



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**UNA INTRODUCCIÓN A LAS APLICACIONES DEL
MODELO DE MARKOV OCULTO GAUSSIANO EN
LA COMERCIALIZACIÓN DE ACTIVOS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

P R E S E N T A :

Alan Martínez Sánchez



**DIRECTOR DE TESIS:
Dr. Frank Patrick Murphy Hernández**

**Cd. Universitaria, CD. MX.
2022**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Al Dr. Frank Patrick Murphy Hernández
por todo el apoyo y confianza.

A mis sinodales,
la Dra. Lizbeth Naranjo Albarrán,
la Dra. Sofía Villers Gómez,
el Lic. Fernando Díaz López
y el Lic. Jimmy Hernández Morales,
por sus valiosos comentarios.

A mi familia y amigos
por acompañarme en todo momento.

Índice general

Introducción	7
Capítulo 1. Antecedentes	9
Capítulo 2. Modelos de Markov Ocultos	15
1. Introducción	15
2. Modelos de Markov Ocultos Discretos	18
Capítulo 3. Estimación de Parámetros	33
1. Algoritmo de Baum-Welch	34
Capítulo 4. Estimación de Meta-parámetros	39
1. Entropía	39
2. Criterios de Información	40
Capítulo 5. Modelos de Markov Ocultos Gaussianos	43
1. Algoritmos	45
2. Estimación de Parámetros	48
Capítulo 6. Modelos de Markov Ocultos en el Comercio de Activos	53
1. Selección del Modelo	53
2. Predicción	55
3. Comercio de Activos	55
4. Otros activos	59
Conclusiones	67
Apéndices	69
Código en R	69
Bibliografía	83

Introducción

El objetivo de un inversionista al adquirir un activo es maximizar su ganancia a través de las transacciones que realice con él, por lo que un desafío al que se enfrenta es conocer el momento oportuno para comprar o venderlo. Para ello, los modelos de predicción que se utilicen deben ajustarse a la naturaleza no estacionaria de las series de tiempo financieras, esto es, deben ser dependientes del tiempo y de estados anteriores.

En varios problemas financieros de este tipo los estados de un sistema pueden modelarse como una cadena de Markov en la cual cada estado depende del anterior en una forma no determinista. En 1966 Leonard Baum y Ted Petrie desarrollaron los fundamentos de lo que serían los modelos de Markov ocultos, en donde se asume que la probabilidad de lo que se observa está determinada por los estados del sistema, que no son visibles o conocidos. A partir de ello, múltiples investigadores han encontrado aplicación de los modelos de Markov ocultos en el análisis de tendencias económicas, en la predicción de precios de activos o para la estructurar portafolios de acuerdo con la variabilidad esperada en los retornos.

Este trabajo tiene el propósito de exponer los fundamentos de los modelos de Markov ocultos con el fin de introducir una aplicación del modelo de Markov oculto gaussiano en el intercambio de activos con base en la predicción del precio de cierre del índice S&P 500.

En el primer capítulo se presentan algunos conceptos básicos de los mercados de valores y sus antecedentes para establecer el marco dentro del cual se desarrollará la aplicación del modelo de Markov oculto gaussiano. Se hace también un breve recorrido por la historia del análisis financiero y los eventos que dieron paso al sentamiento de las bases de los modelos de Markov Ocultos.

En el segundo capítulo se introducen los conceptos básicos de los modelos de Markov ocultos en el caso discreto y algunas de sus aplicaciones en distintos campos. Se plantean también las tres preguntas principales asociadas a estos modelos y se

desarrollan los algoritmos necesarios para resolver las primeras dos de ellas.

En el tercer capítulo se desarrollan métodos para la estimación de parámetros a partir de la información que se tiene disponible. Se desarrolla un método de estimación frecuentista y un método apto para los modelos de Markov ocultos para resolver la tercera pregunta que quedaba pendiente.

En el cuarto capítulo se expone la necesidad de estimar también los meta-parámetros de un modelo y se establecen algunos criterios de información que permitan elegir los meta-parámetros más adecuados.

En el quinto capítulo se define el modelo de Markov oculto gaussiano y se desarrollan nuevamente los conceptos y algoritmos asociados vistos en el caso discreto.

En el sexto capítulo se presenta una aplicación de un modelo de Markov oculto gaussiano en una estrategia de comercialización de activos sobre el índice bursátil S&P 500 con base en la predicción de los precios de cierre mensuales. Se desarrolla la metodología a seguir desde la estimación de meta-parámetros y parámetros, la predicción de los precios y la estrategia de comercialización, así como se comparan los resultados con los que se obtienen a partir de algunas estrategias convencionales. Para ampliar los resultados obtenidos se replica el proceso anterior en los índices Nikkei 225 y FTSE 100.

Capítulo 1

Antecedentes

Un mercado de valores es un espacio físico o digital donde los inversionistas pueden intercambiar acciones de compañías públicas, y funge como intermediario en dichos intercambios haciendo coincidir las ofertas de acuerdo al precio más bajo que esté dispuesto a recibir el vendedor y el precio más alto que esté dispuesto a pagar el comprador por una acción. El propósito de los mercados de valores es ofrecer liquidez en un entorno de igualdad de condiciones con la finalidad de permitir la libre competencia entre toda clase de inversionistas.

Los mercados de valores representan una parte fundamental en la economía de un país en términos de gasto e inversión. Por un lado contribuyen a que los individuos generen una ganancia sobre sus ingresos cuando invierten en ellos y permiten a las compañías recaudar capital de manera más sencilla. Por el otro lado, también permiten a los gobiernos incrementar el gasto público a través de los impuestos que reciben de las corporaciones que cotizan en los mercados, con lo que incentivan la reinversión y aumentan la capacidad de empleo.

Cuando una compañía quiere recaudar capital puede conseguir un préstamo de una institución financiera o puede emitir acciones que representan la propiedad de una parte de la compañía. Para que se emitan acciones a través de la cuales sea posible invertir en una compañía ésta debe estar listada en una bolsa de valores y cumplir una serie de criterios para volverse pública. Al ofrecer acciones en vez de pedir un préstamo la compañía evita incurrir en deuda y en el pago de intereses. Esta es la función principal de los mercados de valores y la que hace de ellos la parte más importante en el crecimiento de la industria y comercio de un país.

La segunda función de los mercados de valores es la de actuar como punto de encuentro entre vendedores y compradores de las acciones listadas en la bolsa. Para poner su capital en una compañía los inversionistas tienen que pasar por el mercado de valores donde los corredores de bolsa se encargan de encontrar un punto de equilibrio entre las ofertas de compradores y vendedores y ejecutar las órdenes. Todo esto toma lugar en el mercado de valores y es la oferta y demanda la que determina

el precio de las acciones de una compañía. Gracias a los mercados de valores es posible comprar o vender una acción en cualquier momento. Esto hace de ellas un bien más líquido que, por ejemplo, los bienes raíces, teniendo como resultado que sea más atractivo para los inversionistas poner su dinero en este mercado.

Actualmente los mercados de valores existen en la mayoría de los países, no obstante, su origen se remonta a Amberes, en Bélgica, el principal centro de comercio internacional a finales del siglo XV. El ascenso de Amberes como centro económico fue resultado de la elección de Portugal de utilizarla como pueblo portuario para el comercio de especias en 1499. El primer envío de pimienta y otros bienes llegó de Lisboa a Amberes en 1501, y tan sólo algunos años más tarde se comenzaron a celebrar los primeros contratos por miles de toneladas de especias. Los portugueses usaban parcialmente los rendimientos de estos envíos en granos, pero también estaban en busca de plata y otros metales, los cuales Amberes podía ofrecer gracias a los caminos comerciales que la conectaban con el sur de Alemania y sus minas. Toda esta actividad atrajo eventualmente a comerciantes de todas partes de Europa.

Mercaderes y prestamistas se reunían en Amberes para tratar negocios, asuntos de gobierno o de deuda individual. Los comerciantes iban a comprar mercancías anticipando que su precio incrementaría y podrían generar una ganancia. De igual manera, prestamistas acudían a Amberes a realizar préstamos a altas tasas de interés para luego vender bonos respaldados por estas deudas. Mercados similares surgieron en países como Inglaterra, Italia o los Países Bajos, sin embargo, aunque sus sistemas e infraestructuras se asemejaban a los de los mercados de valores modernos carecían del elemento principal: acciones.

El primer mercado de valores moderno se formó en Amsterdam a partir de la creación de las compañías de las Indias Orientales. Cuando se descubrió que las Indias Orientales eran un paraíso de riquezas y oportunidades de comercio, los exploradores y comerciantes zarparon en masa, pero no muchos lograban regresar. Para mitigar el riesgo de quiebra causado por los barcos perdidos, los dueños de los navíos reunían fondos con la promesa de otorgar un porcentaje de las ganancias si la expedición resultaba exitosa. Estas asociaciones solían durar una única expedición, se disolvían al terminar ésta y una nueva se creaba para el siguiente viaje. Cuando las compañías de las Indias Orientales se formaron, cambiaron la forma en que los financiamientos se hacían emitiendo acciones que pagarían dividendos sobre todas las expediciones que las compañías emprendieran en el futuro dejando de lado el modelo de viajes individuales. En 1602, la Compañía Holandesa de las Indias Orientales se convirtió en la primera compañía intercambiada públicamente al listar acciones de la compañía en la Bolsa de Valores de Amsterdam, haciendo de esta oficialmente el primer mercado

de valores moderno.

En Estados Unidos, el título del primer mercado de valores se le atribuye a la Bolsa de Valores de Filadelfia, fundada en 1790. Dos años más tarde, 24 corredores de bolsa firmaron en Nueva York el Acuerdo de Buttonwood en donde establecieron las reglas para el intercambio de títulos, dando origen a la Bolsa de Valores de Nueva York (NYSE) con el Banco de Nueva York como su primera compañía listada. Su ubicación en el centro económico y comercial de Estados Unidos y la falta de competencia local le permitió convertirse en poco tiempo en el mayor mercado de valores a nivel global. Después de casi 200 años de dominar la economía, en 1971 entró a la competencia la Asociación Nacional de Corredores de Valores Automatizado de Cotización (NASDAQ), sostenida en un sistema de computadoras en las cuales los inversionistas realizan la compra y venta de las acciones, en vez de en una ubicación física. Al llevarse a cabo todas las operaciones de manera electrónica, se pudo reducir el margen entre los precios de compra y venta, dotándole una ventaja.

Una parte esencial de los mercados de valores son los índices bursátiles: grupos de acciones, derivados u otros instrumentos financieros que representan y miden el rendimiento de un mercado específico o de cierto sector, de manera que conforme el valor de los componentes del índice suba o baje el nivel del índice se mueve para reflejar esos cambios. Un cambio positivo en el valor del índice es señal de que los inversionistas ven valor en el mercado. Por el contrario, un cambio negativo significa la pérdida de confianza en la habilidad de las compañías de generar dinero. De este modo, los mercados de valores son representativos de la economía.

Standard & Poor's 500 (S&P 500) es un índice compuesto por 500 de las empresas más importantes que cotizan en los mercados estadounidenses, seleccionadas por su tamaño de mercado, liquidez y representación de un sector industrial. El valor de S&P 500 está calculado como un promedio ponderado de acuerdo a la capitalización de mercado de sus componentes, lo que significa que las compañías con mayor capitalización de mercado (número de acciones \times precio por acción) tiene mayor influencia en el índice, contrario a índices como el Dow Jones cuyo valor está ponderado de acuerdo al precio de las acciones, es decir, compañías con mayor precio por acción tienen mayor influencia en el índice. Debido a la forma en que se calcula y su gran cantidad de componentes se ha convertido en uno de los principales indicadores financieros entre los inversionistas.

El precio de una acción es el resultado de las expectativas a futuro que tienen los inversionistas con base en el análisis de cierta información. Gráficas, historiales de precios y reportes financieros son algunos de las fuentes que los inversionistas han

usado para analizar el valor de las compañías. Un acercamiento distinto es el análisis cuantitativo.

El análisis cuantitativo es el estudio estadístico de cómo se correlacionan ciertas variables con el comportamiento, siendo este componente estadístico la parte importante, ya que implica que el análisis se hace en muestras grandes permitiendo al analista probar ciertas tendencias en los mercados. De esta manera el analista tiene la habilidad de comparar distintos métodos de valuación y atributos, y optimizar el *trade-off*, el beneficio que se obtiene en un aspecto al sacrificar otro, que se obtiene por a través de la validación estadística. Este proceso puede automatizarse de forma que el analista pueda observar cientos de compañías simultáneamente y en tiempo real comparar su atractivo relativo antes de que los precios cambien.

Los primeros trabajos en esta línea se realizaban en computadores centrales y la única información disponible eran los precios y volumen de transacciones. Además, estos procesos eran lentos, caros y limitados a automatizar el análisis técnico. Se requirió de tres eventos principales para que el análisis cuantitativo llegara a ser lo que es actualmente. El primero de ellos tuvo que ver con la recolección de información. Servicios como el Sistema de Estimación de Corredores Institucionales comenzaron a recolectar estimaciones de rendimientos en intervalos regulares para que se pudieran ver sus cambios en el tiempo. Compustat, por su parte, reunía información financiera de miles de compañías y las hacía disponibles públicamente. Comenzó a haber más información que sólo precios. Todo esto en conjunto con el desarrollo que tuvieron las bases de datos que se iban haciendo cada vez más complejas y más precisas. El segundo evento fue que esta nueva información necesitaba poder manipularse fácilmente para elaborar modelos, validarlos y crear nuevas variables a partir de las existentes sin requerir de habilidades avanzadas de programación. El análisis estadístico se volvió una rama fundamental del software y su desarrollo permitió que los modelos se implementaran más rápido y el análisis obtenido fuera más valioso. Finalmente, el crecimiento acelerado del poder computacional fue uno de los factores que más influyó en el establecimiento del análisis cuantitativo. Con la introducción de la computadora personal se permitió a los analistas trabajar de manera individual e independiente a las computadoras centrales.

El replanteamiento de las bases de los mercados financieros ha guiado a la demanda urgente de modelos innovadores para entender el comportamiento de los instrumentos financieros. En las últimas décadas se han propuesto distintos sistemas con base en acercamientos clásicos tales como modelos autorregresivos, de medias móviles o suavizado exponencial. A pesar de su eficacia estos modelos presentan varias desventajas debido a su bajo desempeño frente a grandes cantidades de información

con alta complejidad intrínseca y alta dimensionalidad. Además, estos métodos no resultan aptos para entender las relaciones y dependencias ocultas que hay en los datos. Gracias a la automatización de varios procesos y el rápido desarrollo de la tecnología se han podido superar estos obstáculos en el uso de modelos basados en el aprendizaje de máquina.

El matemático ruso Andrey Markov dedicó su trabajo a estudiar secuencias de variables aleatorias mutuamente dependientes, y en 1906 sentó las bases de los procesos estocásticos al introducir por primera vez el concepto de cadena. Por varios años sus aportaciones no pudieron ser usadas prácticamente debido a la complejidad de los cálculos, pero con la llegada de las computadoras modernas esta limitación se redujo dando paso a gran cantidad de aplicaciones, entre ellas los modelos de Markov ocultos.

La teoría básica de los modelos de Markov ocultos fue establecida a finales de los años 1960s y a inicios de los 1970s por Leonard Baum y sus colaboradores a lo largo de varias publicaciones y fue desarrollado por distintos grupos durante los años 1970s. Desde sus primeras aplicaciones en el reconocimiento del habla los modelos de Markov ocultos se han ido popularizando en los últimos años. Esto puede atribuirse primero a que los modelos son ricos en estructura matemática y por lo tanto se pueden formar las bases teóricas para su uso en un gran rango de aplicaciones, y segundo a que aplicados adecuadamente se desempeñan bastante bien en la práctica.

Modelos de Markov Ocultos

1. Introducción

La predicción de eventos es esencial para ajustarse a los cambios que pueda haber en el futuro. Los gobiernos pueden encontrarla útil en la planeación de políticas públicas, las empresas para maximizar su utilidad o los servicios de transporte en el pronóstico de la cantidad de pasajeros. Estas predicciones pueden hacerse a partir de información histórica, la cual es estocástica y no determinista, por lo que es importante utilizar el modelo que mejor se ajuste.

En el modelado de sistemas que manejen información financiera es importante considerar un componente temporal. Este problema se ha resuelto con el diseño de una representación conveniente para la información financiera como series de tiempo: puntos de datos observados secuencialmente a lo largo del tiempo.

Los modelos de regresión lineal se utilizan principalmente para determinar el efecto de una o más variables sobre otra o calcular una tendencia, y a partir de ello establecer una relación lineal para predecir el valor de la variable dependiente en función de los valores de las demás. El modelo en su forma simple se suele representar por la ecuación

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X,$$

donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son coeficientes ajustados a partir de un conjunto de observaciones (X_i, Y_i) . Al aplicarlos sobre una serie de tiempo se toma al componente temporal como la variable independiente. Sin embargo, debe tenerse en consideración que puesto que la regresión lineal supone el cumplimiento de ciertas hipótesis sobre las observaciones (linealidad, homocedasticidad, independencia y normalidad) el modelo puede resultar poco efectivo en conjuntos de observaciones que no se ajusten a estas propiedades.

EJEMPLO 2.1. *Considérese el tipo de cambio USD/MXN durante la semana del 04 al 08 de Mayo de 2020. El modelo que más se ajusta es aquel con coeficientes $\hat{\beta}_0 = 24.4031$ y $\hat{\beta}_1 = -0.1179$, es decir,*

$$\hat{Y} = 24.4031 - 0.1179X.$$

En la tabla 2.1 puede verse una muestra de las observaciones y sus respectivos valores ajustados de acuerdo al modelo, y en la figura 2.1 se ilustra gráficamente.

Fecha	Observación	Valor Ajustado
04/05/2020	\$24.38	\$24.28
05/05/2020	\$23.83	\$24.16
06/05/2020	\$24.29	\$24.04
07/05/2020	\$24.05	\$23.93
08/05/2020	\$23.68	\$23.81

TABLA 2.1. Modelo de regresión lineal para el tipo de cambio USD/MXN en el período 04/05/2020 - 08/05/2020.

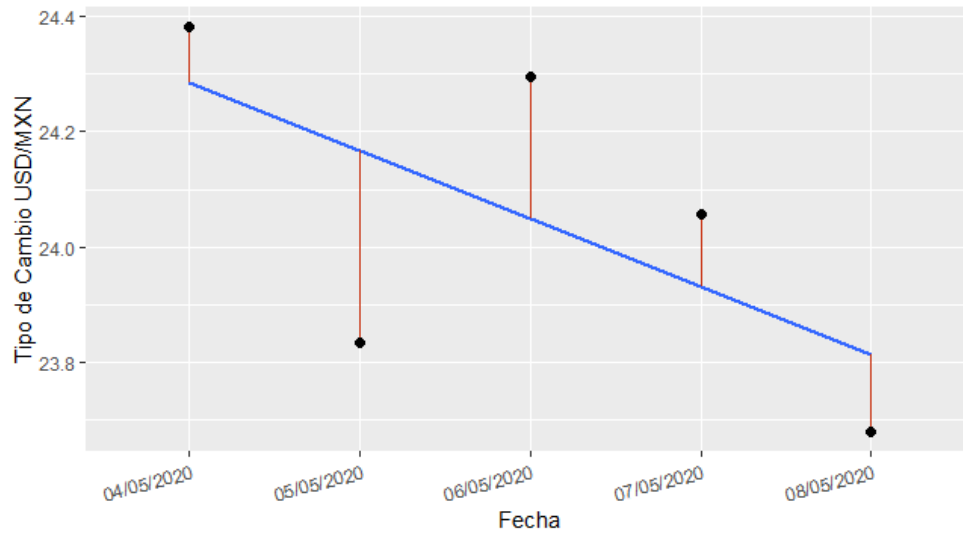


FIGURA 2.1. Modelo de regresión lineal para el tipo de cambio USD/MXN en el período 04/05/2020 - 08/05/2020.

Otra forma de hacer predicciones sobre series de tiempo es a través del análisis de sus componentes. Una serie de tiempo se puede ver como la suma de tres elementos

$$Y_t = T_t + S_t + R_t,$$

donde T_t es la tendencia, S_t es la estacionalidad y R_t un componente aleatorio. Cada uno de estos componentes determina un comportamiento en la serie de tiempo, y al estimarlos se puede ajustar un modelo predictivo para la serie.

EJEMPLO 2.2. *Considérese ahora el tipo de cambio USD/MXN durante el año 2019. Modelándola como una serie de tiempo con frecuencia mensual, puede descomponerse en las series representadas en la figura 2.2.*

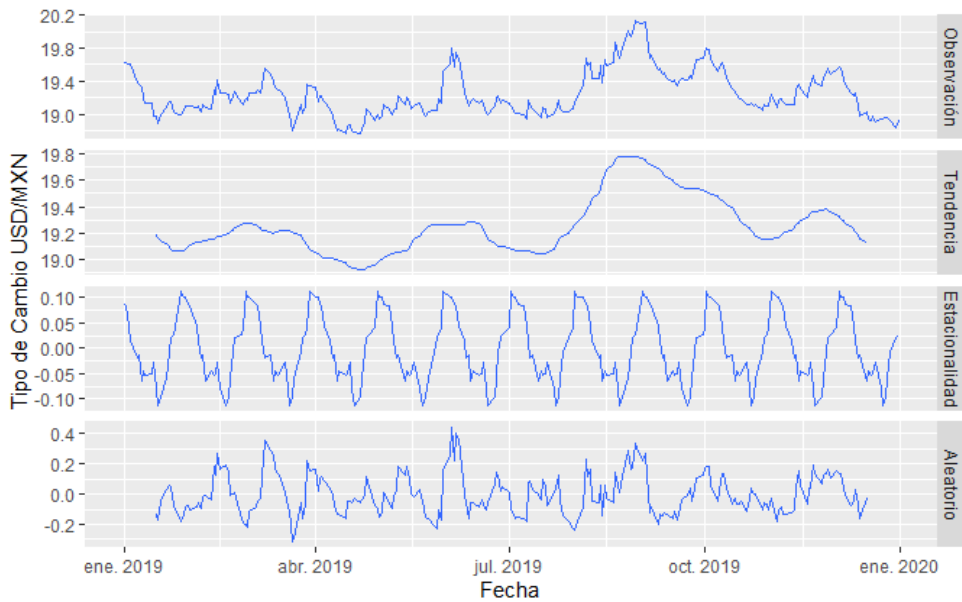


FIGURA 2.2. Descomposición de la serie de tiempo del tipo de cambio USD/MXN durante el año 2019.

Algunos estudios [CSB] han señalado la dificultad de estudiar con precisión las series de tiempo debido a sus patrones no lineales y no estacionarios, y han demostrado la insuficiencia de métodos tradicionales como los anteriores en capturar dichos patrones. Así mismo existen trabajos comparativos [RTS] que muestran cómo algunos métodos con base en aprendizaje de máquina se desempeñan mejor en el campo de las finanzas cuantitativas.

A continuación se profundizará en una clase de modelos de este tipo: los modelos de Markov ocultos.

2. Modelos de Markov Ocultos Discretos

DEFINICIÓN 2.1. *Un modelo de Markov oculto o HMM (Hidden Markov Model) es un proceso estocástico bivariado $\{Q_t, O_t\}_{t \geq 1}$ que consiste en una cadena de Markov $\{Q_t\}_{t \geq 1}$, llamada estado, la cual no es observada directamente, sino a través de las observaciones que emite otro proceso estocástico $\{O_t\}_{t \geq 1}$ ligado al primero de manera que el estado Q_t determina la distribución de la variable O_t . De esta manera, la observación O_t depende únicamente del estado actual Q_t , siendo independiente tanto de cualquier otro estado como de cualquier otra observación. Esta relación de dependencia se ejemplifica en la figura 2.3.*

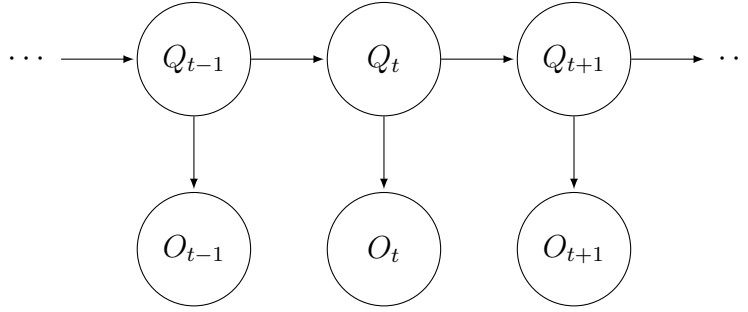


FIGURA 2.3. Estructura de dependencia de un modelo de Markov oculto.

En un modelo de Markov oculto $\{Q_t, O_t\}_{t \geq 1}$ cada una de las variables Q_t y O_t toma valores respectivamente en un espacio de estados S y en un espacio de símbolos V , y es a partir de ellos y de las probabilidades asociadas que los modelos se clasifican en discretos y continuos.

Un modelo de Markov oculto discreto puede definirse a partir de las probabilidades de transición entre estados $a_{ij} = P(Q_t = s_j | Q_{t-1} = s_i)$, las probabilidades condicionales de emitir cada símbolo $b_{ik} = P(O_t = v_k | Q_t = s_i)$ y la distribución inicial $p_i = P(Q_1 = s_i)$, donde los espacios de estados $S = \{s_i\}$ y de símbolos $V = \{v_k\}$ son conjuntos de cardinalidad finita N y M , respectivamente. Si la cantidad de símbolos para cada estado oculto es infinita (pero numerable), el símbolo v_k será omitido del modelo y la probabilidad condicional b_{ik} será escrita como $b_{ik} = b_i(O_t) = P(O_t | Q_t = s_i)$.

NOTACIÓN 2.1. *Los elementos básicos de un modelo de Markov oculto discreto son:*

Elemento	Notación
Longitud de las observaciones	T
Cantidad de estados ocultos	N
Cantidad de símbolos por estado oculto	M
Sucesión de observaciones	$O = \{O_t, t = 1, 2, \dots, T\}$
Sucesión de estados ocultos	$Q = \{Q_t, t = 1, 2, \dots, T\}$
Valores posibles de cada estado oculto	$S = \{s_i, i = 1, 2, \dots, N\}$
Símbolos posibles por estado observado	$V = \{v_k, k = 1, 2, \dots, M\}$
Matriz de transición	$A = (a_{ij}), a_{ij} = P(Q_t = s_j Q_{t-1} = s_i), i, j = 1, 2, \dots, N$
Distribución inicial de los estados ocultos	$p = (p_i), p_i = P(Q_1 = s_i), i = 1, 2, \dots, N$
Matriz de probabilidad de las observaciones	$B = (b_{ik}), b_{ik} = P(O_t = v_k Q_t = s_i), i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, M$

Un modelo de Markov oculto discreto es parametrizado por su matriz de transición A , su matriz de probabilidades de las observaciones B y su vector de probabilidades iniciales p , y puede caracterizarse como $\lambda \equiv \{A, B, p\}$.

EJEMPLO 2.3. [*St*] *Supóngase que se quiere determinar la temperatura anual promedio de una región a lo largo de una serie de años tan remota que los termómetros aún no habían sido inventados, por lo que se necesita recurrir a métodos indirectos.*

Para hacer el problema más sencillo, considérense únicamente dos temperaturas, cálido (H) y frío (C), y que existe evidencia que dicta que la probabilidad de pasar de un año cálido a otro igual es 0.7 y la probabilidad de que a un año frío le siga otro frío es 0.6. Supóngase también que estas probabilidades se mantienen durante toda la serie de años de interés. Esta información puede resumirse en

$$A = \begin{matrix} & \begin{matrix} H & C \end{matrix} \\ \begin{matrix} H \\ C \end{matrix} & \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \end{matrix}.$$

Ahora, supóngase que se prueba que existe una fuerte correlación entre la temperatura anual y el tamaño de los anillos de crecimiento de los árboles. Considérese únicamente tres posibles tamaños para los anillos, pequeño (S), mediano (M) y grande (L), cuya relación con la temperatura está dada por las probabilidades

$$B = \begin{matrix} & \begin{matrix} S & M & L \end{matrix} \\ \begin{matrix} H \\ C \end{matrix} & \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} \end{matrix}.$$

Entonces, la temperatura anual promedio puede verse como un proceso de Markov con estados $S = \{H, C\}$ y probabilidades de transición dadas por la matriz A . Como se mencionó anteriormente, no puede observarse directamente la temperatura, por lo que la serie de estados estará oculta. Sin embargo, lo que sí se observa es el tamaño de los anillos de los árboles, a través de los cuales, dada la relación dada por la matriz

B , será posible inferir la temperatura. Completando con una distribución inicial p a conveniencia puede definirse un modelo de Markov oculto.

EJEMPLO 2.4. [CMR] Los modelos de Markov ocultos pueden utilizarse para modelar la ocurrencia de errores de transmisión en algunos canales de comunicación digitales. Una señal digital es una serie de bits 0 y 1, y puede representarse como una sucesión $\{B_t\}_{t \geq 0}$ de Bernoullis independientes idénticamente distribuidas. El modelo más sencillo de cómo esta sucesión se distorsiona durante la transmisión es el Canal Simétrico Binario o BSC (Binary Symmetric Channel) en el cual se asume que cada bit puede ser volteado aleatoriamente por una sucesión independiente de errores $\{V_t\}_{t \geq 0}$ también Bernoulli con $P(V_t = 1) = b$. El bit recibido es entonces la suma módulo dos $O_t = B_t \oplus V_t$, donde el bit de salida O_t es igual al de entrada B_t si $V_t = 0$; de otro modo un error ocurre.

Si ahora se considera que el canal puede estar en un estado de error alto o en otro de error bajo, esto se puede modelar con una cadena de Markov $\{S_t\}_{t \geq 0}$ de dos estados con matriz de transición determinada por las probabilidades de cambiar de estado $a_0 = P(S_{t+1} = 1 | S_t = 0)$ (transición al estado de error alto) y $a_1 = P(S_{t+1} = 0 | S_t = 1)$ (transición al estado de error bajo). En cada estado el modelo actúa como el BSC con probabilidades de error $b_0 = P(O_t \neq B_t | S_t = 0)$ y $b_1 = P(O_t \neq B_t | S_t = 1)$, con $b_0 < b_1$. Recuperando la estructura de los modelos de Markov ocultos, defínase la sucesión de los estados ocultos como el proceso conjunto de los bits emitidos y la sucesión de los estados, $q_t = (B_t, S_t)$. Ésta es una cadena de Markov con matriz de transición:

$$A = \begin{matrix} & \begin{matrix} (0,0) & (0,1) & (1,0) & (1,1) \end{matrix} \\ \begin{matrix} (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{matrix} & \begin{pmatrix} (1-a_0)/2 & a_0/2 & (1-a_0)/2 & a_0/2 \\ a_1/2 & (1-a_1)/2 & a_1/2 & (1-a_1)/2 \\ (1-a_0)/2 & a_0/2 & (1-a_0)/2 & a_0/2 \\ a_1/2 & (1-a_1)/2 & a_1/2 & (1-a_1)/2 \end{pmatrix} \end{matrix}.$$

Ni el bit enviado B_t ni el régimen de error S_t son observados directamente, pero el modelo asegura que condicional a $\{q_t\}_{t \geq 0}$, las observaciones distribuyen Bernoulli con $P(O_t = v | B_t = v, S_t = s) = 1 - b_s$.

$$B = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{matrix} & \begin{pmatrix} 1-b_0 & b_0 \\ 1-b_1 & b_1 \\ b_0 & 1-b_0 \\ b_1 & 1-b_1 \end{pmatrix} \end{matrix}.$$

EJEMPLO 2.5. [CMR] *En Biología los modelos de Markov ocultos pueden emplearse para encontrar genes en el ADN. El ADN puede verse como una sucesión sobre un alfabeto de cuatro letras A, C, G y T (correspondientes a las cuatro bases que lo conforman: adenina, citosina, guanina y timina). Esta sucesión se divide en regiones codificantes, llamadas genes, y regiones no codificantes. Desafortunadamente, no existe manera fácil de distinguir las regiones codificantes de las no codificantes.*

Una forma de abordar este problema consiste en modelar la sucesión de bases $\{O_t\}_{t \geq 0} \in \{A, C, G, T\}$ a través de un modelo de Markov oculto de dos estados correspondientes a las regiones codificantes y no codificantes, respectivamente. En la forma más simple del modelo, la distribución condicional de la variable O_t dada q_t está parametrizada por el vector de probabilidades de observar A, C, G o T cuando se está en cada región. A pesar de su simplicidad, el modelo es efectivo al estimar los parámetros de este modelo en secuencias genómicas reales y determinar la sucesión estimada de estados q_t , es decir, el tipo de región.

Este modelo, suponiendo una distribución inicial uniforme, tiene parámetros

$$A = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 1 - a_0 & a_0 \\ a_1 & 1 - a_1 \end{pmatrix} \end{matrix}, \quad B = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} b_{0A} & b_{0C} & b_{0G} & b_{0T} \\ b_{1A} & b_{1C} & b_{1G} & b_{1T} \end{pmatrix} \end{matrix}$$

$$y \quad p = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ & \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \end{matrix}.$$

EJEMPLO 2.6. [CMR] *Jerome Dupuis estudió a una población de lagartijas y su movimiento entre tres zonas interconectadas. Para una cierta lagartija, la sucesión de zonas en las que ha estado puede modelarse con una cadena de Markov. En un momento dado, varias lagartijas no son observadas, por lo que este es un modelo de Markov parcialmente oculto. El experimento se realiza como sigue. A tiempo $t = 0$, una cantidad aleatoria de lagartijas son capturadas, marcadas y liberadas. Esto se repite para los tiempos siguientes $t = 1, \dots, n$ marcando a las nuevas lagartijas capturadas y registrando la zona en que se recapturó a las anteriores. Considerando que las lagartijas pueden morir o dejar la región de observación, un cuarto estado † es añadido a las otras tres zonas, el cual desde el punto de vista de la cadena de Markov subyacente es un estado absorbente, mientras que desde el punto de vista del HMM siempre está oculto.*

La probabilidad de cada lagartija de moverse de una zona a otra o salir del grupo de observación está determinada por la matriz

$$A = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & \dagger \\ \begin{array}{l} 1 \\ 2 \\ 3 \\ \dagger \end{array} & \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{1\dagger} \\ a_{21} & a_{22} & a_{23} & a_{2\dagger} \\ a_{31} & a_{32} & a_{33} & a_{3\dagger} \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \end{array}.$$

Las observaciones pueden entonces representarse por la sucesión $\{O_t^{(m)}\}_{0 \leq t \leq n}$ del registro de recapturas de cada lagartija m . Por ejemplo, se puede registrar

$$\{O_t^{(m)}\}_{0 \leq t \leq n} = (0, \dots, 0, 1, 1, 2, 0, 2, 0, 0, 3, 0, 0, 0, 1, 0, \dots, 0),$$

donde 0 significa que la lagartija no fue capturada en ese momento. A cada serie observada le corresponde otra parcialmente oculta $\{q_t^{(m)}\}_{0 \leq t \leq n}$. Por ejemplo, para la observación anterior,

$$\{q_t^{(m)}\}_{0 \leq t \leq n} = (1, \dots, 2, \mathbf{1}, \mathbf{1}, \mathbf{2}, 2, \mathbf{2}, 3, 2, \mathbf{3}, 3, 2, 2, \mathbf{1}, \dagger, \dots, \dagger),$$

indicando que la lagartija desapareció después de la última captura, y donde los valores obtenidos determinísticamente de las observaciones están resaltados. La probabilidad de cada lagartija de ser capturada o no dependiendo la zona en que se encuentre está dada por

$$B = \begin{array}{c} \begin{array}{cccc} & 0 & 1 & 2 & 3 \\ \begin{array}{l} 1 \\ 2 \\ 3 \\ \dagger \end{array} & \begin{pmatrix} 1 - b_1 & b_1 & 0 & 0 \\ 1 - b_2 & 0 & b_2 & 0 \\ 1 - b_3 & 0 & 0 & b_3 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \end{array}.$$

Dado un modelo de Markov oculto existen tres problemas esenciales que deben resolverse para que el modelo pueda tener aplicaciones reales. **[R]**

1. Dados los parámetros del modelo $\lambda = \{A, B, p\}$ y la sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$, ¿cómo se calcula la probabilidad de las observaciones $P(O | \lambda)$?
2. Dados los parámetros del modelo $\lambda = \{A, B, p\}$ y la sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$, ¿cómo estimar la mejor sucesión de estados $Q = \{Q_t, | t = 1, 2, \dots, T\}$?
3. Dada la sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$, ¿cómo estimar los parámetros $\lambda = \{A, B, p\}$ que maximicen $P(O | \lambda)$?

El primer problema es el problema de la evaluación del modelo, es decir, dados los parámetros del modelo y la sucesión de observaciones, cómo se calcula la probabilidad

de que la sucesión observada haya sido producida por el modelo. Para resolverlo, primero véase lo siguiente.

NOTACIÓN 2.2. Sea $p^{(t)}$ un vector de probabilidades tal que $p_i^{(t)} = P(Q_t = s_i)$.

LEMA 2.1. Sea $\{Q_t\}_{t \geq 1}$ una cadena de Markov homogénea. Entonces

$$p^{(t)} = p A^{t-1}.$$

PROPOSICIÓN 2.1. Sea $\lambda \equiv \{A, B, p\}$ un modelo de Markov oculto discreto. La probabilidad de que en el tiempo t se observe el símbolo v_k es

$$P(O_t = v_k) = p A^{t-1} B e_k,$$

donde e_k es un vector de dimensión N con la k -ésima entrada igual a 1 y el resto nulas.

DEMOSTRACIÓN. La probabilidad de observar v_k en el tiempo t depende del estado en que se encuentre la cadena de Markov en ese momento. Para cada estado s , la probabilidad de observar v_k es igual a la probabilidad de encontrarse en ese estado $P(Q_t = s)$ por la probabilidad condicional de emitir la observación $P(O_t = v_k | Q_t = s)$. Entonces, por el Teorema de Probabilidad Total,

$$\begin{aligned} P(O_t = v_k) &= \sum_{i=1}^N P(O_t = v_k | Q_t = s_i) P(Q_t = s_i) \\ &= \sum_{i=1}^N p_i^{(t)} b_{ik}. \end{aligned}$$

La expresión anterior puede verse matricialmente como

$$\begin{aligned} P(O_t = v_k) &= \begin{pmatrix} p_1^{(t)} & p_2^{(t)} & \dots & p_N^{(t)} \end{pmatrix} \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{Nk} \end{pmatrix} \\ &= p^{(t)} B e_k \\ &= p A^{t-1} B e_k. \end{aligned}$$

Por lo tanto,

$$P(O_t = v_k) = p A^{t-1} B e_k.$$

□

COROLARIO 2.1. Si la distribución inicial p es estacionaria, entonces

$$P(O_t = v_k) = p B e_k.$$

EJEMPLO 2.7. *Considérese el modelo de Markov oculto descrito en el ejemplo 2.3 con parámetros*

$$A = \begin{matrix} & \begin{matrix} H & C \end{matrix} \\ \begin{matrix} H \\ C \end{matrix} & \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} \end{matrix} \quad B = \begin{matrix} & \begin{matrix} S & M & L \end{matrix} \\ \begin{matrix} H \\ C \end{matrix} & \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} \end{matrix} \quad p = \begin{matrix} & \begin{matrix} H & C \end{matrix} \\ & \begin{pmatrix} 4/7 & 3/7 \end{pmatrix} \end{matrix}.$$

La distribución inicial p es estacionaria, por lo que las probabilidades de observar v_k a tiempo t son

$$\begin{aligned} P(O_t = v_k) &= (4/7 \quad 3/7) \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} e_k \\ &= (41/70 \quad 22/70 \quad 7/70) e_k. \end{aligned}$$

De esta manera,

$$P(O_t = S) = 41/70 \quad P(O_t = M) = 22/70 \quad P(O_t = L) = 7/70.$$

Ahora, para resolver el primer problema, el cálculo de la probabilidad $P(O | \lambda)$ puede realizarse a través del algoritmo Forward.

DEFINICIÓN 2.2. *Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. Defínase la función de probabilidad conjunta $\alpha_t(i) = P(O_1, O_2, \dots, O_t, Q_t = s_i | \lambda)$.*

PROPOSICIÓN 2.2 (Algoritmo Forward). *Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. La función $\alpha_t(i)$ puede calcularse recursivamente y la probabilidad $P(O | \lambda)$ se obtiene como sigue.*

1. *Para $i = 1, 2, \dots, N$, sea*

$$\alpha_1(i) = p_i b_i(O_1).$$

2. *Para $j = 1, 2, \dots, N$ y $t = 2, 3, \dots, T$, defínase*

$$\alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right] b_j(O_t).$$

3. *Entonces*

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

EJEMPLO 2.8. *Considérese el modelo de Markov oculto λ parametrizado por*

$$A = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{matrix}, \quad B = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} \end{matrix}, \quad p = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ & \begin{pmatrix} 1/3 & 2/3 \end{pmatrix} \end{matrix},$$

y la sucesión de observaciones $O = (O_1, O_2, O_3) = (1, 1, 1)$. Calcúlense las probabilidades $P(O_1 | \lambda)$, $P(O_1, O_2 | \lambda)$ y $P(O_1, O_2, O_3 | \lambda)$.

Con el algoritmo Forward, se pueden obtener las probabilidades parciales a tiempo t en el estado i siguientes:

$$\begin{aligned} \alpha_1(1) &= 1/3 \cdot 1/2 = 1/6. \\ \alpha_1(2) &= 2/3 \cdot 1 = 2/3. \end{aligned}$$

$$\begin{aligned} \alpha_2(1) &= (1/6 \cdot 1/2 + 2/3 \cdot 1/4) \cdot 1/2 = 1/8. \\ \alpha_2(2) &= (1/6 \cdot 1/2 + 2/3 \cdot 3/4) \cdot 1 = 7/12. \\ \alpha_3(1) &= (1/8 \cdot 1/2 + 7/12 \cdot 1/4) \cdot 1/2 = 5/48. \\ \alpha_3(2) &= (1/8 \cdot 1/2 + 7/12 \cdot 3/4) \cdot 1 = 1/2. \end{aligned}$$

Por lo tanto

$$\begin{aligned} P(O_1 | \lambda) &= 1/6 + 2/3 = 5/6. \\ P(O_1, O_2 | \lambda) &= 1/8 + 7/12 = 17/24. \\ P(O_1, O_2, O_3 | \lambda) &= 5/48 + 1/2 = 29/48. \end{aligned}$$

EJEMPLO 2.9. *Considérese de nuevo el modelo de Markov oculto λ descrito en el ejemplo 2.3 y la sucesión de observaciones $O = (O_1, O_2, O_3) = (S, M, L)$. Calcúlense las probabilidades $P(O_1 | \lambda)$, $P(O_1, O_2 | \lambda)$ y $P(O_1, O_2, O_3 | \lambda)$.*

Con el algoritmo Forward, se pueden obtener las probabilidades parciales a tiempo t en el estado i siguientes:

$$\begin{aligned} \alpha_1(H) &= 4/7 \cdot 1/2 = 2/7. \\ \alpha_1(C) &= 3/7 \cdot 7/10 = 3/10. \end{aligned}$$

$$\begin{aligned} \alpha_2(H) &= (2/7 \cdot 7/10 + 3/10 \cdot 2/5) \cdot 2/5 = 16/125. \\ \alpha_2(C) &= (2/7 \cdot 3/10 + 3/10 \cdot 3/5) \cdot 1/5 = 93/1750. \end{aligned}$$

$$\alpha_3(H) = (16/125 \cdot 7/10 + 93/1750 \cdot 2/5) \cdot 1/10 = 97/8750.$$

$$\alpha_3(C) = (16/125 \cdot 3/10 + 93/1750 \cdot 3/5) \cdot 1/10 = 123/17500.$$

Por lo tanto

$$P(O_1 | \lambda) = 2/7 + 3/10 = 41/70.$$

$$P(O_1, O_2 | \lambda) = 16/125 + 93/1750 = 317/1750.$$

$$P(O_1, O_2, O_3 | \lambda) = 97/8750 + 123/17500 = 317/17500.$$

De manera similar, la probabilidad $P(O | \lambda)$ puede ser calculada definiendo otra función $\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T, Q_t = s_i | \lambda)$.

DEFINICIÓN 2.3. Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. Defínase la función de probabilidad conjunta $\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T, Q_t = s_i | \lambda)$.

PROPOSICIÓN 2.3 (Algoritmo Backward). Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. La función $\beta_t(i)$ puede calcularse recursivamente y la probabilidad $P(O | \lambda)$ se obtiene como sigue.

1. Para $i = 1, 2, \dots, N$, sea

$$\beta_T(i) = 1.$$

2. Para $i = 1, 2, \dots, N$ y $t = 1, 2, \dots, T - 1$, defínase

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j).$$

3. Entonces

$$P(O | \lambda) = \sum_{i=1}^N p_i b_i(O_1) \beta_1(i).$$

EJEMPLO 2.10. Considérese de nuevo el modelo λ y la sucesión de observaciones O dados en el ejemplo 2.8 y calcúlese $P(O_1, O_2, O_3 | \lambda)$. Con el algoritmo Backward se obtienen

$$\beta_3(1) = \beta_3(2) = 1.$$

$$\begin{aligned}\beta_2(1) &= (1/2 \cdot 1/2 \cdot 1) + (1/2 \cdot 1 \cdot 1) = 3/4. \\ \beta_2(2) &= (1/4 \cdot 1/2 \cdot 1) + (3/4 \cdot 1 \cdot 1) = 7/8.\end{aligned}$$

$$\begin{aligned}\beta_1(1) &= (1/2 \cdot 1/2 \cdot 3/4) + (1/2 \cdot 1 \cdot 7/8) = 5/8. \\ \beta_1(2) &= (1/4 \cdot 1/2 \cdot 3/4) + (3/4 \cdot 1 \cdot 7/8) = 3/4.\end{aligned}$$

Por lo tanto

$$P(O_1, O_2, O_3 | \lambda) = (1/3 \cdot 1/2 \cdot 5/8) + (2/3 \cdot 1 \cdot 3/4) = 29/8,$$

que es lo que se obtuvo con el algoritmo Forward.

EJEMPLO 2.11. Siguiendo con el modelo de Markov oculto λ y la sucesión de observaciones O dados en el ejemplo 2.9, calcúlese $P(O_1, O_2, O_3 | \lambda)$.

Con el algoritmo Backward se obtienen

$$\beta_3(H) = \beta_3(C) = 1.$$

$$\beta_2(H) = (7/10 \cdot 1/10 \cdot 1) + (3/10 \cdot 1/10 \cdot 1) = 1/10.$$

$$\beta_2(C) = (2/5 \cdot 1/10 \cdot 1) + (3/5 \cdot 1/10 \cdot 1) = 1/10.$$

$$\beta_1(H) = (7/10 \cdot 2/5 \cdot 1/10) + (3/10 \cdot 1/5 \cdot 1/10) = 17/500.$$

$$\beta_1(C) = (2/5 \cdot 2/5 \cdot 1/10) + (3/5 \cdot 1/5 \cdot 1/10) = 7/250.$$

Por lo tanto,

$$P(O_1, O_2, O_3 | \lambda) = (4/7 \cdot 1/2 \cdot 17/500) + (3/7 \cdot 7/10 \cdot 7/250) = 317/17500,$$

que es lo que se obtuvo con el algoritmo Forward.

PROPOSICIÓN 2.4. Sea $\lambda \equiv \{A, B, p\}$ un modelo de Markov oculto discreto homogéneo. La observación esperada a tiempo t es

$$E(O_t) = p A^{t-1} B V,$$

donde $V = (v_k)$ es el vector de símbolos.

DEMOSTRACIÓN. El primer momento de O_t está dado por la ponderación de las esperanzas condicionales

$$E(O_t) = \sum_{i=1}^N P(q_t = S_i) E(O_t | q_t = s_i).$$

Esto puede verse matricialmente como

$$\begin{aligned}
E(O_t = v_k) &= \begin{pmatrix} p_1^{(t)} & p_2^{(t)} & \dots & p_N^{(t)} \end{pmatrix} \begin{pmatrix} E(O_t | q_t = s_1) \\ E(O_t | q_t = s_2) \\ \vdots \\ E(O_t | q_t = s_N) \end{pmatrix} \\
&= \begin{pmatrix} p_1^{(t)} & p_2^{(t)} & \dots & p_N^{(t)} \end{pmatrix} \begin{pmatrix} \sum_{k=1}^M b_{1k} v_k \\ \sum_{k=1}^M b_{2k} v_k \\ \vdots \\ \sum_{k=1}^M b_{Nk} v_k \end{pmatrix} \\
&= \begin{pmatrix} p_1^{(t)} & p_2^{(t)} & \dots & p_N^{(t)} \end{pmatrix} \begin{pmatrix} b_{11} & \dots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{N1} & \dots & b_{NM} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{pmatrix} \\
&= p^{(t)} B V \\
&= p A^{t-1} B V.
\end{aligned}$$

Por lo tanto,

$$E(O_t) = p A^{t-1} B V.$$

□

COROLARIO 2.2. Si la distribución inicial p es estacionaria, entonces

$$E(O_t) = p B V.$$

COROLARIO 2.3. Los siguientes momentos de O_t , y en general la esperanza de $g(O_t)$, son

$$E(g(O_t)) = p A^{t-1} B g(V),$$

donde $g(V) = (g_k)$ con $g_k = g(v_k)$.

EJEMPLO 2.12. Considérese el ejemplo 2.3. Dado que los símbolos S , M y L pertenecen a una escala ordinal, se les puede asignar valores que conserven el orden. Sea entonces $V = (S, M, L) = (1, 2, 3)$. La esperanza de O_t es

$$\begin{aligned}
E(O_t = v_k) &= \begin{pmatrix} 4/7 & 3/7 \end{pmatrix} \begin{pmatrix} .5 & .4 & .1 \\ .7 & .2 & .1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= 53/35 \approx 1.5142.
\end{aligned}$$

Esto significa que el tamaño esperado a tiempo t de los anillos de los árboles está en un punto medio entre S y M .

En general, los modelos de Markov ocultos no satisfacen la pérdida de memoria de las cadenas de Markov. Para probarlo considérese el modelo λ y las observaciones O del ejemplo 2.8.

Anteriormente se obtuvo que $P(O_1 = 1, O_2 = 1, O_3 = 1 | \lambda) = 29/48$. A su vez, puede obtenerse que $P(O_1 = 1, O_2 = 1, | \lambda) = P(O_2 = 1, O_3 = 1 | \lambda) = 17/24$ y $P(O_2 = 1 | \lambda) = 5/6$.

De esto se sigue que

$$\begin{aligned} P(O_3 = 1 | O_1 = 1, O_2 = 1, \lambda) &= \frac{P(O_1 = 1, O_2 = 1, O_3 = 1 | \lambda)}{P(O_1 = 1, O_2 = 1 | \lambda)} \\ &= \frac{29/48}{17/24} = \frac{29}{34} \end{aligned}$$

y

$$\begin{aligned} P(O_3 = 1 | O_2 = 1, \lambda) &= \frac{P(O_2 = 1, O_3 = 1 | \lambda)}{P(O_2 = 1 | \lambda)} \\ &= \frac{17/24}{5/6} = \frac{17}{20}. \end{aligned}$$

Por lo tanto, $P(O_3 = 1 | O_1 = 1, O_2 = 1, \lambda) \neq P(O_3 = 1 | O_2 = 1, \lambda)$, es decir, no cumple la propiedad de Markov.

Regresando a los problemas de los modelos de Markov ocultos que nos interesa resolver, es el segundo problema aquel que intenta descubrir la parte oculta del modelo, o lo que es lo mismo, el que intenta encontrar la sucesión de estados ocultos correcta. Cabe destacar que este problema no tiene una solución única, y que el decir que una sucesión es la correcta depende del criterio que se aplique. En nuestro caso, dados un modelo $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$ la mejor sucesión de estados $Q = \{Q_t, | t = 1, 2, \dots, T\}$, y para la cual se propondrá a continuación una forma de encontrarla, será aquella que maximice la probabilidad $P(Q | O, \lambda)$, es decir, el camino de estados más probable.

PROPOSICIÓN 2.5. *Sea $\lambda = \{A, B, p\}$ y $O = \{O_t | t = 1, 2, \dots, T\}$. Defínase la función $\delta_t(i) = \max_{Q_1, Q_2, \dots, Q_{t-1}} P(Q_1, Q_2, \dots, Q_t = s_i, O_1, O_2, \dots, O_t | \lambda)$. Entonces $\delta_{t+1}(i) = \max_j [\delta_t(j) a_{ji}] b_i(O_{t+1})$.*

PROPOSICIÓN 2.6 (Algoritmo de Viterbi). Sea $\lambda = \{A, B, p\}$ y $O = \{O_t \mid t = 1, 2, \dots, T\}$. La sucesión de estados más probable se obtiene como sigue.

1. Para $i = 1, 2, \dots, N$, sean

$$\delta_1(i) = p_i b_i(O_1);$$

$$\phi_1(i) = 0.$$

2. Para $i = 1, 2, \dots, N$ y $t = 2, 3, \dots, T$, defínase

$$\delta_t(i) = \max_j [\delta_{t-1}(j) a_{ji}] b_i(O_t);$$

$$\phi_t(i) = \arg \max_j [\delta_{t-1}(j) a_{ji}].$$

3. Entonces, para $t = 1, 2, \dots, T - 1$

$$Q_T^* = \arg \max_i [\delta_T(i)];$$

$$Q_t^* = \phi_{t+1}(Q_{t+1}^*).$$

EJEMPLO 2.13. Considérese una vez más el modelo λ y la sucesión de observaciones O dados en el ejemplo 2.8. Siguiendo el algoritmo de Viterbi para estimar la sucesión de estados ocultos más probable se tiene lo siguiente.

$$\delta_1(1) = 1/3 \cdot 1/2 = 1/6.$$

$$\delta_1(2) = 2/3 \cdot 1 = 2/3.$$

$$\phi_1(1) = \phi_1(2) = 0.$$

$$\delta_2(1) = \max[1/6 \cdot 1/2, \mathbf{2/3} \cdot \mathbf{1/4}] \cdot 1/2 = 1/12.$$

$$\delta_2(2) = \max[1/6 \cdot 1/2, \mathbf{2/3} \cdot \mathbf{3/4}] \cdot 1 = 1/2.$$

$$\phi_2(1) = \phi_2(2) = 2.$$

$$\delta_3(1) = \max[1/12 \cdot 1/2, \mathbf{1/2} \cdot \mathbf{1/4}] \cdot 1/2 = 1/16.$$

$$\delta_3(2) = \max[1/12 \cdot 1/2, \mathbf{1/2} \cdot \mathbf{3/4}] \cdot 1 = 3/8.$$

$$\phi_3(1) = \phi_3(2) = 2.$$

Por lo tanto

$$Q_3^* = 2.$$

$$Q_2^* = \phi_3(2) = 2.$$

$$Q_1^* = \phi_2(2) = 2.$$

Entonces, la sucesión más probable de estados ocultos es $Q^* = (2, 2, 2)$.

EJEMPLO 2.14. Considerando el modelo λ y la sucesión de observaciones O dados en el ejemplo 2.9 y siguiendo el algoritmo de Viterbi se tiene lo siguiente.

$$\delta_1(H) = 4/7 \cdot 1/2 = 2/7.$$

$$\delta_1(C) = 3/7 \cdot 7/10 = 3/10.$$

$$\phi_1(H) = \phi_1(C) = 0.$$

$$\delta_2(H) = \text{máx}[2/7 \cdot 7/10, 3/10 \cdot 2/5] \cdot 2/5 = 2/25.$$

$$\delta_2(C) = \text{máx}[2/7 \cdot 3/10, 3/10 \cdot 3/5] \cdot 1/5 = 9/250.$$

$$\phi_2(H) = H.$$

$$\phi_2(C) = C.$$

$$\delta_3(H) = \text{máx}[2/25 \cdot 7/10, 9/250 \cdot 2/5] \cdot 1/10 = 7/1250.$$

$$\delta_3(C) = \text{máx}[2/25 \cdot 3/10, 9/250 \cdot 3/5] \cdot 1/10 = 3/1250.$$

$$\phi_3(H) = H.$$

$$\phi_3(C) = H.$$

Por lo tanto

$$Q_3^* = H.$$

$$Q_2^* = \phi_3(H) = H.$$

$$Q_1^* = \phi_2(H) = H.$$

Entonces, la sucesión más probable de estados ocultos es $Q^* = (H, H, H)$.

Capítulo 3

Estimación de Parámetros

Para terminar con los problemas de los modelos de Markov ocultos planteados en el capítulo anterior hace falta resolver el tercero y último. Una vez dada la clase de modelo que se quiere utilizar para un proceso es esencial dar los parámetros que mejor lo representen. Dar los parámetros reales es una tarea complicada y en general imposible, sin embargo existen varios métodos mediante los cuales pueden estimarse, como en el ejemplo siguiente.

Considérese una serie de 50 observaciones de una cadena binaria con estados en el espacio $\{X, Y\}$

X X Y X Y X Y Y Y X X Y Y Y X X X Y Y Y Y Y Y Y X Y
X Y Y Y X X Y Y X X Y Y Y Y X Y X X X Y X Y X Y.

Una forma sencilla de calcular las probabilidades de transición es contar cuántas veces se pasó del estado i al estado j , representado por la matriz

$$\begin{matrix} & \begin{matrix} X & Y \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{pmatrix} 7 & 13 \\ 12 & 17 \end{pmatrix}, \end{matrix}$$

y calcular la proporción que representa del total de transiciones que parten del estado i . La matriz de transición es entonces

$$A^* = \begin{matrix} & \begin{matrix} X & Y \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{pmatrix} 7/20 & 13/20 \\ 12/29 & 17/29 \end{pmatrix}. \end{matrix}$$

De igual manera podría estimarse la distribución inicial como la proporción de veces que se observó cada estado

$$p^* = \begin{matrix} & \begin{matrix} X & Y \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{pmatrix} 2/5 & 3/5 \end{pmatrix}. \end{matrix}$$

Supóngase ahora que cada uno de los estados X y Y puede emitir símbolos 0 y 1, y que para la cadena de estados se observó la cadena de símbolos

1 1 1 0 0 1 0 0 1 1 0 0 0 1 1 0 1 0 0 0 1 0 0 1 0
0 0 1 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 1 1 0 0 0 0 0.

Aplicando un procedimiento similar puede obtenerse la matriz de conteo de veces que se observó el símbolo k estando en el estado i

$$\begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{pmatrix} 7 & 13 \\ 23 & 7 \end{pmatrix}, \end{matrix}$$

y obtener las proporciones para estimar la matriz de probabilidades

$$B^* = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{pmatrix} 7/20 & 13/20 \\ 23/30 & 7/30 \end{pmatrix}. \end{matrix}$$

1. Algoritmo de Baum-Welch

Nótese que en el método de estimación de parámetros anterior se supone el conocimiento de la cadena de estados ocultos, la cual en un modelo de Markov oculto no es observable. Afortunadamente, existe un método propio de los modelos de Markov ocultos para la estimación de sus parámetros a partir de una serie de observaciones: el algoritmo de Baum-Welch. Antes de desarrollarlo, defínanse las siguientes funciones.

PROPOSICIÓN 3.1. *Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. Defínase $\gamma_t(i) = P(Q_t = s_i | O, \lambda)$ como la probabilidad de estar en el estado i a tiempo t . Entonces*

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(O | \lambda)}.$$

PROPOSICIÓN 3.2. *Sea un modelo de Markov oculto con parámetros $\lambda = \{A, B, p\}$ y una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$. Defínase $\xi_t(i, j) = P(Q_t = s_i, Q_{t+1} = s_j | O, \lambda)$ como la probabilidad de estar en el estado i a tiempo t y en el estado j a tiempo $t + 1$. Entonces*

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)}.$$

De esto puede deducirse que

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

PROPOSICIÓN 3.3 (Algoritmo de Baum-Welch). *Sea una sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$ y los espacios $\{s_i\}$ y $\{v_k\}$ de N de estados y M de símbolos, respectivamente. Los parámetros del modelo de Markov oculto $\lambda = \{A, B, p\}$ pueden estimarse como sigue.*

1. Comiencese dando parámetros λ , la tolerancia τ y un número real Δ .
2. Repítase hasta que $\Delta < \tau$.
 - Calcúlese $P(O | \lambda)$ a través del algoritmo Forward.
 - Calcúlense los nuevos parámetros λ^* : para $i = 1, 2, \dots, N$.

$$p_i^* = \gamma_1(i).$$

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, j = 1, 2, \dots, N.$$

$$b_{ik}^* = \frac{\sum_{t=1, O_t=v_k}^T \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)}, k = 1, 2, \dots, M.$$

- Calcúlese $\Delta = |P(O | \lambda) - P(O | \lambda^*)|$.
- Actualícese $\lambda = \lambda^*$.

EJEMPLO 3.1. Considérese la sucesión de observaciones $O = (1, 1, 1)$, una tolerancia $\tau = 0.1$ y $\Delta = 1$, e iníciase el algoritmo con los parámetros $\lambda = \{A, B, p\}$ dados en el ejemplo 2.8.

Anteriormente se obtuvieron los valores de $\alpha_t(i)$ y $\beta_t(i)$ para cada $i = 1, 2, \dots, N$ y $t = 1, 2, \dots, T$, y de la probabilidad $P(O | \lambda)$.

A partir de ello, los valores de $\gamma_t(i)$ y $\xi_t(i, j)$ son

$$\gamma_1(1) = 5/29.$$

$$\gamma_1(2) = 24/29.$$

$$\gamma_2(1) = 9/58.$$

$$\gamma_2(2) = 49/58.$$

$$\gamma_3(1) = 5/29.$$

$$\gamma_3(2) = 24/29.$$

y

$$\xi_1(1, 1) = 3/58.$$

$$\xi_1(1, 2) = 7/58.$$

$$\xi_1(2, 1) = 3/29.$$

$$\xi_1(2, 2) = 21/29.$$

$$\xi_2(1, 1) = 3/58.$$

$$\xi_2(1, 2) = 3/29.$$

$$\xi_2(2, 1) = 7/58.$$

$$\xi_2(2, 2) = 21/29.$$

Entonces, para la primera iteración se tiene que

$$p^* = \begin{matrix} & 1 & 2 \\ \begin{matrix} 1 \\ 2 \end{matrix} & \left(\begin{array}{cc} 5/29 & 24/29 \end{array} \right) \end{matrix}.$$

$$A^* = \begin{matrix} & 1 & 2 \\ \begin{matrix} 1 \\ 2 \end{matrix} & \left(\begin{array}{cc} 6/19 & 13/19 \\ 13/97 & 84/97 \end{array} \right) \end{matrix}.$$

$$B^* = \begin{matrix} & 0 & 1 \\ \begin{matrix} 1 \\ 2 \end{matrix} & \left(\begin{array}{cc} 0 & 1 \\ 0 & 1 \end{array} \right) \end{matrix}.$$

$$\Delta = 19/48.$$

Como $\Delta > \tau$, se debe realizar una segunda iteración, ahora con los parámetros λ^* como entrada. Sin embargo, al finalizar la segunda iteración el valor de los parámetros estimados no cambia, por lo que resultará que $\Delta = 0$ dando fin al algoritmo.

Por lo tanto, dados los parámetros con que se inició el algoritmo, los parámetros estimados λ^* que maximizan la probabilidad de las observaciones O son

$$A^* = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 6/19 & 13/19 \\ 13/97 & 84/97 \end{pmatrix} \end{matrix}.$$

$$B^* = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \end{matrix}.$$

$$p^* = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 5/29 & 24/29 \end{pmatrix} \end{matrix}.$$

Estimación de Meta-parámetros

Como se mencionó anteriormente, un problema fundamental del análisis estadístico es la selección de un modelo que mejor describa a un conjunto de datos dado. Esto no se limita a determinar la clase de modelo o a estimar los parámetros correspondientes, sino que también incluye la estimación de su orden o dimensión, así como definir qué significa que un modelo sea el 'mejor'.

Este problema puede reducirse a cómo dar un criterio adecuado que permita comparar en una clase de modelos con distinto número de parámetros y elegir el que mejor aproxime dado un conjunto de datos. Además de criterios que representen cómo describe el modelo a los datos, es también importante tener en consideración criterios sobre la parsimonia de los parámetros (es decir, que modelos más sencillos pueden devolver mejores resultados), entre ellos el costo de medir los parámetros requeridos para implementar el modelo y la complejidad de éste [B].

Antes de introducir los criterios que se utilizarán para la selección del modelo, véase primero el concepto principal detrás de algunos de ellos.

1. Entropía

Shannon propone en *A Mathematical Theory of Communication* [Sha] una cantidad para medir la incertidumbre asociada para el caso de una fuente de información discreta. Dada una variable aleatoria X con eventos x_1, x_2, \dots, x_n y probabilidades asociadas p_1, p_2, \dots, p_n Shannon define la información de un evento como $I(x_i) = -\log(p_i)$, y a su medida como

$$H(X) = E(I(X)) = - \sum_{i=1}^n p_i \log(p_i),$$

donde $p_i \log(p_i) = 0$ si $p_i = 0$ en consistencia con $\lim_{x \rightarrow 0} x \log(x) = 0$, y, tomando el concepto de la Física Estadística, la llama entropía (también entropía de Shannon).

Con base en las ideas de Shannon, en su artículo *On Information Sufficiency* [KL] Kullback y Leibler derivan una forma de calcular la entropía de una distribución estimada g relativa a la distribución 'real' f de un proceso. La información (o divergencia) de Kullback-Leibler se define como

$$I(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x)} \right) dx.$$

Para que la divergencia de Kullback-Leibler esté bien definida (es decir, sea finita) es necesario que la distribución f sea absolutamente continua respecto a la distribución g , lo que implicará que cuando $g(x) = 0$ entonces $f(x) = 0$. De esta manera, $f(x) \log \left(\frac{f(x)}{g(x)} \right) = 0$ cuando $g(x) = 0$ o $f(x) = 0$. En un contexto de selección de modelos, aquel que se ajuste más al proceso real será aquel que minimice $I(f, g)$ sobre g .

2. Criterios de Información

Akaike encuentra una relación formal entre los resultados de Kullback y Leibler y la teoría de la verosimilitud, y en *A New Look at the Statistical Model Identification* [A] introduce su criterio como un estimador asintóticamente insesgado de la divergencia de Kullback-Leibler entre la distribución 'real' y el modelo candidato. El **criterio de información de Akaike** se define como

$$AIC = -2 \ln(L) + 2k,$$

donde L es la función de verosimilitud del modelo dadas las observaciones y k la cantidad de parámetros. Puede interpretarse [ABW] el primer término de la expresión como una medida del ajuste del modelo a la información y el segundo término como una penalización sobre la complejidad del modelo, consiguiendo de esta manera un equilibrio entre el ajuste del modelo y la simplicidad del mismo.

Schwarz propone en *Estimating the Dimension of a Model* [Sc] una solución distinta a la de Akaike. Partiendo de un enfoque bayesiano, Schwarz deriva el **criterio de información bayesiano** en términos de las probabilidades posteriores como

$$BIC = -2 \ln(L) + k \ln(M),$$

donde M es la cantidad de observaciones, y L y k representan lo mismo que en AIC. De esta manera, dado un conjunto de modelos candidatos este criterio se inclinará por elegir aquel que maximice la probabilidad posterior de que las observaciones hayan sido generadas por el modelo.

Debido a que BIC es un criterio consistente [W] suele preferirse sobre AIC que no lo es. Esto es que si se considera al modelo 'real' dentro de los modelos candidatos la probabilidad de seleccionar el modelo correcto con BIC tiende a 1 cuando el tamaño de las observaciones tiende a ∞ . Por otro lado, AIC es asintóticamente eficiente y BIC no en el sentido de que el error cuadrático medio del modelo elegido por el criterio es asintóticamente equivalente al menor posible de entre los modelos candidatos [Shi], por lo que cuando no se considera al modelo 'real' dentro de los candidatos AIC tiende a elegir de mejor manera el modelo que mejor lo aproxime. Así mismo, dado que el factor de penalización sobre los parámetros en AIC no depende del tamaño de las observaciones, se corre el riesgo de elegir un modelo sobreajustado, mientras que, al contrario, BIC penaliza más rigurosamente el exceso de parámetros corriendo el riesgo de elegir un modelo subajustado.

Propuesta inicialmente como una manera de determinar el orden de una autorregresión, Hannan y Quinn dan otra alternativa a AIC que, a diferencia de éste, es consistente gracias a la ley del logaritmo iterado. En su artículo *The Determination of the Order of an Autoregression* [HQ] el **criterio de información de Hannan-Quinn** se define como

$$HQIC = -2\ln(L) + k \ln(\ln(M)).$$

En *Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions* [B] Bozdogan alude a cuestionamientos sobre la consistencia de AIC y sobre la suficiencia del factor de penalización. Para corregir esto, sin violar los principios de Akaike, Bozdogan propone una extensión en el proceso de derivación de AIC obteniendo un criterio consistente y con una penalización más rigurosa como

$$CAIC = -2\ln(L) + k(\ln(M) + 1),$$

y, al ser una variante de AIC, lo llama el **criterio de consistencia de información de Akaike**.

Modelos de Markov Ocultos Gaussianos

En el capítulo 2 se desarrollaron los conceptos principales de los modelos de Markov ocultos en el caso discreto. Si ahora el espacio de símbolos del modelo fuera no numerable y las probabilidades de observarlas se distribuyen continuamente se tiene un modelo de Markov oculto continuo. En particular, si las probabilidades de las observaciones pertenecen a una distribución gaussiana, se le llamará modelo de Markov oculto gaussiano.

NOTACIÓN 5.1. *Los elementos básicos de un modelo de Markov oculto gaussiano son:*

Elemento	Notación
Longitud de las observaciones	T
Cantidad de estados ocultos	N
Sucesión de observaciones	$O = \{O_t, t = 1, 2, \dots, T\}$
Sucesión de estados ocultos	$Q = \{Q_t, t = 1, 2, \dots, T\}$
Valores posibles de cada estado oculto	$S = \{s_i, i = 1, 2, \dots, N\}$
Matriz de transición	$A = (a_{ij}), a_{ij} = P(Q_t = s_j Q_{t-1} = s_i), i, j = 1, 2, \dots, N$
Distribución inicial de los estados ocultos	$p = (p_i), p_i = P(Q_1 = s_i), i = 1, 2, \dots, N$
Vector de medias	$\mu = (\mu_i), \mu_i = E(O_t Q_t = s_i), i = 1, 2, \dots, N$
Vector de desviaciones estándar	$\sigma = (\sigma_i), \sigma_i^2 = Var(O_t Q_t = s_i), i = 1, 2, \dots, N$

Un modelo de Markov oculto gaussiano es parametrizado por su matriz de transición A , los vectores de medias μ y de desviaciones estándar σ de las distribuciones asociadas a cada estado, y su distribución inicial p , y puede caracterizarse como $\lambda \equiv \{A, \mu, \sigma, p\}$.

EJEMPLO 5.1. *Supóngase que se desea conocer la temperatura promedio diaria que hubo durante una serie de días en una región y que, para simplificar, sólo se requiere saber si la temperatura fue Baja (B), Media (M) o Alta (A). Además, se sabe que entre estas categorías hay probabilidades de transición*

$$A = \begin{matrix} & \begin{matrix} B & M & A \end{matrix} \\ \begin{matrix} B \\ M \\ A \end{matrix} & \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.15 & 0.7 & 0.15 \\ 0 & 0.6 & 0.4 \end{pmatrix} \end{matrix}.$$

Supóngase también que se tiene registro de la precipitación pluvial diaria en la región y que se sabe que la temperatura de un día determina los milímetros de precipitación 30 días después con una distribución normal condicional a la temperatura con parámetros

$$\mu = \begin{matrix} & \begin{matrix} B & M & A \end{matrix} \\ \begin{matrix} B \\ M \\ A \end{matrix} & \begin{pmatrix} 1 & 2.5 & 3.5 \end{pmatrix} \end{matrix} \quad y \quad \sigma = \begin{matrix} & \begin{matrix} B & M & A \end{matrix} \\ \begin{matrix} B \\ M \\ A \end{matrix} & \begin{pmatrix} 1 & 2 & 1.75 \end{pmatrix}.$$

Agregando una distribución inicial p adecuada, puede darse un modelo de Markov oculto gaussiano para determinar la temperatura promedio diaria a partir de las observaciones de la precipitación.

PROPOSICIÓN 5.1. Sea $\lambda \equiv \{A, \mu, \sigma, p\}$ un modelo de Markov oculto gaussiano. La función de densidad marginal de la variable O_t es una combinación convexa de N variables aleatorias Z_i donde $Z_i \sim \mathcal{N}(\mu_i, \sigma_i)$ y pesos $p_i^{(t)}$

$$f(O_t) = \sum_{i=1}^N p_i^{(t)} f_i(O_t).$$

PROPOSICIÓN 5.2. Sea $\lambda \equiv \{A, \mu, \sigma, p\}$ un modelo de Markov oculto gaussiano. La observación esperada a tiempo t es

$$E(O_t) = p A^{t-1} \mu.$$

DEMOSTRACIÓN. Por definición, el primer momento de O_t está dado por

$$\begin{aligned} E(O_t) &= \int O_t \sum_{i=1}^N p_i^{(t)} f_i(O_t) \\ &= \sum_{i=1}^N p_i^{(t)} \int O_t f_i(O_t) \\ &= \sum_{i=1}^N p_i^{(t)} E(Z_i) \\ &= \sum_{i=1}^N p_i^{(t)} \mu_i. \end{aligned}$$

Esto puede representarse como el producto de matrices

$$\begin{aligned} E(O_t) &= p^{(t)} \mu \\ &= p A^{t-1} \mu. \end{aligned}$$

□

COROLARIO 5.1. *Si la distribución inicial es estacionaria, entonces*

$$E(O_t) = p \mu.$$

EJEMPLO 5.2. *Considérese el modelo de Markov oculto $\lambda \equiv \{A, \mu, \sigma, p\}$ del ejemplo 5.1 y la distribución inicial $p = \begin{matrix} & B & M & A \\ \left(\begin{matrix} .12 & .24 & .64 \end{matrix} \right)$. Como la distribución inicial es estacionaria, entonces*

$$\begin{aligned} E(O_t) &= p \mu \\ &= (.12 \quad .24 \quad .64) \begin{pmatrix} 1 \\ 2.5 \\ 3.5 \end{pmatrix} \\ &= 2.96. \end{aligned}$$

1. Algoritmos

Anteriormente se describieron los tres problemas básicos para los modelos de Markov ocultos y se definieron algoritmos para resolverlos en el caso discreto. A excepción del algoritmo de Baum-Welch (el cual se discutirá más adelante) todos los algoritmos se definen de la misma manera y pueden implementarse sin mayor complicación en el caso continuo utilizando las funciones de probabilidad b_i correspondientes. En el caso de los modelos de Markov ocultos gaussianos esta probabilidad estará dada por $b_i(O_t) = \mathcal{N}(O_t; \mu_i, \sigma_i)$.

EJEMPLO 5.3 (Algoritmo Forward). *Continuando con el modelo λ del ejemplo 5.2 considérese la sucesión de observaciones $O = (1, 2.5, 3.5)$. Calcúlense las probabilidades $P(O_1 | \lambda)$, $P(O_1, O_2 | \lambda)$ y $P(O_1, O_2, O_3 | \lambda)$.*

Para cada observación O_t las probabilidades condicionales al estado i son

$$\begin{aligned} b_B(O_1) &= \mathcal{N}(1; 1, 1) = 0.3989. \\ b_B(O_2) &= \mathcal{N}(2.5; 1, 1) = 0.1295. \\ b_B(O_3) &= \mathcal{N}(3.5; 1, 1) = 0.0175. \end{aligned}$$

$$b_M(O_1) = \mathcal{N}(1; 2.5, 2) = 0.1505.$$

$$b_M(O_2) = \mathcal{N}(2.5; 2.5, 2) = 0.1994.$$

$$b_M(O_3) = \mathcal{N}(3.5; 2.5, 2) = 0.176.$$

$$b_A(O_1) = \mathcal{N}(1; 3.5, 1.75) = 0.0821.$$

$$b_A(O_2) = \mathcal{N}(2.5; 3.5, 1.75) = 0.1936.$$

$$b_A(O_3) = \mathcal{N}(3.5; 3.5, 1.75) = 0.2279.$$

Siguiendo el algoritmo Forward se tiene

$$\alpha_1(B) = 0.12 \cdot 0.3989 = 0.0478.$$

$$\alpha_1(M) = 0.24 \cdot 0.1505 = 0.0361.$$

$$\alpha_1(A) = 0.64 \cdot 0.0821 = 0.0525.$$

$$\alpha_2(B) = (0.0478 \cdot 0.7 + 0.0361 \cdot 0.15 + 0.0525 \cdot 0) \cdot 0.1295 = 0.005.$$

$$\alpha_2(M) = (0.0478 \cdot 0.3 + 0.0361 \cdot 0.7 + 0.0525 \cdot 0.6) \cdot 0.1994 = 0.0142.$$

$$\alpha_2(A) = (0.0478 \cdot 0 + 0.0361 \cdot 0.15 + 0.0525 \cdot 0.4) \cdot 0.1936 = 0.0051.$$

$$\alpha_3(B) = (0.005 \cdot 0.7 + 0.0142 \cdot 0.15 + 0.0051 \cdot 0) \cdot 0.0175 = 0.0001.$$

$$\alpha_3(M) = (0.005 \cdot 0.3 + 0.0142 \cdot 0.7 + 0.0051 \cdot 0.6) \cdot 0.176 = 0.0026.$$

$$\alpha_3(A) = (0.005 \cdot 0 + 0.0142 \cdot 0.15 + 0.0051 \cdot 0.4) \cdot 0.2279 = 0.0009.$$

Por lo tanto

$$P(O_1 | \lambda) = 0.0479 + 0.0361 + 0.0525 = 0.1365.$$

$$P(O_1, O_2 | \lambda) = 0.005 + 0.0142 + 0.0051 = 0.0243.$$

$$P(O_1, O_2, O_3 | \lambda) = 0.0001 + 0.0026 + 0.0009 = 0.0036.$$

EJEMPLO 5.4 (Algoritmo Backward). *Ahora calcúlese la probabilidad $P(O_1, O_2, O_3 | \lambda)$ con el algoritmo Backward.*

$$\beta_3(B) = \beta_3(M) = \beta_3(A) = 1.$$

$$\beta_2(B) = (0.7 \cdot 0.0175 \cdot 1) + (0.3 \cdot 0.176 \cdot 1) + (0 \cdot 0.2279 \cdot 1) = 0.0651.$$

$$\beta_2(M) = (0.15 \cdot 0.0175 \cdot 1) + (0.7 \cdot 0.176 \cdot 1) + (0.15 \cdot 0.2279 \cdot 1) = 0.1601.$$

$$\beta_2(A) = (0 \cdot 0.0175 \cdot 1) + (0.6 \cdot 0.176 \cdot 1) + (0.4 \cdot 0.2279 \cdot 1) = 0.1968.$$

$$\beta_1(B) = (0.7 \cdot 0.1295 \cdot 0.0651) + (0.3 \cdot 0.1994 \cdot 0.1601) + (0 \cdot 0.1936 \cdot 0.1968) = 0.0155.$$

$$\beta_1(M) = (0.15 \cdot 0.1295 \cdot 0.0651) + (0.7 \cdot 0.1994 \cdot 0.1601) + (0.15 \cdot 0.1936 \cdot 0.1968) = 0.0293.$$

$$\beta_1(A) = (0 \cdot 0.1295 \cdot 0.0651) + (0.6 \cdot 0.1994 \cdot 0.1601) + (0.4 \cdot 0.1936 \cdot 0.1968) = 0.0344.$$

Por lo tanto

$$\begin{aligned} P(O_1, O_2, O_3 | \lambda) &= (0.12 \cdot 0.3989 \cdot 0.0155) + (0.24 \cdot 0.1505 \cdot 0.0293) + (0.64 \cdot 0.0821 \cdot 0.0344) \\ &= 0.0036. \end{aligned}$$

EJEMPLO 5.5 (Algoritmo de Viterbi). *Siguiendo con el mismo modelo λ y las mismas observaciones $O = (1, 2.5, 3.5)$, estímese ahora la cadena de estados ocultos asociada.*

$$\delta_1(B) = 0.12 \cdot 0.3989 = 0.0478.$$

$$\delta_1(M) = 0.24 \cdot 0.1505 = 0.0361.$$

$$\delta_1(A) = 0.64 \cdot 0.0821 = 0.0525.$$

$$\phi_1(B) = \phi_1(M) = \phi_1(A) = 0.$$

$$\delta_2(B) = \text{máx}[0.0478 \cdot 0.7, 0.0361 \cdot 0.15, 0.0525 \cdot 0] \cdot 0.1295 = 0.0043.$$

$$\delta_2(M) = \text{máx}[0.0478 \cdot 0.3, 0.0361 \cdot 0.7, 0.0525 \cdot 0.6] \cdot 0.1994 = 0.0063.$$

$$\delta_2(A) = \text{máx}[0.0478 \cdot 0, 0.0361 \cdot 0.15, 0.0525 \cdot 0.4] \cdot 0.1936 = 0.0041.$$

$$\phi_2(B) = B.$$

$$\phi_2(M) = A.$$

$$\phi_2(A) = A.$$

$$\delta_3(B) = \text{máx}[0.0043 \cdot 0.7, 0.0063 \cdot 0.15, 0.0041 \cdot 0] \cdot 0.0175 = 0.00005.$$

$$\delta_3(M) = \text{máx}[0.0043 \cdot 0.3, 0.0063 \cdot 0.7, 0.0041 \cdot 0.6] \cdot 0.176 = 0.00077.$$

$$\delta_3(A) = \text{máx}[0.0043 \cdot 0, 0.0063 \cdot 0.15, 0.0041 \cdot 0.4] \cdot 0.2279 = 0.00037.$$

$$\phi_3(B) = B.$$

$$\phi_3(M) = M.$$

$$\phi_3(A) = A.$$

Entonces

$$\begin{aligned} Q_3^* &= M. \\ Q_2^* &= \phi_3(M) = M. \\ Q_1^* &= \phi_2(M) = A. \end{aligned}$$

Por lo tanto, la cadena más probable de estados ocultos es $Q^* = (A, M, M)$.

2. Estimación de Parámetros

En el capítulo 3 se describió un método para estimar los parámetros $\lambda \equiv \{A, B, p\}$ de un modelo de Markov oculto discreto de manera frecuentista. En el caso en que el modelo sea un modelo de Markov oculto gaussiano $\lambda = \{A, \mu, \sigma, p\}$ los parámetros A y p pueden estimarse de la misma manera, y μ y σ pueden estimarse como los vectores de medias y desviaciones estándar muestrales, respectivamente, de las observaciones agrupadas por el estado que las emitió.

Claramente la serie de observaciones binaria que se utilizó para ejemplificar el proceso de estimación de la matriz B en el capítulo 3 no pertenece a una distribución gaussiana, pero para fines ilustrativos supóngase que sí. En ese capítulo se contaron las veces que se observó cada símbolo k estando en cada estado i . Usando esos valores pueden estimarse

$$\begin{aligned} \mu_X^* &= \frac{7 \cdot 0 + 13 \cdot 1}{20} = 0.65, \\ \mu_Y^* &= \frac{23 \cdot 0 + 7 \cdot 1}{30} = 0.233, \\ \sigma_X^* &= \sqrt{\frac{7 \cdot (0 - 0.65)^2 + 13 \cdot (1 - 0.65)^2}{20 - 1}} = .489, \\ \sigma_Y^* &= \sqrt{\frac{23 \cdot (0 - 0.233)^2 + 7 \cdot (1 - 0.233)^2}{30 - 1}} = .43. \end{aligned}$$

Entonces los vectores de medias y desviaciones estimados serían

$$\begin{aligned} \mu^* &= \begin{pmatrix} X & Y \\ .65 & .233 \end{pmatrix} \\ \sigma^* &= \begin{pmatrix} X & Y \\ .489 & .43 \end{pmatrix}. \end{aligned}$$

EJEMPLO 5.6. *Supóngase que no se conocen los parámetros $\lambda \equiv \{A, \mu, \sigma, p\}$ del modelo descrito en el ejemplo 5.1 y que para estimarlos se decide hacerlo con los datos de temperatura y precipitación de otra región. Considérense entonces los*

registros diarios de temperatura máxima y precipitación de la Ciudad de México durante el año 2019 y clasifíquese la temperatura como Baja (B), Media (M) o Alta (A).

La matriz de conteo de veces que se pasó del estado i al estado j es

$$\begin{matrix} & B & M & A \\ \begin{matrix} B \\ M \\ A \end{matrix} & \begin{pmatrix} 17 & 12 & 0 \\ 11 & 208 & 12 \\ 0 & 12 & 92 \end{pmatrix} \end{matrix}.$$

y la matriz de transición es entonces

$$A^* = \begin{matrix} & B & M & A \\ \begin{matrix} B \\ M \\ A \end{matrix} & \begin{pmatrix} 17/29 & 12/29 & 0 \\ 11/231 & 208/231 & 12/231 \\ 0 & 3/26 & 23/26 \end{pmatrix} \end{matrix}.$$

Ahora, para estimar las medias y desviaciones estándar correspondientes a cada estado, clasifíquese primero las precipitaciones de acuerdo a la temperatura que hubo 30 días antes. Una vez separados los conjuntos de datos, puede procederse a estimar los parámetros μ^* y σ^* . Estos parámetros resultan ser

$$\begin{aligned} \mu^* &= \begin{pmatrix} B & M & A \\ .1625 & 2.5669 & 2.8903 \end{pmatrix} \\ \sigma^* &= \begin{pmatrix} B & M & A \\ .1784 & 2.5634 & 2.2345 \end{pmatrix}. \end{aligned}$$

Como se mencionó en el capítulo 3, el procedimiento anterior sólo es útil cuando se conoce la cadena de estados ocultos y para solucionar ese problema se definió el algoritmo de Baum-Welch para el caso discreto. Este algoritmo también es aplicable en el caso continuo, pero a diferencia de los algoritmos Forward, Backward y de Viterbi no se define completamente de la misma manera. Esto es obvio dado que en el modelo continuo no se tiene el parámetro B , y en su lugar, en el caso de un modelo gaussiano, se tienen los parámetros μ y σ .

Entonces, si las probabilidades b_{ik} son gaussianas se utilizarán las siguientes fórmulas para actualizar los parámetros del modelo $\lambda = \{A, \mu, \sigma, p\}$:

$$\mu_i^* = \frac{\sum_{t=1}^{T-1} \gamma_t(i) O_t}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\sigma_i^* = \frac{\sum_{t=1}^T \gamma_t(i) (O_t - \mu_i) (O_t - \mu_i)'}{\sum_{t=1}^T \gamma_t(i)}.$$

EJEMPLO 5.7. *Considérese la sucesión de observaciones $O = (1, 2.5, 3.5)$, una tolerancia $\tau = .1$ y $\Delta = 1$, e iníciase el algoritmo con los parámetros $\lambda \equiv \{A, \mu, \sigma, p\}$ dados en el ejemplo 5.2.*

Anteriormente se obtuvieron los valores de $\alpha_t(i)$ y $\beta_t(i)$ para cada $i = B, M, A$ y $t = 1, 2, 3$, y la probabilidad $P(O | \lambda)$.

A partir de ello, los valores de $\gamma_t(i)$ y $\xi_t(i, j)$ son

$$\begin{array}{lll} \gamma_1(B) = 0.2053. & \gamma_2(B) = 0.0909. & \gamma_3(B) = 0.0275. \\ \gamma_1(M) = 0.2936. & \gamma_2(M) = 0.6298. & \gamma_3(M) = 0.7085. \\ \gamma_1(A) = 0.5011. & \gamma_2(A) = 0.2793. & \gamma_3(A) = 0.264. \end{array}$$

y

$$\begin{array}{ll} \xi_1(B, B) = 0.0783. & \xi_2(B, B) = 0.017 . \\ \xi_1(B, M) = 0.127 . & \xi_2(B, M) = 0.0733. \\ \xi_1(B, A) = 0 . & \xi_2(B, A) = 0 . \\ \xi_1(M, B) = 0.0126. & \xi_2(M, B) = 0.0104. \\ \xi_1(M, M) = 0.2237. & \xi_2(M, M) = 0.4859. \\ \xi_1(M, A) = 0.0572. & \xi_2(M, A) = 0.1348. \\ \xi_1(A, B) = 0 . & \xi_2(A, B) = 0 . \\ \xi_1(A, M) = 0.2791. & \xi_2(A, M) = 0.1499. \\ \xi_1(A, A) = 0.2221. & \xi_2(A, A) = 0.1294. \end{array}$$

Entonces, para la primera iteración se tiene que

$$p^* = \begin{pmatrix} B & M & A \\ 0.2053 & 0.2936 & 0.5011 \end{pmatrix},$$

$$A^* = \begin{matrix} & B & M & A \\ B & \begin{pmatrix} 0.3221 & 0.6779 & 0 \\ 0.0249 & 0.7674 & 0.2077 \\ 0 & 0.5496 & 0.4504 \end{pmatrix} \\ M & \\ A & \end{matrix},$$

$$\mu^* = \begin{pmatrix} B & M & A \\ 1.4604 & 2.0231 & 1.5368 \end{pmatrix}$$

$$\sigma^* = \begin{pmatrix} B & M & A \\ 0.8895 & 1.1059 & 1.1664 \end{pmatrix},$$

$$\Delta = 0.0078.$$

Como $\Delta < \tau$, no se necesita de una segunda iteración y el algoritmo termina.

Por lo tanto, dados los parámetros con que se inició el algoritmo, los parámetros estimados λ^* que maximizan la probabilidad de las observaciones O son

$$A^* = \begin{matrix} & B & M & A \\ B & \begin{pmatrix} 0.3221 & 0.6779 & 0 \\ 0.0249 & 0.7674 & 0.2077 \\ 0 & 0.5496 & 0.4504 \end{pmatrix} \\ M & \\ A & \end{matrix},$$

$$\mu^* = \begin{pmatrix} B & M & A \\ 1.4604 & 2.0231 & 1.5368 \end{pmatrix},$$

$$\sigma^* = \begin{pmatrix} B & M & A \\ 0.8895 & 1.1059 & 1.1664 \end{pmatrix},$$

$$p^* = \begin{pmatrix} B & M & A \\ 0.2053 & 0.2936 & 0.5011 \end{pmatrix}.$$

Modelos de Markov Ocultos en el Comercio de Activos

Una vez establecidos los conceptos principales de los modelos de Markov ocultos se procederá a desarrollar una aplicación de ellos en la comercialización de activos.

En este capítulo se implementará un modelo de Markov oculto gaussiano sobre los precios de cierre mensuales del índice accionario estadounidense Standard & Poor's 500 (S&P 500). Con este modelo se intercambiarán acciones a través del fondo de inversión SPDR S&P 500 ETF Trust (SPY) en distintos períodos y se compararán los resultados obtenidos con los que se obtendrían con otros métodos tradicionales.

1. Selección del Modelo

Para la selección del mejor modelo entre un conjunto de modelos de Markov ocultos con distinta cantidad de estados se utilizarán los criterios AIC, BIC, HQC y CAIC para evaluar el desempeño de cada uno en distintos períodos de la manera que se explica a continuación. Para mantener el modelo simple sólo se considerarán modelos de 2 a 6 estados.

Se comenzará por calibrar los parámetros de cada modelo candidato utilizando los precios mensuales en períodos de 10 años (120 observaciones) y el algoritmo de Baum-Welch. Luego, se calculará la probabilidad L de observar cada secuencia con el algoritmo Forward para calcular junto con el número de parámetros N de cada modelo el valor de los criterios de información.

La primera calibración se realizará con los precios mensuales de Diciembre de 1996 a Noviembre de 2006 y parámetros iniciales

$$\begin{aligned}A &= (a_{ij}), a_{ij} = 1/N, \\p &= (1, 0, \dots, 0), \\ \mu_i &= \mu^{(0)} + Z, Z \sim \mathcal{N}(0, 1), \\ \sigma_i &= \sigma^{(0)}.\end{aligned}$$

En la siguiente calibración el período de 10 años se recorrerá un mes hacia adelante tomando ahora los precios de Enero de 1997 a Diciembre de 2006 y tomando como parámetros el modelo λ resultante de la calibración anterior. Este proceso se repetirá hasta calibrar los parámetros para el período de Noviembre de 2006 a Diciembre de 2016 calculando en cada paso AIC, BIC, HQC y CAIC de cada modelo. Los resultados de las 120 calibraciones realizadas pueden verse en la figura 6.1, pudiendo concluir de ellos que el modelo de 4 estados será el óptimo al ser el que consistentemente se encuentra entre los tres modelos con menor valor en cada índice a lo largo de las calibraciones.

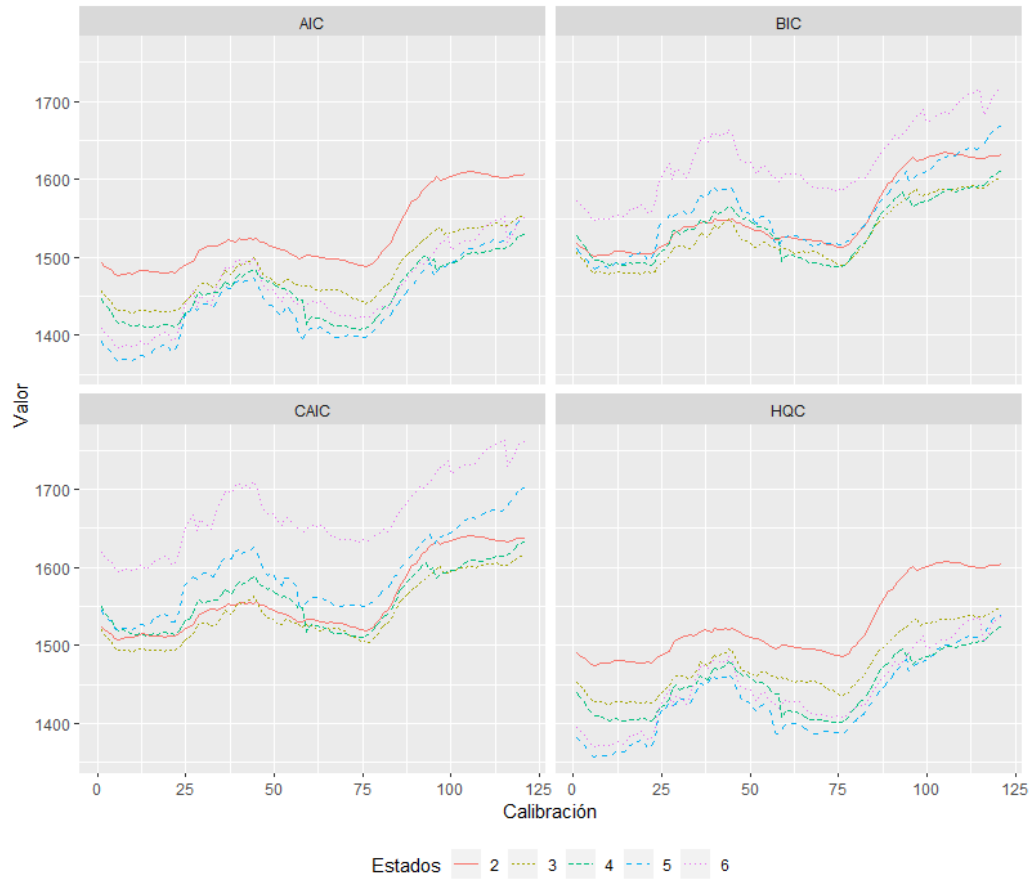


FIGURA 6.1. Criterios de información en 120 calibraciones de los modelos de Markov ocultos gaussianos de 2 a 6 estados con los precios de cierre mensuales de S&P 500.

2. Predicción

Una vez elegido el modelo, éste se utilizará para predecir los precios mensuales del índice S&P 500 de Diciembre de 2006 a Noviembre de 2016 a través del método propuesto por Hassan & Nath [HN] y adaptado por Nguyen [N] como se describe a continuación.

Supóngase que se está parado en el tiempo T y que se conoce la sucesión de observaciones $O = \{O_t | t = 1, 2, \dots, T\}$, y que se quiere predecir la observación a tiempo $T + 1$. Primero se elegirá una ventana de tiempo D para calibrar los parámetros del modelo λ con las observaciones $O^T = \{O_t | t = T - D + 1, T - D + 2, \dots, T\}$ y las fórmulas dadas para la primera iteración del algoritmo de Baum-Welch, y luego calcular la probabilidad $P(O | \lambda)$. Posteriormente, se moverá el bloque de observaciones una observación hacia atrás resultando en una nueva sucesión $O^{T-1} = \{O_t | t = T - D, T - D + 1, \dots, T - 1\}$ con la cual se calculará $P(O^{T-1} | \lambda)$. Este paso se repetirá hasta encontrar una sucesión de observaciones $O^{T^*} = \{O_t | t = T^* - D + 1, T^* - D + 2, \dots, T^*\}$ tal que $P(O^T | \lambda) \simeq P(O^{T^*} | \lambda)$. Finalmente, la observación pronosticada se obtendrá mediante la fórmula

$$\hat{O}_{T+1} = O_T + (O_{T^*+1} - O_{T^*}) \times \text{signo}(P(O^T | \lambda) - P(O^{T^*} | \lambda)).$$

Para continuar y calcular el precio de cierre a tiempo $T + 2$, estando parados ahora en el tiempo $T + 1$ y conociendo el valor real de O_{T+1} se toman ahora las observaciones $O^{T+1} = \{O_t | t = T - D + 2, T - D + 3, \dots, T + 1\}$ y, utilizando esta sucesión y los parámetros λ de la predicción anterior como parámetros iniciales, se volverán a calibrar los parámetros del modelo λ mediante el algoritmo de Baum-Welch. Con estos nuevos parámetros estimados se repetirán los pasos que se siguieron en la primera predicción para obtener \hat{O}_{T+2} . Este proceso se realizará para las predicciones restantes.

Los resultados de 120 predicciones de los precios de cierre mensuales de S&P 500 (Diciembre de 2006 a Noviembre de 2016) pueden observarse en la figura 6.2. Las predicciones se realizaron comenzando en Noviembre de 2006 utilizando los precios de cierre mensuales de Enero de 1950 a la fecha para encontrar las cadenas similares y con una ventana de tiempo $D = 120$.

3. Comercio de Activos

Finalmente, se utilizarán dos estrategias para invertir en S&P 500 a través de acciones del ETF SPY y se compararán los retornos obtenidos al final de distintos períodos. En ambas estrategias se comenzará con un capital de \$10,000 USD distribuidos en \$5,000 USD en efectivo y el equivalente a \$5,000 USD en acciones al precio

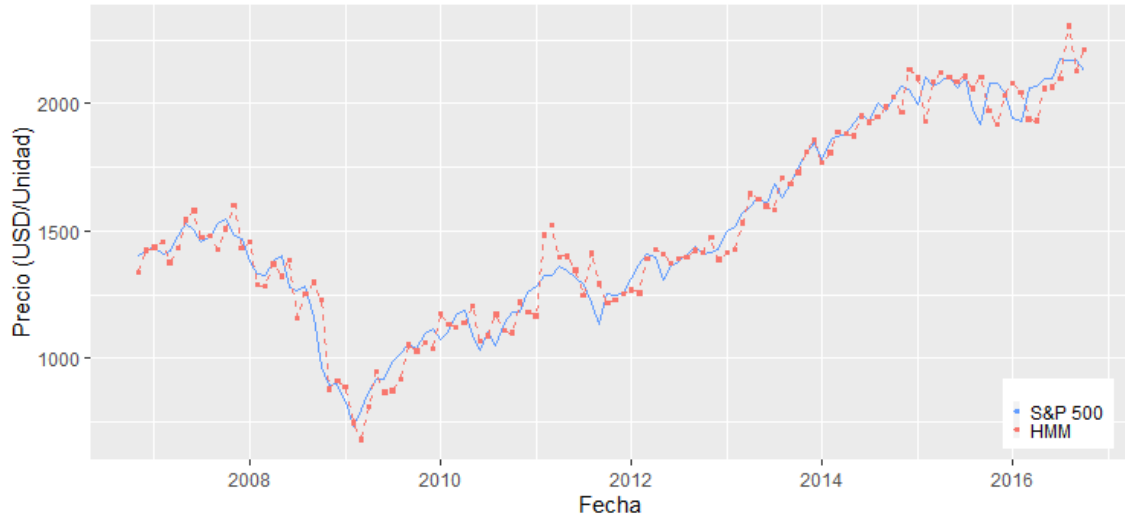


FIGURA 6.2. Precios de cierre mensuales de S&P 500 contra precios pronosticados por el modelo de Markov oculto.

del mes de inicio.

En la primera estrategia se utilizarán los precios pronosticados por el modelo de Markov oculto gaussiano de 4 estados y se comerciará cada mes. Si el modelo predice que el siguiente mes el precio subirá, se usará todo el efectivo disponible para comprar acciones este mes y venderlas al siguiente. En caso contrario, si el modelo predice que el precio bajará, se venderán todas las acciones que se tengan este mes para recomprar al mes siguiente.

La segunda estrategia consistirá en el uso de bandas de Bollinger para determinar el momento en el cuál vender o comprar. Se tomarán dos posiciones: la primera posición se abrirá cuando se cruce la banda inferior comprando todas las acciones posibles con el efectivo disponible y se cerrará al cruzar la banda media revendiendo las acciones compradas anteriormente, y la segunda se abrirá cuando se cruce la banda superior vendiendo las acciones que se tengan en el momento y se cerrará al cruzar la banda media comprando todas las acciones posibles con el efectivo recibido de la venta anterior. Esta estrategia se llevará a cabo con medias móviles simples (fig. 6.3) y exponenciales (fig. 6.4).



FIGURA 6.3. Bandas de Bollinger calculadas con medias móviles simples para los precios de cierre mensuales de S&P 500.



FIGURA 6.4. Bandas de Bollinger calculadas con medias móviles exponenciales para los precios de cierre mensuales de S&P 500.

Al finalizar el período de transacciones en cada una de las estrategias las acciones restantes en el portafolio serán convertidas a dólares al precio vigente.

Los resultados de las transacciones realizadas con cada una de las estrategias anteriores se muestran en la tabla 6.1.

Período	Estrategia	Ganancia	Rendimiento en el Período	Rendimiento Anual	Volatilidad en el Período	Inflación en el Período	Inflación Anual
40 meses (08/2013 -11/2016)	HMM	\$3,082.72	30.82%	8.39%	2.48%	3.32%	0.98%
	SMA	\$1,252.78	12.52%	3.60%	1.12%	3.32%	0.98%
	EMA	\$1,556.42	15.56%	4.43%	1.10%	3.32%	0.98%
	B&H	\$3,043.67	30.43%	8.30%	3.14%	3.32%	0.98%
60 meses (12/2011 -11/2016)	HMM	\$5,683.00	56.83%	9.42%	2.40%	6.68%	1.30%
	SMA	\$2,445.32	24.45%	4.47%	1.21%	6.68%	1.30%
	EMA	\$2,783.35	27.83%	5.03%	1.20%	6.68%	1.30%
	B&H	\$7,484.18	74.84%	11.82%	2.97%	6.68%	1.30%
80 meses (04/2010 -11/2016)	HMM	\$7,864.81	78.64%	9.09%	3.26%	10.16%	1.46%
	SMA	\$5,862.17	58.62%	7.17%	1.58%	10.16%	1.46%
	EMA	\$7,653.67	76.53%	8.90%	1.64%	10.16%	1.46%
	B&H	\$8,528.93	85.28%	9.69%	3.70%	10.16%	1.46%
100 meses (08/2008 -11/2016)	HMM	\$6,989.90	69.89%	6.57%	3.77%	10.19%	1.17%
	SMA	\$5,076.87	50.76%	5.05%	1.95%	10.19%	1.17%
	EMA	\$5,170.35	51.70%	5.13%	1.98%	10.19%	1.17%
	B&H	\$7,140.31	71.40%	6.68%	4.56%	10.19%	1.17%
120 meses (12/2006 -11/2016)	HMM	\$8,313.74	83.13%	6.24%	3.54%	19.60%	1.80%
	SMA	\$4,032.96	40.32%	3.45%	3.14%	19.60%	1.80%
	EMA	\$4,119.97	41.19%	3.51%	3.02%	19.60%	1.80%
	B&H	\$5,503.14	55.03%	4.48%	4.40%	19.60%	1.80%

TABLA 6.1. Resultados de las transacciones mensuales realizadas con unidades de S&P 500.

En la tabla 6.1 se observa como la estrategia en que se emplearon las predicciones hechas por el modelo de Markov oculto obtuvo rendimientos mayores que aquellas que consistieron en el uso de bandas de Bollinger, llegando incluso a tener el doble de rendimiento en algunos períodos.

Adicionalmente a los resultados de las estrategias anteriores se muestran también los rendimientos resultantes de comprar las acciones al inicio del período y venderlas únicamente hasta que éste finalice (Buy & Hold). En este caso la estrategia B&H se beneficia de la tendencia creciente de los precios de S&P 500 en los períodos de 40 a 100 meses obteniendo mejores resultados que con el modelo de Markov oculto, sin embargo, la diferencia es poca. Es en cambio en el período de 120 meses donde se puede apreciar el desempeño del modelo de Markov oculto, pues es el que comprende los meses en que ocurrió la crisis financiera de 2008. Mientras que el desplome de los precios afectó negativamente el B&H, el modelo de Markov oculto logra sacar provecho.

Como criterio extra puede contrastarse el rendimiento obtenido con la inflación habida en el período. De esta manera puede observarse que no sólo hubo ganancias,

sino que tampoco se perdió poder adquisitivo. Aún si el rendimiento anual obtenido utilizando la estrategia de bandas de Bollinger en ambas variantes fue mayor que la inflación anual en cada período, el rendimiento resultante de usar el modelo de Markov oculto lo fue en un factor mayor.

4. Otros activos

Para comprobar la eficacia de los modelos de Markov ocultos en el comercio de activos se aplicaron las mismas estrategias de intercambio anteriores en dos activos distintos, el índice bursátil japonés Nikkei 225 (a través del fondo de inversión iShares Core Nikkei 225 ETF (1329.T)) y el índice británico FTSE 100 (a través del fondo de inversión iShares Core FTSE 100 UCITS ETF (ISF.L)), replicando todo el proceso de selección de modelo y predicción de precios que se hizo para S&P 500.

Para ambos índices se utilizaron los precios mensuales correspondientes de Diciembre de 1996 a Noviembre de 2016 para realizar las calibraciones correspondientes con cadenas de 120 observaciones para la selección del mejor modelo. Los valores de los criterios de información para cada calibración de Nikkei 225 y FTSE 100 pueden verse en las figuras 6.5 y 6.6, respectivamente, de los cuales puede concluirse que el modelo de 3 estados es el mejor entre los candidatos para el índice japonés y el modelo de 4 estados el mejor para el índice británico.

Las predicciones de los precios mensuales de Diciembre de 2006 a Noviembre de 2016 (Nikkei 225: fig. 6.7; FTSE 100: fig. 6.8) se realizaron utilizando los precios de cierre mensuales de Febrero de 1965 a Noviembre de 2016 para buscar cadenas similares para Nikkei 225. En el caso de FTSE 100, se utilizaron los precios de Enero de 1986 a Noviembre de 2016. De igual manera se calcularon las Bandas de Bollinger con medias móviles simples y exponenciales para ambos índices (Nikkei 225: figs. 6.9 y 6.10; FTSE 100: figs. 6.11 y 6.12).

Finalmente, los resultados de obtenidos de hacer transacciones con las predicciones y bandas anteriores pueden verse en las tablas 6.2 y 6.3.

Como se observa en las tablas 6.2, al aplicarlo en Nikkei 225 el modelo de Markov oculto vuelve a dar rendimientos superiores a los que se obtendrían con el uso de bandas de Bollinger, los cuales llegan incluso a tener pérdidas considerables en el período de 120 meses. Al igual que con S&P 500 la estrategia B&H vuelve a tener rendimientos similares o mayores que el modelo de Markov oculto en los períodos de 40 a 80 meses, sin embargo, el período de 120 meses vuelve a dar una razón para confiar de esta estrategia. Tras perder Nikkei 225 más del 50% de su valor en 2008,

Período	Estrategia	Ganancia	Rendimiento en el Período	Rendimiento Anual	Volatilidad en el Período	Inflación en el Período	Inflación Anual
40 meses	HMM	\$3,786.81	37.86 %	10.11 %	3.86 %	3.32 %	0.98 %
(08/2013	SMA	\$2,039.92	20.39 %	5.72 %	2.17 %	3.32 %	0.98 %
-11/2016)	EMA	\$2,130.04	21.30 %	5.96 %	3.06 %	3.32 %	0.98 %
	B&H	\$3,674.41	36.74 %	9.48 %	5.13 %	3.32 %	0.98 %
60 meses	HMM	\$7,702.09	77.02 %	12.09 %	3.67 %	6.68 %	1.30 %
(12/2011	SMA	\$2,595.38	25.95 %	4.72 %	1.89 %	6.68 %	1.30 %
-11/2016)	EMA	\$2,689.65	27.83 %	4.87 %	2.78 %	6.68 %	1.30 %
	B&H	\$11,683.13	116.83 %	16.74 %	5.28 %	6.68 %	1.30 %
80 meses	HMM	\$6,688.01	66.88 %	7.98 %	3.46 %	10.16 %	1.46 %
(04/2010	SMA	\$702.03	7.02 %	1.02 %	1.91 %	10.16 %	1.46 %
-11/2016)	EMA	\$551.62	5.51 %	0.80 %	2.89 %	10.16 %	1.46 %
	B&H	\$6,557.67	65.57 %	7.85 %	5.36 %	10.16 %	1.46 %
100 meses	HMM	\$10,790.66	107.90 %	9.18 %	3.57 %	10.19 %	1.17 %
(08/2008	SMA	\$286.48	2.86 %	0.03 %	2.93 %	10.19 %	1.17 %
-11/2016)	EMA	-\$953.34	-9.53 %	-1.19 %	3.02 %	10.19 %	1.17 %
	B&H	\$4,004.94	40.04 %	4.12 %	6.16 %	10.19 %	1.17 %
120 meses	HMM	\$6,757.86	67.57 %	5.23 %	3.63 %	19.60 %	1.80 %
(12/2006	SMA	-\$3,568.16	-35.68 %	-4.31 %	3.98 %	19.60 %	1.80 %
-11/2016)	EMA	-\$2,834.94	-28.34 %	-3.27 %	4.33 %	19.60 %	1.80 %
	B&H	\$628.50	6.28 %	0.61 %	5.98 %	19.60 %	1.80 %

TABLA 6.2. Resultados de las transacciones mensuales realizadas con unidades de Nikkei 225.

Período	Estrategia	Ganancia	Rendimiento en el Período	Rendimiento Anual	Volatilidad en el Período	Inflación en el Período	Inflación Anual
40 meses	HMM	-\$274.13	-2.74 %	-0.83 %	2.33 %	3.32 %	0.98 %
(08/2013	SMA	\$1,046.89	10.46 %	3.03 %	1.90 %	3.32 %	0.98 %
-11/2016)	EMA	\$1,034.62	10.34 %	2.99 %	1.41 %	3.32 %	0.98 %
	B&H	\$351.93	3.51 %	1.04 %	3.12 %	3.32 %	0.98 %
60 meses	HMM	\$974.91	9.74 %	1.87 %	2.15 %	6.68 %	1.30 %
(12/2011	SMA	\$2,797.47	27.97 %	5.05 %	1.74 %	6.68 %	1.30 %
-11/2016)	EMA	\$3,229.13	32.29 %	5.75 %	1.42 %	6.68 %	1.30 %
	B&H	\$2,601.04	26.01 %	4.73 %	3.23 %	6.68 %	1.30 %
80 meses	HMM	\$2,228.15	22.28 %	3.06 %	2.80 %	10.16 %	1.46 %
(04/2010	SMA	\$2,690.76	26.90 %	3.63 %	1.88 %	10.16 %	1.46 %
-11/2016)	EMA	\$3,579.98	35.79 %	4.69 %	1.66 %	10.16 %	1.46 %
	B&H	\$2,040.51	20.40 %	2.82 %	3.74 %	10.16 %	1.46 %
100 meses	HMM	\$1,515.46	15.15 %	1.70 %	3.28 %	10.19 %	1.17 %
(08/2008	SMA	\$2,924.99	29.24 %	3.12 %	3.29 %	10.19 %	1.17 %
-11/2016)	EMA	\$2,978.82	29.78 %	3.17 %	3.20 %	10.19 %	1.17 %
	B&H	\$2,917.96	29.17 %	3.12 %	4.53 %	10.19 %	1.17 %
120 meses	HMM	\$3,090.63	30.90 %	2.72 %	3.09 %	19.60 %	1.80 %
(12/2006	SMA	\$2,377.56	23.77 %	2.15 %	3.26 %	19.60 %	1.80 %
-11/2016)	EMA	\$2,177.50	21.77 %	1.98 %	3.16 %	19.60 %	1.80 %
	B&H	\$1,487.33	14.87 %	1.39 %	4.39 %	19.60 %	1.80 %

TABLA 6.3. Resultados de las transacciones mensuales realizadas con unidades de FTSE 100.

aún si se recuperó, después de 10 años logró terminar apenas 6.28 % arriba de su valor al inicio del período, no llegando siquiera a contrarrestar los efectos de la inflación.

Por otro lado, FTSE 100 resultó ser un activo que no se benefició tanto del modelo de Markov oculto como los dos anteriores. Si en los dos índices anteriores el rendimiento anual mínimo que se obtenía con esta estrategia era mayor a 5%, en FTSE 100 sólo se obtuvo poco más de 3% como máximo e incluso se llegó a tener pérdidas en uno de los períodos. Esto no significa que el modelo de Markov oculto sea una mala estrategia, pues cabe notarse que, con excepción del período de 40 meses, el rendimiento obtenido siempre estuvo por encima de la inflación, conservando al menos el poder adquisitivo.

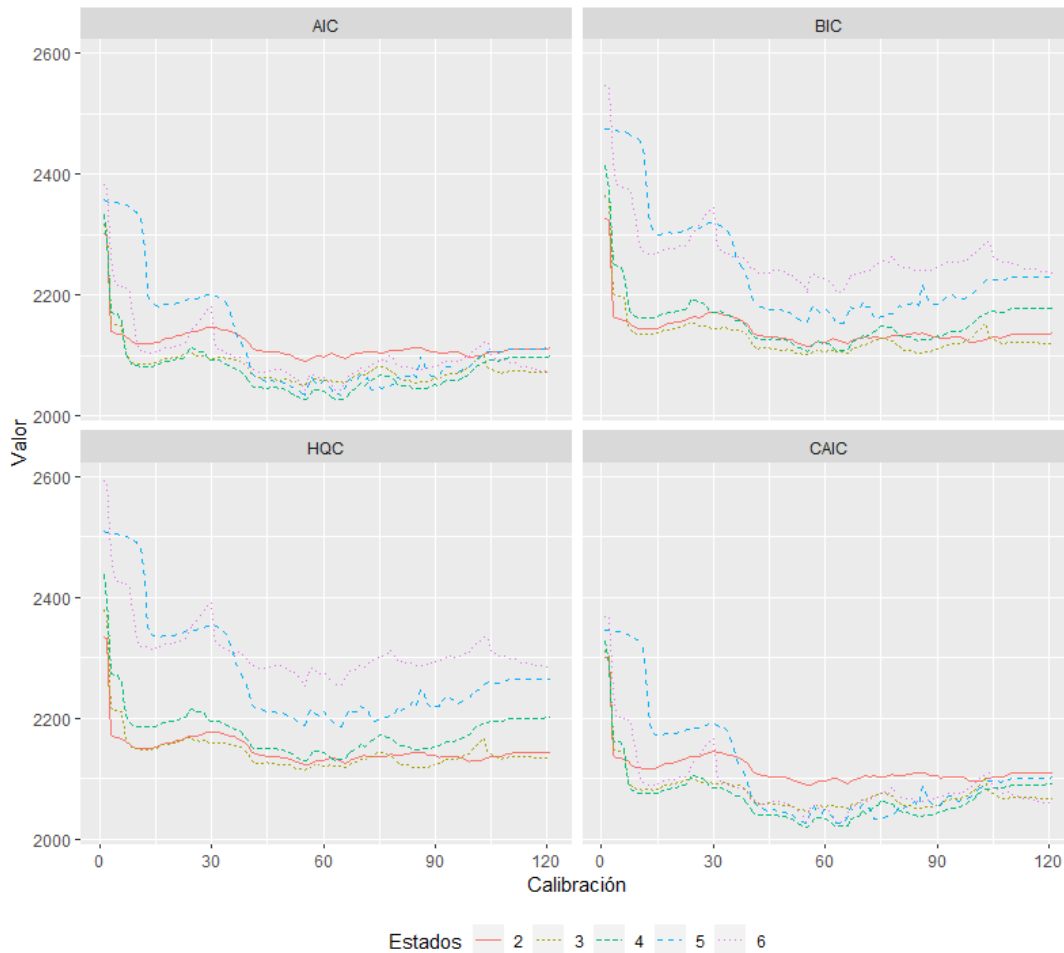


FIGURA 6.5. Criterios de información en 120 calibraciones de los modelos de Markov ocultos gaussianos de 2 a 6 estados con los precios de cierre mensuales de Nikkei 225.

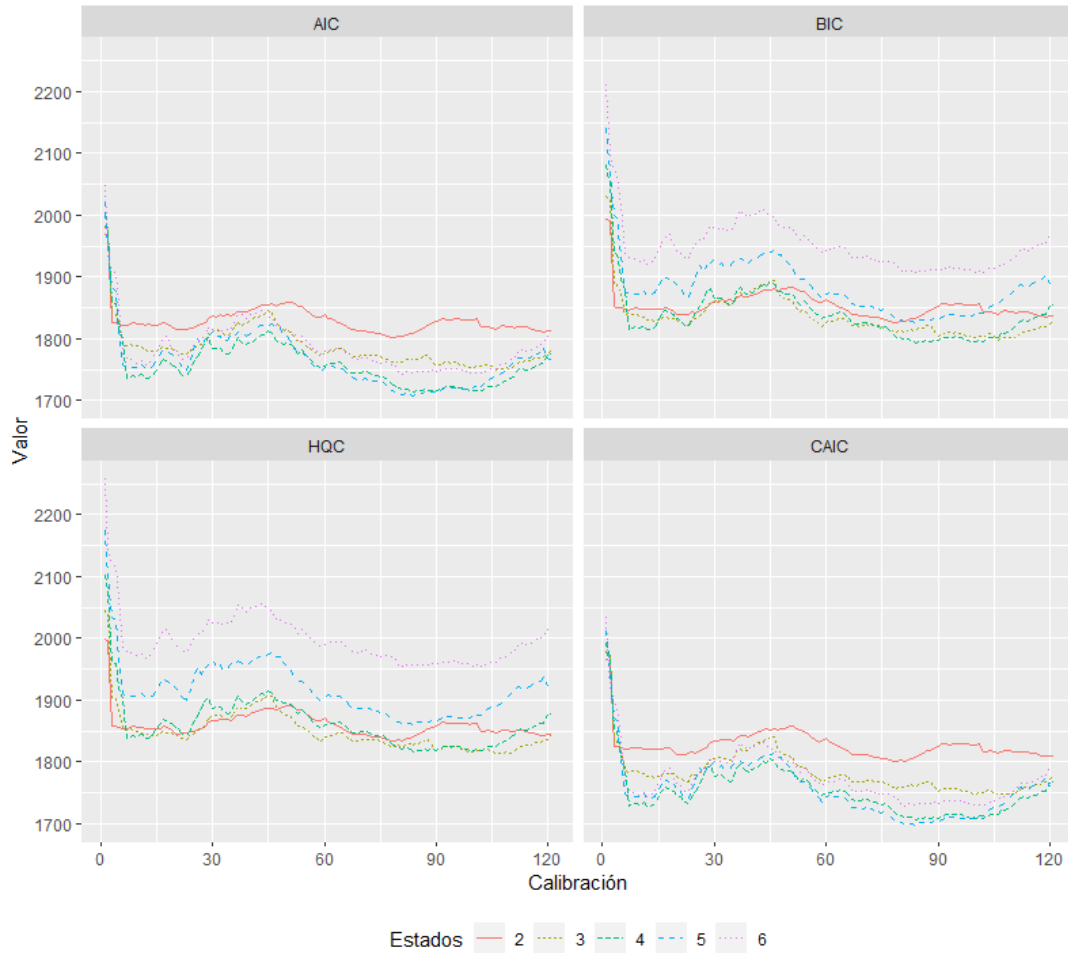


FIGURA 6.6. Criterios de información en 120 calibraciones de los modelos de Markov ocultos gaussianos de 2 a 6 estados con los precios de cierre mensuales de FTSE 100.

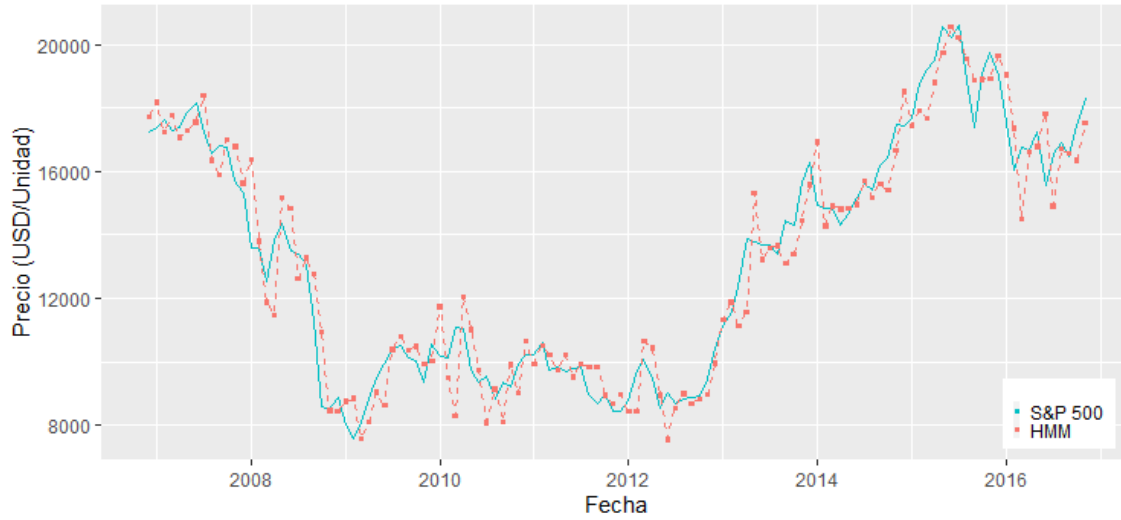


FIGURA 6.7. Precios de cierre mensuales de Nikkei 225 contra precios pronosticados por el modelo de Markov oculto.

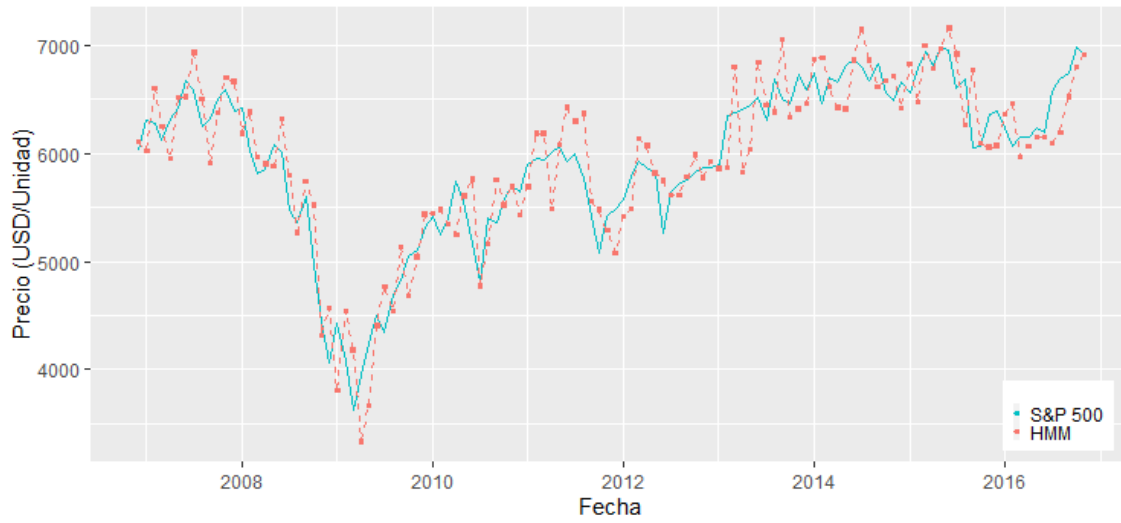


FIGURA 6.8. Precios de cierre mensuales de FTSE 100 contra precios pronosticados por el modelo de Markov oculto.

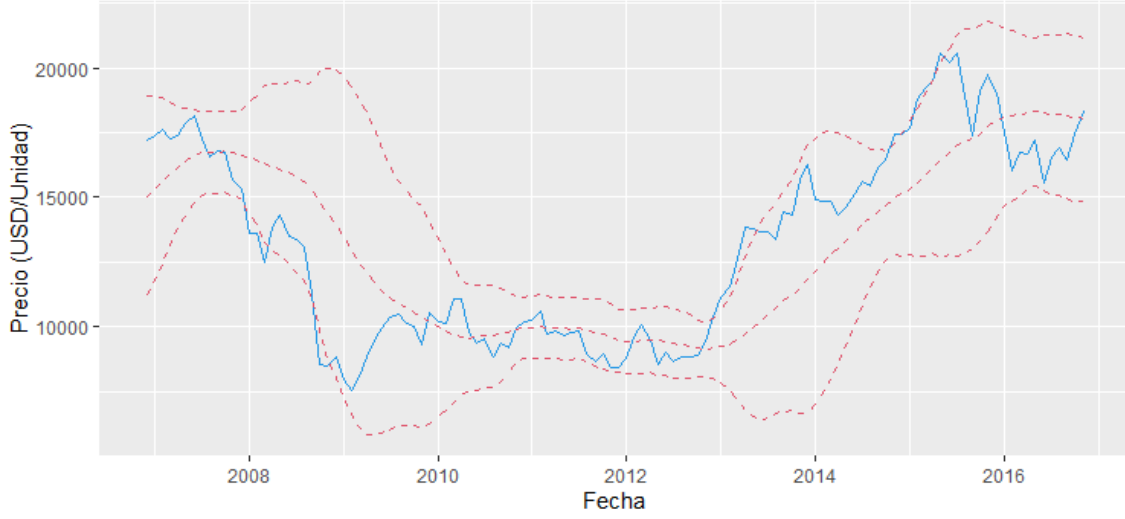


FIGURA 6.9. Bandas de Bollinger calculadas con medias móviles simples para los precios de cierre mensuales de Nikkei 225.

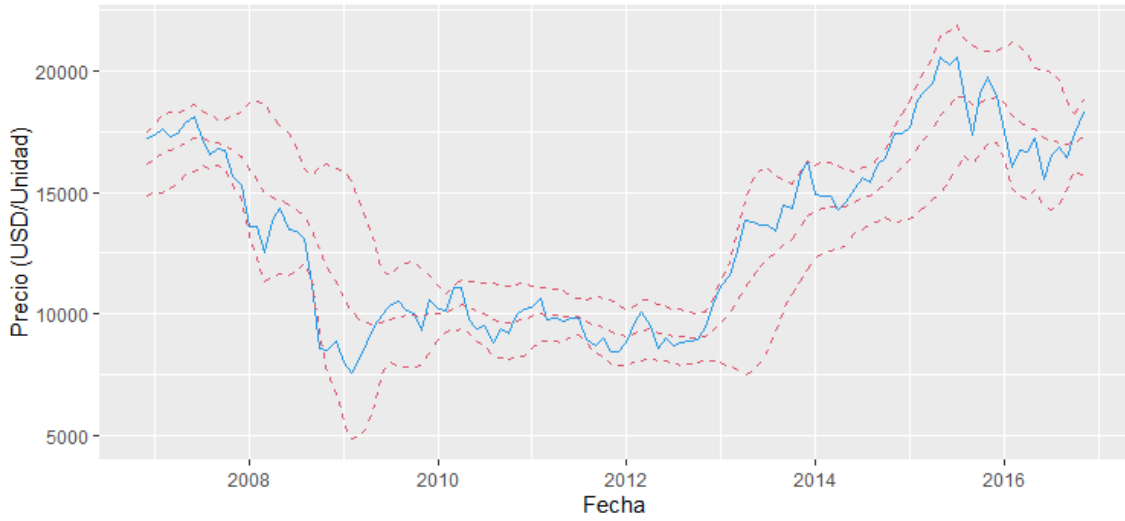


FIGURA 6.10. Bandas de Bollinger calculadas con medias móviles exponenciales para los precios de cierre mensuales de Nikkei 225.

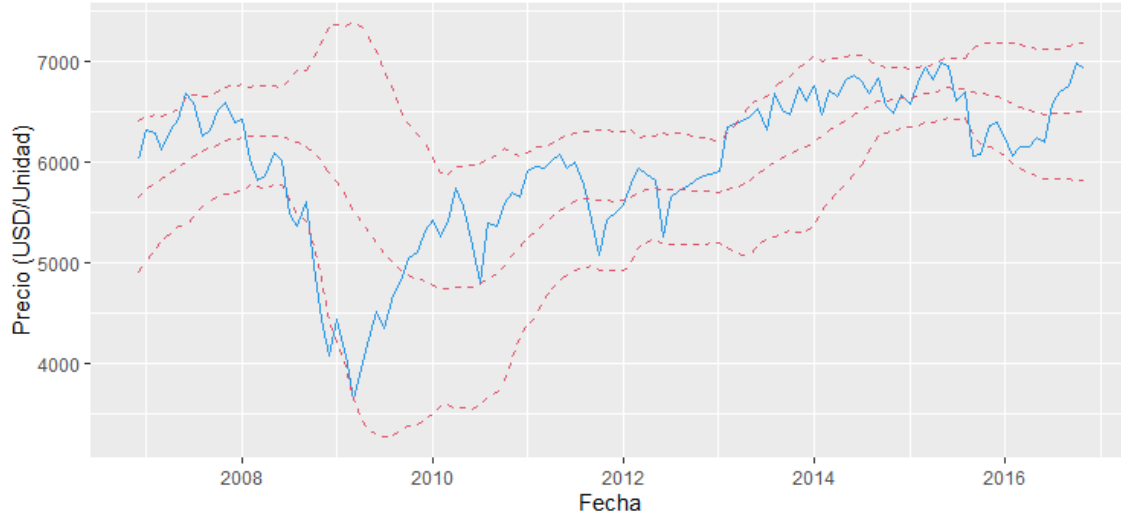


FIGURA 6.11. Bandas de Bollinger calculadas con medias móviles simples para los precios de cierre mensuales de FTSE 100.

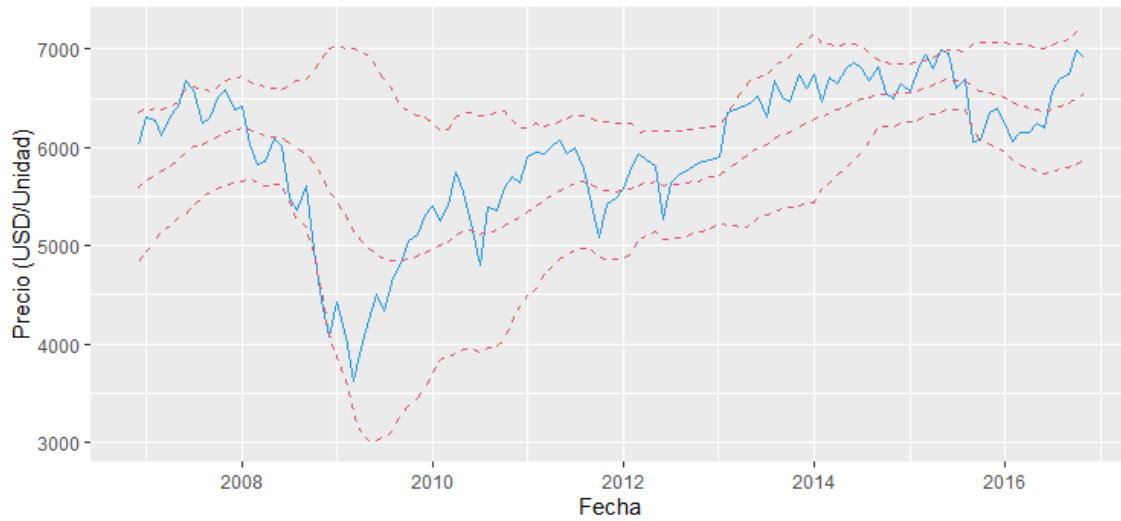


FIGURA 6.12. Bandas de Bollinger calculadas con medias móviles exponenciales para los precios de cierre mensuales de FTSE 100.

Conclusiones

En el capítulo anterior se aplicaron algunas estrategias tradicionales (bandas de Bollinger y Buy & Hold) para invertir en tres índices accionarios en distintos períodos arrojando buenos resultados en términos de rendimiento y contrarrestar los efectos de la inflación. Sin embargo, se presentó un contraste entre estas estrategias y otra con base en las predicciones hechas por un modelo de Markov oculto para resaltar como estrategias alternativas pueden obtener un mejor desempeño.

A pesar de haber observado que cada estrategia se desempeñó mejor que las otras dos en distintos escenarios, dejando claro así que ninguna estrategia de inversión es universal, puede afirmarse que el modelo de Markov oculto en conjunto con la estrategia implementada resultó ser el más consistente.

Un resultado importante que se obtuvo fue observar cómo el modelo de Markov oculto demuestra una capacidad de adaptarse a distintas series de tiempo y desempeñarse bien en situaciones adversas como es la caída de precios.

Cabe destacar que a lo largo de los distintos períodos e índices la volatilidad asociada a la estrategia con base en bandas de Bollinger fue menor que en la estrategia que utiliza las predicciones hechas con el modelo de Markov oculto. Esto naturalmente por el hecho de que con las bandas de Bollinger a momentos el capital está parcial o totalmente en efectivo e inmóvil, contrario a la segunda estrategia en la que cada mes se realiza una transacción y el valor del portafolio se encuentra en constante cambio. Se tiene entonces una estrategia que resulta en un mayor retorno a mayor riesgo y otra con un menor retorno a menor riesgo, dos buenas opciones que se adecúan a los distintos perfiles de inversionistas.

En cuanto a la estrategia Buy & Hold y la del modelo de Markov oculto ambas obtuvieron rendimientos similares en varios periodos, sin embargo la volatilidad asociada a la segunda estrategia fue considerablemente menor a la de la primera. De lo anterior puede concluirse que la estrategia con base en las predicciones hechas por el

modelo de Markov oculto ofrece el mejor rendimiento ajustado por riesgo.

Aunque los resultados obtenidos demuestran que los modelos de Markov ocultos son un algoritmo robusto y eficiente para la aplicación que se eligió no se debe caer en el error de pensarlos como una solución general. Una forma de evaluar si los modelos de Markov ocultos son la herramienta ideal para modelar un fenómeno es considerar los problemas que presentan frente a ciertas situaciones. Los dos principales son las hipótesis que suponen estos modelos: la propiedad de Markov y la independencia condicional de las observaciones, es decir, que el estado actual sólo depende del estado anterior y que lo que se observa depende únicamente del estado actual. Esto no siempre se ajusta a la realidad del fenómeno y en tal caso es conveniente elegir otro tipo de modelo.

Como posible extensión al trabajo realizado pueden considerarse diferentes períodos de tiempo que representen distintos escenarios económicos en los cuáles replicar la estrategia propuesta y verificar que los resultados se mantengan. Así mismo, pueden desarrollarse nuevas estrategias con base en los modelos de Markov ocultos, pudiendo aumentar la complejidad del modelo o cambiar la forma en que se compran y venden los activos.

Apéndices

Código en R

Para la elaboración de los ejemplos que se incluyen se escribió el siguiente código en R.

Algoritmos para modelos de Markov ocultos discretos.

Para calcular las probabilidades de observar la sucesión de observaciones O bajo un modelo de Markov oculto discreto con parámetros $\lambda \equiv \{A, B, p\}$ se escribieron las siguientes funciones para los algoritmos Forward y Backward:

```
forward <- function(A, B, p, O, j = 1)
{
  T <- length(O)
  a <- data.frame()
  for(i in colnames(A))
  {
    a[i,1] <- p[1,i] * B[i,O[1]]
  }

  aT <- c(sum(a[1]))

  if(length(O) > 1)
  {
    for(t in 2:T)
    {
      for(i in colnames(A))
      {
        a[i,t] <- sum(a[,t-1] %*% A[,i]) * B[i,O[t]]
      }

      aT[t] <- sum(a[t])
    }
  }
  return(list(aT[T], aT, a)[[j]])
}

backward <- function(A, B, p, O, j = 1)
{
```

```

T <- length(O)
b <- as.data.frame(matrix(rep(0, ncol(A)*T), ncol = T,
                          dimnames = list(colnames(A), 1:T)))
b[,T] <- 1

if(T > 1)
{
  for (t in (T-1):1)
  {
    for(i in colnames(A))
    {
      b[i,t] <- sum(A[i,] * B[,0[t+1]] * b[,t+1])
    }
  }
}

b0 <- sum(p * B[,0[1]] * b[,1])

return(list(b0, b)[[j]])
}

```

Para encontrar la cadena de estados ocultos más probable dado un modelo de Markov oculto con parámetros $\lambda \equiv \{A, B, p\}$ y una sucesión de observaciones O se utilizó la siguiente función del algoritmo de Viterbi:

```

viterbi <- function(A, B, p, O, j = 1)
{
  T <- length(O)
  d <- data.frame()
  phi <- data.frame()
  Q <- c()

  for(i in colnames(A))
  {
    d[i,1] <- p[1,i] * B[i,0[1]]
    phi[i,1] <- "0"
  }

  if(T > 1)
  {
    for(t in 2:T)
    {
      for(i in colnames(A))
      {
        d[i,t] <- max(d[,t-1] * A[,i]) * B[i,0[t]]
        phi[i,t] <- names(d[,t-1] * A[,i])[which.max((d[,t-1] * A[,i]))]
      }
    }
  }
}

```

```

Q[T] = row.names(d)[which.max(d[,T])]

if(T > 1)
{
  for(t in (T-1):1)
  {
    Q[t] <- phi[Q[t+1],t+1]
  }
}

return(list(Q = Q, d = d, phi = phi)[[j]])
}

```

Para estimar los parámetros $\lambda^* \equiv \{A^*, B^*, p^*\}$ de un modelo de Markov oculto discreto a partir de una sucesión de observaciones O y parámetros iniciales $\lambda \equiv \{A, B, p\}$ se utilizó la siguiente función para el algoritmo de Baum-Welch:

```

baum_welch <- function(O, A, B, p, tau, delta = 1)
{
  r <- 0

  while(delta > tau)
  {
    prob <- forward(A, B, p, O)
    a <- forward(A, B, p, O, 3)
    b <- backward(A, B, p, O, 2)
    g <- data.frame()
    x <- list()
    T <- length(O)

    for(t in 1:T)
    {
      x[[t]] <- data.frame()

      for(i in colnames(A))
      {
        g[i,t] = (a[i,t] * b[i,t]) / prob

        if(t < T)
        {
          for(j in colnames(A))
          {
            x[[t]][i,j] <- (a[i,t] * A[i,j] * B[j, O[t+1]] * b[j,t+1]) / prob
          }
        }
      }
    }

    for(i in colnames(p))
    {

```



```

p[1,i] <- g[i,1]

for(j in colnames(A))
{
  A[i,j] <- Reduce('+', x[1:(T-1)])[i,j] / sum(g[i,1:(T-1)])
}

for(k in colnames(B))
{
  B[i,k] <- ifelse(length(g[i, 0 == k]) > 0,
                  sum(g[i, 0 == k]), 0) / sum(g[i,])
}
}

delta <- abs(prob - forward(A, B, p, 0))
r <- r+1

}

return(list(A = A, B = B, p = p, r = r))
}

```

Algoritmos para modelos de Markov ocultos gaussianos.

Para calcular las probabilidades de observar la sucesión de observaciones O bajo un modelo de Markov oculto gaussiano con parámetros $\lambda \equiv \{A, \mu, \sigma, p\}$ se escribieron las siguientes funciones para los algoritmos forward y backward:

```

Forward <- function(A, mu, sigma, p, O, j = 1)
{
  T <- length(O)
  a <- data.frame()
  for(i in colnames(A))
  {
    a[i,1] <- p[1,i] * dnorm(O[1], mu[i], sigma[i])
  }

  aT <- c(sum(a[1]))

  if(length(O) > 1)
  {
    for(t in 2:T)
    {
      for(i in colnames(A))
      {
        a[i,t] <- sum(a[,t-1] %*% A[,i]) * dnorm(O[t], mu[i], sigma[i])
      }
    }

    aT[t] <- sum(a[t])
  }
}

```

```

    }
  }
  return(list(aT[T], aT, a)[[j]])
}

Backward <- function(A, mu, sigma, p, O, j = 1)
{
  T <- length(O)
  b <- as.data.frame(matrix(rep(0, ncol(A)*T), ncol = T,
                           dimnames = list(colnames(A), 1:T)))
  b[,T] <- 1

  if(T > 1)
  {
    for (t in (T-1):1)
    {
      for(i in colnames(A))
      {
        b[i,t] <- sum(A[i,] * dnorm(O[t+1], mu, sigma) * b[,t+1])
      }
    }
  }

  b0 <- sum(p * dnorm(O[1], mu, sigma) * b[,1])

  return(list(b0, b)[[j]])
}

```

Para encontrar la cadena de estados ocultos más probable dado un modelo de Markov oculto con parámetros $\lambda \equiv \{A, \mu, \sigma, p\}$ y una sucesión de observaciones O se utilizó la siguiente función del algoritmo de Viterbi:

```

Viterbi <- function(A, mu, sigma, p, O, j = 1)
{
  T <- length(O)
  d <- data.frame()
  phi <- data.frame()
  Q <- c()

  for(i in colnames(A))
  {
    d[i,1] <- p[1,i] * dnorm(O[1], mu[i], sigma[i])
    phi[i,1] <- "0"
  }

  if(T > 1)
  {
    for(t in 2:T)
    {
      for(i in colnames(A))

```

```

    {
      d[i,t] <- max(d[,t-1] * A[,i]) * dnorm(O[t], mu[i], sigma[i])
      phi[i,t] <- names(d[,t-1] * A[,i])[which.max((d[,t-1] * A[,i]))]
    }
  }
}

Q[T] = row.names(d)[which.max(d[,T])]

if(T > 1)
{
  for(t in (T-1):1)
  {
    Q[t] <- phi[Q[t+1],t+1]
  }
}

return(list(Q = Q, d = d, phi = phi)[[j]])
}

```

Para estimar los parámetros $\lambda^* \equiv \{A^*, \mu^*, \sigma^*, p^*\}$ de un modelo de Markov oculto gaussiano a partir de una sucesión de observaciones O y parámetros iniciales $\lambda \equiv \{A, \mu, \sigma, p\}$ se utilizó la siguiente función para el algoritmo de Baum-Welch:

```

Baum_Welch <- function(O, A, mu, sigma, p, tau, delta)
{
  r <- 0

  while(delta > tau)
  {
    prob <- Forward(A, mu, sigma, p, O)
    a <- data.frame()
    b <- as.data.frame(matrix(rep(0, ncol(A)*T), ncol = T,
                             dimnames = list(colnames(A), 1:T)))

    z <- c()
    g <- data.frame()
    x <- list()
    T <- length(O)

    for(i in colnames(A))
    {
      a[i,1] <- p[1,i] * dnorm(O[1], mu[i], sigma[i])
      z[1] <- sum(a[,1])
      a[,1] <- a[,1] / z[1]
    }

    if(length(O) > 1)
    {
      for(t in 2:T)
      {

```

```

    for(i in colnames(A))
    {
      a[i,t] <- sum(a[,t-1] %*% A[,i]) * dnorm(0[t], mu[i], sigma[i])
    }
    z[t] <- sum(a[,t])
    a[,t] <- a[,t] / z[t]
  }
}

b[,T] <- 1

if(T > 1)
{
  for (t in (T-1):1)
  {
    for(i in colnames(A))
    {
      b[i,t] <- sum(A[i,] * dnorm(0[t+1], mu, sigma) * b[,t+1])
    }
    b[,t] <- b[,t] / z[t]
  }
}

for(t in 1:T)
{
  x[[t]] <- data.frame()

  for(i in colnames(A))
  {
    g[i,t] = (a[i,t] * b[i,t]) / prob

    if(t < T)
    {
      for(j in colnames(A))
      {
        x[[t]][i,j] <- (a[i,t]*A[i,j]*dnorm(0[t+1],mu[j],sigma[j]) * b[j,t+1]) / prob
      }
    }
  }
}

for(i in colnames(p))
{
  p[1,i] <- g[i,1]

  for(j in colnames(A))
  {
    A[i,j] <- Reduce('+', x[1:(T-1)])[i,j] / sum(g[i,1:(T-1)])
  }
}

```

```

    mu[i] <- sum(g[i,1:(T-1)] * 0[1:(T-1)]) / sum(g[i,1:(T-1)])

    sigma[i] <- sqrt(sum(g[i,1:T]*(0[1:T]-mu[i])*(0[1:T]-mu[i])) / sum(g[i,1:T]))
  }

  delta <- abs(prob - Forward(A, mu, sigma, p, 0))
  r <- r+1
}

return(list(A = A, mu = mu, sigma = sigma, p = p, r = r))
}

```

Selección del modelo, predicción de precios e intercambio de acciones.

Para la selección del modelo adecuado se utilizó la siguiente función para obtener el calor de los criterios de información para cada iteración de cada número distinto de estados:

```

criterios <- function(data, N, size)
{
  set.seed(N)
  T <- length(data)
  k <- N^2 + 2*N - 1
  info <- data.frame()

  lambda <- dthmm(x = NULL,
                 Pi = matrix(rep(1/N, N^2), ncol = N),
                 delta = c(1, rep(0, N-1)),
                 distn = "norm",
                 pm = list(mean = mean(data[1:size]) + rnorm(N),
                          sd = rep(sd(data[1:size]), N)))

  j <- 0

  while((size+j) < (T+1))
  {
    lambda$x <- data[(j+1):(size+j)]
    lambda <- BaumWelch(lambda)

    i <- j + 1

    info[i,"i"] <- i
    info[i,"AIC"] <- -2*lambda$LL + 2*k
    info[i,"BIC"] <- -2*lambda$LL + k*log(T)
    info[i,"HQC"] <- -2*lambda$LL + k*log(log(T))
    info[i,"CAIC"] <- -2*lambda$LL + k*(log(T)+1)
  }
}

```

```

j <- j + 1

print(j)
}

return(info)
}

```

Para la predicción de los precios con el modelo de Markov de Markov Oculto gaussiano de N estados se utilizó el siguiente código:

```

predictHMM <- function(O, N, n, s, D, dif, tol)
{
  set.seed(N)
  pred <- c()
  lambda <- dthmm(x = O[(s-D+1):s],
                 Pi = matrix(rep(1/N, N^2), ncol = N),
                 delta = c(1, rep(0, N-1)),
                 distn = "norm",
                 pm = list(mean = mean(x) + rnorm(N, sd = sd(x)),
                          sd = rep(sd(x), N)))

  for(i in 1:n)
  {
    T <- length(O[1:(s+i-1)])

    lambda$x <- O[1:(s+i-1)][(T-D+1):T]
    lambda <- BaumWelch(lambda, bwcontrol(prt = FALSE, tol = tol))

    p <- lambda$LL
    Tp <- T - 1
    pp <- forwardback(O[1:(s+i-1)][(Tp-D+1):Tp],
                    lambda$Pi,
                    lambda$delta,
                    "norm",
                    lambda$pm)$LL

    d <- c(abs(p - pp))

    while((abs(p - pp) > dif) && (Tp > D))
    {
      Tp <- Tp-1
      pp <- forwardback(O[1:(s+i-1)][(Tp-D+1):Tp],
                      lambda$Pi,
                      lambda$delta,
                      "norm",
                      lambda$pm)$LL
      d <- c(d, c(abs(p - pp)))
    }
  }
}

```

```

if(Tp == D)
{
  Tp <- T - which.min(d)
}

On <- O[1:(s+i-1)][T] + ((O[1:(s+i-1)][Tp+1] - O[1:(s+i-1)][Tp]) * sign(p - pp))

pred[i] <- On
}

return(pred)
}

```

Para el cálculo de las Bandas de Bollinger con medias móviles simples y exponenciales se usaron las siguientes funciones:

```

BollingerSMA <- function(data, n, s, r)
{
  T <- length(data)
  bands <- data.frame()

  for(i in 1:(T-s+1))
  {
    bands[i, "Lower"] <- mean(data[(s-n+i):(s+i-1)]) - r*sd(data[(s-n+i):(s+i-1)])
    bands[i, "Middle"] <- mean(data[(s-n+i):(s+i-1)])
    bands[i, "Upper"] <- mean(data[(s-n+i):(s+i-1)]) + r*sd(data[(s-n+i):(s+i-1)])
  }

  return(bands)
}

```

```

BollingerEMA <- function(data, n, s, r, f)
{
  SMA <- mean(data[(s-n):(s-1)])
  a <- f/(n+1)

  T <- length(data)
  bands <- data.frame()

  bands[1, "Lower"] <- mean(data[(s-n):(s-1)]) - r*sd(data[(s-n+1):(s)])
  bands[1, "Middle"] <- mean(data[(s-n):(s-1)])
  bands[1, "Upper"] <- mean(data[(s-n):(s-1)]) + r*sd(data[(s-n+1):(s)])

  for(i in 2:(T-s+1))
  {
    bands[i, "Middle"] <- (a * data[(s+i-1)]) + ((1-a) * bands$Middle[i-1])
    bands[i, "Lower"] <- bands$Middle[i] - r*sd(data[(s-n+i):(s+i-1)])
    bands[i, "Upper"] <- bands$Middle[i] + r*sd(data[(s-n+i):(s+i-1)])
  }
}

```

```

return(bands)
}

```

Para calcular el capital y rendimiento al final de las transacciones realizadas con los precios predichos por el modelo de Markov oculto se utilizó el siguiente código:

```

hmm_todo <- function(pred, real, ini)
{
  dollars <- ini / 2
  shares <- (ini / 2) / real[1]

  for(i in 1:length(pred))
  {
    if((pred[i] - real[i]) > 0)
    {
      shares <- shares + (dollars / real[i])
      dollars <- 0
    }

    else
    {
      dollars <- dollars + (shares * real[i])
      shares <- 0
    }
  }

  dollars <- dollars + (shares * tail(real,1))
  shares <- 0

  return(c(dollars, dollars / ini))
}

```

Para calcular el capital y rendimiento al final de las transacciones realizadas con las bandas de Bollinger calculadas se utilizó el siguiente código:

```

tradeBollinger <- function(real, bands, ini)
{
  dollarsL <- ini / 2
  sharesL <- 0
  dollarsU <- 0
  sharesU <- (ini / 2) / real[1]

  real <- tail(real,-1)
  pos_a <- "Mu"
  pos_b <- "Mu"

  for(i in 1:length(real))
  {
    if(real[i] > bands$Middle[i])
    {

```



```

if(real[i] < bands$Upper[i])
  { pos_a <- "Mu" }
else
  { pos_a <- "U" }
}
else
{
  if(real[i] > bands$Lower[i])
    { pos_a <- "Ml" }
  else
    { pos_a <- "L" }
}

if(pos_b == "Mu")
{
  if(pos_a == "U")
  {
    dollarsU <- dollarsU + (sharesU * real[i])
    sharesU <- 0
  }

  if(pos_a == "Ml")
  {
    sharesU <- sharesU + (dollarsU / real[i])
    dollarsU <- 0
  }
}

if(pos_b == "Ml")
{
  if(pos_a == "L")
  {
    sharesL <- sharesL + (dollarsL / real[i])
    dollarsL <- 0
  }

  if(pos_a == "Mu")
  {
    dollarsL <- dollarsL + (sharesL * real[i])
    sharesL <- 0
  }
}

pos_b <- pos_a
}

dollars <- dollarsU + dollarsL + ((sharesU + sharesL) * real[length(real)])

return(dollars / ini)
}

```


Bibliografía

- [A] Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–23.
- [ABW] Anderson, David R., Kenneth P. Burnham, y Gary C. White. 1998. Comparison of AIC and CAIC for Model Selection and Statistical Inference from Capture-Recapture Studies. *Journal of Applied Statistics* 25: 263-82.
- [B] Bozdogan, Hamparsum. 1987. Model selection and Akaike’s Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 52: 345–70.
- [BA] Burnham, Kenneth P., y David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33: 261–304.
- [CMR] Cappé, Olivier, Eric Moulines, y Tobias Rydén. 2005. *Inference in Hidden Markov Models*. Nueva York: Springer.
- [CSB] Cheng, Changqing, Akkarapol Sa-Ngasoongsong, Omer Beyca, Trung Le, Hui Yang, Zhenyu Kong, y Satish T.S. Bukkapatnam. 2015. Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. *IIE Transactions*. 47: 1053-1071.
- [DCL] Dziak, John J., Donna L. Coffman, Stephanie T. Lanza, Runze Li, y Lars S. Jermiin. 2020. Sensitivity and Specificity of Information Criteria. *Briefings in Bioinformatics* 21: 553–565.
- [HQ] Hannan, Edward J., y Barry G. Quinn. 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)* 41: 190–95.

- [HN] Hassan, Md Rafiul, y Baikunth Nath. 2005. Stock Market Forecasting Using Hidden Markov Models: A New Approach. Artículo presentado en la 5ta Conferencia Internacional sobre Diseño de Sistemas Inteligentes y Aplicaciones (ISDA'05), Breslavia, Polonia, Septiembre 8-10, pp. 192–96.
- [KMV] Karagrigoriou, Alex, Kyriacos Mattheou y Ilia Vonta. 2011. On Asymptotic Properties of AIC Variants with Applications. *Open Journal of Statistics* 1: 105-109.
- [KL] Kullback, Solomon, y Richard A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* 22: 79–86.
- [N] Nguyen, Nguyet. 2018. Hidden Markov Model for Stock Trading. *Int. J. Financial Stud.* 6: 36.
- [NN] Nguyen, Nguyet, y Dung Nguyen. 2015. Hidden Markov Model for Stock Selection. *Risks* 3: 455-73.
- [R] Rabiner, Lawrence R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE* 77: 257–86.
- [Sc] Schwarz, Gideon. 1978. Estimating the Dimension of A Model. *The Annals of Statistics* 6: 461–64.
- [Sha] Shannon, Claude E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27: 379-423.
- [Shi] Shibata, Ryutaro. 1980. Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. *The Annals of Statistics* 8: 147–164.
- [St] Stamp, Mark. 2017. *Introduction to Machine Learning with Applications in Information Security*. Florida: Chapman & Hall/CRC Press.
- [RTS] Rundo, Francesco, Francesca Trenta, Agatino Luigi di Stallo, y Sebastiano Battiato. 2019. Machine Learning for Quantitative Finance Applications: A Survey. *Applied Sciences*. 9: 1-20.

- [W] Wei, C. Z. 1992. On Predictive Least Squares Principles. *The Annals of Statistics* 20: 1-42.
- [ZM] Zucchini, Walter, e Ian MacDonald. 2009. *Hidden Markov Models for Time Series: An Introduction Using R*. Florida: Chapman & Hall/CRC Press.