



MODELOS OCULTOS DE MARKOV: UNA APLICACIÓN DE ESTIMACIÓN BAYESIANA PARA SERIES DE TIEMPO FINANCIERAS

LIZBETH NARANJO ALBARRÁN & LUZ JUDITH RODRÍGUEZ ESPARZA

RESUMEN. Los Modelos Ocultos de Markov (HMM) han sido muy importantes últimamente debido a su gran cantidad de aplicaciones prácticas. Una aplicación que ha cobrado relevancia es en el modelado de series de tiempo financieras, pues muchas de éstas presentan el llamado “agrupamiento de volatilidad”, es decir, periodos con actividad significativa tienden a ocurrir cercanamente. En este trabajo incrustamos un HMM -para cambiar entre dos o más variables latentes- dentro del modelo econométrico Markov-Switching GARCH, que permite el “cambio de régimen”, y lo aplicamos a una serie de tiempo de rendimientos bursátiles. Para la estimación de los parámetros se utilizó el paradigma Bayesiano, la cual se pudo realizar fácil y eficientemente usando *Stan*.

1. INTRODUCCIÓN

Los modelos ocultos de Markov (HMM por sus siglas en inglés *Hidden Markov Models*) son una herramienta para representar la distribución de probabilidad de una secuencia de observaciones, donde se supone que cada una de ellas fue generada por algún proceso cuyo estado está oculto para el observador. Los HMM tienen una estructura matemática robusta que forma una base sólida para usarse en distintas aplicaciones. Cuando estos modelos son aplicados propiamente, los resultados de modelación pueden ser significativos.

Los HMM fueron desarrollados por el matemático estadounidense Leonard E. Baum y sus compañeros del Instituto de Análisis de Defensa en Princeton, Nueva Jersey, a fines de los años sesenta y principios de los 70 (Baum y Petrie, 1966; Baum *et al.*, 1970).

Sin embargo, hasta varios años después se publicó la primera aplicación del modelo y fue en reconocimiento de voz, escrito por Rabiner (1989). A partir de este trabajo, se han utilizado ampliamente los HMM en varias otras aplicaciones, como por ejemplo: búsqueda de genes (Lukashin y Borodovsky, 1998); rastreo de gestos manuales (Chen *et al.*, 2003); reconocimiento de actividad de video (Niu y Abdel-Mottaleb, 2005); para modelar datos de series temporales (Zucchini *et al.*, 2016); para predecir el comportamiento de acciones del mercado de valores (Gupta y Dhingra, 2012; Hassan y Nath, 2005), entre otras aplicaciones.

Muchas series de tiempo financieras no se pueden modelar a través de procesos lineales clásicos, como los autorregresivos (AR) o los autorregresivos de promedios móviles (ARMA), que tienen la premisa de tener una varianza constante; sin embargo, dichas series tienen heterocedasticidad condicional, es decir, una variabilidad condicional no constante a lo largo de la serie. Algunas series de tiempo financieras exhiben el llamado “agrupamiento de volatilidad”, es decir, periodos de actividad significativa tienden a ocurrir muy juntos, lo que sugiere que hay alguna forma de memoria a corto plazo en la volatilidad (desviación estándar de los rendimientos). Para solucionar el problema de la volatilidad condicional surgen los modelos con heterocedasticidad condicional autorregresiva (ARCH) en los años 80 propuestos por Robert F. Engle (Engle *et al.*,

2010 *Mathematics Subject Classification.* 60J10, 62F15, 62M05, 62M10, 65C05.

Palabras clave. Modelos ocultos de Markov, Markov-switching GARCH, estadística Bayesiana, series de tiempo, *Stan*.

1987); y a su vez, éstos se extienden en otro tipo de modelos como son los modelos de heterocedasticidad condicional autorregresiva generalizada (GARCH), que se utilizan comúnmente para capturar este fenómeno y han sido ampliamente estudiados en la literatura econométrica (Bollerslev, 1986, 2023). Otro fenómeno importante que se presenta en series financieras es el denominado “cambio de régimen” -la observación que los mercados financieros atraviesan periodos alternos de auges de baja volatilidad y caídas de alta volatilidad-. El modelo Markov-switching GARCH (MS-GARCH) de Haas *et al.* (2004) utiliza un HMM para cambiar entre dos procesos GARCH latentes. En Ardia *et al.* (2019) se presenta un estudio sobre la estimación de los MS-GARCH utilizando el paquete estadístico R.

Este artículo está organizado de la siguiente manera. En Sección 2 se presenta una breve introducción de estadística Bayesiana. En Sección 3 se definen los modelos ocultos de Markov, y los problemas y algoritmos necesarios para la estimación se discuten en Sección 4. En Sección 5 se muestra la aplicación de los HMM en series de tiempo financieras. Finalmente, en Sección 6 concluimos. Los códigos de R y Stan se encuentran disponibles en el Apéndice A.

2. ESTADÍSTICA BAYESIANA

En los últimos años ha habido un mayor interés en el desarrollo de métodos de inferencia Bayesiana. La razón principal es que la metodología Bayesiana proporciona un paradigma completo para la inferencia estadística bajo incertidumbre, que permite combinar información derivada de observaciones con información obtenida de expertos (Berger, 1985; Bernardo y Smith, 1994; Robert, 1994).

2.1. Paradigma Bayesiano. Sea $\mathbf{X} = (X_1, \dots, X_n)$ variables aleatorias (v.a.) de una función de densidad de probabilidad (f.d.p.) $\mathbb{P}(\mathbf{X}|\boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbb{R}^d$, donde $\boldsymbol{\theta}$ es el vector de parámetros y Θ es el espacio paramétrico, cuya función de densidad conjunta o verosimilitud es $\mathbb{P}(\mathbf{X}|\boldsymbol{\theta})$. El problema en estadística se reduce a hacer inferencias sobre el supuesto valor de $\boldsymbol{\theta}$.

Particularmente, en estadística Bayesiana (Robert, 1994; Berger, 1985; Bernardo y Smith, 1994) se modela la incertidumbre de $\boldsymbol{\theta}$ usando métodos probabilísticos, y se considera a $\boldsymbol{\theta}$ como aleatoria, en específico, se usa el Teorema de Bayes:

$$\mathbb{P}(\boldsymbol{\theta}|\mathbf{X}) = \frac{\mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\mathbf{X}|\boldsymbol{\theta})}{\mathbb{P}(\mathbf{X})},$$

donde $\mathbb{P}(\mathbf{X}) = \int_{\Theta} \mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\mathbf{X}|\boldsymbol{\theta})d\boldsymbol{\theta}$ no depende de $\boldsymbol{\theta}$, por lo que es común escribir:

$$\mathbb{P}(\boldsymbol{\theta}|\mathbf{X}) \propto \mathbb{P}(\boldsymbol{\theta})\mathbb{P}(\mathbf{X}|\boldsymbol{\theta}).$$

La distribución de probabilidad inicial (*a priori*) $\mathbb{P}(\boldsymbol{\theta})$ describe la información inicial que se tiene de $\boldsymbol{\theta}$. Esta distribución se basa en experiencia previa (información histórica, experiencia de expertos en los datos u otra información adicional). Entonces, la distribución de probabilidad final (*a posteriori*) $\mathbb{P}(\boldsymbol{\theta}|\mathbf{X})$ permite incorporar información contenida en la muestra a través de $\mathbb{P}(\mathbf{X}|\boldsymbol{\theta})$ e información inicial a través de $\mathbb{P}(\boldsymbol{\theta})$. Por tanto, en estadística Bayesiana las inferencias sobre el parámetro $\boldsymbol{\theta}$ se basan en calcular la distribución final $\mathbb{P}(\boldsymbol{\theta}|\mathbf{X})$.

En ocasiones el propósito de un análisis estadístico es hacer predicciones de una observación futura X^* . La distribución de probabilidad que describe el comportamiento de una observación futura X^* es $\mathbb{P}(X^*|\boldsymbol{\theta})$, sin embargo $\boldsymbol{\theta}$ es desconocido, así que es necesario estimarlo.

Desde la perspectiva Bayesiana, la distribución marginal de X^* :

$$\mathbb{P}(X^*) = \int_{\Theta} \mathbb{P}(X^*|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta})d\boldsymbol{\theta},$$

conocida como distribución predictiva inicial (*predictiva a priori*), describe la información acerca de X^* dada la información inicial disponible por $\mathbb{P}(\boldsymbol{\theta})$.

Una vez obtenida la muestra \mathbf{X} , la distribución final del parámetro $\mathbb{P}(\boldsymbol{\theta}|\mathbf{X})$ junto con el modelo $\mathbb{P}(X^*|\boldsymbol{\theta})$, inducen una distribución conjunta de $(X^*, \boldsymbol{\theta})$ condicional a los valores observados:

$$\mathbb{P}(X^*, \boldsymbol{\theta}|\mathbf{X}) = \mathbb{P}(X^*|\boldsymbol{\theta}, \mathbf{X})\mathbb{P}(\boldsymbol{\theta}|\mathbf{X}) = \mathbb{P}(X^*|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta}|\mathbf{X}),$$

donde la igualdad $\mathbb{P}(X^*|\boldsymbol{\theta}, \mathbf{X}) = \mathbb{P}(X^*|\boldsymbol{\theta})$ se cumple siempre que haya independencia condicional entre la observación futura X^* y la muestra observada \mathbf{X} , dado el parámetro $\boldsymbol{\theta}$. Por tanto, la distribución de probabilidad:

$$\mathbb{P}(X^*|\mathbf{X}) = \int_{\Theta} \mathbb{P}(X^*|\boldsymbol{\theta})\mathbb{P}(\boldsymbol{\theta}|\mathbf{X})d\boldsymbol{\theta},$$

conocida como distribución predictiva final (*predictiva a posteriori*), describe el comportamiento de X^* dada toda la información disponible de \mathbf{X} y de $\boldsymbol{\theta}$.

2.2. Algoritmos computacionales en estadística Bayesiana. El paradigma Bayesiano ha sido reconocido desde hace mucho tiempo como conceptualmente atractivo, sin embargo, en la práctica su implementación puede ser compleja debido a las dificultades de cálculo, ya que se requiere la integración en altas dimensiones para calcular distribuciones de probabilidad. Recientemente, los métodos Bayesianos se han vuelto populares con la aparición de nuevos algoritmos computacionales que abordan esta integración de manera directa. El desarrollo de métodos, como los métodos de Monte Carlo vía cadenas de Markov (MCMC, *Markov chain Monte Carlo*) (Chen *et al.*, 2000; Gamerman y Lopes, 2006), el algoritmo de aproximación integral anidada de Laplace (INLA, *Integrated Nested Laplace Approximation*) (Rue *et al.*, 2009, 2017), o los métodos de Monte Carlo Hamiltoniano (HMC, *Hamiltonian Monte Carlo*) (Betancourt, 2018), han permitido dar soluciones numéricas para problemas basados en modelos complejos.

La implementación de los algoritmos juega un papel fundamental para la metodología Bayesiana. Para esto, el software debe ser lo suficientemente flexible y potente para realizar los cálculos necesarios. El software R (R Core Team, 2021; Albert, 2009) es uno de los lenguajes de programación más utilizados para el cálculo y gráficas en estadística. Otro proyecto interesante es BUGS (*Bayesian inference Using Gibbs Sampling*) (Lunn *et al.*, 2012), el cual sirve para análisis Bayesiano de modelos complejos utilizando métodos MCMC; WinBUGS (Lunn *et al.*, 2000; Ntzoufras, 2009) y su versión de código abierto OpenBUGS son parte de este proyecto. Recientemente los programas como INLA (Gómez-Rubio, 2020), JAGS (*Just Another Gibbs Sampling*) (Plummer, 2003) y Stan (*Sampling Through Adaptive Neighborhoods*, también acrónimo en honor a Stanislaw Ulam, 1909-1984) (Stan Development Team, 2023, 2020; Kruschke, 2010), se han desarrollado para la estimación Bayesiana.

Stan¹ es una plataforma de última generación para el modelado estadístico y el cálculo estadístico de alto rendimiento. En Stan se especifican funciones de densidad en el lenguaje de programación probabilístico y lo que se obtiene es inferencia estadística Bayesiana. La biblioteca matemática de Stan proporciona funciones de probabilidad diferenciables y álgebra lineal (C++ autodiff). Los paquetes de R adicionales proporcionan modelado lineal basado en expresiones, visualización posterior y validación cruzada.

Ejemplo 1 (Poisson). Los datos de conteo resultan de un número de ensayos y su rango son los números enteros no negativos. En estos casos, y bajo ciertas condiciones, un modelo que puede utilizarse es la distribución Poisson. Sea X una v.a. con distribución Poisson con parámetro $\theta > 0$, y f.d.p.:

$$\mathbb{P}(X = x|\theta) = e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \dots, \text{ para } \theta > 0.$$

¹<http://mc-stan.org>

Suponga que $\mathbf{X} = \{X_1, \dots, X_n\}$ son n variables aleatorias con distribución Poisson(θ). La distribución inicial de θ se puede tomar como una distribución gama con parámetros α_0 y β_0 , cuya f.d.p. está dada por:

$$\mathbb{P}(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0 \theta}, \quad \theta > 0, \alpha_0 > 0, \beta_0 > 0.$$

La verosimilitud asociada a la distribución Poisson es:

$$\mathbb{P}(x_1, \dots, x_n | \theta) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} \propto e^{-n\theta} \theta^{\sum_{i=1}^n x_i}.$$

Luego, la distribución final de θ está dada por:

$$\begin{aligned} (1) \quad \mathbb{P}(\theta | \mathbf{x}) &= \frac{\mathbb{P}(\mathbf{x} | \theta) \mathbb{P}(\theta)}{\mathbb{P}(\mathbf{x})} = \frac{\mathbb{P}(\mathbf{x} | \theta) \mathbb{P}(\theta)}{\int \mathbb{P}(\mathbf{x} | \theta) \mathbb{P}(\theta) d\theta} \\ &= \frac{\left[\prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} \right] \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0 \theta}}{\int \left[\prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} \right] \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0 \theta} d\theta} \\ &= \frac{\frac{1}{\prod_{i=1}^n x_i!} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\beta_0 \theta - n\theta}}{\frac{1}{\prod_{i=1}^n x_i!} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \int \theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\beta_0 \theta - n\theta} d\theta} \\ &= \frac{\theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\beta_0 \theta - n\theta}}{\int \theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\theta(\beta_0 + n)} d\theta} = \frac{\theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\beta_0 \theta - n\theta}}{\frac{\Gamma(\alpha_0 + \sum_{i=1}^n x_i)}{(\beta_0 + n)^{\alpha_0 + \sum_{i=1}^n x_i}}}, \end{aligned}$$

que es igual a una distribución gama con parámetros $\alpha_1 = \alpha_0 + \sum_{i=1}^n x_i$ y $\beta_1 = \beta_0 + n$, específicamente, $p(\theta | \mathbf{x}) = \text{Gama}(\theta | \alpha_1, \beta_1)$.

Note en la antepenúltima igualdad de (1) que los productos que no dependen del parámetro θ , es decir $\frac{1}{\prod_{i=1}^n x_i!} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}$, se cancelan por estar simultáneamente en el numerador y denominador. También note en la última igualdad de (1) que el término en el numerador es una función de θ , es decir $\theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} e^{-\beta_0 \theta - n\theta}$, que es el núcleo (o *kernel*) de una variable aleatoria θ con distribución gama con parámetros α_1 y β_1 . Por último, note que el resultado de la integral en el denominador es $\frac{\Gamma(\alpha_0 + \sum_{i=1}^n x_i)}{(\beta_0 + n)^{\alpha_0 + \sum_{i=1}^n x_i}}$, el cual es la constante de integración de la distribución $\text{Gama}(\theta | \alpha_1, \beta_1)$.

Esto conduce a que podemos calcular la distribución final de θ , omitiendo el producto que no depende del parámetro θ , $\mathbb{P}(\mathbf{x})$, es decir:

$$\begin{aligned} \mathbb{P}(\theta | \mathbf{x}) \propto \mathbb{P}(\mathbf{x} | \theta) \mathbb{P}(\theta) &= \left[\prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!} \right] \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} e^{-\beta_0 \theta} \\ &\propto \theta^{\alpha_0 + \sum_{i=1}^n x_i - 1} \exp[-\theta(\beta_0 + n)] \end{aligned}$$

que corresponde al núcleo de la distribución $\text{Gama}(\theta | \alpha_1, \beta_1)$. Debido a que la distribución inicial y la final pertenecen a la misma familia de distribuciones, se dice que la familia de distribuciones gama es conjugada para la familia de distribuciones Poisson.

La distribución predictiva final está dada por:

$$\begin{aligned} p(x^* | x_1, \dots, x_n) &= \int_0^\infty p(x^* | \theta) p(\theta | x_1, \dots, x_n) d\theta \\ &= \int_0^\infty e^{-\theta} \frac{\theta^{x^*}}{x^*!} \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \theta^{\alpha_1-1} e^{-\beta_1 \theta} d\theta \\ &= \frac{1}{x^*!} \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \int_0^\infty e^{-\theta - \beta_1 \theta} \theta^{x^* + \alpha_1 - 1} d\theta \\ &= \frac{1}{x^*!} \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_1 + x^*)}{(\beta_1 + 1)^{(\alpha_1 + x^*)}}. \end{aligned}$$

Esta distribución se conoce como *Poisson-Gamma*($\alpha_1, \beta_1, 1$).

Considerando un conjunto de $n = 50$ datos simulados a partir de una distribución Poisson($\theta = 3$), se obtienen las estimaciones con métodos Bayesianos usando Stan. Con 2000 iteraciones y un calentamiento de 1000, la Tabla 1 presenta los resultados de estimación de θ y x^* ; como la media, el error estándar de la media (SE media), la desviación estandar (SD), varios cuantiles, el $neff$ es una medida cruda del tamaño efectivo de la muestra, y $Rhat$ es el factor de reducción de escala potencial en cadenas divididas (en convergencia $Rhat=1$). En la Figura 1 se muestran las distribuciones final de θ y predictiva final de x^* . En Robert y Casella (2000) se pueden encontrar más detalles acerca de los métodos de Monte Carlo y del diagnóstico de convergencia.

TABLA 1. Poisson: resultados de la distribución final de θ , y distribución predictiva final de x^* usando Stan.

	Media	SE media	SD	2.5 %	25 %	50 %	75 %	97.5 %	neff	Rhat
θ	2.91	0.01	0.23	2.48	2.76	2.9	3.07	3.36	719	1
x^*	2.90	0.05	1.71	0.00	2.00	3.0	4.00	6.00	1051	1

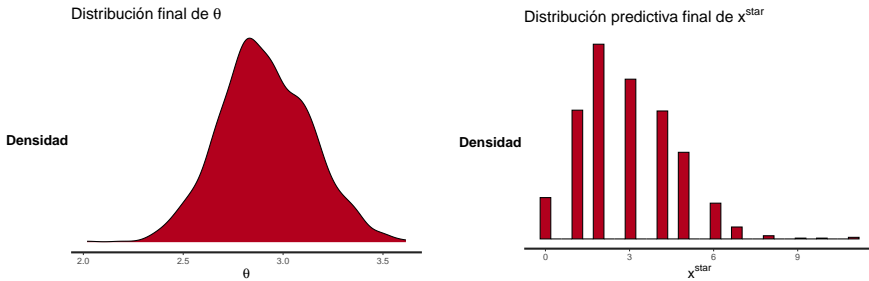


FIGURA 1. Poisson: distribución final de θ , y distribución predictiva final de x^* usando Stan.

3. MODELOS OCULTOS DE MARKOV

Los modelos ocultos de Markov (HMM) son modelos gráficos probabilísticos que representan procesos dinámicos; en particular, modelan cómo el estado de un proceso cambia con el tiempo. Estos modelos se utilizan para describir datos secuenciales o series de tiempo, tales como datos financieros. La idea principal es que un HMM es un modelo estadístico en el cual el sistema a estimar es un proceso de Markov con parámetros desconocidos, en donde la ocurrencia de los estados está asociada con una distribución de probabilidad y las transiciones entre los estados están gobernadas por un conjunto de probabilidades de transición.

Un HMM es un caso particular de las redes Bayesianas dinámicas de primer orden (Sucar, 2015). En Zucchini *et al.* (2016) se aborda el tema de estimación de parámetros, la selección del modelo y la predicción utilizando el lenguaje R. Mientras que en Damiano *et al.* (2018) se utiliza Stan.

Para presentar la teoría básica de los HMM es indispensable proveer primero la teoría de cadenas de Markov y mostrar el concepto de estados ocultos, donde la observación es una función probabilística del estado.

3.1. Cadenas de Markov. Vamos a considerar cadenas de Markov (CM) a tiempo discreto. Una sucesión de variables aleatorias discretas $\{C_t : t \in \mathbb{N}\}$ se dice que es una cadena de Markov (a tiempo discreto) si para todo $t \in \mathbb{N}$ satisface la propiedad de Markov:

$$\mathbb{P}(C_{t+1}|C_t, \dots, C_1) = \mathbb{P}(C_{t+1}|C_t).$$

Esto significa que la probabilidad del siguiente estado solo depende del estado actual.

Otro concepto importante es el de probabilidad de transición. Consideremos un espacio de estados $\{1, \dots, N\}$, y sean i y j dos estados cualesquiera, la probabilidad de transición entre estos dos estados está dada por:

$$\mathbb{P}(C_{s+t} = j | C_s = i).$$

Si estas probabilidades no dependen de s entonces se dice que la CM es homogénea, de otra forma es heterogénea. En este trabajo utilizaremos CM homogéneas y las probabilidades de transición las denotaremos por $\gamma_{ij}(t)$, es decir, $\gamma_{ij}(t) = \mathbb{P}(C_{s+t} = j | C_s = i)$. Sea $\mathbf{\Gamma}(t)$ la matriz que tiene como ij -ésimo elemento a $\gamma_{ij}(t)$. Entonces, si aplicamos la definición anterior sobre todo el conjunto de estados obtendríamos la matriz de transición de un paso $\mathbf{\Gamma}(1)$, denotada abreviadamente como $\mathbf{\Gamma}$, dada por:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1N} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \dots & \gamma_{NN} \end{pmatrix}.$$

Esta matriz satisface lo siguiente:

- (i) $\gamma_{ij} \geq 0$ para todo $i, j \in \{1, \dots, N\}$.
- (ii) $\sum_{j=1}^N \gamma_{ij} = 1, \forall i \in \{1, \dots, N\}$.
- (iii) Ecuación de Chapman-Kolmogorov: $\mathbf{\Gamma}(t+s) = \mathbf{\Gamma}(t)\mathbf{\Gamma}(s)$, lo que implica que $\mathbf{\Gamma}(t) = \mathbf{\Gamma}(1)^t = \mathbf{\Gamma}^t$.

Como los renglones de la matriz $\mathbf{\Gamma}$ suman 1, entonces $\mathbf{\Gamma}\mathbf{1}' = \mathbf{1}'$, es decir, el vector columna $\mathbf{1}'$ es un eigenvector de $\mathbf{\Gamma}$ correspondiente al eigenvalor 1.

Las probabilidades no condicionales $\mathbb{P}(C_t = j)$ de una cadena de Markov que está en un cierto estado j en un tiempo dado t , también son de interés, y se definen como:

$$\mathbf{u}(t) = (\mathbb{P}(C_t = 1), \dots, \mathbb{P}(C_t = N)); \quad t \in \mathbb{N},$$

que cumple que, para el tiempo $t+1$, $\mathbf{u}(t+1) = \mathbf{u}(t)\mathbf{\Gamma}$.

También necesitamos saber la probabilidad de comenzar desde un estado determinado. A esto se le conoce como la distribución de probabilidad inicial y se define de la siguiente manera $\delta_i = \mathbb{P}(C_1 = i)$, y que cumple con $\sum_{i=1}^N \delta_i = 1$. Note que coincide con $\mathbf{u}(1)$.

Ejemplo 2. Supongamos que queremos modelar el comportamiento de un conductor al volante. Los $N = 4$ posibles comportamientos (estados) son:

- Estado 1: Acelerar.
- Estado 2: Velocidad constante.
- Estado 3: Reposo (el motor funciona lentamente pero el vehículo no se mueve).
- Estado 4: Freno.

En la Figura 2, los estados se representan como nodos y las transiciones como flechas. Vemos que a veces el conductor cambia de un estado el vehículo a otro estado, pero otras veces se queda en el mismo estado.

Por tanto, a partir de este modelo, si queremos predecir el estado futuro, lo único que importa es el estado actual del vehículo. Los estados pasados no tienen relación con el estado futuro excepto a través del estado actual (propiedad de Markov).

Dado que no estamos seguros del comportamiento del conductor en un momento dado. Por lo tanto, para modelar esta incertidumbre, el modelo se convierte en un modelo probabilístico. Ver las probabilidades de transición en la Figura 2 marcadas con las flechas naranjas. Por ejemplo, considerando el estado denominado “reposo”, la probabilidad de que el automóvil pase de este estado al estado “acelerar” es de 0.7, y escribimos $\mathbb{P}(\text{“acelerar”} | \text{“reposo”}) = 0.7$. Note que la suma de las probabilidades de transición que salen de cualquier estado dado es 1. Por ejemplo, la suma de las probabilidades de transición que salen del estado “freno”: $\mathbb{P}(\text{“acelerar”} | \text{“freno”}) + \mathbb{P}(\text{“constante”} | \text{“freno”}) + \mathbb{P}(\text{“reposo”} | \text{“freno”}) + \mathbb{P}(\text{“freno”} | \text{“freno”}) = 0.4 + 0.1 + 0.4 + 0.1 = 1$.

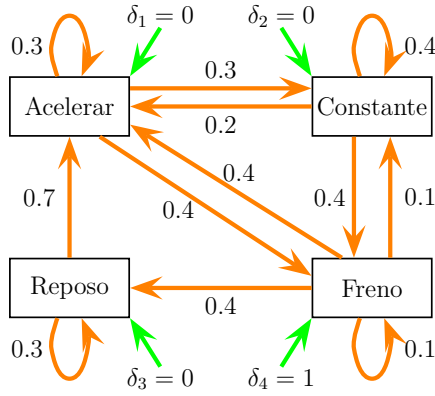


FIGURA 2. Estados del conductor: cadena de Markov que modela el comportamiento de un conductor al volante.

Como sabemos, en este ejemplo, el conductor no puede encender el automóvil en ningún estado, por ejemplo, es imposible encender el automóvil en el estado de “velocidad constante”. Solo puede arrancar el automóvil desde el estado de frenado, luego el vector de probabilidades iniciales está dado por $\delta = (0, 0, 0, 1)$. En la Figura 2, las probabilidades iniciales se indican con flechas verdes.

Supongamos que la actividad del conductor al tiempo t está caracterizada por un solo estado (de los cuatro anteriores). Luego, la matriz de transición está dada por:

$$\Gamma = \begin{matrix} & \begin{matrix} Acelerar & Constante & Reposo & Freno \end{matrix} \\ \begin{matrix} Acelerar \\ Constante \\ Reposo \\ Freno \end{matrix} & \begin{pmatrix} 0.3 & 0.3 & 0 & 0.4 \\ 0.2 & 0.4 & 0 & 0.4 \\ 0.7 & 0 & 0.3 & 0 \\ 0.4 & 0.1 & 0.4 & 0.1 \end{pmatrix} \end{matrix}.$$

Dado que al tiempo $t = 1$ el conductor está en *Freno*, una pregunta que nos haríamos sería: ¿cuál es la probabilidad que el conductor realice el siguiente comportamiento en los próximos instantes $O = \text{“freno, freno, freno, acelerar, acelerar, freno, constante, freno”}$. Esta probabilidad se puede expresar como:

$$\begin{aligned} \mathbb{P}(O|Modelo) &= \mathbb{P}(freno) \cdot \mathbb{P}(freno|freno) \cdot \mathbb{P}(freno|freno) \\ &\cdot \mathbb{P}(acelerar|freno) \cdot \mathbb{P}(acelerar|acelerar) \cdot \mathbb{P}(freno|acelerar) \\ &\cdot \mathbb{P}(constante|freno) \cdot \mathbb{P}(freno|constante) \\ &= \delta_4 \cdot \gamma_{44} \cdot \gamma_{44} \cdot \gamma_{41} \cdot \gamma_{11} \cdot \gamma_{14} \cdot \gamma_{42} \cdot \gamma_{24} \\ &= 1 \cdot 0.1 \cdot 0.1 \cdot 0.4 \cdot 0.3 \cdot 0.4 \cdot 0.1 \cdot 0.4 \\ &= 0.0000192. \end{aligned}$$

Todos los estados son recurrentes, esto significa que con probabilidad 1 se pueden visitar estos estados toda vez que la cadena comenzó en ellos. Esta cadena de Markov no tiene estados transitorios, ni tampoco estados absorbentes. En la Tabla 2 se presentan la distribución estacionaria (distribución que no cambia al transcurrir el tiempo, $\pi = \pi\Gamma$) y el tiempo medio de recurrencia $1/\pi$. Note que aproximadamente el 36 % del tiempo el conductor acelera, el 26 % frena, el 23 % se mantiene constante y un 15 % está en reposo.

TABLA 2. Estados del conductor: algunos datos de la Cadena de Markov.

	Acelerar	Constante	Reposo	Freno
Distribución estacionaria	0.3634476	0.2253375	0.1495327	0.2616822
Tiempo medio de recurrencia	2.7514290	4.4377880	6.6875000	3.8214290

Ahora considere que tiene una muestra observada de los estados realizados por el conductor, ver Figura 3, y como ocurre en la vida real, tenemos los datos observados pero desconocemos la matriz de transición. Usando los métodos de estadística Bayesiana y Stan podemos estimar los parámetros del modelo. La Figura 4 muestra las estimaciones de los parámetros que representan las probabilidades en la matriz de transición. Note que, además del valor medio estimado, podemos obtener otras cantidades, por ejemplo, los intervalos de probabilidad del 95 % correspondientes a los cuantiles 0.025 y 0.975 de la distribución posterior. Note que por la naturaleza del método, y por el tamaño de muestra, los estimadores de los parámetros no son exactamente iguales a los valores reales, pero buscamos que sean lo más similares.

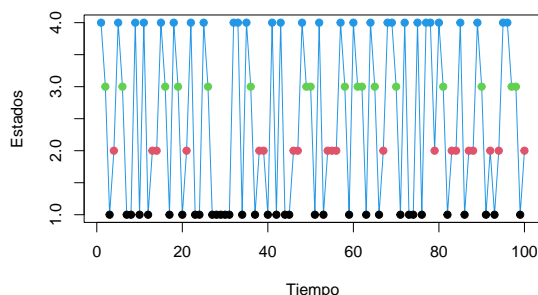


FIGURA 3. Estados del conductor: cadena de Markov simulada.

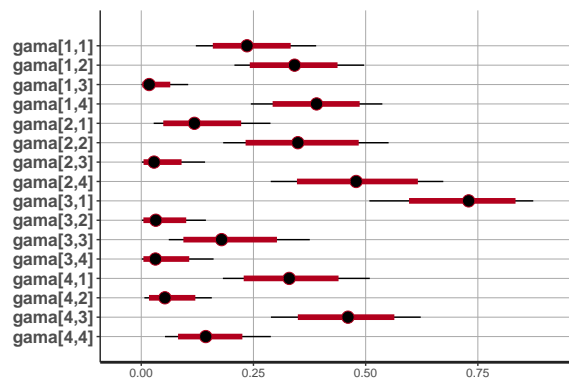


FIGURA 4. Estados del conductor: estimaciones medias e intervalos de probabilidad del 95 % de las probabilidades de la matriz de transición.

3.2. Definición de HMM. Un modelo oculto de Markov (HMM, *Hidden Markov Model*) consiste en un proceso doblemente estocástico. El primer proceso estocástico es una cadena de Markov que se caracteriza por estados y probabilidades de transición, llamado *proceso paramétrico*. Los estados de la cadena no son observables, por lo tanto, están “ocultos” y los denotaremos por $\{C_t : t = 1, 2, \dots\}$. El segundo proceso estocástico se define con las observaciones que tenemos, a partir de la distribución de probabilidad que depende del estado oculto, lo denotaremos por $\{X_t : t = 1, 2, \dots\}$, y se le llama *proceso dependiente del estado*. Si la cadena de Markov $\{C_t\}$ tiene N estados, decimos que $\{X_t\}$ es un N -estado HMM.

La Figura 5 muestra la estructura de los HMM usando una gráfica dirigida acíclica. Los nodos cuadrados representan variables observadas y los óvalos representan variables latentes². La dirección de las flechas indica dependencia condicional.

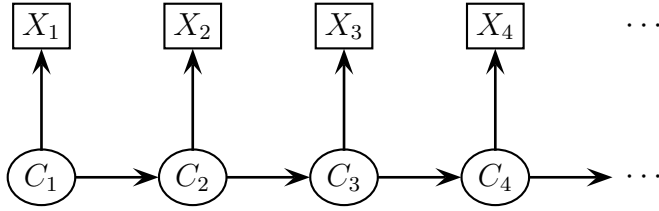


FIGURA 5. Representación gráfica de un modelo oculto de Markov.

La propiedad de Markov sólo se asume para los estados ocultos, y las observaciones son condicionalmente independientes dados los estados ocultos. Mientras que las observaciones pueden no exhibir un comportamiento Markoviano, la estructura Markoviana de los estados ocultos es suficiente para facilitar su inferencia estadística.

3.2.1. *Elementos del HMM.* Los elementos que constituyen un HMM con espacio de observaciones discreto son cinco, y se muestran a continuación:

1. Espacio de estados: $\mathcal{C} = \{1, 2, \dots, N\}$. El estado en el tiempo t se denota por C_t .
2. Espacio de observaciones: $\{X_t : t = 1, \dots, T\}$.
3. Vector de probabilidades iniciales: $\boldsymbol{\delta}$, con i -ésimo elemento:

$$\delta_i = \mathbb{P}(C_1 = i), \quad i \in \mathcal{C}.$$

4. Matriz de transición a un paso: $\boldsymbol{\Gamma} = \{\gamma_{ij}\}_{i,j \in \mathcal{C}}$, definidas por:

$$\gamma_{ij} = \mathbb{P}(C_{t+1} = j | C_t = i).$$

5. Distribución de probabilidad de las observaciones: $\boldsymbol{P} = \{p_i(x)\}_{i \in \mathcal{C}}$, donde $p_i(x)$ es la probabilidad de que la observación x sea emitida por el estado i , i.e.:

$$p_i(x) = \mathbb{P}(X_t = x | C_t = i).$$

Por lo que un HMM se representa por $\boldsymbol{\lambda} = \{\boldsymbol{\Gamma}, \boldsymbol{P}, \boldsymbol{\delta}\}$.

Para el caso univariado, y teniendo observaciones discretas con $u_i(t) = \mathbb{P}(C_t = i)$, para $t = 1, \dots, T$,

$$\begin{aligned} \mathbb{P}(X_t = x) &= \sum_{i=1}^N \mathbb{P}(C_t = i) \mathbb{P}(X_t = x | C_t = i) \\ &= \sum_{i=1}^N u_i(t) p_i(x) \\ &= \mathbf{u}(t) \boldsymbol{P}(x) \mathbf{1}', \end{aligned}$$

con $\boldsymbol{P}(x)$ la matriz diagonal con i -ésimo elemento $p_i(x)$ y $\mathbf{1}$ el vector fila de unos de dimensión apropiada. Como $\mathbf{u}(t) = \mathbf{u}(1) \boldsymbol{\Gamma}^{t-1}$, entonces:

$$\mathbb{P}(X_t = x) = \mathbf{u}(1) \boldsymbol{\Gamma}^{t-1} \boldsymbol{P}(x) \mathbf{1}'.$$

Más aún, si consideramos la cadena de Markov estacionaria, con distribución estacionaria $\boldsymbol{\delta}$, i.e., $\boldsymbol{\delta} \boldsymbol{\Gamma}^{t-1} = \boldsymbol{\delta}$, y considerando que $\mathbf{u}(1)$ coincide con la distribución estacionaria $\boldsymbol{\delta}$, entonces:

$$\mathbb{P}(X_t = x) = \boldsymbol{\delta} \boldsymbol{P}(x) \mathbf{1}'.$$

²Las variables latentes son variables que no se observan directamente, pero que son inferidas a partir de otras variables observadas.

La función de verosimilitud está dada por:

$$\begin{aligned} L_T &= \sum_{c_1, \dots, c_T=1}^N (\delta_{c_1} \gamma_{c_1, c_2} \cdots \gamma_{c_{T-1}, c_T}) (p_{c_1}(x_1) p_{c_2}(x_2) \cdots p_{c_T}(x_T)) \\ &= \sum_{c_1, \dots, c_T=1}^N \delta_{c_1} p_{c_1}(x_1) \gamma_{c_1, c_2} p_{c_2}(x_2) \cdots \gamma_{c_{T-1}, c_T} p_{c_T}(x_T) \\ &= \boldsymbol{\delta} \mathbf{P}(x_1) \boldsymbol{\Gamma} \mathbf{P}(x_2) \boldsymbol{\Gamma} \mathbf{P}(x_3) \cdots \boldsymbol{\Gamma} \mathbf{P}(x_T) \mathbf{1}' \end{aligned}$$

Como $\boldsymbol{\delta}$, la distribución de C_1 , es la distribución estacionaria de la cadena de Markov, entonces $\boldsymbol{\delta} \mathbf{P}(x_1) = \boldsymbol{\delta} \boldsymbol{\Gamma} \mathbf{P}(x_1)$, y por lo tanto, la verosimilitud es:

$$L_T = \boldsymbol{\delta} \boldsymbol{\Gamma} \mathbf{P}(x_1) \boldsymbol{\Gamma} \mathbf{P}(x_2) \boldsymbol{\Gamma} \mathbf{P}(x_3) \cdots \boldsymbol{\Gamma} \mathbf{P}(x_T) \mathbf{1}'.$$

Puede encontrar más información del modelo HMM en Zucchini *et al.* (2016).

Ejemplo 3. En la universidad, Karina y Daniel fueron compañeros, ahora ambos están graduados y viven en distintas ciudades, Karina por su parte vive en Ciudad de México y Daniel en Monterrey; por lo que su comunicación es vía telefónica.

Resulta que el estado de ánimo de Daniel cambia según el estado del tiempo. Si está soleado (S), Daniel está mayormente feliz (F), pero si está lluvioso (L), entonces Daniel está mayormente enojado (E).

Cierto día, hablando por teléfono, Daniel le comenta a Karina que estaba feliz, por lo que ella infirió que hacía sol en Monterrey. El clima en Monterrey está oculto para Karina y la única información que tiene de Daniel es si está feliz o enojado (observación). Así pues, el espacio de estados ocultos está dado por $\mathcal{C} = \{S, L\}$ y el espacio de observaciones tomará valores del conjunto $\{F, E\}$.

Supongamos que el vector de probabilidades iniciales está dado por:

$$\delta_S = \mathbb{P}(C_1 = S) = 0.83, \quad \delta_L = \mathbb{P}(C_1 = L) = 0.17.$$

Además, supongamos que si hoy es un día soleado, la probabilidad de que mañana también lo sea es de 0.9, mientras que existe una probabilidad de 0.1 de que sea un día lluvioso. Por otro lado, si hoy es lluvioso, la probabilidad de que mañana también lo sea es de 0.6, por lo que la probabilidad de que el día siguiente sea soleado se convierte en 0.4. Luego, las transiciones a un paso (de un día a otro) están dadas por:

$$\begin{aligned} \gamma_{SS} &= \mathbb{P}(C_{t+1} = S | C_t = S) = 0.9, & \gamma_{SL} &= \mathbb{P}(C_{t+1} = L | C_t = S) = 0.1, \\ \gamma_{LS} &= \mathbb{P}(C_{t+1} = S | C_t = L) = 0.4, & \gamma_{LL} &= \mathbb{P}(C_{t+1} = L | C_t = L) = 0.6. \end{aligned}$$

Considerando los estados de ánimo de Daniel dependiendo del estado del tiempo, se traducen en las probabilidades de las observaciones dadas por:

$$\begin{aligned} p_S(F) &= \mathbb{P}(F|S) = 0.8, & p_L(F) &= \mathbb{P}(F|L) = 0.3, \\ p_S(E) &= \mathbb{P}(E|S) = 0.2, & p_L(E) &= \mathbb{P}(E|L) = 0.7. \end{aligned}$$

La Figura 6 muestra los elementos del HMM usando una representación gráfica. Los nodos cuadrados representan los estados de ánimo de Daniel (variables observadas) y los óvalos representan los estados del tiempo de Monterrey (variables latentes), que son los estados no observados para Karina. Las flechas verdes denotan las probabilidades iniciales, las flechas naranjas indican las probabilidades de transición, y las flechas azules denotan las probabilidades de las observaciones.

Si durante cinco días Daniel estuvo feliz, enojado, enojado, feliz, feliz, entonces una cuestión interesane es saber cuál fue el clima (estado oculto) en estos cinco días en Monterrey.

4. LOS TRES PROBLEMAS BÁSICOS PARA LOS HMM

La dificultad de los HMM es estimar los parámetros $\{\boldsymbol{\Gamma}, \mathbf{P}, \boldsymbol{\delta}\}$; además debemos asegurar que estamos tomando la mejor sucesión de estados y tener los parámetros que maximicen la verosimilitud del modelo. Entonces podemos hacer uso de al menos

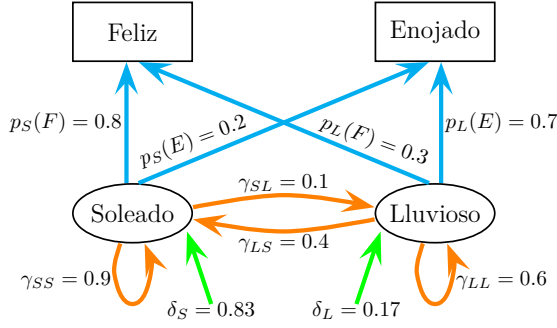


FIGURA 6. Elementos del HMM: representa la relación entre los estados de ánimo y los estados del tiempo.

tres algoritmos para solucionar estos problemas. Los problemas básicos para aplicar un HMM son:

1. Problema 1: *Evaluación*. Dada una secuencia de observaciones y un modelo, ¿cuál es la probabilidad de la secuencia observada dado el modelo?
2. Problema 2: *Decodificación*. Dada una secuencia de observaciones y el modelo, ¿cómo elegimos su correspondiente secuencia de estados ocultos?
3. Problema 3: *Aprendizaje*. ¿Cómo ajustamos los parámetros del modelo para maximizar la probabilidad de las observaciones dado el modelo?

Estos problemas se han podido resolver con los algoritmos *forward-backward*, de Viterbi y de Baum-Welch, respectivamente. Para detalles ver Rabiner (1989) y Zucchini *et al.* (2016).

Es importante enfatizar que lo que se describe a continuación es una explicación general de estos tres problemas, y de los algoritmos que dan soluciones. Para la estimación de los parámetros se pueden utilizar técnicas de estadística Bayesiana, como lo ejemplificaremos en la sección 5 para HMM en series de tiempo financieras.

4.1. El problema de la evaluación. Suponga que se tiene un modelo oculto de Markov $\lambda = \{\Gamma, P, \delta\}$ y un serie de observaciones $X = \{x_1, \dots, x_T\}$; se desea calcular la probabilidad de que la serie observada sea producida por el modelo, es decir, calcular $\mathbb{P}(X|\lambda)$. Note que X se puede generar por una serie diferente de estados $C = \{c_1, \dots, c_T\}$:

$$\begin{aligned} \mathbb{P}(X|\lambda) &= \sum_{\text{todo } C} \mathbb{P}(X, C|\lambda) \\ &= \sum_{\text{todo } C} \mathbb{P}(X|C, \lambda)\mathbb{P}(C|\lambda) \\ &= \sum_{c_1} \dots \sum_{c_T} \mathbb{P}(x_1, \dots, x_T|c_1, \dots, c_T, \lambda)\mathbb{P}(c_1, \dots, c_T|\lambda), \end{aligned}$$

donde, suponiendo independencia condicional de las observaciones y de los estados ocultos:

$$\begin{aligned} \mathbb{P}(x_1, \dots, x_T|c_1, \dots, c_T, \lambda) &= \mathbb{P}(x_1|c_1, \lambda)\mathbb{P}(x_2|c_2, \lambda) \dots \mathbb{P}(x_T|c_T, \lambda) \\ &= p_{c_1}(x_1)p_{c_2}(x_2) \dots p_{c_T}(x_T), \end{aligned}$$

y

$$\begin{aligned} \mathbb{P}(c_1, \dots, c_T|\lambda) &= \mathbb{P}(c_1|\lambda)\mathbb{P}(c_2|c_1, \lambda) \dots \mathbb{P}(c_T|c_{T-1}, \lambda) \\ &= \delta_{c_1}\gamma_{c_1c_2} \dots \gamma_{c_{T-1}c_T}, \end{aligned}$$

por lo tanto:

$$\mathbb{P}(X, C|\lambda) = \delta_{c_1}p_{c_1}(x_1)\gamma_{c_1c_2}p_{c_2}(x_2) \dots \gamma_{c_{T-1}c_T} \dots p_{c_T}(x_T).$$

Entonces, para un modelo con N estados y T observaciones, se tendrían N^T series de estados posibles. Esta evaluación requiere una serie de operaciones del orden de $2T \cdot N^T$, lo cual lo hace ineficiente de calcular directamente, por lo que conviene recurrir a un algoritmo que optimice el proceso. Para resolver el problema de evaluación se puede utilizar el algoritmo forward-backward.

4.1.1. Algoritmo Forward-Backward. El algoritmo forward-backward se basa en utilizar variables auxiliares que pueden obtenerse de manera recursiva para agilizar el procedimiento para calcular $\mathbb{P}(X|\boldsymbol{\lambda})$.

Considere la variable *forward* $\alpha_t(i)$ definida como:

$$\alpha_t(i) = \mathbb{P}(x_1, x_2, \dots, x_t, c_t = i | \boldsymbol{\delta}).$$

Se puede resolver $\alpha_t(i)$ por inducción, como se muestra a continuación:

1. Inicializar: para $i = 1, \dots, N$,

$$\alpha_1(i) = \mathbb{P}(c_1 = i | \boldsymbol{\delta}) \mathbb{P}(x_1 | c_1 = i, \boldsymbol{\delta}) = \delta_i p_i(x_1).$$

2. Inducción: para $j = 1, \dots, N$ y $t = 1, \dots, T - 1$,

$$\begin{aligned} \alpha_{t+1}(j) &= \sum_{i=1}^N \mathbb{P}(x_1, x_2, \dots, x_{t+1}, c_{t+1} = j, c_t = i | \boldsymbol{\delta}) \\ &= \sum_{i=1}^N \mathbb{P}(x_1, x_2, \dots, x_t, c_t = i, \boldsymbol{\delta}) \mathbb{P}(c_{t+1} = j | c_t = i, \boldsymbol{\delta}) \mathbb{P}(x_{t+1} | c_{t+1} = j, \boldsymbol{\delta}) \\ &= \sum_{i=1}^N \alpha_t(i) \gamma_{ij} p_j(x_{t+1}). \end{aligned}$$

3. Terminación:

$$\mathbb{P}(X|\boldsymbol{\lambda}) = \sum_{i=1}^N \mathbb{P}(x_1, \dots, x_T, c_T = i | \boldsymbol{\lambda}) = \sum_{i=1}^N \alpha_T(i).$$

Esto requiere una serie de operaciones del orden de N^2T , lo que lo hace más eficiente.

De manera análoga se puede considerar la variable *backward* $\beta_t(i)$, definida como:

$$\beta_t(i) = \mathbb{P}(x_{t+1}, x_{t+2}, \dots, x_T | c_t = i, \boldsymbol{\lambda}),$$

y se puede resolver de manera inductiva:

1. Inicializar: para $i = 1, \dots, N$,

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

2. Inducción: para $i = 1, \dots, N$ y $t = T - 1, T - 2, \dots, 1$,

$$\begin{aligned} \beta_t(i) &= \mathbb{P}(x_{t+1}, x_{t+2}, \dots, x_T, c_{t+1} = j | c_t = i, \boldsymbol{\lambda}) \\ &= \sum_{j=1}^N \mathbb{P}(c_{t+1} = j | c_t = i, \boldsymbol{\lambda}) \mathbb{P}(x_{t+1} | c_{t+1} = j, \boldsymbol{\lambda}) \mathbb{P}(x_{t+2}, x_{t+3}, \dots, x_T | c_{t+1} = j, \boldsymbol{\lambda}) \\ &= \sum_{j=1}^N \gamma_{ij} p_j(x_{t+1}) \beta_{t+1}(j). \end{aligned}$$

Sin embargo, se puede tener una solución óptima al problema operando ambos resultados:

$$\mathbb{P}(X|\boldsymbol{\lambda}) = \sum_{i=1}^N \alpha_t(i) \beta_t(i).$$

4.2. El problema de la decodificación. Suponga que se tiene un HMM λ y una serie de observaciones X , la pregunta es: ¿cuál es la serie de estados ocultos óptima asociada con la serie de observaciones?

La dificultad radica en la definición de la serie óptima, debido a que hay varios criterios de optimización posibles. Una forma de elegirla es utilizar la serie de estados más probable; para esto es necesario saber en cada instante del tiempo t , $t = 1, \dots, T$, cuál es el estado más probable. Para esto se define la variable auxiliar que $\gamma_t(i)$, para $i = 1, \dots, N$:

$$\gamma_t(i) = \mathbb{P}(c_t = i | X, \lambda),$$

y usando las variables del algoritmo forward-backward:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\mathbb{P}(X|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}.$$

Entonces, de manera individual, el estado oculto óptimo al tiempo t , $t = 1, \dots, T$, es:

$$c_t = \arg \max_{i \in \{1, \dots, N\}} \gamma_t(i).$$

La forma de encontrar el estado más probable i_t^* para cada observación al tiempo t está dada por:

$$i_t^* = \arg \max_{i \in \{1, \dots, N\}} \xi_t(i).$$

Una vez obtenida la decodificación para cada observación, se puede establecer la sucesión (trayectoria) de estados ocultos óptima. Sin embargo, tiene dos inconvenientes: se requiere hacer N^T operaciones, y no toma en cuenta las probabilidades de transición, lo que implica que no necesariamente sea sucesión óptima. La solución para el problema de decodificación es utilizar el algoritmo de Viterbi.

4.2.1. Algoritmo de Viterbi. El algoritmo de Viterbi hace una recursión entre un tiempo t y el tiempo $t - 1$ para optimizar el proceso de encontrar la ruta más probable de estados ocultos. Se define $\xi_t(i)$ como:

$$\xi_t(i) = \max_{c_1, c_2, \dots, c_{t-1}} \mathbb{P}(c_1, c_2, \dots, c_{t-1}, c_t = i, x_1, x_2, \dots, x_t | \lambda),$$

que es la trayectoria con mayor probabilidad al tiempo t , para el estado i . Por inducción se tiene que para $t = 1$:

$$\begin{aligned} \xi_1(i) &= \mathbb{P}(c_1 = i, x_1) = \delta_i p_i(x_1), \\ \xi_2(j) &= \max_{c_1} \mathbb{P}(c_1, c_2 = j, x_1, x_2 | \lambda) \\ &= [\max_i \mathbb{P}(c_1 = i, x_1, \lambda) \mathbb{P}(c_2 = j | c_1 = i, \lambda)] \mathbb{P}(x_2 | c_2 = j, \lambda) \\ &= [\max_i \xi_1(i) \gamma_{ij}] p_j(x_2), \end{aligned}$$

y en general, para $t = 2, \dots, T$:

$$\xi_{t+1}(j) = \left[\max_i \xi_t(i) \gamma_{ij} \right] p_j(x_{t+1}).$$

El algoritmo consta de los siguientes pasos:

1. Inicializar: para $i = 1, \dots, N$,

$$\xi_1(i) = \delta_i p_i(x_1) \quad \text{y} \quad \psi_1(i) = 0.$$

2. Recursión: para $t = 2, \dots, T$ y $j = 1, \dots, N$,

$$\begin{aligned} \xi_t(j) &= \left[\max_i \xi_{t-1}(i) \gamma_{ij} \right] p_j(x_t), \\ \psi_t(j) &= \arg \max_i \xi_t(i) \gamma_{ij}. \end{aligned}$$

3. Terminación:

$$\begin{aligned} p^* &= \max_i \xi_T(i), \\ q_T^* &= \arg \max_i \xi_T(i). \end{aligned}$$

4. Trayectoria de retroceso (secuencia de estados): para $t = T - 1, T - 2, \dots, 2, 1$,

$$q_t^* = \psi_{t+1}(q_{t+1}^*).$$

4.3. El problema del aprendizaje. El problema del aprendizaje es determinar un método adecuado para la estimación de los parámetros $\boldsymbol{\lambda}$ del modelo HMM, para maximizar la probabilidad de las observaciones, es decir:

$$\arg \max_{\boldsymbol{\lambda}} \mathbb{P}(X|\boldsymbol{\lambda}).$$

Note que se supone que el número de estados ocultos del modelo ya está dado, por lo que se puede probar el modelo con diferente número de estados y determinar cuál es el mejor tomando en cuenta diferentes criterios.

Para resolver el problema de aprendizaje no existe una fórmula analítica. Entonces, el problema se resuelve usando un procedimiento iterativo, como el algoritmo de Baum-Welch, que se basa en una estructura del algoritmo EM (*expectation-maximization*).

El algoritmo EM es un método iterativo que calcula estimadores máximo verosímiles considerando datos faltantes. El algoritmo EM se define en dos pasos:

1. Paso E (*expectation*): se calcula la esperanza de la log-verosimilitud con respecto a la distribución de los datos faltantes, dada una aproximación de la solución.
2. Paso M (*maximization*): se calcula una nueva aproximación maximizando la log-verosimilitud que resultó del paso E.

Estos pasos se realizan hasta que el algoritmo converge. Ver detalles en Dempster *et al.* (1977).

4.3.1. Algoritmo de Baum-Welch. El algoritmo de Baum-Welch está diseñado para estimar los parámetros de un HMM. Se define la variable auxiliar $v_t(i, j)$:

$$v_t(i, j) = \mathbb{P}(c_t = i, c_{t+1} = j | X, \boldsymbol{\lambda}),$$

y usando las variables forward y backward se obtiene que:

$$\begin{aligned} v_t(i, j) &= \frac{\alpha_t(i) \gamma_{ij} p_j(x_{t+1}) \beta_{t+1}(j)}{\mathbb{P}(X|\boldsymbol{\lambda})} \\ &= \frac{\alpha_t(i) \gamma_{ij} p_j(x_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^N \sum_{l=1}^N \alpha_t(k) \gamma_{kl} p_l(x_{t+1}) \beta_{t+1}(l)}, \end{aligned}$$

y defina $u_t(j)$ como la probabilidad de estar en el estado i al tiempo t dadas las observaciones y el modelo, es decir:

$$u_t(i) = \mathbb{P}(c_t = i | X, \boldsymbol{\lambda}) = \sum_{j=1}^N v_t(i, j).$$

Note que se pueden interpretar algunas cantidades: $\sum_{t=1}^{T-1} u_t(i)$ es el número esperado de visitar al estado oculto i , y $\sum_{t=1}^{T-1} v_t(i, j)$ es el número esperado de transiciones del estado i al estado j . Esto implica que se pueden estimar los parámetros de interés como:

$$(2) \quad \hat{\delta}_i = u_1(i),$$

$$(3) \quad \hat{\gamma}_{ij} = \frac{\sum_{t=1}^{T-1} v_t(i, j)}{\sum_{t=1}^{T-1} u_t(i)},$$

$$(4) \quad \hat{p}_i(x) = \frac{\sum_{t=1}^T u_t(i) \mathbb{I}_{(x_t=x)}}{\sum_{t=1}^T u_t(i)}.$$

Se estiman los parámetros del modelo HMM como $\hat{\lambda} = \{\hat{\Gamma}, \hat{\mathbf{P}}, \hat{\delta}\}$, tal que maximizan la log-verosimilitud a partir de las observaciones dadas y el número de estados ocultos.

Note que (4) depende de la distribución que se suponga para los datos observados, por ejemplo, si se supone una distribución Poisson para el valor observado x :

$$p_j(x) = \exp(-\theta_j)\theta_j^x \frac{1}{x!},$$

y el estimador máximo verosímil de θ_j es:

$$(5) \quad \hat{\theta}_j = \frac{\sum_{t=1}^T \hat{u}_t(j)x_t}{\sum_{t=1}^T \hat{u}_t(j)}.$$

Algoritmo 1 Algoritmo de Baum-Welch

- 1: Iniciar aleatoriamente los parámetros del HMM $\lambda = \{\Gamma, \mathbf{P}, \delta\}$.
 - 2: Calcular los estimadores máximo verosímiles $\hat{\delta}_i$, $\hat{\gamma}_{ij}$ y $\hat{p}_i(x)$ usando (2),(3) y (4), respectivamente.
 - 3: Re-estimar usando los estimadores máximo verosímiles de los parámetros de la distribución $p_j(x)$, por ejemplo, para la distribución Poisson calcular $\hat{\theta}_j$ usando (5).
 - 4: Se obtienen las estimaciones $\hat{\lambda} = \{\hat{\Gamma}, \hat{\mathbf{P}}, \hat{\delta}\}$ hasta que el algoritmo converge, tal que maximicen la log-verosimilitud bajo las condiciones especificadas.
-

5. SERIES DE TIEMPO FINANCIERAS CON HMM

En la práctica cuando observamos una serie de tiempo, por ejemplo, el índice de precios y cotizaciones, el proceso “verdadero” que rige dicha serie no se conoce. La idea detrás de la teoría de series temporales es el estudio y construcción de modelos, $\{X_t\}_{t=1}^{\infty}$ que nos ayuden a entender las características de dependencia (la dinámica) de la serie observada.

5.1. Series de tiempo financieras. Sea \mathcal{T} un conjunto de índices. Consideremos el proceso estocástico $\{X_t\}_{t \in \mathcal{T}}$ definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y con valores en (E, \mathcal{E}) . Una serie temporal es una sucesión de observaciones ordenadas y equidistantes cronológicamente sobre una característica de una unidad observable en diferentes momentos. A dicha sucesión de observaciones se le puede ver como una realización de un proceso estocástico a tiempo discreto, i.e., $\{x_t : t = 1, \dots, n\}$ donde $x_t := X_t(\omega)$ con $\omega \in \Omega$ fijo.

Los modelos para series temporales se pueden definir a través de una representación estocástica o mediante suposiciones distribucionales acerca de las variables aleatorias que representan una serie dada. Por ejemplo, el modelo autorregresivo de orden 1, AR(1), se puede definir por $X_t = \theta X_{t-1} + \epsilon_t$ donde $\epsilon_t \sim N(0, \sigma^2)$; y de manera distribucional, se podría definir el mismo modelo AR(1) mediante una sucesión de variables aleatorias $\{X_t\}_{t=1}^{\infty}$, con distribuciones condicionales $X_t|X_{t-1} \sim N(\theta X_{t-1}, \sigma^2)$.

En el caso particular de datos financieros, muchas series de tiempo financieras exhiben los efectos llamados “agrupamiento de volatilidad” (*volatility clustering*) y “cambio de régimen” (*regime-switching*). La volatilidad es la desviación estándar de los rendimientos de la serie financiera, por tanto, en muchas series de tiempo financieras lo que interesa es modelar la volatilidad. El agrupamiento de volatilidad ocurre cuando los periodos de actividad significativa tienden a ocurrir muy juntos, lo que sugiere que existe algún tipo de memoria a corto plazo para la volatilidad. El cambio de régimen ocurre cuando se observa que los mercados financieros atraviesan periodos alternos entre auge de baja volatilidad y caídas de alta volatilidad.

Los modelos GARCH (modelos de heterocedasticidad condicional autorregresiva generalizada, *generalized autoregressive conditional heterocedasticity*) (Bollerslev, 1986,

2023) se usan generalmente para capturar el fenómeno de agrupamiento de volatilidad y han sido ampliamente estudiados en la literatura econométrica. Los modelos Markov-switching GARCH (MS-GARCH) (Haas *et al.*, 2004; Ardia *et al.*, 2019) permiten capturar el fenómeno de cambio de régimen en la dinámica de la varianza condicional de series de tiempo, usando un HMM para cambiar entre dos procesos latentes GARCH. Mostraremos cómo ambos modelos se pueden estimar usando **Stan**.

5.2. Modelo GARCH. Un proceso $\{X_t\}_{t \in \mathcal{T}}$ es un ARCH(1) (modelo con heterocedasticidad condicional autorregresiva de orden 1, *autoregressive conditional heteroscedasticity*) si:

$$\begin{aligned}\mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0, \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2,\end{aligned}$$

donde \mathcal{F}_{t-1} denota a la σ -álgebra generada por las realizaciones hasta el instante $t-1$; con $\alpha_0 > 0$, $\alpha_1 \geq 0$, y $Z_t = X_t/\sigma_t$ donde $\mathbb{E}(Z_t) = 0$ y $\text{Var}(Z_t) = 1$, y la restricción $\alpha_1 < 1$. Generalmente, el modelo ARCH(1) se representa como:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2, \\ X_t &= Z_t \sigma_t.\end{aligned}$$

Si $Z_t \sim N(0, 1)$ entonces $X_t | \mathcal{F}_{t-1} \sim N(0, \sigma_t^2)$.

De manera general, el proceso $\{X_t\}_{t \in \mathcal{T}}$ es un ARCH(q) si:

$$\begin{aligned}\mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0, \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= \sigma_t^2 = \alpha_0 + \sum_{j=1}^q \alpha_j X_{t-j}^2,\end{aligned}$$

con $\alpha_0 > 0$, $\alpha_1 \geq 0, \dots, \alpha_q \geq 0$, y $Z_t = X_t/\sigma_t$ donde $\mathbb{E}(Z_t) = 0$ y $\text{Var}(Z_t) = 1$. Generalmente, el modelo ARCH(p) se representa como:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_q X_{t-q}^2, \\ X_t &= Z_t \sigma_t.\end{aligned}$$

Así, una generalización de los modelos ARCH son los modelos GARCH, en los cuales la varianza condicional no sólo depende de los cuadrados de las perturbaciones, sino que además también depende de las varianzas condicionales de periodos anteriores, es decir, de σ_t^2 pasados.

El proceso $\{X_t\}_{t \in \mathcal{T}}$ es un GARCH(p, q) si cumple con:

$$\begin{aligned}\mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0, \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2,\end{aligned}$$

con $\alpha_0 > 0$, $\alpha_1 \geq 0, \dots, \alpha_q \geq 0$, $\beta_1 \geq 0, \dots, \beta_p \geq 0$ y $Z_t = X_t/\sigma_t$ con $\mathbb{E}(Z_t) = 0$ y $\text{Var}(Z_t) = 1$. Usualmente, el modelo GARCH(p, q) se escribe como:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_q X_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2, \\ X_t &= Z_t \sigma_t.\end{aligned}$$

Por ejemplo, el caso específico de los modelos GARCH(1,1), está definido por un proceso estocástico $\{X_t\}_{t \in \mathcal{T}}$ que cumple con:

$$\begin{aligned}\mathbb{E}(X_t | \mathcal{F}_{t-1}) &= 0, \\ \text{Var}(X_t | \mathcal{F}_{t-1}) &= \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,\end{aligned}$$

con $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ y $Z_t = X_t/\sigma_t$ con $\mathbb{E}(Z_t) = 0$ y $\text{Var}(Z_t) = 1$, y la restricción $\alpha_1 + \beta_1 < 1$. De manera equivalente, el modelo se describe como:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\ X_t &= Z_t \sigma_t,\end{aligned}$$

donde usualmente X_t sigue una distribución gaussiana $X_t|\mathcal{F}_{t-1} \sim N(0, \sigma^2)$, o equivalentemente, $Z_t \sim N(0, 1)$; pero también se podrían considerar otras distribuciones tal que $E(Z_t) = 0$ y $\text{Var}(Z_t) = 1$.

Ejemplo 4. Presentamos el ajuste de un modelo GARCH para la serie de tiempo de los precios de IBM del 1 de enero de 2019 al 31 de diciembre de 2022. La Figura 7 presenta las inversiones de IBM y sus rendimientos.

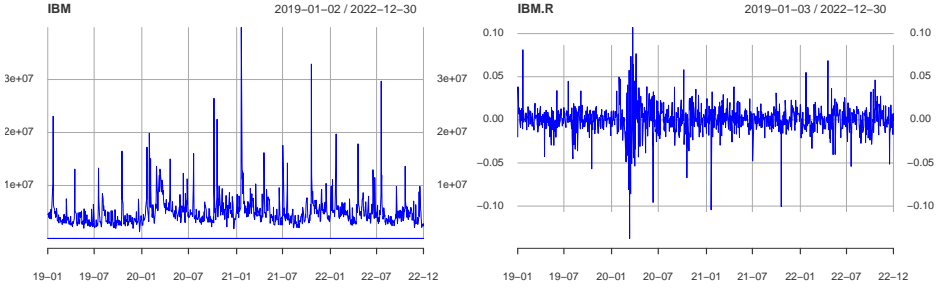


FIGURA 7. GARCH: inversiones de IBM (izquierda) y sus rendimientos (derecha).

Ajustando a esta serie financiera un modelo GARCH(1,1), con distribución gaussiana $X_t|\mathcal{F}_{t-1} \sim N(\mu, \sigma^2)$, y usando **Stan** se obtienen las estimaciones de los parámetros.

Para las distribuciones iniciales de los parámetros del modelo se utilizan $\mu \sim \text{Normal}(0, 100)$, $\alpha_0 \sim \text{Normal}(0, 100)$, $\alpha_1 \sim \text{Beta}(2, 2)$, $\beta_1 \sim \text{Beta}(2, 2)$ y el valor inicial $\sigma_0 \sim \text{Gama}(0.1, 0.1)$, con las restricciones de que $\alpha_0 > 0$, $0 < \alpha_1 < 1$, $\beta_1 > 0$ y $\alpha_1 + \beta_1 < 1$. La Figura 8 muestra las densidades posterior estimadas para los parámetros del modelo GARCH(1,1), así como las estimaciones puntuales de éstos y sus intervalos de probabilidad del 95 %, y se resumen en Tabla 3.

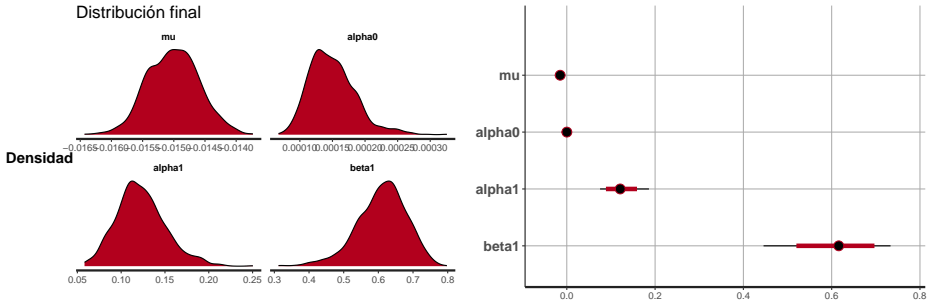


FIGURA 8. GARCH: distribución final de los parámetros, y sus estimaciones puntuales con intervalos del 95 % de probabilidad.

TABLA 3. GARCH: resumen de las estimaciones de los parámetros.

	Media	SE media	SD	2.5 %	25 %	50 %	75 %	97.5 %	neff	Rhat
μ	-0.01	0	0.00	-0.02	-0.02	-0.01	-0.01	-0.01	1054	1
α_0	0.00	0	0.00	0.00	0.00	0.00	0.00	0.00	534	1
α_1	0.12	0	0.03	0.07	0.10	0.12	0.14	0.19	687	1
β_1	0.61	0	0.07	0.45	0.57	0.62	0.66	0.72	544	1

5.3. Modelo Markov-switching GARCH. El modelo Markov-switching GARCH (MS-GARCH), se define a partir de dos procesos GARCH, que son procesos latentes paralelos con diferentes parámetros, y la desviación estándar del retorno se dibuja de acuerdo a uno de los procesos, la elección de los procesos se determina por medio de un HMM.

Un proceso estocástico $\{X_t\}_{t \in \mathcal{T}}$ es un Markov-switching GARCH(1,1) si cumple con:

$$\begin{aligned} (\sigma_t^{(k)})^2 &= \alpha_0^{(k)} + \alpha_1^{(k)} X_{t-1}^2 + \beta_1^{(k)} (\sigma_t^{(k-1)})^2, \\ S_t | S_{t-1} = k &\sim \text{Categorica}(p_k), \\ X_t | S_t = k &= Z_t \sigma_t^{(k)}, \end{aligned}$$

para $k = 1, 2, \dots, K$, usualmente $K = 2$, con $\alpha_0^{(k)} > 0$, $\alpha_1^{(k)} \geq 0$, y $\beta_1^{(k)} \geq 0$, con la restricción $\alpha_1^{(k)} + \beta_1^{(k)} < 1$. Donde,

$$X_t | \mathcal{F}_{t-1}, S_t = k \sim \mathcal{D}\left(0, (\sigma_t^{(k)})^2\right),$$

sigue una distribución continua con media cero y varianza variable en el tiempo $(\sigma_t^{(k)})^2$, usualmente se usa la distribución gaussiana, $X_t | S_t = k \sim N(0, (\sigma_t^{(k)})^2)$, es decir, $Z_t \sim N(0, 1)$. El proceso estocástico $\{S_t\}_{t \in \mathcal{T}}$ definido en el espacio de estados discreto $\{1, \dots, K\}$, caracteriza el modelo MS-GARCH. Note que $\{S_t\}$ es una cadena de Markov homogénea de primer orden, latente (no observada), con matriz de probabilidades de transición:

$$\mathbf{P} = \begin{pmatrix} p_{11} \\ \vdots \\ p_{K1} \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \cdots & p_{KK} \end{pmatrix},$$

donde $p_{ij} = \mathbb{P}(S_t = j | S_{t-1} = i)$ es la probabilidad de transición del estado $S_{t-1} = i$ al estado $S_t = j$, con las restricciones de $0 \leq p_{ij} \leq 1 \forall i, j \in \{1, \dots, K\}$ y $\sum_{j=1}^K p_{ij} = 1 \forall i \in \{1, \dots, K\}$.

En los modelos MS-GARCH, las varianzas condicionales $(\sigma_t^{(k)})^2$, $k = 1, \dots, K$, se supone que siguen K procesos separados tipo GARCH que evolucionan en paralelo.

Los modelos MS-GARCH permiten abordar los cambios de régimen en los procesos de retorno, donde el proceso estocástico que gobierna los precios del mercado es diferente durante los tiempos de estrés en comparación con los tiempos normales. Los MS-GARCH permiten diferenciar los procesos subyacentes del cambio durante la crisis y la no crisis del mercado.

Ejemplo 5. Presentamos el ajuste de un modelo MS-GARCH para la serie de tiempo de los precios de IBM del 1 de enero de 2019 al 31 de diciembre de 2022. Se ajusta un modelo MS-GARCH(1,1) con $K = 2$ procesos latentes GARCH, y usando **Stan** se obtienen las estimaciones de los parámetros.

La Figura 9 muestra las densidades posterior estimadas para los parámetros del modelo MS-GARCH(1,1). Mientras que la Figura 10 muestra las estimaciones puntuales de los parámetros y sus intervalos de probabilidad del 95%, y se resumen en la Tabla 4.

También se obtienen las estimaciones medias posterior de las volatilidades condicionales de los dos procesos latentes GARCH. La Figura 11 muestra las volatilidades condicionales de los estados 1 y 2, es decir, de los dos procesos GARCH estimados. Se pueden estimar las probabilidades posteriores de pertenecer a los estados ocultos 1 y 2 del proceso de Markov oculto, que se pueden interpretar como los estados de alta volatilidad y baja volatilidad. Ver Figura 12.

Finalmente, podría ser de interés estimar la secuencia más probable de estados ocultos. En la Figura 13 notamos que el estado 1 (de baja volatilidad) se presentó con mayor frecuencia, en comparación al estado 2 (de alta volatilidad). Note también que podríamos identificar las fechas en donde se presentaron estos eventos, y por ejemplo,

relacionarlos a eventos que hayan ocurrido de manera general en la bolsa de valores, o durante la pandemia de COVID-19.

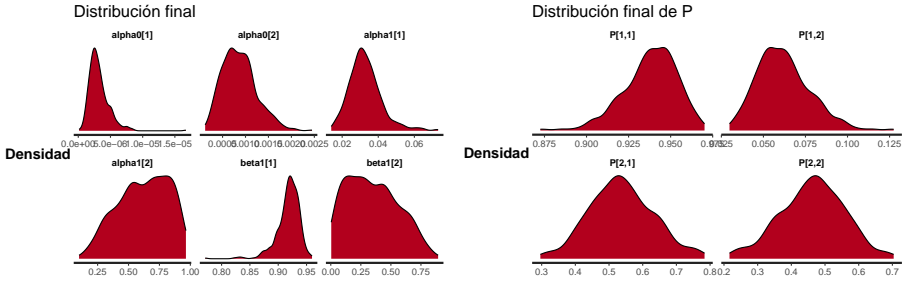


FIGURA 9. MS-GARCH: distribuciones finales de los parámetros.

TABLA 4. MS-GARCH: resumen de las estimaciones de los parámetros.

	Media	SE Media	SD	2.5 %	25 %	50 %	75 %	97.5 %	neff	Rhat
$\alpha_0^{(1)}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	293.615	0.999
$\alpha_0^{(2)}$	0.001	0.000	0.000	0.000	0.001	0.001	0.001	0.002	250.704	1.002
$\alpha_1^{(1)}$	0.033	0.000	0.009	0.018	0.027	0.032	0.037	0.055	340.355	0.999
$\alpha_1^{(2)}$	0.621	0.013	0.202	0.222	0.476	0.639	0.793	0.928	252.491	1.007
$\beta_1^{(1)}$	0.918	0.001	0.021	0.871	0.909	0.921	0.931	0.947	233.346	0.999
$\beta_1^{(2)}$	0.347	0.013	0.212	0.024	0.160	0.323	0.500	0.764	247.781	1.006
$P_{1,1}$	0.938	0.001	0.016	0.902	0.927	0.939	0.950	0.963	361.554	0.998
$P_{1,2}$	0.062	0.001	0.016	0.037	0.050	0.061	0.073	0.098	361.554	0.998
$P_{2,1}$	0.547	0.005	0.090	0.378	0.480	0.549	0.609	0.740	315.908	0.998
$P_{2,2}$	0.453	0.005	0.090	0.260	0.391	0.451	0.520	0.622	315.908	0.998

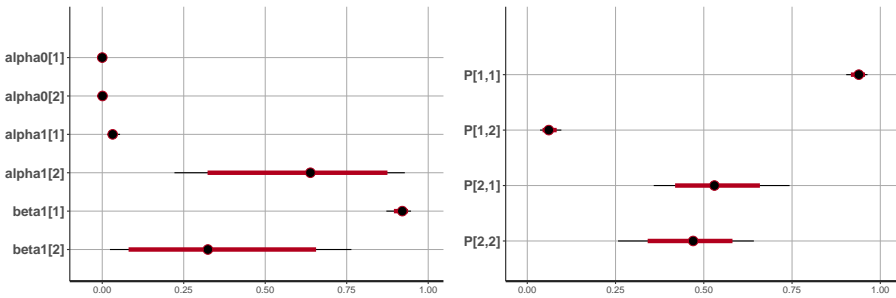


FIGURA 10. MS-GARCH: estimaciones puntuales e intervalos del 95% de probabilidad para los parámetros.

6. CONCLUSIONES

En este trabajo se ha abordado la teoría de los modelos ocultos de Markov, describiendo sus principales características. También, hemos ilustrado un ejemplo simple de aplicación de los HMM, como lo son las series temporales financieras.

Hoy en día la aplicación de los HMM no se limita al reconocimiento de voz, sino que áreas de investigación como la robótica móvil, ingeniería genética, finanzas, mantenimiento predictivo industrial, psicología, nanotecnología, comunicaciones, entre muchas otras, hacen uso de las herramientas matemáticas de estos modelos para analizar eventos que ocurren en forma aleatoria a través del tiempo.

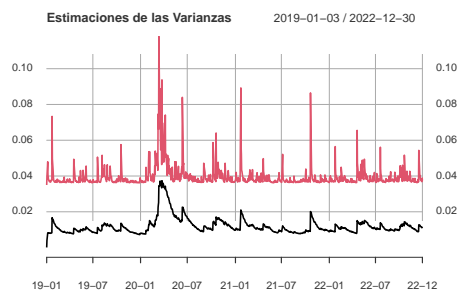


FIGURA 11. MS-GARCH: estimaciones medias finales de las varianzas instantáneas.

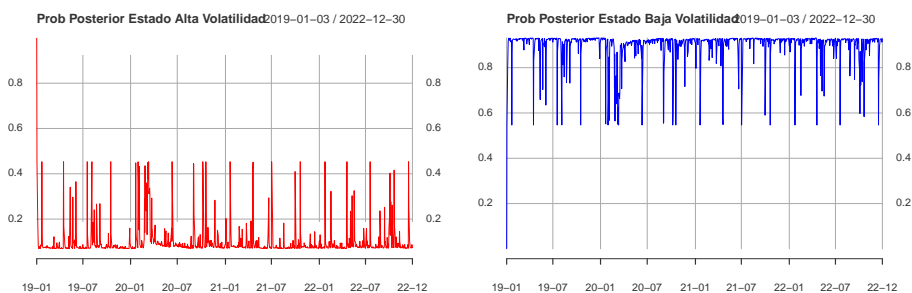


FIGURA 12. MS-GARCH: probabilidades estimadas finales de los estados de alta y baja volatilidad.

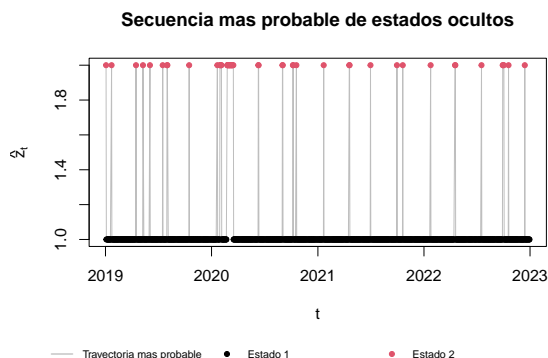


FIGURA 13. MS-GARCH: secuencia estimada más probable de los estados ocultos.

El análisis de series de tiempo financieras es importante hoy en día para gestionar y proyectar resultados. Estas series se componen por variables de mercado que incorporan incertidumbre, de ahí que se necesitan pronosticar. Muchas de las series financieras han sido estudiadas a través de distintos modelos como son: AR, MA, ARCH, GARCH, etc., sin embargo, ninguno de ellos captura la volatilidad cuando hay cambios de régimen, de ahí surgen los modelos MS-GARCH, los cuales actualmente son muy usados en finanzas.

En este trabajo incrustamos los HMM en los MS-GARCH para cambiar entre dos procesos latentes tipo GARCH, esto debido al agrupamiento de volatilidad presente en la mayoría de las series financieras. Estimamos los parámetros de estos modelos a través de la estadística Bayesiana y usando *Stan*, pues resulta ser una herramienta fácil y eficaz en la implementación de estos modelos.

Debido a su naturaleza Markoviana de los HMM, éstos no toman en cuenta la secuencia de estados que conducen a un estado determinado, sólo depende del estado anterior. Además, no se captura explícitamente el tiempo que la cadena está en un estado dado. En resumen, el HMM básico proporciona un modelo útil para muchas aplicaciones. Sin embargo, éstas pueden estar limitadas por la estructura intrínseca del modelo. Existen muchas extensiones que aumentan el poder de representación del HMM, aunque los parámetros adicionales requeridos a menudo complican las tareas de aprendizaje e inferencia.

AGRADECIMIENTOS

Agradecemos a los revisores anónimos que brindaron comentarios y sugerencias perspicaces que mejoraron en gran medida tanto la legibilidad como el contenido de este artículo.

APÉNDICE A. CÓDIGOS DE R Y STAN

Los códigos de R y Stan están disponibles en GitHub [lizbethna/HMMBayes.git](https://github.com/lizbethna/HMMBayes.git)

REFERENCIAS

- Albert, J. (2009). *Bayesian Computation with R*. Springer, New York, second edición.
- Ardia, D., Bluteau, K., Boudt, K., Catania, L., y Trottier, D.-A. (2019). Markov-switching GARCH models in R: The MSGARCH package. *Journal of Statistical Software*, 91:1–38.
- Baum, L. E. y Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, 37(6):1554–1563.
- Baum, L. E., Petrie, T., Soules, G., y Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- Bernardo, J. M. y Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons, Chichester.
- Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv: Methodology*.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Bollerslev, T. (2023). The story of GARCH: A personal odyssey. *Journal of Econometrics*.
- Chen, F.-S., Fu, C.-M., y Huang, C.-L. (2003). Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and vision computing*, 21(8):745–758.
- Chen, M.-H., Shao, Q.-M., e Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Series in Statistics. Springer, New York.
- Damiano, L., Peterson, B., y Weylandt, M. (2018). A tutorial on hidden Markov models using Stan. En *Stan conference*.
- Dempster, A. P., Laird, N. M., y Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Engle, R. F., Lilien, D. M., y Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica: Journal of the Econometric Society*, pp. 391–407.
- Gamerman, D. y Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, Boca Raton, Florida, second edición.
- Gómez-Rubio, V. (2020). *Bayesian Inference with INLA*. CRC Press.

- Gupta, A. y Dhingra, B. (2012). Stock market prediction using hidden Markov models. En *2012 Students Conference on Engineering and Systems*, pp. 1–4. IEEE.
- Haas, M., Mittnik, S., y Paoletta, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.
- Hassan, M. R. y Nath, B. (2005). Stock market forecasting using hidden Markov model: a new approach. En *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, pp. 192–196. IEEE.
- Kruschke, J. (2010). *Doing Bayesian data analysis: A tutorial introduction with R, JAGS and Stan*. Academic Press.
- Lukashin, A. V. y Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic acids research*, 26(4):1107–1115.
- Lunn, D., Jackson, C., Best, N., Thomas, A., y Spiegelhalter, D. (2012). *The BUGS Book - A Practical Introduction to Bayesian Analysis*. CRC Press / Chapman & Hall, Boca Raton, Florida.
- Lunn, D. J., Thomas, A., Best, N., y Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.
- Niu, F. y Abdel-Mottaleb, M. (2005). HMM-based segmentation and recognition of human activities from video sequences. En *2005 IEEE International Conference on Multimedia and Expo*, pp. 804–807. IEEE.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, New Jersey.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. En *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Viena, Austria.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.
- Robert, C. P. (1994). *The Bayesian Choice*. Springer-Verlag, New York.
- Robert, C. P. y Casella, G. (2000). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer.
- Rue, H., Martino, S., y Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society*, 71(2):319–392.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., y Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4(1):395–421.
- Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2.
- Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual 2.31*.
- Sucar, L. E. (2015). Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition. London: Springer London*. doi, 10(978):1.
- Zucchini, W., MacDonald, I. L., y Langrock, R. (2016). *Hidden Markov models for time series: an introduction using R*. Número 150 en Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, 2a. edición.

Lizbeth Naranjo Albarrán,
Departamento de Matemáticas,
Facultad de Ciencias,
Universidad Nacional Autónoma de México.
Círculo Exterior s/n, Ciudad Universitaria, Coyoacán,
C.P. 04510, Ciudad de México, México.
e-mail: lizabethna@ciencias.unam.mx

Luz Judith Rodríguez Esparza,
Departamento de Matemáticas y Física,
Cátedra CONACyT, Universidad Autónoma de Aguascalientes.
Av. Universidad 940, Ciudad Universitaria,
C.P. 20100, Aguascalientes, Aguascalientes, México.
e-mail: judithr19@gmail.com