si se conoce el estado inicial $X_0 = i$ y se desconoce el estado final $X_n = j$, los valores de la opción call $C_i(S_0, t)$ se muestran en el cuadro 4.4.

$i \rightarrow ?$	$C_i(S_0, t)$
1 $\rightarrow ?$	1.217084
2 $\rightarrow ?$	1.243362
3 $\rightarrow ?$	1.357154
4 $\rightarrow ?$	1.468955
5 $\rightarrow ?$	1.640366

Cuadro 4.4: Valores de la opción call sin conocimiento del estado j .

Cuando se conoce el estado final $X_n = j$ y se desconoce el estado inicial $X_0 = i$, los valores de la opción call $C^j(S_0, t)$ se muestran en el cuadro 4.5.

$? \rightarrow j$	$C^j(S_0, t)$
? $\rightarrow 1$	1.157841
? $\rightarrow 2$	1.216306
? $\rightarrow 3$	1.357367
? $\rightarrow 4$	1.481893
? $\rightarrow 5$	1.656786

Cuadro 4.5: Valores de la opción call sin conocimiento del estado i .

Cuando no se conoce tanto el estado inicial X_0 como el estado final X_n , se obtuvo que el valor de la opción call es

$$C(S_0, t) = 1.494789.$$

Caso r calculado con promedios en los estados

Si consideramos que las tasas de interés no necesariamente satisfacen un modelo de regresión lineal con los precios de las acciones, pero que si son afectadas por los cambios en los estados del proceso de Markov. Entonces, como se hizo para los datos de la variable observable, los datos de las tasas de interés $\{r_0, r_1, \dots, r_n\}$ se separan en m grupos G_1, G_2, \dots, G_m correspondientes

a los m estados de la cadena de Markov, y con los datos que quedaron en cada grupo se obtiene un promedio, este promedio se utiliza como tasa de interés del grupo correspondiente.

Los promedios r_i son iguales a:

$$\left(0.036433 \quad 0.033653 \quad 0.041129 \quad 0.042725 \quad 0.044238 \right)$$

La matriz $[r_{ij}]$ se estimó con la fórmula $r_{ij} = r_{ji} = \frac{r_i + r_j}{2}$. De donde se obtiene la matriz

$$\mathbf{r} = \begin{pmatrix} 0.036433 & 0.035043 & 0.038781 & 0.039579 & 0.040335 \\ 0.035043 & 0.033653 & 0.037391 & 0.038189 & 0.038946 \\ 0.038781 & 0.037391 & 0.041129 & 0.041927 & 0.042684 \\ 0.039579 & 0.038189 & 0.041927 & 0.042725 & 0.043481 \\ 0.040335 & 0.038946 & 0.042684 & 0.043481 & 0.044238 \end{pmatrix}$$

Al usar la relación $\rho_{ij} = \log(1 + r_{ij})$, se obtiene la matriz $[\rho_{ij}]$.

$$[\rho_{ij}] = \begin{pmatrix} 0.035785 & 0.034443 & 0.038048 & 0.038816 & 0.039543 \\ 0.034443 & 0.033099 & 0.036709 & 0.037478 & 0.038206 \\ 0.038048 & 0.036709 & 0.040306 & 0.041072 & 0.041798 \\ 0.038816 & 0.037478 & 0.041072 & 0.041838 & 0.042563 \\ 0.039543 & 0.038206 & 0.041798 & 0.042563 & 0.043287 \end{pmatrix}$$

Con estos nuevos datos se aplicó el modelo de Markov de Black-Scholes para $S(0) = 24.8$, $K = 29.8$ y $t = 6$ meses. Con estos datos se obtuvieron los siguientes resultados:

Al conocer el estado inicial $X_0 = i$ y el estado final $X_n = j$, los valores de la opción $C_{ij}(S_0, t)$ para los diferentes estados se muestra en el cuadro 4.6.

$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$
1 → 1	0.957094	2 → 1	0.749279	3 → 1	1.201241
1 → 2	0.794181	2 → 2	0.544141	3 → 2	0.937818
1 → 3	1.251193	2 → 3	0.967617	3 → 3	1.405812
1 → 4	1.390413	2 → 4	1.120100	3 → 4	1.531517
1 → 5	1.523632	2 → 5	1.241323	3 → 5	1.643080
4 → 1	1.318656	5 → 1	1.458992		
4 → 2	1.066473	5 → 2	1.187319		
4 → 3	1.522319	5 → 3	1.630782		
4 → 4	1.635215	5 → 4	1.733688		
4 → 5	1.739685	5 → 5	1.835432		

Cuadro 4.6: Valores de la opción call conociendo el estado i y el estado j .

Si se conoce el estado inicial $X_0 = i$ y se desconoce el estado final $X_n = j$, los valores de la opción call $C_i(S_0, t)$ se muestran en el cuadro 4.7.

$i \rightarrow ?$	$C_i(S_0, t)$
1 → ?	1.337144
2 → ?	1.071799
3 → ?	1.483585
4 → ?	1.588226
5 → ?	1.693833

Cuadro 4.7: Valores de la opción call sin conocimiento del estado j .

Cuando se conoce el estado final $X_n = j$ y se desconoce el estado inicial $X_0 = i$, los valores de la opción call $C^j(S_0, t)$ se muestran en el cuadro 4.8.

$? \rightarrow j$	$C^j(S_0, t)$
$? \rightarrow 1$	1.282631
$? \rightarrow 2$	1.038076
$? \rightarrow 3$	1.485375
$? \rightarrow 4$	1.600936
$? \rightarrow 5$	1.710001

Cuadro 4.8: Valores de la opción call sin conocimiento del estado i .

Cuando no se conoce el estado inicial X_0 ni el estado final X_n , se obtuvo que el valor de la opción call es

$$C(S_0, t) = 1.556431.$$

Ahora, con los mismos valores de $S(0)$, K y t , se aplicó el modelo de Black-Scholes, con lo cual se obtuvo:

$$C(S, t) = 3.207977.$$

Recordemos que para aplicar el modelo de Black-Scholes se debe conocer la tasa de interés anual y la volatilidad, por lo cual con los datos históricos de la Bolsa Mexicana de Valores, se calculó la volatilidad histórica, la cual fue de 12.6495%; y para la tasa de interés anual se consideró una tasa de 3.2950%.

Podemos observar que la tasa de interés que se considero para calcular el valor de la opción con el modelo de Black-Scholes es del mismo orden que los valores que se obtuvieron para ρ_{ij} con el método presentado en esta tesis; sin embargo la volatilidad si difiere mucho. Esta puede ser la razón por la que los precios calculados de las acciones difieren tanto.

4.2.2. Resultados con los datos en la segunda ventana de tiempo

En la figura 4.4 se muestra la gráfica del precio de las acciones (observaciones) y la gráfica de las tasas de interés.

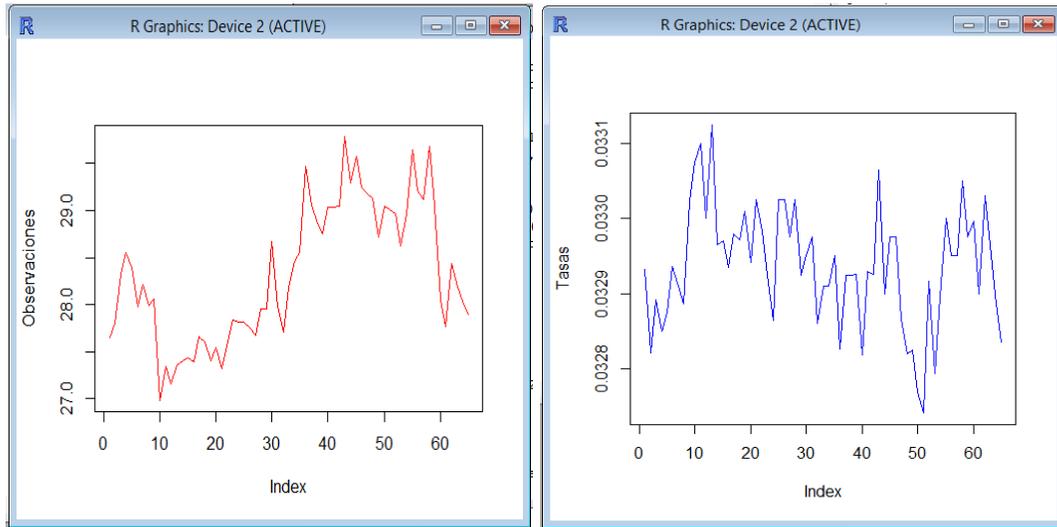


Figura 4.4: Gráfica de datos hitóricos y tasas de interés.

En el cuadro 4.9 se presentan los valores de los criterios AIC y BIC asociados a los modelos obtenidos mediante el algoritmo EM para los distintos valores de m .

	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
AIC	77.30705	63.54448	70.75265	84.83684	85.12092
BIC	88.17899	87.46274	112.066	147.8941	174.2708

Cuadro 4.9: AIC y BIC

Obsérvese que el AIC y el BIC menor ocurrió cuando $m = 3$, al revisar el histograma de los datos, vemos como se manifiestan 3 distintas distribuciones normales, que estarán relacionadas con los estados del proceso de Markov. Por lo que decidimos considerar 3 estados en el proceso de Markov para seguir con el análisis, ver figura 4.5.

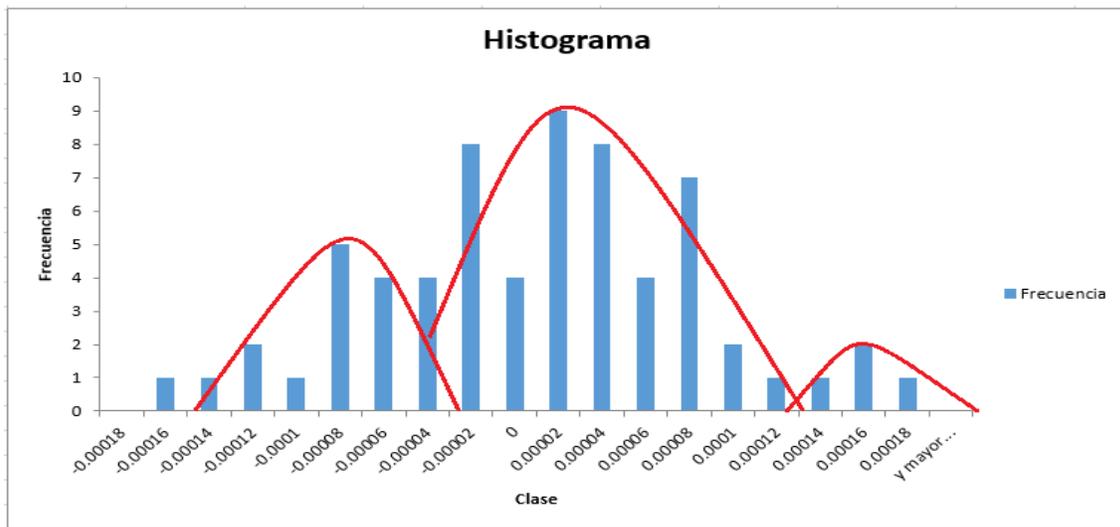


Figura 4.5: Histograma de los datos históricos diarios de la bolsa mexicana de valores. Se puede ver las 3 distribuciones normales. La segunda distribución tienen mayor probabilidad de ocurrir.

En el cuadro 4.10 se muestran los parámetros obtenidos para las distribuciones Normales. Estos parámetros se obtuvieron con el algoritmo EM desarrollado en el capítulo 2.

Parametros	Estado 1	Estado 2	Estado 3
μ	27.3988	28.0473	29.1331
σ	0.1788	0.2849	0.3078

Cuadro 4.10: Parámetros de las distribuciones Normales

La matriz de transición P se estimó con el algoritmo forward-backward y el algoritmo EM, ambos presentados en el capítulo 2. La estimación resultante fue la siguiente

$$\mathbf{P} = \begin{pmatrix} 0.9212 & 0.0787 & 0.0000 \\ 0.0386 & 0.9230 & 0.0382 \\ 0.0000 & 0.0410 & 0.9589 \end{pmatrix}$$

esta matriz puede ayudar a los inversionistas a tomar decisiones al predecir los cambios en el mercado de valores. Esta matriz es regular y su distribución límite es:

$$\pi = \left(0.2017804 \quad 0.4119248 \quad 0.3862948 \right)$$

Solo mostraremos los resultados de la matriz $\sigma = [\sigma_{ij}]$ y la matriz $[\rho_{ij}]$, ya que anteriormente se ha explicado el método para obtenerlas con 523 datos.

Matriz $\sigma = [\sigma_{ij}]$

$$\sigma = \begin{pmatrix} 0.008099 & 0.010522 & 0.015552 \\ 0.011797 & 0.011867 & 0.014341 \\ 0.016379 & 0.013947 & 0.012728 \end{pmatrix}.$$

Caso r calculado con regresión lineal

Haciendo regresión lineal:

$$r = -(3.268E - 05 * O) + 3.386239E - 02.$$

$$r_i = \begin{pmatrix} 0.032966 & 0.032945 & 0.032909 \end{pmatrix}$$

Matriz $[r_{ij}]$

$$\mathbf{r} = \begin{pmatrix} 0.032966 & 0.032956 & 0.032938 \\ 0.032956 & 0.032945 & 0.032927 \\ 0.032938 & 0.032927 & 0.032909 \end{pmatrix}.$$

Matriz $[\rho_{ij}]$

$$[\rho_{ij}] = \begin{pmatrix} 0.032434 & 0.032424 & 0.032407 \\ 0.032424 & 0.032414 & 0.032397 \\ 0.032407 & 0.032397 & 0.032379 \end{pmatrix}.$$

En este caso se aplicó el modelo de Markov de Black-Scholes para $S(0) = 28.4$, $K = 32$ y $t = 6$ meses. Se obtuvieron los siguientes resultados.

Cuando se conoce el estado inicial $X_0 = i$ y el estado final $X_n = j$, $C_{ij}(S_0, t)$ se muestra en el cuadro 4.11.

$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$
1 \rightarrow 1	2.059008	2 \rightarrow 1	2.058525	3 \rightarrow 1	2.067710
1 \rightarrow 2	2.057736	2 \rightarrow 2	2.056977	3 \rightarrow 2	2.057655
1 \rightarrow 3	2.064211	2 \rightarrow 3	2.058664	3 \rightarrow 3	2.052621

Cuadro 4.11: Valores de la opción call conociendo el estado i y el estado j .

Si se conoce el estado i y se desconoce el estado j , los valores de la opción call, $C_i(S_0, t)$, se muestran en el cuadro 4.12.

$i \rightarrow ?$	$C_i(S_0, t)$
1 $\rightarrow ?$	2.060494
2 $\rightarrow ?$	2.057941
3 $\rightarrow ?$	2.057739

Cuadro 4.12: Valores de la opción call sin conocimiento del estado j .

Si se conoce el estado j y se desconoce el estado i , los valores de la opción call, $C^j(S_0, t)$, se muestran en el cuadro 4.13.

$? \rightarrow j$	$C^j(S_0, t)$
? \rightarrow 1	2.062171
? \rightarrow 2	2.057392
? \rightarrow 3	2.057449

Cuadro 4.13: Valores de la opción call sin conocimiento del estado i .

Si no se conocen ninguno de los dos estados, se obtuvo

$$C(S_0, t) = 2.058378.$$

Caso r calculado con promedios en los estados

Si consideramos que las tasas de interés no necesariamente satisfacen un modelo de regresión lineal con los precios de las acciones, pero que si son afectadas por los cambios en los estados del proceso de Markov. Entonces, como se hizo para los datos de la variable observable, los datos de las tasas de interés $\{r_0, r_1, \dots, r_n\}$ se separan en m grupos G_1, G_2, \dots, G_m correspondientes a los m estados de la cadena de Markov, y con los datos que quedaron en cada grupo se obtiene un promedio, este promedio se utiliza como tasa de interés del grupo correspondiente.

Los promedios r_i son iguales a:

$$r_i = \begin{pmatrix} 0.033006 & 0.032930 & 0.032905 \end{pmatrix}$$

La matriz $[r_{ij}]$ se estimó con la fórmula $r_{ij} = r_{ji} = \frac{r_i + r_j}{2}$. De donde se obtiene la matriz

$$\mathbf{r} = \begin{pmatrix} 0.033006 & 0.032968 & 0.032956 \\ 0.032968 & 0.032930 & 0.032918 \\ 0.032956 & 0.032918 & 0.032905 \end{pmatrix}$$

Al usar la relación $\rho_{ij} = \log(1 + r_{ij})$, se obtiene la matriz $[\rho_{ij}]$.

$$[\rho_{ij}] = \begin{pmatrix} 0.032473 & 0.032436 & 0.032424 \\ 0.032436 & 0.032400 & 0.032388 \\ 0.032424 & 0.032388 & 0.032376 \end{pmatrix}$$

Se aplico nuevamente el modelo de Markov de Black-Scholes para $S(0) = 28.4$, $K = 32$ y $t = 6$ meses. Se obtuvieron los siguientes resultados.

Cuando se conoce el estado $X_0 = i$ y el estado $X_n = j$, $C_{ij}(S_0, t)$ se muestra en el cuadro 4.14.

$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$	$i \rightarrow j$	$C_{ij}(S_0, t)$
1 \rightarrow 1	2.065069	2 \rightarrow 1	2.060415	3 \rightarrow 1	2.070341
1 \rightarrow 2	2.059631	2 \rightarrow 2	2.054722	3 \rightarrow 2	2.056230
1 \rightarrow 3	2.066861	2 \rightarrow 3	2.057242	3 \rightarrow 3	2.051999

Cuadro 4.14: Valores de la opción call conociendo el estado i y el estado j.

Si se conoce el estado i y se desconoce el estado j , los valores de la opción call, $C_i(S_0, t)$, se muestran en el cuadro 4.15.

$i \rightarrow ?$	$C_i(S_0, t)$
1 $\rightarrow ?$	2.063521
2 $\rightarrow ?$	2.056844
3 $\rightarrow ?$	2.057443

Cuadro 4.15: Valores de la opción call sin conocimiento del estado j .

Si se conoce el estado j y se desconoce el estado i , los valores de la opción call, $C^j(S_0, t)$, se muestran en el cuadro 4.16.

$? \rightarrow j$	$C^j(S_0, t)$
? $\rightarrow 1$	2.065189
? $\rightarrow 2$	2.056295
? $\rightarrow 3$	2.057158

Cuadro 4.16: Valores de la opción call sin conocimiento del estado i .

Si no se conocen ninguno de los dos estados, se obtuvo

$$C(S_0, t) = 2.058423.$$

Con los mismos valores de $S(0)$, K y t , se aplicó el modelo de Black-Scholes, con lo cual se obtuvo:

$$C(S, t) = 3.81915.$$

Se consideró una tasa de interés anual de 3.2975% y la volatilidad histórica calculada fue de 9.7944%.

4.2.3. Comparación de precios futuros calculados con los métodos revisados

En esta sección se presenta los cálculos de $C(S_0, t)$ para distintos valores de t , calculados con el método de Black-Scholes, y con la extensión de Markov de Black-Scholes, para hacer la comparación, únicamente se consideró el caso donde se desconocen el estado inicial X_0 y el estado X_n , ya que es el único caso en que se obtiene un solo valor. Para los cálculos con la extensión de Markov se consideró la tasa de interés obtenida con regresión lineal y con promedios en los estados. El tiempo está medido en meses.

En el cuadro 4.17 se muestran los resultados obtenidos de la primera ventana de tiempo (523 datos). Por M-B-S nos referimos al modelo de Markov de Black-Scholes.

Tiempo	Black-Scholes	M-B-S (regresion)	M-B-S (promedios)
1 mes	0.192858	4.368e-9	5.304e-9
3 meses	1.327904	0.014837	0.016907
6 meses	3.207977	1.494789	1.556431
12 meses	6.693199	6.432508	6.542614
24 meses	12.23925	13.46978	13.6040
100 meses	23.70811	24.25829	24.28272

Cuadro 4.17: Calculo de $C(S_0, t)$ para t en meses, con los distintos metodos para los 523 datos.

En el cuadro 4.18 se muestran los resultados obtenidos de la segunda ventana de tiempo (65 datos).

Tiempo	Black-Scholes	M-B-S (regresion)	M-B-S (promedios)
1 mes	0.301300	4.49e-10	4.52e-10
3 meses	1.660593	0.058446	0.058441
6 meses	3.81915	2.058378	2.058423
12 meses	7.832234	6.709523	6.709601
24 meses	14.24668	13.6976	13.6977
100 meses	27.21759	27.14741	27.14744

Cuadro 4.18: Calculo de $C(S_0, t)$ para t en meses, con los distintos metodos para los 65 datos.

Se observa en ambos casos que conforme aumenta el tiempo, los precios estimados de las acciones tienden a ser similares. La razón de este resultado puede deberse a que conforme más tiempo pasa, los estados del proceso de Markov influyen menos en los precios de las acciones. La razón de este hecho creemos que debería ser investigado con más detalle para identificar sus causas.

Capítulo 5

Conclusiones

Dado que a los inversionistas les interesa conocer los posibles valores de una acción en un tiempo futuro, se han desarrollado métodos que permiten determinarlos, uno de estos métodos es la fórmula de Black-Scholes. En este sentido, Tuyen (2010) propone utilizar los modelos ocultos de Markov para determinar los parámetros asociados a los posibles estados de la extensión de Markov del modelo de Black-Scholes para conocer el precio de las acciones en un tiempo futuro. Este autor basó su investigación en considerar que el ambiente económico donde se desarrollan los productos financieros es cambiante, los cambios se pueden catalogar como “bueno”, “regular”, “malo”, etc., y que pasar de una situación a otra se puede modelar usando una cadena de Markov.

Es importante señalar que si el modelo propuesto por Janssen y Manca (2006) seguido por Tuyen (2010) se ajusta satisfactoriamente a la realidad, entonces con este modelo se debería estimar con mayor precisión los precios de las acciones.

En esta tesis se retomó la propuesta de Tuyen, y se hizo una aplicación de su modelo a datos históricos diarios de la bolsa mexicana de valores (BMV) (<http://mx.investing.com/equities/bolsa-mexicana-de-valores-a-historical-data>), y tasas de interés interbancaria de equilibrio (TIIE) a 28 días, del Banco de México. Consideramos que sería interesante probar el modelo con los precios de algunas acciones en particular y usando distintas tasas de interés, sin embargo no tuvimos disponibilidad de datos de este tipo.

Para sondear el efecto del tamaño de la muestra en los resultados finales, se probó el modelo con

dos ventanas de tiempo: primero se tomaron 523 datos, correspondientes del dos de enero del dos mil trece al treinta de enero del dos mil quince, se hizo el mismo análisis con 65 datos, del dieciocho de julio del dos mil catorce al diecisiete de octubre del dos mil catorce. El resultado más sobresaliente de esta comparación es que con 65 datos se identificaron tres estados, y con 523 datos se identificaron cinco estados. Estos resultados nos hacen creer que con más datos se determina mejor el número de estados del ambiente económico. Quedaría pendiente determinar si esta creencia es cierta o falsa.

Los valores calculados con la fórmula de Black-Scholes y la extensión de Markov de Black-Scholes, en el tiempo $t = 6$ meses, difieren. Una posibilidad es que en la extensión de Markov de Black-Scholes, la volatilidad y las tasas de interés se suponen variables aleatorias, pero también podría deberse a que la volatilidad en la fórmula de Black-Scholes es mayor, casi 10 veces mayor, que los mismos valores que la extensión de Markov de Black-Scholes. La diferencia en la volatilidad se debe a que los datos dentro de cada estado son más homogéneos, y su variación es menor que la variación con todos los datos juntos. Se espera en un trabajo futuro determinar la razón de este hecho observado.

Se hizo un estudio comparativo de la fórmula de Black-Scholes y la extensión de Markov de Black-Scholes, en el caso que se desconoce los estados en el tiempo inicial y en el tiempo n , se presentaron los cálculos de $C(S_0, t)$ para distintos valores de t , calculados con el método de Black-Scholes, únicamente se consideró el caso donde se desconocen el estado inicial X_0 y el estado en n , X_n , ya que es el único caso en que se obtiene un solo valor y es posible hacer la comparación. Para los cálculos con la extensión de Markov se consideró la tasa de interés obtenida con regresión lineal y con promedios en los estados. El tiempo está medido en meses. Se puede observar que conforme aumenta el valor de t , los valores obtenidos con los tres métodos parecen converger en un valor similar, queda pendiente investigar porque ocurre esto.

En resumen, la metodología de los modelos ocultos de Markov permite encontrar el número de estados más probables dados los datos observados. Se obtiene la secuencia más probable de esos estados y las estimaciones tanto del modelo de Markov, como de las distribuciones asociadas a las observaciones.

Programas en R

En esta sección se presentan los programas realizados en R y utilizados para resolver la valuación de opciones aplicando el algoritmo antes presentado, solo se presentan los programas que realizan los tres problemas básicos de los HMM.

Programa para encontrar las probabilidades forward-backward.

```
norm.HMM.lalphabeta <-function(x,m,mu,sigma,gamma,delta=NULL)
{
  if(is.null(delta))delta <-solve(t(diag(m)-gamma +1),rep(1,m))
  n <- length(x)
  lalpha <- lbeta <-matrix(NA ,m,n)
  allprobs <- matrix(0,n,m)
  for (i in 1:length(x))
  {
    for (j in 1:length(mu))
    {
      allprobs[i,j]=dnorm(x[i],mu[j],sigma[j])
    }
  }
  foo <- delta*allprobs[1,]
  sumfoo <- sum(foo)
  lscale <- log(sumfoo)
  foo <- foo/sumfoo
  lalpha [,1] <- log(foo)+lscale
```

```
for (i in 2:n)
  {
  foo <- foo %*% gamma*allprobs[i,]
  sumfoo <- sum(foo)
  lscale <- lscale+log(sumfoo)
  foo <- foo/sumfoo
  lalpha[,i] <- log(foo)+lscale
  }
lbeta[,n] <- rep(0,m)
foo <- rep (1/m,m)
lscale <- log(m)
for (i in (n-1) :1)
  {
  foo <- gamma %*%( allprobs[i+1,]*foo)
  lbeta[,i] <- log(foo)+lscale
  sumfoo <- sum(foo)
  foo <- foo/sumfoo
  lscale <- lscale+log(sumfoo)
  }
list(la=lalpha ,lb=lbeta)
}
```

Programa para resolver el algoritmo de viterbi.

```
norm.HMM.viterbi <-function(x,m,mu,sigma ,gamma ,delta=NULL ,...)
{
  if(is.null(delta))delta <-solve(t(diag(m)-gamma +1),rep(1,m))
  n <- length(x)
  normprobs <- matrix(0,n,m)
  for (i in 1:length(x))
  {
    for (j in 1:m)
```

```
    {
      normprobs[i,j]=dnorm(x[i],mu[j],sigma[j])
    }
  }
xi <- matrix(0,n,m)
foo <- delta*normprobs[1,]
xi[1,] <- foo/sum(foo)
for (i in 2:n)
  {
    foo <- apply(xi[i-1,]*gamma ,2,max)*normprobs[i,]
    xi[i,] <- foo/sum(foo)
  }
iv <-numeric(n)
iv[n] <-which.max(xi[n,])
for (i in (n-1) :1)
  iv[i] <- which.max(gamma[,iv[i+1]]* xi[i,])
iv
}
```

Programa para estimar los parámetros del HMM.

```
norm.HMM.EM <- function(x,m,mu,sigma,gamma,delta,maxiter=10000,tol=1e-6,...)
{
  mu.next <- mu
  sigma.next <- sigma
  gamma.next <- gamma
  delta.next <- delta
  for (iter in 1: maxiter)
    {
      lallprobs <- matrix(NA,length(x),length(mu))
      for (i in 1:length(x))
        {
```

```
for (j in 1:length(mu))
  {
    lallprobs[i,j]=log(dnorm(x[i],mu[j],sigma[j]))
  }
}
fb <- norm.HMM.lalphabet(x,m,mu,sigma,gamma,delta=delta)
la <- fb$la
lb <- fb$lb
n<-length(x)
c <- max(la[,n])
llk <- c+log(sum(exp(la[,n]-c)))
for (j in 1:m)
  {
    for (k in 1:m)
      {
        gamma.next[j,k] <- gamma[j,k]*sum(exp(la[j,1:(n-1)]+
        lallprobs [2:n,k]+lb[k,2:n]-llk))
      }
    mu.next[j] <- sum(exp(la[j,]+lb[j,]-llk)*x)/sum(exp(la[j,]+lb[j,]-llk))
    sigma.next[j] <- sqrt(sum(exp(la[j,]+lb[j,]-llk)*(x-(sum(exp(la[j,]+
        lb[j,]-llk)*x)/sum(exp(la[j,]+lb[j,]-llk))))^2)/sum(exp(la[j,]+
        lb[j,]-llk)))
  }
gamma.next <- gamma.next/apply(gamma.next ,1,sum)
delta.next <- exp(la[,1]+lb[,1]-llk)
delta.next <- delta.next/sum(delta.next)
crit <- sum(abs(mu -mu.next))+sum(abs(gamma -gamma.next))+
        sum(abs(delta -delta.next))+ sum(abs(sigma-sigma.next))
if(crit < tol)
  {
    np <- m*m+m-1
```

```
AIC <- -2*(llk -np)
BIC <- -2*llk+np*log(n)
return(list(mu=mu,sigma=sigma ,gamma=gamma ,delta=delta ,mllk=-llk ,
           AIC=AIC ,BIC=BIC))
}
mu <- mu.next
sigma <- sigma.next
gamma <- gamma.next
delta <- delta.next
}
print(paste ("No convergence after",maxiter ," iterations "))
NA
}
```

Para elaborar los programas de esta tesis se contacto a Luc Try Tuyen, autor del artículo en el cual basamos nuestra investigación. Luc Try Tuyen nos ayudó y aconsejó sobre la elaboración de los programas y la interpretación de los mismos. Sus consejos fueron muy útiles e interesantes.

Bibliografía

- [1] Edit for Dymarski P. *Hidden Markov Models, Theory and Applications*. INTECH Open Access Publisher. 2011.
- [2] Janssen J. and Manca R. *Semi-markov Risk Models for Finance, Insurance and Reliability*. Springer Verlag. First edition, 2010.
- [3] Janssen J. and Manca R. *Applied semi-markov Processes*. Springer Verlag. First edition, 2006.
- [4] Knab B., Schliep A., Steckmetz B. and Wichern B. *Model-based clustering with Hidden Markov Models and its application to financial time-series data*. Article. 2003.
- [5] Mirás C. *Matemáticas en Wall Street: la fórmula de Black-Scholes*. Notas. Facultad de economía.
- [6] Rincon L. *Reconocimiento de Firmas Mediante Modelos Ocultos de Markov*. Tesis de Maestría de la Universidad de los Andes. 2008.
- [7] Ross S. *Introduction Probability Models*. Academic press. Ninth edition, 2009.
- [8] S. Karlin and H.A.M. Taylor. *A First Course in Stochastic Processes*. Academic Press. Second Edition, 1975.
- [9] Shreve, S. *Stochastic Calculus and Finance*. Versión Preliminar, no publicada, 1997.

- [10] Tuyen L. *Markov Financial Model Using Hidden Markov Model*. Vietnam Academy of Science and Technology. Article. 2010.
-