



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

MODELOS MARKOVIANOS DE CAMBIO DE
RÉGIMEN Y SUS APLICACIONES A SERIES DE
TIEMPO GARCH.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

PRESENTA:

ITZEL ABIGAIL RAMÍREZ GONZÁLEZ

DIRECTOR DE TESIS:

M. EN C. DANIEL CERVANTES FILOTEO

CIUDAD UNIVERSITARIA, CDMX, 2025





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.

Índice general

Introducción	iv
1. Cadenas de Markov	1
1.1. Definiciones básicas	2
1.1.1. Definición de cadenas de Markov	2
1.1.2. Matriz de probabilidades de transición	5
1.1.3. Distribuciones	8
1.2. Ecuaciones de Chapman–Kolmogorov	9
1.3. Clases de comunicación	11
1.4. Ejemplos	15
1.5. Estimación de probabilidades de transición	18
1.6. Cadenas de Markov en \mathbb{R}	21
1.6.1. Librería “markovchain” en \mathbb{R}	21
1.6.2. Simulación y estimación estadística	24
2. Series de tiempo	28
2.1. Motivación	28
2.2. Definiciones básicas	29
2.2.1. Características clásicas de una serie de tiempo.	30
2.2.2. Operadores	32
2.3. Descomposición clásica de una serie de tiempo	34
2.3.1. Estimación y eliminación de la tendencia	36
2.3.2. Modelo de Holt-Winters.	40
2.4. Modelos ARMA	42
2.4.1. Modelo autorregresivo (AR)	43
2.4.2. Modelo de medias móviles (MA)	45
2.4.3. Modelo autorregresivo de medias móviles (ARMA)	46
2.5. Procesos GARCH	48

2.5.1.	Hechos estilizados	49
2.5.2.	Proceso Autorregresivo de Heterocedasticidad Condicional (ARCH)	49
2.5.3.	Proceso Autorregresivo Generalizado de Heterocedasticidad Condicional (GARCH)	50
2.5.4.	Simulación de series de tiempo GARCH	51
2.6.	Estimación de parámetros	60
2.6.1.	Métodos de estimación	60
2.6.2.	Propiedades de los estimadores	61
3.	Modelos markovianos	63
3.1.	Modelo autorregresivo de media condicional	63
3.1.1.	Modelo básico	64
3.1.2.	Extensiones del modelo	67
3.1.3.	Estimación estadística	70
3.2.	Modelo GARCH con switch markoviano	78
3.2.1.	Descripción del modelo	78
3.2.2.	Simulación del modelo	80
4.	Aplicación al S&P/BMV IPC	87
4.1.	Análisis descriptivo de los datos	87
4.1.1.	Descripción de los datos	87
4.1.2.	Estadística descriptiva	91
4.2.	Estimación estadística	96
4.2.1.	Selección y ajuste de modelo	97
4.2.2.	Simulaciones con el modelo ajustado	101
4.3.	Conclusiones	105
5.	Conclusiones	106
A.	Otras distribuciones	108
A.1.	Distribución de Error Generalizada (GED)	108
A.2.	Distribuciones Sesgadas Estandarizadas	109
B.	Código en R	110
B.1.	Serie de tiempo del S&P/BMV IPC	110
B.2.	Análisis de las diferencias entre las fechas reportadas	111
B.3.	Cálculo del retorno sobre la inversión y sus gráficas	111
B.4.	Ajuste de los datos a modelo MSGARCH y comparación de parámetros	112

<i>ÍNDICE GENERAL</i>	iii
B.5. Descripción y estimación del modelo MSGARCH	113
B.6. Simulaciones del modelo ajustado MSGARCH	114
C. Índices AIC y BIC	117
Bibliografía	119

Introducción

Los fenómenos financieros desempeñan un papel crucial en la sociedad actual, de manera que su estudio ha motivado el desarrollo de diversos modelos estadísticos y probabilísticos a lo largo del tiempo. Por ejemplo, se ha profundizado en el estudio de las series de tiempo y su predicción. En particular, han surgido ciertos modelos conocidos como modelos GARCH (heterocedasticidad condicional autorregresiva generalizada), en los cuales la volatilidad del proceso no es constante en el tiempo, sino que depende de la volatilidad y errores que tuvo en tiempos previos. Estos modelos fueron desarrollados en 1986 por Tim Bollerslev como una manera de predecir la volatilidad en los precios de los activos, lo cual a su vez fue una generalización de los modelos ARCH (heterocedasticidad condicional autorregresiva) propuestos por Robert Engle en 1982. Actualmente son muy utilizados para describir fenómenos financieros, pues concuerdan con los hechos estilizados y, en la mayoría de los casos, basta con considerar el modelo más sencillo GARCH(1,1) para obtener una muy buena explicación de los datos.

Sin embargo, conforme se han ido estudiando y modelando fenómenos más complejos, los modelos también han incrementado en complejidad. Por ejemplo, algunos conjuntos de datos parecen comportarse como series de tiempo GARCH distintas de acuerdo a los intervalos de tiempo en los que se encuentra la serie. Es decir, si observamos a la serie de datos por pedazos, podemos notar que se comportan como series GARCH(1,1) de parámetros distintos. Esto se conoce como un cambio de régimen y puede ser regido por distintos procesos deterministas o probabilísticos. En este trabajo, consideraremos que el cambio de régimen está dado por una cadena de Markov.

El objetivo principal de esta tesis es intentar describir cierta serie de datos financiera mediante un modelo GARCH con cambio de régimen markoviano, así como realizar simulaciones con este modelo ajustado. Para esto, presen-

taremos una variación del modelo GARCH asumiendo que los parámetros no son constantes, sino que pueden cambiar a lo largo del tiempo de acuerdo con un switch markoviano específico. Así, otro objetivo es explicar la idea general y algunos ejemplos de modelos con cambio de régimen.

En el primer capítulo, se introducirán los conceptos básicos de las cadenas de Markov, ya que estas determinarán el comportamiento del cambio de régimen. Se hablará acerca de las definiciones básicas, como son la matriz de probabilidad de transición y las distribuciones, además de que se presentarán las ecuaciones de Chapman-Kolmogorov y la partición de los estados de la cadena en clases de comunicación. Posteriormente, se incluirán ejemplos detallados de cadenas de Markov y se describirá un método para realizar la estimación estadística de las probabilidades de transición asociadas a la cadena. Finalmente, se realizará un estudio de las cadenas de Markov en R, considerando una librería conocida, así como simulación y estimación estadística mediante funciones definidas por el usuario.

En el segundo capítulo, se introducirán los conceptos básicos de las series de tiempo, ya que éstas serán utilizadas para la descripción de fenómenos financieros. Para esto, se presentará una introducción teórica a las series de tiempo, considerando las definiciones básicas y la descomposición clásica. Posteriormente, se explicarán los modelos clásicos ARMA y los modelos GARCH, en los cuales se profundizará mediante la explicación de los hechos estilizados y la simulación de series de tiempo GARCH en R. Para finalizar, se describirán las bases de la estimación estadística paramétrica.

En el tercer capítulo, se introducirán dos modelos markovianos de cambio de régimen. Para comenzar, se estudiará el modelo autorregresivo de media condicional con cambio de régimen markoviano, considerando el modelo básico, algunas extensiones y dos métodos para su estimación estadística. Posteriormente, se describirá y simulará el modelo principal de este trabajo: el modelo GARCH con switch markoviano MSGARCH, para lo cual se presentará su descripción básica y su simulación en R.

Para concluir el trabajo, en el cuarto capítulo se aplicará el modelo markoviano de cambio de régimen MSGARCH descrito en el capítulo previo. Para esto, se estudiará la serie de datos del Índice de Precios y Cotizaciones de S&P Dow Jones Indices y la Bolsa Mexicana de Valores, realizando un análisis descriptivo de los datos y estimación estadística incluyendo su ajuste y simulaciones.

Capítulo 1

Cadenas de Markov

En el presente capítulo, comenzaremos con los conceptos básicos de las cadenas de Markov. Estas son procesos estocásticos en los cuales la probabilidad de que ocurra un evento depende únicamente de la ocurrencia del evento inmediatamente anterior. Existe una gran cantidad de fenómenos aleatorios que se pueden modelar bajo esa característica, de manera que su estudio resulta de gran importancia en una gran variedad de campos, como son la economía, la física, la química e incluso la epidemiología.

En este caso, nos limitaremos al estudio de las cadenas de Markov discretas, presentando la definición de dicho proceso y algunos conceptos relacionados, como son la matriz de probabilidades de transición, las distribuciones asociadas y las ecuaciones de Chapman-Kolmogorov. Adicionalmente, veremos la partición en clases de comunicación que se puede generar a partir de cierta relación de equivalencia específica relativa a las probabilidades positivas, así como algunos ejemplos detallados. Para concluir el capítulo, se describirá la estimación estadística de probabilidades de transición para una cadena de Markov.

El contenido de este capítulo está basado en las ideas de Ross (2010) [1]; Suhov (2008) [2]; y Craig y Sendi (2002) [3].

1.1. Definiciones básicas

1.1.1. Definición de cadenas de Markov

Definición 1.1.1. Un **proceso estocástico** es un conjunto de variables aleatorias indexadas, lo cual se denota como $\{X_t\}_{t \in I}$.

Un proceso estocástico es **a tiempo discreto** cuando I es un conjunto a lo más numerable y **a tiempo continuo** cuando I es un conjunto no numerable.

Definición 1.1.2. El **soporte de un proceso estocástico**, también conocido como **espacio de estados**, es la unión del soporte de cada una de las variables aleatorias en el conjunto, es decir,

$$\text{sop } \{X_t\}_{t \in I} = \bigcup_{t \in I} \text{sop } X_t.$$

Un proceso estocástico se denomina **discreto** cuando $\text{sop}\{X_t\}_{t \in I}$ es a lo más numerable y **continuo** cuando no es discreto.

Definición 1.1.3. Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso estocástico discreto y a tiempo discreto con espacio de estados denotado por ε . Decimos que $\{X_n\}_{n \in \mathbb{N}}$ es una **cadena de Markov** si para cualquier entero $n \geq 0$ y para todos los $x_0, x_1, \dots, x_n \in \varepsilon$ se cumple con la propiedad de Markov, es decir, si

$$\mathbb{P}[X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}] = \mathbb{P}[X_n = x_n \mid X_{n-1} = x_{n-1}].$$

Esto se asocia con la *propiedad de pérdida de memoria*, ya que nos indica que el estado de la cadena a cierto tiempo futuro depende únicamente de lo que sucedió en el tiempo actual, sin importar lo sucedido en el pasado. En la siguiente proposición generalizaremos dicha propiedad.

Proposición 1.1.1. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov y $n_1, n_2, \dots, n_m \in \mathbb{Z}$ tales que $0 \leq n_1 < n_2 < \dots < n_m$. Entonces, para todos los $x_1, x_2, \dots, x_m \in \varepsilon$, se cumple lo siguiente:

1. La probabilidad de la trayectoria con inicio dado se calcula como

$$\begin{aligned} & \mathbb{P}[X_{n_2} = x_2, \dots, X_{n_m} = x_m \mid X_{n_1} = x_1] \\ &= \prod_{i=1}^{m-1} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i]. \end{aligned}$$

2. La probabilidad de la trayectoria se calcula como

$$\begin{aligned} & \mathbb{P}[X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_m} = x_m] \\ &= \mathbb{P}[X_{n_1} = x_1] \prod_{i=1}^{m-1} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i]. \end{aligned}$$

3. Se cumple la propiedad de Markov generalizada, es decir,

$$\begin{aligned} & \mathbb{P}[X_{n_m} = x_m \mid X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_{m-1}} = x_{m-1}] \\ &= \mathbb{P}[X_{n_m} = x_m \mid X_{n_{m-1}} = x_{m-1}]. \end{aligned}$$

Demostración. Bajo las hipótesis dadas:

1. Aplicando el teorema de la probabilidad total inductivamente,

$$\begin{aligned} & \mathbb{P}[X_{n_2} = x_2, \dots, X_{n_m} = x_m \mid X_{n_1} = x_1] \\ &= \mathbb{P}[X_{n_3} = x_3, \dots, X_{n_m} = x_m \mid X_{n_1} = x_1, X_{n_2} = x_2] \mathbb{P}[X_{n_2} = x_2 \mid X_{n_1} = x_1] \\ & \vdots \\ &= \mathbb{P}[X_{n_m} = x_m \mid X_{n_1} = x_1, \dots, X_{n_{m-1}} = x_{m-1}] \mathbb{P}[X_{n_2} = x_2 \mid X_{n_1} = x_1] \\ & \quad \mathbb{P}[X_{n_3} = x_3 \mid X_{n_1} = x_1, X_{n_2} = x_2] \mathbb{P}[X_{n_4} = x_4 \mid X_{n_1} = x_1, \dots, X_{n_3} = x_3] \\ & \quad \dots \mathbb{P}[X_{n_{m-1}} = x_{m-1} \mid X_{n_1} = x_1, \dots, X_{n_{m-2}} = x_{m-2}]. \end{aligned}$$

Aplicando la propiedad de Markov, podemos quitar las variables intermedias en las condicionales, con lo cual se sigue que

$$\begin{aligned} & \mathbb{P}[X_{n_2} = x_2, \dots, X_{n_m} = x_m \mid X_{n_1} = x_1] \\ &= \mathbb{P}[X_{n_m} = x_m \mid X_{n_{m-1}} = x_{m-1}] \mathbb{P}[X_{n_2} = x_2 \mid X_{n_1} = x_1] \\ & \quad \mathbb{P}[X_{n_3} = x_3 \mid X_{n_2} = x_2] \mathbb{P}[X_{n_4} = x_4 \mid X_{n_3} = x_3] \\ & \quad \dots \mathbb{P}[X_{n_{m-1}} = x_{m-1} \mid X_{n_{m-2}} = x_{m-2}] \\ &= \prod_{i=1}^{m-1} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i]. \end{aligned}$$

□

2. Aplicando el teorema de la probabilidad total,

$$\begin{aligned} & \mathbb{P}[X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_m} = x_m] \\ &= \mathbb{P}[X_{n_2} = x_2, \dots, X_{n_m} = x_m \mid X_{n_1} = x_1] \mathbb{P}[X_{n_1} = x_1] \\ &= \mathbb{P}[X_{n_1} = x_1] \prod_{i=1}^{m-1} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i], \end{aligned}$$

donde la última igualdad se sigue de 1. \square

3. Utilizando la definición de probabilidad condicional,

$$\begin{aligned} & \mathbb{P}[X_{n_m} = x_m \mid X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_{m-1}} = x_{m-1}] \\ &= \frac{\mathbb{P}[X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_m} = x_m]}{\mathbb{P}[X_{n_1} = x_1, X_{n_2} = x_2, \dots, X_{n_{m-1}} = x_{m-1}]} \\ &= \frac{\mathbb{P}[X_{n_1} = x_1] \prod_{i=1}^{m-1} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i]}{\mathbb{P}[X_{n_1} = x_1] \prod_{i=1}^{m-2} \mathbb{P}[X_{n_{i+1}} = x_{i+1} \mid X_{n_i} = x_i]} \\ &= \mathbb{P}[X_{n_m} = x_m \mid X_{n_{m-1}} = x_{m-1}], \end{aligned}$$

donde la segunda igualdad se sigue de 2. \square

Con la proposición previa tenemos más información respecto a cómo se comportan las probabilidades de trayectorias específicas en cadenas de Markov, en congruencia con la propiedad de pérdida de memoria.

A continuación, veremos una proposición para demostrar que cierto proceso estocástico es una cadena de Markov, lo cual es útil pues en muchas ocasiones resulta complicado probar la propiedad de Markov directamente y resulta más sencillo demostrarlo por este medio.

Proposición 1.1.2. *Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso estocástico tal que $X_{n+1} = g(X_n, Y_{n+1})$ para g una función real y $\{Y_n\}_{n \in \mathbb{N}}$ variables aleatorias con Y_{n+1} independiente de (X_0, X_1, \dots, X_n) para todo $n \in \mathbb{N}$. Entonces $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov.*

Demostración. Bajo las hipótesis anteriores, veamos que $\{X_n\}_{n \in \mathbb{N}}$ cumple con la propiedad de Markov.

$$\begin{aligned} & \mathbb{P}[X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{P}[g(X_n, Y_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] \end{aligned}$$

y, como $X_n = x_n$ ya es un valor conocido en la condicional, se sigue que

$$\begin{aligned} & \mathbb{P}[g(X_n, Y_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{P}[g(x_n, Y_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n]. \end{aligned}$$

De esta manera, el único componente aleatorio en $g(x_n, Y_{n+1})$ es la variable Y_{n+1} . Además, por hipótesis tenemos que Y_{n+1} es independiente de las variables X_0, \dots, X_n , entonces,

$$\begin{aligned} & \mathbb{P}[g(x_n, Y_{n+1}) = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{P}[g(x_n, Y_{n+1}) = x_{n+1} \mid X_n = x_n] = \mathbb{P}[g(X_n, Y_{n+1}) = x_{n+1} \mid X_n = x_n] \\ &= \mathbb{P}[X_{n+1} = x_{n+1} \mid X_n = x_n]. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}[X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] = \mathbb{P}[X_{n+1} = x_{n+1} \mid X_n = x_n]$$

y $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov. □

1.1.2. Matriz de probabilidades de transición

Definición 1.1.4. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov. Sean $n, m \in \mathbb{N}$. Definimos una **probabilidad de transición** como

$$P_{i,j}^{(n,n+m)} := \mathbb{P}[X_{n+m} = j \mid X_n = i], \text{ para todos los } i, j \in \varepsilon,$$

lo cual representa la probabilidad de pasar del estado i en tiempo n al estado j en tiempo $n + m$.

Definición 1.1.5. Una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ es **homogénea en el tiempo** si para cada $n, m \in \mathbb{N}$, se cumple que

$$P_{i,j}^{(n,n+1)} = P_{i,j}^{(m,m+1)}, \text{ para todos los } i, j \in \varepsilon.$$

En este caso, dicha probabilidad se denota como $P_{i,j}$.

Esto quiere decir que la probabilidad es la misma para cualquier tiempo, en cuyo caso existe una única probabilidad de transición de i a j en 1 paso. En lo que resta de este escrito, trabajaremos únicamente con cadenas homogéneas, pues esto bastará para plantear los modelos de interés.

Definición 1.1.6. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov. Utilizando $\text{Mat}_{|\varepsilon| \times |\varepsilon|}(\mathbb{R})$ como la notación para el conjunto de matrices de tamaño $|\varepsilon| \times |\varepsilon|$ sobre el campo \mathbb{R} , definimos a la **matriz de probabilidades de transición** como la matriz $\mathbb{P} \in \text{Mat}_{|\varepsilon| \times |\varepsilon|}(\mathbb{R})$ con entradas

$$\mathbb{P} = (P_{i,j})_{i,j \in \varepsilon} .$$

Observación. $P_{i,j}$ con i fija es la función de densidad de la variable $(X_{n+1} | X_n = i)$ para cualquier $n \in \mathbb{N}$, de manera que la suma sobre su soporte es 1 y $f_{X_{n+1} | X_n = i}(j) \geq 0$ para todo $j \in \varepsilon$.

Proposición 1.1.3. La matriz de probabilidades de transición \mathbb{P} es una matriz estocástica, es decir, satisface las siguientes propiedades:

1. $P_{i,j} \geq 0$ para todos los $i, j \in \varepsilon$.
2. $\sum_{j \in \varepsilon} P_{i,j} = 1$ para todos los $i, j \in \varepsilon$.

Demostración. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov y $n \in \mathbb{N}$.

1. Sean $i, j \in \varepsilon$. Entonces $P_{i,j} = \mathbb{P}[X_{n+1} = j | X_n = i] \geq 0$, pues es una probabilidad.
2. Sea $i \in \varepsilon$. Entonces,

$$\sum_{j \in \varepsilon} P_{i,j} = \sum_{j \in \varepsilon} \mathbb{P}[X_{n+1} = j | X_n = i] = \sum_{j \in \varepsilon} f_{X_{n+1} | X_n = i}(j).$$

Esta es la función de densidad de $(X_{n+1} | X_n = i)$ sobre todo el espacio de estados. Adicionalmente, $\text{sop}_{X_{n+1} | X_n = i} \subseteq \text{sop}_{X_{n+1}} \subseteq \text{sop}_{\{X_n\}_{n \in \mathbb{N}}} = \varepsilon$, entonces dicha suma incluye el soporte de $(X_{n+1} | X_n = i)$ y, por consiguiente, $\sum_{j \in \varepsilon} P_{i,j} = 1$.

□

Definición 1.1.7. Sean $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov homogénea en el tiempo y $m \in \mathbb{N}$.

- La **probabilidad de transición en m pasos** se define como sigue.

Si $m = 0$,

$$P_{i,j}^{(0)} := \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad \text{para todos los } i, j \in \varepsilon.$$

En cualquier otro caso,

$$P_{i,j}^{(m)} := \mathbb{P}[X_m = j \mid X_0 = i], \quad \text{para todos los } i, j \in \varepsilon.$$

Como es homogénea, se cumple que

$$P_{i,j}^{(m)} = \mathbb{P}[X_{n+m} = j \mid X_n = i] \quad \text{para todo } n \in \mathbb{N}.$$

- La **matriz de probabilidades de transición a m pasos** está definida como la matriz $\mathbb{P}^{(m)} \in \text{Mat}_{|\varepsilon| \times |\varepsilon|}(\mathbb{R})$ con entradas

$$\mathbb{P}^{(m)} = (P_{i,j}^{(m)})_{i,j \in \varepsilon}.$$

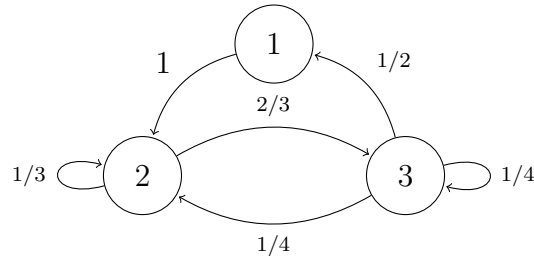
La matriz $\mathbb{P}^{(m)}$ también es una matriz estocástica; la prueba resulta análoga a la demostración para la matriz \mathbb{P} en la **Proposición 1.1.3**.

Ejemplo 1.1.1. A continuación presentaremos un ejemplo numérico de una cadena de Markov sencilla y su matriz de probabilidades de transición.

Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov homogénea con espacio de estados $\varepsilon = \{1, 2, 3\}$ y con las siguientes probabilidades de transición:

$$\begin{aligned} P_{1,2} &= \mathbb{P}[X_1 = 2 \mid X_0 = 1] = 1 & P_{3,1} &= \mathbb{P}[X_1 = 1 \mid X_0 = 3] = \frac{1}{2} \\ P_{2,2} &= \mathbb{P}[X_1 = 2 \mid X_0 = 2] = \frac{1}{3} & P_{3,2} &= \mathbb{P}[X_1 = 2 \mid X_0 = 3] = \frac{1}{4} \\ P_{2,3} &= \mathbb{P}[X_1 = 3 \mid X_0 = 2] = \frac{2}{3} & P_{3,3} &= \mathbb{P}[X_1 = 3 \mid X_0 = 3] = \frac{1}{4}. \end{aligned}$$

Esto se puede representar mediante el siguiente esquema:



En este caso, la matriz de probabilidades de transición es

$$\mathbb{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Notemos que esta matriz en efecto cumple con tener entradas positivas y con que la suma de cada uno de sus renglones es igual a 1, de manera de que cada renglón i representa la función de densidad de la variable $(X_1 \mid X_0 = i)$.

1.1.3. Distribuciones

Definición 1.1.8. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados $\varepsilon = \{a_1, a_2, a_3, \dots\}$ (puede ser un conjunto finito o infinito). Definimos las siguientes distribuciones.

- La **distribución inicial** está dada por

$$\pi^{(0)} = (\pi_{a_1}^{(0)}, \pi_{a_2}^{(0)}, \pi_{a_3}^{(0)}, \dots),$$

donde

$$\pi_{a_i}^{(0)} = \mathbb{P}[X_0 = a_i], \text{ para cada } a_i \in \varepsilon.$$

Este vector representa la distribución de la variable aleatoria X_0 y puede ser finito o infinito, dependiendo de ε .

- Sea $n \in \mathbb{N}$. La **distribución a tiempo n** está dada por

$$\pi^{(n)} = (\pi_{a_1}^{(n)}, \pi_{a_2}^{(n)}, \pi_{a_3}^{(n)}, \dots),$$

donde

$$\pi_{a_i}^{(n)} = \mathbb{P}[X_n = a_i], \text{ para cada } a_i \in \varepsilon.$$

Este vector representa la distribución de la variable aleatoria X_n y puede ser finito o infinito, dependiendo de ε .

Proposición 1.1.4. Para todo $n \in \mathbb{N}$, el vector $\pi^{(n)}$ es un vector estocástico, es decir, satisface las siguientes propiedades:

1. $\pi_i^{(n)} \geq 0$ para todo $i \in \varepsilon$.

2. $\sum_{i \in \varepsilon} \pi_i^{(n)} = 1$.

Demostración. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov y $n \in \mathbb{N}$.

1. Sea $i \in \varepsilon$. Entonces $\pi_i^{(n)} = \mathbb{P}[X_n = i] \geq 0$, pues es una probabilidad.
2. $\sum_{i \in \varepsilon} \pi_i^{(n)} = \sum_{i \in \varepsilon} \mathbb{P}[X_n = i] = \sum_{i \in \varepsilon} f_{X_n}(i)$.

Esta es la función de densidad de X_n sobre todo el espacio de estados. Adicionalmente, $\text{sop}_{X_n} \subseteq \text{sop}_{\{X_n\}_{n \in \mathbb{N}}} = \varepsilon$, entonces dicha suma incluye el soporte de X_n y, por consiguiente, $\sum_{i \in \varepsilon} \pi_i^{(n)} = 1$.

□

1.2. Ecuaciones de Chapman–Kolmogorov

Teorema 1.2.1. (*Ecuaciones de Chapman–Kolmogorov*) Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con matriz de probabilidades de transición \mathbb{P} y $n \in \mathbb{N}$, entonces para todo $r \in \{0, 1, \dots, n-1\}$ se cumple que

$$P_{i,j}^{(n)} = \sum_{k \in \varepsilon} P_{i,k}^{(r)} P_{k,j}^{(n-r)} \text{ para todos los } i, j \in \varepsilon.$$

De forma matricial, dicho sistema de ecuaciones se puede expresar como

$$\mathbb{P}^{(n)} = \mathbb{P}^{(r)} \mathbb{P}^{(n-r)}.$$

Demostración. Sean $n \in \mathbb{N}$, $r \in \{0, 1, \dots, n-1\}$ y $i, j \in \varepsilon$.

Si $r = 0$ se cumple lo requerido, pues $P_{i,j}^{(0)} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$.

Si $r \neq 0$, por definición $P_{i,j}^{(n)} = \mathbb{P}[X_n = j \mid X_0 = i]$.

Sean $\mathbb{P}_i[X_n = j] := \mathbb{P}[X_n = j \mid X_0 = i]$, aplicando el teorema de la probabilidad total,

$$\begin{aligned} P_{i,j}^{(n)} &= \mathbb{P}_i[X_n = j] = \sum_{k \in \varepsilon} \mathbb{P}_i[X_n = j \mid X_r = k] \mathbb{P}_i[X_r = k] \\ &= \sum_{k \in \varepsilon} \mathbb{P}[X_n = j \mid X_0 = i, X_r = k] \mathbb{P}[X_r = k \mid X_0 = i]. \end{aligned}$$

Utilizando la propiedad de Markov generalizada,

$$\begin{aligned} P_{i,j}^{(n)} &= \sum_{k \in \varepsilon} \mathbb{P}[X_n = j \mid X_r = k] \mathbb{P}[X_r = k \mid X_0 = i] \\ &= \sum_{k \in \varepsilon} P_{k,j}^{(n-r)} P_{i,k}^{(r)} = \sum_{k \in \varepsilon} P_{i,k}^{(r)} P_{k,j}^{(n-r)}. \end{aligned}$$

□

Corolario 1.2.1. *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con matriz de probabilidades de transición \mathbb{P} . Entonces para todos los $n \in \mathbb{N}$ y $r \in \{0, 1, \dots, n-1\}$, se cumple que $\mathbb{P}^{(n)} = \mathbb{P}^n$ y $\pi^{(n)} = \pi^{(r)} \mathbb{P}^{(n-r)}$.*

Demostración. Comprobemos que se cumple cada una de las afirmaciones.

- Procedemos por inducción. Para $n = 1$, $\mathbb{P}^{(1)} = \mathbb{P} = \mathbb{P}^1$.

Supongamos que para $n = k$ se cumple que $\mathbb{P}^{(k)} = \mathbb{P}^k$, es decir, que para todos los $i, j \in \varepsilon$, $P_{i,j}^{(k)} = (\mathbb{P}^{(k)})_{i,j} = (\mathbb{P}^k)_{i,j}$. Entonces, para todos los $i, j \in \varepsilon$ se cumple que

$$\begin{aligned} (\mathbb{P}^{(k+1)})_{i,j} &= P_{i,j}^{(k+1)} = \sum_{h \in \varepsilon} P_{i,h} P_{h,j}^{(k)} = \sum_{h \in \varepsilon} P_{i,h} (\mathbb{P}^k)_{h,j} \\ &= \sum_{h \in \varepsilon} (\mathbb{P})_{i,h} (\mathbb{P}^k)_{h,j} = (\mathbb{P} \mathbb{P}^k)_{i,j} = (\mathbb{P}^{k+1})_{i,j}. \end{aligned}$$

Por lo tanto,

$$\mathbb{P}^{(k+1)} = \mathbb{P}^{k+1}.$$

□

- Para todo $i \in \varepsilon$, la entrada i cumple que

$$\begin{aligned} \pi_i^{(n)} &= \mathbb{P}[X_n = i] = \sum_{k \in \varepsilon} \mathbb{P}[X_n = i \mid X_r = k] \mathbb{P}[X_r = k] \\ &= \sum_{k \in \varepsilon} P_{k,i}^{(n-r)} \pi_k^{(r)} = \sum_{k \in \varepsilon} \pi_k^{(r)} P_{k,i}^{(n-r)} = \left(\pi^{(r)} \mathbb{P}^{(n-r)} \right)_i. \end{aligned}$$

Por lo tanto,

$$\pi^{(n)} = \pi^{(r)} \mathbb{P}^{(n-r)}.$$

□

Este corolario resulta muy útil para el cálculo de distribuciones y probabilidades de transición a m pasos, ya que se reduce a realizar operaciones algebraicas con matrices y vectores. Es decir, si tenemos una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ y queremos calcular la distribución $\pi^{(m)}$ para algún $m \in \mathbb{N}$, un posible procedimiento es el siguiente:

1. Diagonalizar la matriz \mathbb{P} mediante cualquier método y expresar a dicha matriz como $\mathbb{P} = ADA^{-1}$ para alguna D matriz diagonal y A matriz invertible.
2. Con esto, tenemos la expresión más sencilla para la matriz de probabilidades de transición a m pasos

$$\mathbb{P}^{(m)} = \mathbb{P}^m = \mathbb{P} \mathbb{P} \cdots \mathbb{P} = ADA^{-1} ADA^{-1} \cdots ADA^{-1} = AD^m A^{-1}.$$

Por lo tanto,

$$\mathbb{P}^{(m)} = AD^m A^{-1}.$$

La matriz D^m se puede calcular fácilmente elevando cada una de las entradas de su diagonal a la m .

3. Finalmente, la distribución a tiempo m está dada por

$$\pi^{(m)} = \pi^{(0)} \mathbb{P}^{(m)}.$$

1.3. Clases de comunicación

Definición 1.3.1. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov y sean $i, j \in \varepsilon$.

- Decimos que i **accede a** j si existe $n \in \mathbb{N}$ tal que $P_{i,j}^{(n)} > 0$. Esto se denota como $i \rightarrow j$.
- Decimos que i **se comunica con** j si se cumple que i accede a j y que j accede a i . Esto se denota como $i \leftrightarrow j$.

Proposición 1.3.1. La comunicación es una relación de equivalencia.

Demostración. Sean $i, j, k \in \varepsilon$. Veamos que cumplen con lo requerido.

- Reflexividad:

$$\text{Como } P_{i,j}^{(0)} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}, \text{ entonces } P_{i,i}^{(0)} = 1 \text{ y } i \rightarrow i.$$

Por lo tanto, $i \leftrightarrow i$.

■ Simetría:

$$i \leftrightarrow j \implies i \rightarrow j \wedge j \rightarrow i \implies j \rightarrow i \wedge i \rightarrow j \implies j \leftrightarrow i.$$

Por lo tanto, $i \leftrightarrow j \implies j \leftrightarrow i$.

■ Transitividad:

Si $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces $i \rightarrow j$ y $j \rightarrow k$. Esto implica que existen $n_1, n_2 \in \mathbb{N}$ tales que $P_{i,j}^{(n_1)} > 0$ y $P_{j,k}^{(n_2)} > 0$ y, por consiguiente, $P_{i,j}^{(n_1)} P_{j,k}^{(n_2)} > 0$. Este es un sumando de las ecuaciones de Chapman-Kolmogorov. Así,

$$\begin{aligned} P_{i,k}^{(n_1+n_2)} &= \sum_{h \in \varepsilon} P_{i,h}^{(n_1)} P_{h,k}^{(n_2)} = P_{i,j}^{(n_1)} P_{j,k}^{(n_2)} + \sum_{h \in \varepsilon \setminus \{j\}} P_{i,h}^{(n_1)} P_{h,k}^{(n_2)} \\ &\geq P_{i,j}^{(n_1)} P_{j,k}^{(n_2)} > 0. \end{aligned}$$

Es decir, para $n_1 + n_2 \in \mathbb{N}$ se tiene que $P_{i,k}^{(n_1+n_2)} > 0$ y $i \rightarrow k$. Análogamente, obtenemos que $k \rightarrow j \wedge j \rightarrow i \implies k \rightarrow i$. Entonces, $i \leftrightarrow k$.

Por lo tanto, $i \leftrightarrow j \wedge j \leftrightarrow k \implies i \leftrightarrow k$.

En conclusión, la relación de comunicación es reflexiva, simétrica y transitiva. Por lo tanto, la comunicación es una clase de equivalencia. \square

Ya que tenemos definida una relación de equivalencia, se puede generar una partición del espacio de estados ε de acuerdo a sus clases de equivalencia.

Definición 1.3.2. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov y $i \in \varepsilon$. La **clase de comunicación del estado i** se define como

$$C(i) := \{j \in \varepsilon : j \leftrightarrow i\}.$$

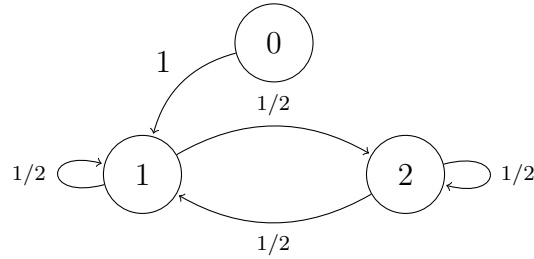
Estas clases de comunicación representan que los elementos en una misma clase se pueden comunicar entre ellos y los elementos que se encuentran en clases distintas no pueden comunicarse. En este último caso, es posible que exista una trayectoria para acceder a otra clase con probabilidad positiva, pero que ya no sea posible regresar por ninguna otra trayectoria.

En muchas ocasiones, las clases de comunicación se pueden identificar fácilmente a partir del esquema, como veremos en el ejemplo siguiente.

Ejemplo 1.3.1. Consideremos la cadena de Markov con espacio de estados $\varepsilon = \{0, 1, 2\}$ y matriz de probabilidad de transición

$$\mathbb{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Lo cual se puede visualizar en el siguiente esquema:



En este caso, contamos con dos clases de comunicación, $C(0) = \{0\}$ y $C(1) = C(2) = \{1, 2\}$. Estas clases tienen una intersección ajena y su unión genera el total, de manera que en efecto generan una partición del espacio de estados.

Definición 1.3.3. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov.

1. Una **clase** C se define como un subconjunto cualquiera de ε .
2. Decimos que el estado $i \in \varepsilon$ es **absorbente** si $P_{i,i} = 1$. Es decir, si en algún punto la trayectoria entra a estado i , ya no puede salir.
3. Decimos que una clase C es **cerrada** si para todos los $i \in C$ y $j \in C^c = \varepsilon \setminus C$ se cumple que

$$P_{i,j}^{(n)} = 0 \text{ para todo } n \geq 0.$$

O equivalentemente, para cada $i \in C$ no existe ningún $j \in C^c$ tal que $i \rightarrow j$. Es decir, si en algún punto la trayectoria entra a la clase C , ya no puede salir.

Una clase C es **abierto** si no es cerrada.

Ejemplo 1.3.1. (Continuación) Retomando el ejemplo anterior, podemos clasificar las clases de comunicación de la siguiente manera:

- Para $1 \in C(0)^c$ se cumple que $P_{0,1}^{(1)} = 1 \neq 0$. Por consiguiente, $C(0)$ es una clase abierta.

- Para todo $i \in \{1, 2\} = C(1)$ se cumple que $P_{i,0}^{(n)} = 0$ para todo $n \geq 0$. Por consiguiente, $C(1)$ es una clase cerrada.

Definición 1.3.4. Decimos que una cadena de Markov es **irreducible** si existe una única clase de comunicación, es decir, cuando $C(i) = C(j)$ para todos los $i, j \in \varepsilon$.

Proposición 1.3.2. Una cadena de Markov es irreducible si se cumple cualquiera de las siguientes condiciones.

- a) $C(i) = C(j)$ para todos los $i, j \in \varepsilon$.
- b) $i \rightarrow j$ para todos los $i, j \in \varepsilon$.
- c) $i \leftrightarrow j$ para todos los $i, j \in \varepsilon$.
- d) $C(i) = \varepsilon$ para todo $i \in \varepsilon$.
- e) $C(i) = \varepsilon$ para alguna $i \in \varepsilon$.

Demostración. A continuación, demostraremos la equivalencia previa.

a \implies c)

Sean $i, j \in \varepsilon$. Por reflexividad, $i \leftrightarrow i$ y así $i \in C(i) = C(j)$. Entonces $i \in C(j)$ y $i \leftrightarrow j$.

c \implies a)

Sean $i, j \in \varepsilon$. Como $i \leftrightarrow j$, para todo $k \in C(i)$ se cumple que $k \leftrightarrow j$ por transitividad. Entonces $k \in C(j)$ y así $C(i) \subseteq C(j)$. Análogamente, $C(j) \subseteq C(i)$. Por lo tanto, $C(i) = C(j)$.

b \implies c)

Sean $i, j \in \varepsilon$, entonces por hipótesis $i \rightarrow j$ y $j \rightarrow i$. Por lo tanto, $i \leftrightarrow j$.

c \implies d)

Sea $i \in \varepsilon$. Se cumple que $\varepsilon \subseteq C(i)$, pues para todo $j \in \varepsilon$, $i \leftrightarrow j$ por hipótesis y entonces $j \in C(i)$. Por otro lado, trivialmente $C(i) \subseteq \varepsilon$, de manera que $C(i) = \varepsilon$.

d \implies e)

Como por hipótesis se cumple para todo $i \in \varepsilon$, en particular se cumple para alguna $i \in \varepsilon$.

e \implies **b**)

Sea k tal que $C(k) = \varepsilon$ (lo cual existe por hipótesis), entonces para todos los $i, j \in \varepsilon$, se cumple que $i, j \in C(k)$. Así, $i \leftrightarrow k$ y $j \leftrightarrow k$. Finalmente, por transitividad, $i \leftrightarrow j$ y, en particular, $i \rightarrow j$.

Con esto quedan demostradas todas las equivalencias. □

1.4. Ejemplos

Ejemplo 1.4.1. (Cadena de dos estados) Uno de los ejemplos más sencillos de una cadena de Markov es la cadena de dos estados, la cual consiste en un proceso estocástico que cumple la propiedad de Markov y que cuenta con un espacio de estados de cardinalidad dos, así como ciertas probabilidades de transición homogéneas en el tiempo.

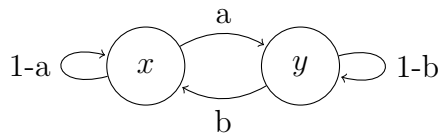
Sea $\{X_n\}_{n \in \mathbb{N}}$ una **cadena de Markov de dos estados**. A continuación, enunciaremos algunas de sus propiedades.

- a) El espacio de estados tiene cardinalidad 2, es decir, $\varepsilon = \{x, y\}$ para algunas x y y .
- b) Cumple con la propiedad de Markov.
- c) Tiene las siguientes probabilidades de transición:

$$\begin{aligned} P_{x,x} &= \mathbb{P}[X_1 = x | X_0 = x] = 1 - a & P_{y,x} &= \mathbb{P}[X_1 = x | X_0 = y] = b \\ P_{x,y} &= \mathbb{P}[X_1 = y | X_0 = x] = a & P_{y,y} &= \mathbb{P}[X_1 = y | X_0 = y] = 1 - b, \end{aligned}$$

para algunos $0 \leq a, b \leq 1$.

Esto se puede representar en un esquema de la siguiente manera:



En este caso, la matriz de probabilidad de transiciones es

$$\mathbb{P} = \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix}.$$

Adicionalmente, podemos escribir su distribución inicial como $\pi^{(0)} = (p_0, p_1)$ con $p_0 + p_1 = 1$.

Para obtener la distribución de la cadena a tiempo n , realizamos el procedimiento expuesto en la **sección 1.2**. Diagonalizando la matriz,

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1 - a - b \end{pmatrix} \quad y \quad A = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \implies A^{-1} = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{1}{a+b} & -\frac{1}{a+b} \end{pmatrix}.$$

Por lo tanto,

$$\mathbb{P}^{(n)} = AD^nA^{-1} = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1 - a - b)^n \end{pmatrix} \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{1}{a+b} & -\frac{1}{a+b} \end{pmatrix}.$$

Ejemplo 1.4.2. (Cadena de Ehrenfest) La Cadena de Ehrenfest fue propuesta por Tatyana y Paul Ehrenfest para describir el intercambio aleatorio de partículas en dos regiones separadas por una membrana porosa, donde solo una partícula puede atravesar la membrana a la vez. A continuación estudiaremos una versión simplificada de dicho modelo.

Planteamiento. Se tienen N partículas en dos contenedores, denotados por A y B. En cada tiempo, una partícula seleccionada de forma equiprobable se cambia de contenedor.

Para todo $n \in \mathbb{N}$, definimos a X_n como el número de partículas en el contenedor A después de n extracciones.

Demostración de que es una cadena de Markov. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov. Para todo $n \in \mathbb{N}$, sea Y_n la variable dada por

$$Y_n = \begin{cases} -1 & \text{si la partícula que se movió a tiempo } n \text{ salió del contenedor A,} \\ 1 & \text{si la partícula que se movió a tiempo } n \text{ salió del contenedor B.} \end{cases}$$

Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ dada por $g(x, y) = x + y$. Entonces, para todo $n \in \mathbb{N}$ se cumple que $X_{n+1} = g(X_n, Y_{n+1})$, donde Y_{n+1} es independiente de X_0, \dots, X_n . Finalmente, por la **Proposición 1.1.2** podemos afirmar que $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov.

Cálculo de la matriz de probabilidades de transición. El espacio de estados de esta cadena es $\varepsilon = \{0, 1, 2, \dots, N\}$, pues en el contenedor A se pueden tener entre 0 y N partículas. Entonces, para todo tiempo $n \in \mathbb{N}$, las probabilidades de transición son las siguientes:

- a) $P_{0,1} = \mathbb{P}[X_{n+1} = 1 \mid X_n = 0] = 1$, pues si no hay ninguna partícula en el compartimiento A a tiempo n , forzosamente la partícula seleccionada pasará del compartimiento B al compartimiento A y este incrementará su número de partículas en 1.
- b) $P_{N,N-1} = \mathbb{P}[X_{n+1} = N - 1 \mid X_n = N] = 1$, pues si las N partículas se encuentran en el compartimiento A a tiempo n , forzosamente la partícula seleccionada pasará al compartimiento B del compartimiento A y este disminuirá su número de partículas en 1.
- c) Para $k \in \{1, \dots, N - 1\}$,

$$\begin{aligned} P_{k,k-1} &= \mathbb{P}[X_{n+1} = k - 1 \mid X_n = k] \\ &= \mathbb{P}[\text{Se seleccionó una partícula de A} \mid X_n = k] \\ &= \frac{k}{N} \quad \text{y} \end{aligned}$$

$$\begin{aligned} P_{k,k+1} &= \mathbb{P}[X_{n+1} = k + 1 \mid X_n = k] \\ &= \mathbb{P}[\text{Se seleccionó una partícula de B} \mid X_n = k] \\ &= \frac{N - k}{N} = 1 - \frac{k}{N}. \end{aligned}$$

- d) Los incisos a)-c) representan todos los casos en los cuales se tienen probabilidades positivas, así que el resto de las probabilidades de transición son 0. Esto se puede observar directamente pues para todos los estados i , tenemos que $\sum_{j \in \mathcal{E}} P_{i,j} = 1$.

Así, la matriz de probabilidades de transición es

$$\mathbb{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{N} & 0 & 1 - \frac{1}{N} & 0 & \cdots & 0 & 0 \\ 0 & \frac{2}{N} & 0 & 1 - \frac{2}{N} & \cdots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Es decir,

$$P_{i,j} = \begin{cases} 1 & \text{si } (i = 0 \wedge j = 1) \vee (i = N \wedge j = N - 1) \\ \frac{i}{N} & \text{si } j = i - 1 \text{ y } i \neq N \\ 1 - \frac{i}{N} & \text{si } j = i + 1 \text{ y } i \neq 0 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Cálculo de distribuciones. Supongamos que X_0 se distribuye Binomial con parámetros N y $\frac{1}{2}$, es decir,

$$X_0 \sim \text{Bin} \left(N, \frac{1}{2} \right) \implies f_{X_0}(x) = \binom{N}{x} \frac{1}{2^N}.$$

En otras palabras,

$$\pi^{(0)} = \frac{1}{2} \left(\binom{N}{0}, \binom{N}{1}, \dots, \binom{N}{N} \right).$$

Calculando la distribución de X_1 ,

$$\begin{aligned} \pi^{(1)} &= \pi^{(0)} \mathbb{P} = \frac{1}{2} \left(\binom{N}{0}, \binom{N}{1}, \dots, \binom{N}{N} \right) \mathbb{P} \\ &= \frac{1}{2} \left(\binom{N}{1} \frac{1}{N}, \binom{N}{0} (1) + \binom{N}{2} \frac{2}{N}, \dots, \binom{N}{N-1} \frac{1}{N} \right) \\ &= \frac{1}{2} \left(\binom{N}{0}, \binom{N}{1}, \dots, \binom{N}{N} \right). \end{aligned}$$

Es decir, $X_1 \sim \text{Bin} \left(N, \frac{1}{2} \right)$ y tiene la misma distribución de X_0 . Si realizamos este proceso inductivamente, obtenemos que

$$X_n \sim \text{Bin} \left(N, \frac{1}{2} \right), \text{ para todo } n \in \mathbb{N}.$$

1.5. Estimación de probabilidades de transición

Ya que se estudió el modelo teórico de las cadenas de Markov discretas y homogéneas, es necesario relacionarlas con los datos muestrales obtenidos. En

este caso, supondremos que se tienen datos de un grupo fijo con observaciones en ciertos intervalos y se utilizarán para estimar la matriz de probabilidades de transición.

Para esto, primero definiremos la **longitud del ciclo** como el intervalo constante entre los puntos de tiempo distintos en los cuales se analizan los estados del grupo. De acuerdo con esta longitud, se cuentan con tres casos para la estimación:

(a) **Los intervalos de observación son constantes y coinciden con la longitud del ciclo.**

Este caso se refiere a cuando se tienen observaciones en ciertos periodos de tiempo constantes y se desea obtener una matriz de transición en los mismos periodos de tiempo. Por ejemplo, cuando se tienen observaciones cada dos años y se desea obtener una matriz de transición de probabilidades a dos años.

Para $\{X_t\}_{t \in I}$ una cadena de Markov homogénea con h estados, la estimación consta del siguiente procedimiento:

- Construimos la **matriz de recuento de observaciones** como

$$N = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1h} \\ n_{21} & n_{22} & \cdots & n_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ n_{h1} & n_{h2} & \cdots & n_{hh} \end{pmatrix},$$

donde para todos los $r, c \in \varepsilon$, n_{rc} representa el número de ocurrencias en las cuales un punto del grupo pasó del estado r al estado c en un ciclo.

- Ya que se tiene dicha matriz, el **estimador** de la matriz de transición de probabilidades se define como las proporciones de filas de N , es decir,

$$\widehat{M} = (\widehat{\theta})_{i,j \in \varepsilon}, \quad \text{donde } \widehat{\theta}_{rc} = \frac{n_{rc}}{\sum_{k=1}^h n_{rk}}, \quad \text{para todos los } r, c \in \varepsilon.$$

(b) **Los intervalos de observación son constantes, pero no coinciden con la longitud del ciclo.**

Este caso se refiere a cuando se tienen observaciones en ciertos periodos de tiempo constantes y se desea obtener una matriz de transición en periodos de tiempo distintos. Sea L_0 la longitud de los intervalos de observación y L_d la longitud deseada. Por ejemplo, si se tienen observaciones cada dos años y se desea obtener una matriz de transición de probabilidades a un año, entonces $L_0 = 2$ y $L_d = 1$.

Para obtener el estimador buscado, primero se necesita obtener la matriz de probabilidades de transición \widehat{M}_0 asociada a la longitud L_0 mediante el método definido en el caso **(a)**. Posteriormente, el **estimador** de la matriz de transición de probabilidades deseada está dada por

$$\widehat{M}_d = \widehat{M}_0^t, \quad \text{donde } t = \frac{L_d}{L_0}.$$

La potencia de la matriz \widehat{M}_0 se obtiene mediante su descomposición espectral, es decir,

$$\widehat{M}_0 = PDP^{-1}, \quad \text{con } D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_h \end{pmatrix},$$

donde λ_i es el i -ésimo valor propio y su vector propio asociado es la i -ésima columna de P . Entonces,

$$\widehat{M}_d = PD^tP^{-1}, \quad \text{con } D^t = \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_h^t \end{pmatrix}.$$

Este método no siempre funciona, ya que las cadenas de Markov discretas no necesariamente son cadenas de Markov en todas las longitudes de ciclo posibles. Sin embargo, este método funciona siempre que la matriz de transición estimada sea positiva semidefinida, es decir, que todos los valores propios sean no negativos. En su defecto, la estimación se puede realizar mediante el tercer método que expondremos a continuación.

(c) Los intervalos de observación no son constantes.

Supongamos que se tienen T intervalos de observación los cuales son enteros $\{k_1, k_2, \dots, k_T\}$ múltiplos de la longitud del ciclo. Para utilizar toda la información disponible se utiliza el algoritmo EM, el cual consiste en lo siguiente:

Se requiere una matriz de probabilidades de transición inicial para comenzar el algoritmo. Dado que la convergencia al estimador no está asegurada, es recomendable realizar el procedimiento con distintas matrices iniciales.

Paso E: Estimar los valores faltantes mediante el cálculo de los valores esperados de transiciones de un ciclo. Para esto, se utiliza la matriz de probabilidades de transición estimada en el paso M (o la matriz inicial en la primera iteración) para calcular la probabilidad de cada trayectoria que un punto del grupo pudo haber tomado para terminar donde lo hizo después de k_t ciclos. Posteriormente, para obtener el número de puntos que se movieron a cada estado en un ciclo, se multiplica esta probabilidad por $n_{rc}^{k_t}$ el número observado de puntos del grupo que se movieron del estado r al estado c después de k_t ciclos. Así se obtiene una matriz de recuento de observaciones para cada ciclo.

Paso M: Tratar los valores esperados de transiciones de un ciclo como si fueran los valores reales y aplicar el método definido en el caso **(a)**.

Estos dos pasos se repiten hasta que la matriz de probabilidades de transición se estabilice.

Este algoritmo puede resultar muy largo debido a que es posible contar con un gran número de caminos posibles, pero es fácil de calcular computacionalmente.

En Craig y Sendi (2002) [3], se pueden consultar ejemplos numéricos y más detalles de esta estimación.

1.6. Cadenas de Markov en R

1.6.1. Librería “markovchain” en R

La manera más directa de trabajar con cadenas de Markov en R es utilizando la librería “markovchain”, tal como se muestra en el siguiente ejemplo.

Ejemplo 1.6.1. Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov que representa la salud de un individuo a lo largo de su vida, medida en días. Su espacio de estados es $\varepsilon = \{\textit{saludable}, \textit{enfermo}, \textit{muerte}\}$ y su matriz de probabilidades de transición está dada por

$$\mathbb{P} = \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.4 & 0.5 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}.$$

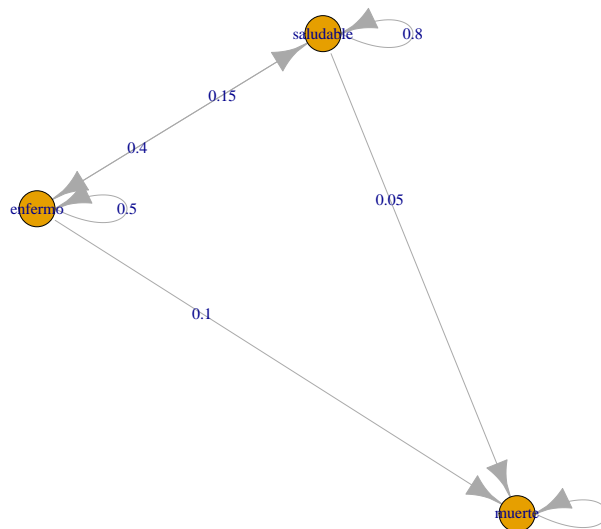
A continuación, utilizaremos dicha librería para obtener el esquema de la cadena, así como la distribución del estado de un individuo después de 2 días y después de 100 días, dado que en tiempo cero inició en estado saludable.

```

1 # LIBRERÍA DE R
2 # install.packages("markovchain")
3 library("markovchain")
4
5 # Creación de un elemento:
6 salud <- new("markovchain",
7             states = c("saludable", "enfermo", "muerte"),
8             transitionMatrix = matrix(data = c(0.8, 0.15, 0.05,
9             0.4, 0.5, 0.1, 0, 0, 1), byrow = TRUE, nrow = 3),
10            name = "Salud")
11
12 # Diagrama de estados:
13 plot(salud)
14
15 # Cálculo de distribuciones:
16 estadoInicial <- c(1, 0, 0) # La persona inicia en estado
17 saludable.
18 despues2dias <- estadoInicial * (salud ^ 2) # Representa la
19 distribución después de 2 días.
20 despues100dias <- estadoInicial * (salud ^ 100) # Representa la
21 distribución después de 100 días.

```

Con lo cual se obtuvo el siguiente esquema:



Y las siguientes distribuciones:

$$\pi^{(2)} = (0.7, 0.195, 0.105),$$

$$\pi^{(100)} = (0.00116408, 0.0003993612, 0.9984366).$$

Esta librería también nos permite clasificar los estados.

```

1 # Obtención de estados absorbentes:
2 absorbingStates(salud)
3 communicatingClasses(salud)
4 summary(salud)

```

Con lo cual se obtiene el siguiente resultado:

```

Console Terminal x Background Jobs x
R 4.4.2 · C:/Users/Itzel Rmz/OneDrive/Escritorio/Tesis/
> absorbingStates(salud)
[1] "muerte"
> communicatingClasses(salud)
[[1]]
[1] "saludable" "enfermo"

[[2]]
[1] "muerte"

> summary(salud)
Salud Markov chain that is composed by:
Closed classes:
muerte
Recurrent classes:
{muerte}
Transient classes:
{saludable,enfermo}
The Markov chain is not irreducible
The absorbing states are: muerte
>
>

```

Finalmente, es posible realizar simulaciones y estimación estadística mediante el siguiente código:

```

1 # Simulación
2 vidaCompleta <- rmarkovchain(n = 120, object = salud, t0 = "
   saludable")
3
4 # Estimación estadística
5 markovchainFit(vidaCompleta)$estimate

```

Sin embargo, esto se estudiará detalladamente mediante funciones definidas por el usuario.

La librería “markovchains” cuenta con una gran variedad de funciones relacionadas a las cadenas de Markov. Se pueden consultar más detalles al respecto en Spedicato et al. (2014) [4].

1.6.2. Simulación y estimación estadística

En esta subsección se presentará un ejemplo teórico de simulación y estimación estadística para cadenas de Markov, basado en los contenidos de este capítulo. En particular, para la estimación estadística utilizaremos el caso (a) de la sección previa.

A continuación, se definen tres funciones:

```

1 # DEFINICIÓN DE FUNCIONES:
2
3 condicionesCM <- function(P, E){
4   # Esta función auxiliar verifica las condiciones básicas y de
5   # formato que deben cumplir la matriz de probabilidades de
6   # transición y el espacio de estados, además de que imprime
7   # los mensajes de error correspondientes.
8
9   c <- FALSE
10  if(sum(rowSums(P) != 1) != 0 || sum(P < 0) != 0){
11    cat("ERROR \nLa matriz no es una matriz estocástica")
12  } else{
13    if(length(E) != nrow(P) || length(E) != ncol(P)){
14      cat("ERROR \nLas dimensiones de la matriz y vector
15      ingresados no concuerdan.")
16    } else{
17      if(is.data.frame(P) == FALSE){
18        cat("ERROR \nLa matriz ingresada no es un data frame.")
19      } else{
20        if(sum(row.names(P) != E, colnames(P) != E) != 0){

```

```

20         cat("ERROR \nLas etiquetas de la matriz no coinciden con
           el espacio de estados.")
21     } else{
22         c <- TRUE
23     }
24 }
25 }
26 }
27 return(c)
28 }
29
30 simulacionCM <- function(P, E, n){
31     # Esta función simula una realización de n pasos de una cadena
32     # de Markov con matriz de probabilidades de transición P y
33     # espacio de estados E.
34
35     # El espacio de estados se tratará como un vector de caracteres
36     # para permitir valores arbitrarios, mientras que P debe ser
37     # un data frame con el espacio de estados como etiquetas.
38
39     c <- condicionesCM(P,E)
40     if(c){
41         E <- as.character(E)
42         X0 <- sample(E, size = 1) # Un estado inicial arbitrario.
43         X <- c(X0, rep(0,n)) # Iniciamos la simulación con X0 y n
           ceros.
44         for(i in 2:(n+1)){
45             X[i] <- sample(E, 1, prob = P[X[i-1], ] )
46         }
47         X <- X[2:(n+1)] # Quitamos el primer punto.
48         return(X)
49     }
50 }
51
52 estimacionCM <- function(X){
53     # Esta función realiza una estimación estadística de la matriz
54     # de probabilidades de transición de una cadena de Markov a
55     # partir de los datos de una realización.
56
57     X <- as.character(X)
58     E <- sort(unique(X)) # Espacio de estados.
59     l <- length(E) # Cardinalidad del espacio de estados.
60     N <- data.frame(matrix(0, l, l) , row.names = E) # Matriz de
           recuento de observaciones.
61     colnames(N) <- E
62     for(i in 2:length(X)){
63         N[X[i-1],X[i]] <- N[X[i-1],X[i]] + 1

```

```

64 }
65 }
66 P <- N/rowSums(N)
67 return(P)
68 }

```

En el código previo, *condicionesCM* es una función auxiliar para llevar a cabo comprobaciones, las cuales incluyen condiciones probabilísticas y de formato para la matriz de probabilidades de transición y el espacio de estados. La segunda función *simulacionCM* realiza una simulación de cadena de Markov, para lo cual es necesario introducir tres argumentos: la matriz de probabilidades de transición P, el espacio de estados E y el número de pasos a simular n. Finalmente, la tercera función *estimacionCM* estima la matriz de probabilidades de transición tomando como único argumento los datos de una realización X de la cadena.

Ejemplo teórico.

Se desea realizar una simulación de trayectoria de tamaño 1000 de una cadena de Markov con espacio de estados $\varepsilon = \{0, 1, 2, 3\}$ y matriz de probabilidades de transición

$$\mathbb{P} = \begin{pmatrix} 0.1 & 0.15 & 0.65 & 0.1 \\ 0.15 & 0.25 & 0.55 & 0.05 \\ 0.1 & 0.05 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}.$$

Para ello, utilizamos el siguiente código:

```

1 # Argumentos iniciales:
2 E <- c(0,1,2,3)
3 P <- matrix(c(0.1, 0.15, 0.65, 0.1,
4             0.15, 0.25, 0.55, 0.05,
5             0.1, 0.05, 0.8, 0.05,
6             0.25, 0.25, 0.25, 0.25),
7           nrow=4,byrow=T)
8 P <- data.frame(P, row.names = E)
9 colnames(P) <- E
10
11 # Simulación y gráfica:
12 X <- simulacionCM(P, E, 1000)
13 X <- as.numeric(X)
14
15 #install.packages("ggplot2")
16 library("ggplot2")
17 t <- 1:1000
18 X <- data.frame(t, X)

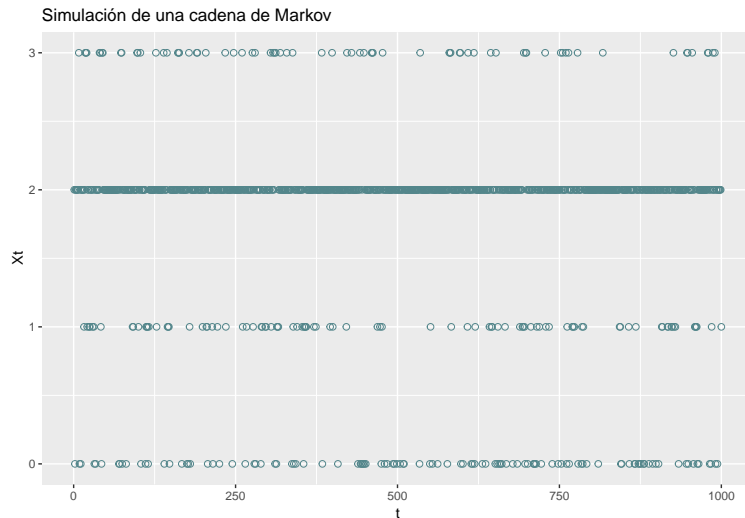
```

```

19 plot(X, type = "p", xlab = "Tiempo")
20 ggplot(data = X, aes(x = t, y = X)) +
21   geom_point(shape = 1, color = 'cadetblue4', size = 2)+
22   labs(x = "t", y = "Xt",
23        title = "Simulación de una cadena de Markov")

```

Con lo cual se obtuvo la siguiente gráfica:



A continuación, se realiza la estimación estadística de la matriz de probabilidades de transición utilizando como único argumento la realización X.

```

1 Ps <- estimacionCM(X$X)
2 error <- sum(abs(P-Ps))

```

Con lo cual se obtuvo el estimador

$$\begin{pmatrix} 0.08035714 & 0.1339286 & 0.6964286 & 0.08928571 \\ 0.14130435 & 0.2608696 & 0.5543478 & 0.04347826 \\ 0.09972299 & 0.0498615 & 0.7922438 & 0.05817175 \\ 0.24657534 & 0.2465753 & 0.2739726 & 0.23287671 \end{pmatrix}.$$

La matriz estimada debe ser muy semejante a la matriz teórica \mathbb{P} . En este caso, obtuvimos que el estimador presenta un error de 0.1875806. Para los fines de este ejercicio el error obtenido se considera suficientemente pequeño, aunque si se desea un estimador más cercano es necesario realizar más simulaciones.

Las cadenas de Markov son un tema muy extenso dentro del campo de la probabilidad, de manera que hay mucho más material al respecto. Si se desea profundizar más en el tema, se sugiere consultar bibliografía como Ross (2010) [1] y Suhov (2008) [2].

Capítulo 2

Series de tiempo

En el presente capítulo, comenzaremos con los conceptos básicos de las series de tiempo, las cuales son procesos estocásticos que describen variables aleatorias que varían a través del tiempo. Esto se utiliza para modelar una gran cantidad de fenómenos de la vida cotidiana, como pueden ser ciertos precios o el tamaño de una población. De tal manera, el principal objetivo de su análisis es la predicción de dichos fenómenos.

A continuación, se presentará una introducción teórica a las series de tiempo, considerando las definiciones básicas y la descomposición clásica. Posteriormente, se explicarán dos modelos (los cuales son especificaciones utilizadas para describir el comportamiento de la serie de tiempo) clásicos: los modelos ARMA y los modelos GARCH. Para finalizar, se describirán las bases de la estimación estadística paramétrica.

El contenido de este capítulo está basado en las ideas de Wei (1990) [5]; Chatfield (2004) [6]; y Larsen y Marx (2001) [7].

2.1. Motivación

De manera intuitiva, las series de tiempo surgen cuando se tiene un conjunto de observaciones de cierto fenómeno registradas secuencialmente en el tiempo. Estas observaciones serán denotadas por $\{x_{t_i}\}_{i \in I}$, donde

- I es un conjunto de índices, el cual puede ser continuo o discreto y puede incluir números negativos. Por ejemplo, I puede ser definido como \mathbb{N} , \mathbb{Z} o

\mathbb{R} . En el presente trabajo, utilizaremos únicamente conjuntos de índices discretos.

- Para toda $i \in I$, $t_i \in \mathbb{R}$ representa un punto en el tiempo en el cual se tiene una observación. Estos números reales deben cumplir con que $\{t_i\}_{i \in I}$ sea una sucesión creciente, pues de esta manera las observaciones estarán ordenadas en el tiempo.
- Para toda $i \in I$, x_{t_i} representa el valor que presentó la observación de interés en el tiempo t_i .

En realidad, la descripción anterior se puede entender como una observación específica del modelo probabilístico de una serie de tiempo.

Ya que se cuenta con los datos muestrales, es posible realizar el análisis de la serie de tiempo, donde generalmente se consideran tres pasos.

1. *Descripción*: Utilización de métodos descriptivos y gráficos sobre los datos que se tienen, con el objetivo de entender la estructura de la serie y tratar los datos atípicos en caso de que existan.
2. *Explicación*: Intentar encontrar un modelo al cual se pueda afirmar que la serie pertenece, con cierta confianza.
3. *Predicción*: Ya que se tiene el modelo, utilizarlo para realizar predicciones a futuro.

2.2. Definiciones básicas

Definición 2.2.1. Una **serie de tiempo** es una colección de variables aleatorias indexadas en el tiempo. Esto se denotará como $\{X_{t_i}\}_{i \in I}$, donde I es el conjunto indexador y $\{t_i\}_{i \in I}$ es una sucesión creciente.

Definición 2.2.2. Una serie de tiempo $\{X_{t_i}\}_{i \in I}$ con $I = \mathbb{Z}$ o \mathbb{N} es **equiespaciada** si la distancia entre tiempos consecutivos es constante, es decir, para todos los i , $i + 1 \in I$ se cumple que $t_{i+1} - t_i = c$ para alguna constante $c \in \mathbb{R}$.

En lo que resta de este escrito trabajaremos con series de tiempo equiespaciadas, pues las observaciones generalmente son registradas de tal manera.

Observación. Una serie de tiempo se puede expresar de manera simplificada como $\{X_t\}_{t \in T}$, donde T es un conjunto ordenado y equiespaciado de índices, el cual podrá ser interpretado como los tiempos en los cuales se tienen observaciones. En lo que resta de este escrito, consideraremos a $T = \mathbb{Z}$ o $T = \mathbb{N}$.

2.2.1. Características clásicas de una serie de tiempo.

Definición 2.2.3. Sea $\{X_t\}_{t \in T}$ una serie de tiempo.

- La **función de medias del proceso** se define como la aplicación

$$t \mapsto \mu_t = \mathbb{E}[X_t], \text{ para todo } t \in T,$$

siempre que esta esperanza exista.

- La **función de varianzas del proceso** se define como la aplicación

$$t \mapsto \sigma_t^2 = \text{Var}[X_t], \text{ para todo } t \in T,$$

siempre que esta varianza exista.

- La **función de autocovarianzas del proceso** se define como la aplicación

$$(t_1, t_2) \mapsto \gamma_{t_1, t_2} = \text{Cov}(X_{t_1}, X_{t_2}), \text{ para todo } (t_1, t_2) \in T \times T,$$

siempre que esta covarianza exista.

- La **función de autocorrelación del proceso** se define como la aplicación

$$(t_1, t_2) \mapsto \rho_{t_1, t_2} = \text{Cor}(X_{t_1}, X_{t_2}) = \frac{\gamma_{t_1, t_2}}{\sqrt{\sigma_{t_1}^2 \sigma_{t_2}^2}}, \text{ para todo } (t_1, t_2) \in T \times T,$$

siempre que esta correlación exista.

Estas definiciones se encuentran en términos de la esperanza, varianza, covarianza y correlación entre variables aleatorias, de manera que heredan todas sus características y propiedades.

Otra propiedad clásica de las series de tiempo es la estacionariedad, la cual se refiere al comportamiento cíclico que pueden presentar ciertas series de tiempo. A continuación, se presentarán algunas definiciones relacionadas.

Definición 2.2.4. Se dice que $\{X_t\}_{t \in T}$ es un **proceso estrictamente estacionario** si para todo $h \geq 0$ y para todos los $t_0, t_0+h, t_1, t_1+h, \dots, t_n, t_n+h \in T$ se cumple que

$$(X_{t_0}, X_{t_1}, \dots, X_{t_n}) \stackrel{D}{=} (X_{t_0+h}, X_{t_1+h}, \dots, X_{t_n+h}).$$

Aquí, $\stackrel{D}{=}$ representa igualdad en distribución, es decir, que se cumple la propiedad de que $F_{X_{t_0}, X_{t_1}, \dots, X_{t_n}}(x) = F_{X_{t_0+h}, X_{t_1+h}, \dots, X_{t_n+h}}(x)$ para todos los $x \in \mathbb{R}^{n+1}$.

Sin embargo, en muchos casos esta es una propiedad muy fuerte y difícil de probar, de manera que utilizaremos definiciones menos restrictivas.

Definición 2.2.5. Se dice que $\{X_t\}_{t \in T}$ es un **proceso estacionario de primer orden** o un **proceso estacionario en media** si la función de medias es constante, es decir, para todo $t \in T$, $\mu_t = \mu$ para alguna μ constante.

Definición 2.2.6. Se dice que $\{X_t\}_{t \in T}$ es un **proceso débilmente estacionario** o un **proceso estacionario de segundo orden** si

1. Las funciones de medias y varianzas existen y son constantes.
2. La función de autocovarianzas cumple con la propiedad de que $\gamma_{t,t+h}$ es independiente de t para todo h , es decir, $\gamma_{t,t+h} = \gamma_{s,s+h}$ para todos los $t, t+h, s, s+h \in T$.

Esto nos indica que la covarianza entre dos puntos en el tiempo depende únicamente de la distancia que hay entre ellos.

Trabajaremos con series de tiempo débilmente estacionarias, de manera que las funciones de autocovarianza y autocorrelación del proceso únicamente dependen de la distancia entre dos puntos, así que las funciones previas se pueden definir de manera simplificada.

Definición 2.2.7. Sea $\{X_t\}_{t \in T}$ una serie de tiempo.

- La **función de autocovarianza**, también conocida como ACVF, se define como

$$\gamma_h = \gamma(h) := \gamma(t, t+h), \text{ para cualquier } t, t+h \in T.$$

Es decir, esta función depende únicamente de la distancia h , la cual se conoce como **rezago**.

- La **función de autocorrelación**, también conocida como ACF, se define como

$$\rho_h = \rho(h) := \frac{\gamma_h}{\gamma_0}, \text{ para cualquier } t, t+h \in T .$$

Si tenemos n datos provenientes de una serie de tiempo, denotados por $\{x_i\}_{i=1}^n$, entonces podemos definir los estimadores de las funciones clásicas de una serie de tiempo.

- *Media muestral:*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

- *Función de autocovarianza (ACVF) muestral:*

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}) (x_t - \bar{x}), \text{ para } -n < h < n .$$

- *Función de autocorrelación (ACF) muestral:*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \text{ para } -n < h < n .$$

2.2.2. Operadores

Para trabajar con las series de tiempo, es útil definir algunos operadores. Estos son ciertas funciones que se les aplican a las series de tiempo con el objetivo de identificar patrones y simplificar los datos, de manera que resultan herramientas útiles para entender su comportamiento y predecirlas. A continuación definiremos tres operadores.

Para comenzar, definimos al operador de retraso, el cual debe su nombre a que al ser aplicado a una variable temporal, la retarda cierto número de periodos. Es decir, sirve para desplazar la serie de tiempo un número específico de periodos hacia atrás. Este operador puede resultar útil para expresar la dependencia temporal de la serie de tiempo.

Definición 2.2.8. (Operador de retraso) Sea $\{X_t\}_{t \in \mathbb{Z}}$ una serie de tiempo. Definimos el **operador de retraso**, denotado por B , como

$$BX_t = X_{t-1}, \text{ para todo } t \in \mathbb{Z}.$$

Aplicando este operador sucesivamente,

$$B^k X_t = X_{t-k}, \text{ para todos los } t \in \mathbb{Z} \text{ y } k \in \mathbb{N}.$$

Observación. B se puede operar con las definiciones de suma y producto de funciones usuales, así como sus propiedades de campo asociadas. Por consiguiente, B^k representa aplicar el operador B sucesivamente k veces. En particular, los operadores de retraso se pueden combinar con polinomios.

Sea A un polinomio real de grado $n \in \mathbb{N} \cup \{\infty\}$, entonces

$$A(x) = \sum_{j=0}^n a_j x^j \text{ para algunos } a_1, \dots, a_n \in \mathbb{R} \implies A(B) = \sum_{j=0}^n a_j B^j.$$

Por lo tanto,

$$A(B)X_t = \sum_{j=0}^n a_j B^j X_t = \sum_{j=0}^n a_j X_{t-j}, \text{ para todo } t \in \mathbb{Z}.$$

El siguiente operador a definir es el operador de diferencia, el cual sirve para calcular la diferencia entre el valor que tomó la serie de tiempo en $t-1$ y t . Este operador es útil pues nos permite estudiar cómo cambió la serie en lugar de estudiar el valor que tomó en cada punto del tiempo.

Definición 2.2.9. (Operador de diferencia) Sea $\{X_t\}_{t \in \mathbb{Z}}$ una serie de tiempo. Definimos el **operador de diferencia**, denotado por ∇ , como $\nabla = 1 - B$. Entonces,

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}.$$

Aplicando este operador inductivamente,

$$\nabla^k X_t = (1 - B)^k X_t = \sum_{j=0}^k \binom{k}{j} (-1)^j X_{t-j}, \text{ para todos los } t \in \mathbb{Z} \text{ y } k \in \mathbb{N}.$$

Finalmente, el operador de diferencia estacional nos permite calcular la diferencia entre el valor que tomó la serie de tiempo en $t-k$ y t . Es decir, es una generalización del operador previo en la cual se puede obtener la diferencia entre k periodos en lugar de únicamente 1 periodo. Este operador es útil pues nos permite estudiar cómo cambió la serie y, de esta manera, también se podrá estudiar el componente no estacional de algunas series de tiempo.

Definición 2.2.10. (*Operador de diferencia estacional*) Sea $\{X_t\}_{t \in \mathbb{Z}}$ una serie de tiempo. Definimos el **operador de diferencia estacional de d periodos**, denotada por ∇_d , como $\nabla_d = 1 - B^d$. Entonces,

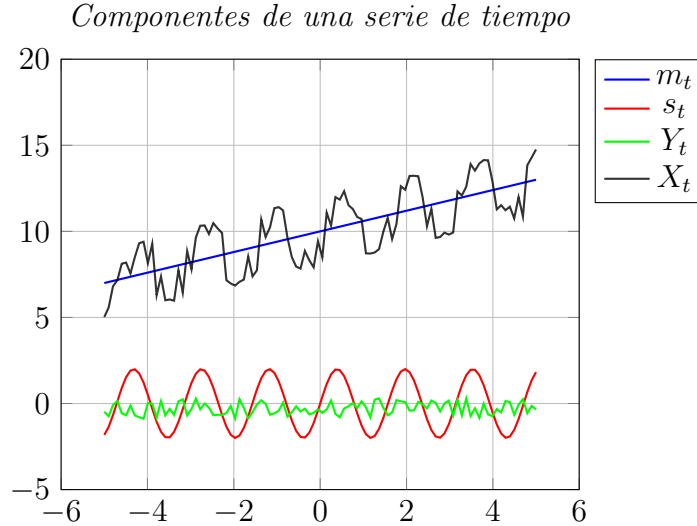
$$\nabla_d X_t = (1 - B^d)X_t = X_t - X_{t-d}.$$

2.3. Descomposición clásica de una serie de tiempo

Sea $\{X_t\}_{t \in T}$ una serie de tiempo, entonces podemos dividir a la serie en tres componentes.

- *Componente de tendencia:* Se refiere al comportamiento general de la serie, de manera que nos representa el promedio de la serie incluyendo sus periodos de crecimiento y decrecimiento. Se denota por $\{m_t\}_{t \in T}$ y generalmente es una función determinista.
- *Componente de estacionariedad:* Se refiere a los patrones predecibles que ocurren regularmente de acuerdo a ciertos ciclos. Por ejemplo, si analizamos el número de vuelos realizados en cada mes, notaremos que existen patrones repetitivos de acuerdo con las temporadas altas y bajas. Este componente se denota como $\{s_t\}_{t \in T}$ y generalmente es una función periódica y determinista.
- *Componente de ruido o de aleatoriedad:* Se refiere a la variabilidad residual que presenta la serie, es decir, aquello que no puede ser explicado por los otros componentes. Este componente se denota como $\{Y_t\}_{t \in T}$.

Analicemos cómo se ven dichos componentes de manera gráfica.



En este caso, el componente de tendencia m_t es la línea azul y representa el comportamiento promedio de la serie; el componente de estacionariedad s_t es la línea roja y representa el comportamiento cíclico de la serie; y la línea verde es el componente de ruido Y_t y representa la variabilidad residual de la serie. Finalmente, la línea negra representa la serie de tiempo original X_t .

De esta manera, ya tenemos la serie de tiempo dividida en sus componentes básicos: la tendencia $\{m_t\}_{t \in T}$, la estacionariedad $\{s_t\}_{t \in T}$ y el ruido $\{Y_t\}_{t \in T}$. A continuación, la serie original se puede representar como una función de tales componentes, es decir,

$$X_t = f(m_t, s_t, Y_t), \text{ para todo } t \in T,$$

donde f es una función específica en concordancia con las observaciones de la serie de tiempo. Convencionalmente se consideran dos tipos de modelos.

$$\text{Modelo aditivo: } X_t = m_t + s_t + Y_t .$$

$$\text{Modelos multiplicativos: } X_t = m_t \cdot s_t \cdot Y_t$$

$$\text{ó } X_t = m_t + s_t \cdot Y_t$$

$$\text{ó } X_t = s_t (m_t + Y_t) .$$

Siempre que $X_t > 0$ para todo $t \in T$, un modelo multiplicativo se puede transformar en un modelo aditivo. En particular, para todo $t \in T$,

$$\begin{aligned} X'_t &= \ln(X_t) = \ln(m_t \cdot s_t \cdot Y_t) = \ln(m_t) + \ln(s_t) + \ln(Y_t) \\ &= m'_t + s'_t + Y'_t . \end{aligned}$$

En el caso de que alguna observación sea menor que cero, se puede sumar una constante. Por consiguiente, el modelo aditivo generalmente bastará.

2.3.1. Estimación y eliminación de la tendencia

Un modelo no estacional es aquel cuya estructura estadística no varía de manera periódica a lo largo del tiempo, es decir, aquel cuya media, varianzas y autocovarianzas no dependen de ninguna periodicidad fija. En lo que resta de esta sección, trabajaremos con un tipo particular de modelo no estacional descrito mediante los componentes de su descomposición clásica, al cual nos referiremos simplemente como el *modelo no estacional aditivo*.

Definición 2.3.1. (*Modelo no estacional aditivo*) El **modelo no estacional aditivo** es un modelo aditivo sin el componente de tendencia, es decir, $X_t = m_t + Y_t$, para todo $t \in T$.

Para la presente sección, consideraremos modelos no estacionales con $\mathbb{E}[Y_t] = 0$ para todo $t \in T$. En realidad, este último supuesto no es muy fuerte, pues si tuviésemos un modelo con $\mathbb{E}[Y_t] = \mu_t \neq 0$, podemos considerar a la serie como $X_t = (m_t + \mu_t) + (Y_t - \mu_t) = m'_t + Y'_t$, con $\mathbb{E}[Y'_t] = 0$.

Después de estimar la tendencia, podemos expresar el componente de ruido como $Y_t = X_t - m_t$ para todo $t \in T$, de manera que podemos analizar su comportamiento a partir de los datos de la serie de tiempo original.

Existen diversos métodos para la estimación de la tendencia y no existe un método que sea mejor que el resto, sino que la selección depende completamente de cada serie de tiempo. A continuación, presentaremos dos métodos clásicos y algunas de sus características importantes.

Suavizado con filtro de medias móviles

Definición 2.3.2. Sea $\{X_t\}_{t \in T}$ una serie de tiempo y sea $q \in \mathbb{N} \setminus \{0\}$. Entonces la tendencia se puede estimar como

$$\hat{m}_t = (2q + 1)^{-1} \sum_{j=-q}^q X_{t-j}, \text{ para todo } t \in \{q, \dots, t - q\}.$$

Este estimador se conoce como el **promedio móvil de orden q** .

En esta definición, para cada tiempo t se está calculando el promedio desde q puntos antes hasta q puntos después, es decir, se están realizando promedios

en ventanas que se van moviendo. Además, en este caso observamos que en la serie se pierden los primeros y últimos datos, ya que no se pueden hacer promedios para los extremos.

Características:

- La estimación es buena para funciones monótonas y es muy buena para funciones lineales.
- Para $q \rightarrow \infty$, \hat{m}_t es constante y, por lo tanto, la estimación es mala.

Para $q \rightarrow 0$, $\hat{m}_t \approx X_t$ y, por lo tanto, la estimación es mala pues no aporta información al modelo.

- El natural q se suele estimar utilizando mínimos cuadrados, es decir, buscamos $q \in \mathbb{N} \setminus \{0\}$ tal que $\sum (X_t - \hat{m}_t)^2$ sea mínima.

Suavizado exponencial

Definición 2.3.3. Sea $\{X_t\}_{t \in \mathbb{N}}$ una serie de tiempo, entonces tenemos variaciones del suavizado exponencial.

a) Simple

Sea $\alpha \in [0, 1]$, entonces definimos

$$\begin{aligned} W_0 &= X_0, \\ W_t &= \alpha X_t + (1 - \alpha)W_{t-1}, \text{ para todo } t \geq 1. \end{aligned}$$

Para todo $t \in T$, W_t depende de todos los valores anteriores de la serie, de manera que se puede expresar recursivamente como

$$W_t = \sum_{j=0}^{t-1} \alpha (1 - \alpha)^j X_{t-j} + (1 - \alpha)^t X_0, \text{ para todo } t \geq 0.$$

Finalmente, el estimador para la tendencia será esta nueva serie W_t , es decir,

$$\hat{m}_t = W_t, \text{ para todo } t \in \mathbb{N}.$$

Este estimador se conoce como el **estimador por suavizado exponencial**.

Características:

- α mide la importancia del presente en la serie, mientras que $(1 - \alpha)$ mide la importancia del pasado de la serie.
- Si $\alpha = 0$, entonces $\hat{m}_t = X_0$ constante y, por lo tanto, la estimación es mala.

Si $\alpha = 1$, entonces $\hat{m}_t = X_t$ y, por lo tanto, la estimación es mala pues no aporta información al modelo.

- La elección de α también se puede hacer por mínimos cuadrados.

b) Brown o doble exponencial

Este suavizado consiste en aplicar el suavizado exponencial simple dos veces.

Sea $\alpha \in [0, 1]$, entonces definimos

$$\begin{aligned} W_0 &= X_0, \\ W_t &= \alpha X_t + (1 - \alpha)W_{t-1}, \text{ para todo } t \geq 1. \end{aligned}$$

Posteriormente, se realiza un suavizado exponencial simple sobre W_t ,

$$\begin{aligned} W'_0 &= W_0 = X_0, \\ W'_t &= \alpha W_t + (1 - \alpha)W'_{t-1}, \text{ para todo } t \geq 1. \end{aligned}$$

Finalmente, el estimador para la tendencia será esta nueva serie W'_t , es decir,

$$\hat{m}_t = W'_t, \text{ para todo } t \in \mathbb{N}.$$

Características:

- Nuevamente, α se puede obtener mediante el método de mínimos cuadrados.
- Para el segundo suavizado se puede tomar un valor de α distinto al primero, sin embargo, resulta usual utilizar el mismo.
- Generalmente, la α del primer suavizado se utiliza en el segundo, aunque pueden existir variaciones, como son obtener una α global o seleccionar una α' como el mínimo cuadrado del segundo suavizado.

- Se puede obtener un muy buen pronóstico para tiempos futuros como

$$\widehat{X}_{t+k} = a_t + kb_t, \text{ con } a_t = 2 W_t - W'_t$$

$$\text{y } b_t = \frac{\alpha}{1-\alpha}(W_t - W'_t).$$

Esto se realiza para el último tiempo t y funciona para $k > 0$.

c) Brown cuadrático o triple exponencial

Este suavizado consiste en aplicar el suavizado exponencial simple tres veces.

Sea $\alpha \in [0, 1]$, entonces definimos

$$W_0 = X_0,$$

$$W_t = \alpha X_t + (1 - \alpha)W_{t-1}, \text{ para todo } t \geq 1.$$

Posteriormente, se realiza un suavizado exponencial simple sobre W_t ,

$$W'_0 = W_0 = X_0,$$

$$W'_t = \alpha W_t + (1 - \alpha)W'_{t-1}, \text{ para todo } t \geq 1.$$

Para finalizar, se realiza un suavizado exponencial simple sobre W'_t ,

$$W''_0 = W'_0 = W_0 = X_0,$$

$$W''_t = \alpha W'_t + (1 - \alpha)W''_{t-1}, \text{ para todo } t \geq 1.$$

Finalmente, el estimador para la tendencia será esta nueva serie W''_t , es decir,

$$\widehat{m}_t = W''_t, \text{ para todo } t \in \mathbb{N}.$$

Características:

- Se puede obtener un muy buen pronóstico para tiempos futuros como

$$\widehat{X}_{t+k} = a_t + kb_t + \frac{1}{2}k^2c_t, \text{ con}$$

$$a_t = 3 W_t - 3 W'_t + W''_t,$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} ((6 - 5\alpha)W_t - (10 - 8\alpha)W'_t + (4 - 3\alpha)W''_t)$$

$$\text{y } c_t = \left(\frac{\alpha}{1-\alpha} \right)^2 (W_t - 2 W'_t + W''_t).$$

Con estos procedimientos, ya podemos realizar una estimación de la tendencia en caso de contar con un modelo no estacional aditivo. En caso de contar con un modelo estacional aditivo descrito mediante los componentes de su descomposición clásica, también es posible aplicar estos métodos mediante una transformación sencilla de la serie:

Observación. Sea $\{X_t\}_{t \in \mathbb{Z}}$ una serie de tiempo con componente estacional de periodo d , es decir, $s_t = s_{t+d}$ para todos los $t, t+d \in T$. Supongamos que $\sum_{i=1}^d s_d = 0$ (es decir, el promedio de puntos en un ciclo es cero). Entonces trabajaremos con la serie

$$Z_t = \nabla_d X_t = X_t - X_{t-d}.$$

Sean $t \in T$ y $X_t = m_t + s_t + Y_t$. La serie se puede expresar como

$$\begin{aligned} Z_t &= \nabla_d X_t = (1 - B^d) m_t + s_t + Y_t \\ &= (m_t + s_t + Y_t) - (m_{t-d} + s_{t-d} + Y_{t-d}) \\ &= (m_t - m_{t-d}) + (s_t - s_{t-d}) + (Y_t - Y_{t-d}) \\ &= m'_t + Y'_t, \end{aligned}$$

donde $m'_t = m_t - m_{t-d}$ es el componente de tendencia, $Y'_t = Y_t - Y_{t-d}$ es el componente aleatorio y $s_t - s_{t-d} = 0$. pues el componente estacional es de periodo d . Esta serie es no-estacional, de manera que podemos trabajar con ella utilizando los métodos expuestos previamente.

Aunque existen más métodos para estimar y eliminar la estacionalidad, únicamente se presentará el anterior, ya que existen modelos que estiman los tres componentes simultáneamente y son más utilizados. A continuación presentaremos uno de dichos modelos.

2.3.2. Modelo de Holt-Winters.

Si tenemos un modelo estacional, podemos realizar estimaciones de los componentes y predicciones de la serie mediante este desarrollo. En este caso, hay tres componentes a considerar.

- *Nivel (Level)*: Se refiere al promedio de la serie y se denota por $\{a_t\}_{t \in T}$.
- *Tendencia (Trend)*: Se refiere a los incrementos o decrementos de la serie y se denota por $\{b_t\}_{t \in T}$.

- *Estacionalidad:* Se refiere a la componente cíclica y se denota por $\{s_t\}_{t \in T}$.

Estos componentes son semejantes a los que habíamos definido previamente, con la estacionalidad definida exactamente como antes y la tendencia m_t dividida en dos componentes distintos que son el nivel a_t y la tendencia b_t definida como incrementos y decrementos.

Definición 2.3.4. Sea $\{X_t\}_{t \in \mathbb{N}}$ una serie de tiempo, entonces tenemos dos variaciones del modelo.

- **Forma aditiva.** Este modelo supone que la componente de estacionariedad afecta de forma aditiva, es decir, $X_t = m_t + s_t + Y_t$.

Entonces la estimación de componentes es la siguiente:

- 1) Para $\alpha \in (0, 1)$,

$$a_t = \alpha(X_t - s_{t-1}) + (1 - \alpha)(a_{t-1} - b_{t-1}) .$$

- 2) Para $\beta \in (0, 1)$,

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} .$$

- 3) Para $\gamma \in (0, 1)$,

$$s_t = \gamma(X_t - a_t) + (1 - \gamma)s_{t-1} .$$

Generalmente se considera $a_0 = b_0 = s_0 = X_0$, aunque también se pueden utilizar otras opciones como $a_0 = b_0 = s_0 = 0$.

Con esto, se puede realizar el pronóstico como

$$\hat{X}_{t+k} = a_t + kb_t + s_{t+(k \bmod p)} .$$

- **Forma multiplicativa.** Este modelo supone que la componente de estacionariedad afecta de forma multiplicativa como $X_t = s_t (m_t + Y_t)$.

Entonces la estimación de componentes es la siguiente:

- 1) Para $\alpha \in (0, 1)$,

$$a_t = \alpha \left(\frac{X_t}{s_{t-1}} \right) + (1 - \alpha)(a_{t-1} - b_{t-1}) .$$

2) Para $\beta \in (0, 1)$,

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} .$$

3) Para $\gamma \in (0, 1)$,

$$s_t = \gamma \left(\frac{X_t}{a_t} \right) + (1 - \gamma)s_{t-1} .$$

Generalmente se considera $b_0 = X_0$ y $a_0 = s_0 = \begin{cases} X_0 & \text{si } X_0 \neq 0 \\ 1 & \text{si } X_0 = 0 \end{cases}$.

Con esto, se puede realizar el pronóstico como

$$\widehat{X}_{t+k} = (a_t + kb_t) s_{t+(k \bmod p)} .$$

Nota. El modelo de Holt-Winters sin estacionalidad se obtiene de considerar el modelo anterior removiendo todas las s_t , es decir, sin considerar la tercera ecuación y quitando la s_{t-1} de la primera ecuación.

2.4. Modelos ARMA

Ya que estudiamos el comportamiento de las series de tiempo, presentaremos algunos modelos clásicos y muy utilizados. Estos modelos describen el presente de la serie en función de lo que ocurrió en tiempos previos.

Definición 2.4.1. Una serie de tiempo $\{\epsilon_t\}_{t \in T}$ es un **proceso de ruido blanco** si cumple con

- 1) Es un proceso débilmente estacionario.
- 2) Su función de autocorrelación (ACF) es

$$\rho_h = \begin{cases} 0 & \text{si } h \neq 0 \\ 1 & \text{si } h = 0 \end{cases} .$$

En ese caso, se denota como $WN(\mu_\epsilon, \sigma_\epsilon^2)$, donde $\mu_\epsilon = \mathbb{E}[\epsilon_t]$ y $\sigma_\epsilon^2 = \text{Var}[\epsilon_t]$.

Este tipo de procesos también se conocen como **procesos estocásticos no-correlacionados**.

Definición 2.4.2. Una serie de tiempo es un **ruido blanco estricto** si es una sucesión de variables aleatorias independientes e idénticamente distribuidas con varianza finita. En ese caso, se denota como $SWN(\mu_\epsilon, \sigma_\epsilon^2)$, donde $\mu_\epsilon = \mathbb{E}[\epsilon_t]$ y $\sigma_\epsilon^2 = Var[\epsilon_t]$.

Corolario 2.4.1. Si $\{\epsilon_t\}_{t \in T}$ es un $SWN(\mu_\epsilon, \sigma_\epsilon^2)$, entonces es un $WN(\mu_\epsilon, \sigma_\epsilon^2)$.

Demostración. Sean $\{\epsilon_t\}_{t \in T}$, $\mu_\epsilon = \mathbb{E}[\epsilon_t]$ y $\sigma_\epsilon^2 = Var[\epsilon_t]$.

- Sean $s, t \in T$. Como ϵ_s y ϵ_t son idénticamente distribuidas se tiene que $\mu_\epsilon = \mathbb{E}[\epsilon_s] = \mathbb{E}[\epsilon_t]$, es decir, la función de medias es constante.

Sean $s, s+h, t, t+h \in T$.

Si $h = 0$, entonces

$$\begin{aligned} \gamma_{s,s+h} &= Cov(\epsilon_s, \epsilon_{s+h}) = Var(\epsilon_s) = \sigma_\epsilon^2 \\ &= Var(\epsilon_t) = Cov(\epsilon_t, \epsilon_{t+h}) = \gamma_{t,t+h} \end{aligned} \cdot$$

Si $h \neq 0$, como son variable independientes se cumple que

$$\gamma_{s,s+h} = Cov(\epsilon_s, \epsilon_{s+h}) = 0 = Cov(\epsilon_t, \epsilon_{t+h}) = \gamma_{t,t+h} \cdot$$

Por lo tanto, $\{\epsilon_t\}_{t \in T}$ es un proceso débilmente estacionario.

- Sean $t, t+h \in T$.

Si $h = 0$, directamente de la definición se tiene que $\rho_h = 1$.

Si $h \neq 0$, se cumple que ϵ_t y ϵ_{t+h} son independientes. Entonces $\gamma_h = Cov(\epsilon_t, \epsilon_{t+h}) = 0$ y $\rho_h = 0$. Por lo tanto,

$$\rho_h = \begin{cases} 0 & \text{si } h \neq 0 \\ 1 & \text{si } h = 0 \end{cases} \cdot$$

□

2.4.1. Modelo autorregresivo (AR)

Definición 2.4.3. Sea $p \in \mathbb{N} \cup \{\infty\}$. Decimos que $\{X_t\}_{t \in T}$ es un **proceso autorregresivo de orden p** , denotado por $AR(p)$, si para $\{\epsilon_t\}_{t \in T}$ un

$WN(\mu_\epsilon, \sigma_\epsilon^2)$, se tiene que

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t.$$

Esta condición también se puede expresar como

$$\Phi(B)X_t = \epsilon_t, \text{ donde } \Phi(x) := 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p .$$

$\Phi(B)$ se conoce como el **operador autorregresivo**.

Observación. Un modelo autorregresivo es débilmente estacionario si las raíces de $\Phi(B)$ deben ser mayores o iguales que 1. Para más detalles, consultar el capítulo 3 de Wei (1990) [5].

Ejemplo 2.4.1. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un $AR(1)$ débilmente estacionario. Entonces $X_t = \phi_1 X_{t-1} + \epsilon_t$ para todo $t \in \mathbb{Z}$ y $|\phi_1| < 1$. A continuación, enunciaremos las propiedades principales de este proceso.

■ **Esperanza.**

Sea $\mu = \mathbb{E}[X_t]$ para cualquier $t \in \mathbb{Z}$, lo cual existe pues es un proceso débilmente estacionario. Entonces,

$$\mu = \mathbb{E}[X_t] = \phi_1 \mathbb{E}[X_{t-1}] + \mathbb{E}[\epsilon_t] = \phi_1 \mu + \mu_\epsilon \implies \mu = \frac{\mu_\epsilon}{1 - \phi_1},$$

lo cual está bien definido pues $\phi_1 \neq 1$.

■ **Varianza.**

Sea $\sigma^2 = Var[X_t]$ para cualquier $t \in \mathbb{Z}$, lo cual existe pues es un proceso débilmente estacionario. Entonces,

$$\sigma^2 = Var[X_t] = \phi_1^2 Var[X_{t-1}] + Var[\epsilon_t] = \phi_1^2 \sigma^2 + \sigma_\epsilon^2 \implies \sigma^2 = \frac{\sigma_\epsilon^2}{1 - \phi_1^2},$$

lo cual está bien definido pues $|\phi_1| < 1$.

■ **ACVF y ACF.**

$$\gamma_h = \begin{cases} \sigma^2 & \text{si } h = 0 \\ \phi_1 \gamma_{h-1} & \text{si } h \neq 0 \end{cases} \quad \text{y} \quad \rho_h = \begin{cases} 1 & \text{si } h = 0 \\ \phi_1 \rho_{h-1} & \text{si } h \neq 0 \end{cases} .$$

El procedimiento para obtener la función de autocovarianza puede ser encontrado en la sección 3.1 de Wei (1990) [5], mientras que la función de autocorrelación se obtiene directamente de la definición.

- **Expresión en términos del ruido blanco.**

Sustituyendo recursivamente en la ecuación,

$$X_t = \phi_1^n X_{t-n} + \sum_{j=0}^{n-1} \phi_1^j \epsilon_{t-j}.$$

Como $|\phi_1| < 1$, entonces $\phi_1^n \xrightarrow{n \rightarrow \infty} 0$ y, en el límite, $X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$.

2.4.2. Modelo de medias móviles (MA)

Definición 2.4.4. Sea $q \in \mathbb{N} \cup \{\infty\}$. Decimos que $\{X_t\}_{t \in T}$ es un **proceso de medias móviles de orden q** , denotado por $MA(q)$, si para $\{\epsilon_t\}_{t \in T}$ un $WN(\mu_\epsilon, \sigma_\epsilon^2)$, se tiene que

$$X_t = \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}.$$

Esta condición también se puede expresar como

$$X_t = \Phi(B)\epsilon_t, \text{ donde } \Phi(x) := 1 + \theta_1 x + \theta_2 x^2 + \dots + \theta_q x^q.$$

$\Phi(B)$ se conoce como el **operador de medias móviles**.

Observación. Este modelo siempre es débilmente estacionario, pues está expresado en términos del ruido blanco que es débilmente estacionario.

Ejemplo 2.4.2. Sea $\{X_t\}_{t \in \mathbb{Z}}$ un $MA(1)$. Entonces $X_t = \epsilon_t + \theta_1 \epsilon_{t-1}$ para todo $t \in \mathbb{Z}$ y es débilmente estacionario. A continuación, enunciaremos las propiedades principales de este proceso.

- **Esperanza.**

Sea $\mu = \mathbb{E}[X_t]$ para cualquier $t \in \mathbb{Z}$, lo cual existe pues es un proceso débilmente estacionario. Entonces,

$$\mu = \mathbb{E}[X_t] = \mathbb{E}[\epsilon_t] + \theta_1 \mathbb{E}[\epsilon_{t-1}] = \mu_\epsilon + \theta_1 \mu_\epsilon = \mu_\epsilon (1 + \theta_1).$$

- **Varianza.**

Sea $\sigma^2 = \text{Var}[X_t]$ para cualquier $t \in \mathbb{Z}$, lo cual existe pues es un proceso débilmente estacionario. Entonces,

$$\sigma^2 = \text{Var}[X_t] = \text{Var}[\epsilon_t] + \theta_1^2 \text{Var}[\epsilon_{t-1}] = \sigma_\epsilon^2 + \theta_1^2 \sigma_\epsilon^2 = \sigma_\epsilon^2 (1 + \theta_1^2).$$

■ **ACVF y ACF.**

$$\gamma_h = \begin{cases} \sigma_\epsilon^2(1 + \theta_1^2) & \text{si } h = 0 \\ \sigma_\epsilon^2\theta_1 & \text{si } h = 1 \\ 0 & \text{en otro caso} \end{cases} \quad \text{y} \quad \rho_h = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta_1}{1+\theta_1^2} & \text{si } h = 1 \\ 0 & \text{en otro caso} \end{cases} .$$

El procedimiento para obtener la función de autocovarianza puede ser encontrado en Nosedal (2019) [8], mientras que la función de autocorrelación se obtiene directamente de la definición.

Notas respecto a los modelos anteriores:

- Para AR(1) débilmente estacionarios, $AR(1) \equiv MA(\infty)$, pues $X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} = \epsilon_t + \sum_{j=1}^{\infty} \theta_j \epsilon_{t-j}$ para $\theta_j = \phi_1^j$, para todo $j \in \{1, 2, 3, \dots\}$.
- Los modelos no son únicos, es decir, puede que dos modelos del mismo tipo sean iguales en distribución aunque tengan parámetros distintos.

2.4.3. Modelo autorregresivo de medias móviles (AR-MA)

Definición 2.4.5. Sean $p, q \in \mathbb{N} \cup \{\infty\}$. Decimos que $\{X_t\}_{t \in T}$ es un **proceso autorregresivo de medias móviles (p, q)**, denotado por $ARMA(p, q)$, si para $\{\epsilon_t\}_{t \in T}$ un $WN(\mu_\epsilon, \sigma_\epsilon^2)$, se tiene que

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=0}^q \theta_j \epsilon_{t-j}, \text{ con } \theta_0 = 1.$$

Esta condición también se puede expresar como

$$\begin{aligned} \Phi(B)X_t &= \Theta(B)\epsilon_t, \text{ donde } \Phi(x) := 1 - \phi_1x - \phi_2x^2 - \dots - \phi_px^p \\ &\text{y } \Theta(x) := 1 + \theta_1x + \theta_2x^2 + \dots + \theta_qx^q. \end{aligned}$$

A continuación, veremos ciertas definiciones que nos brindarán información adicional respecto a los modelos anteriores.

Definición 2.4.6. Un proceso estocástico $\{X_t\}_{t \in T}$ es un **proceso causal** si

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \text{ para todo } t \in T,$$

donde $\{\epsilon_t\}_{t \in T}$ es un proceso de ruido blanco y $\{\psi_i\}_{i=0}^{\infty}$ es una sucesión de constantes reales absolutamente sumables, es decir, $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

Teorema 2.4.1. Sea $\{X_t\}_{t \in T}$ un proceso ARMA(p, q) tal que los polinomios $\Phi(x)$ y $\Theta(x)$ no tienen raíces en común. Entonces el proceso es causal si y solo si $\Phi(x)$ tiene todas sus raíces fuera del círculo unitario.

La demostración de esta propiedad puede ser encontrada en el apéndice B.2 de Shumway y Stoffer (2006) [9].

Proposición 2.4.1. Sea $\{X_t\}_{t \in T}$ un proceso ARMA(p, q) causal con $\mu_\epsilon = 0$, entonces su función de autocovarianzas es

$$\rho_h = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+h}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

Demostración. Sea $\{X_t\}_{t \in T}$ como en las hipótesis, entonces para todo $t \in T$,

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E} \left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \right] = \sum_{i=0}^{\infty} \psi_i \mathbb{E}[\epsilon_{t-i}] = 0 \\ \text{y } \text{Var}[X_t] &= \text{Var} \left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \right] = \sum_{i=0}^{\infty} \psi_i^2 \text{Var}[\epsilon_{t-i}] \\ &= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty, \text{ pues son absolutamente sumables.} \end{aligned}$$

Ahora, sean $h > 0$ y $t, t+h \in T$, entonces

$$\begin{aligned} \gamma_h &= \mathbb{E}[X_t X_{t+h}] = \mathbb{E} \left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \epsilon_{t+h-j} \right] = \mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \epsilon_{t-i} \epsilon_{t+h-j} \right] \\ &= \mathbb{E} \left[\sum_{i=0}^{\infty} \psi_i \psi_{i+h} \epsilon_{t-i}^2 \right] = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}. \end{aligned}$$

Finalmente, $\rho_h = \frac{\gamma_h}{\gamma_0}$ y con ello se obtiene la expresión deseada. \square

A continuación, presentaremos una definición que nos ayudará a identificar los parámetros de un proceso ARMA.

Definición 2.4.7. Sea $\{X_t\}_{t \in T}$ una serie de tiempo. Definimos su **función de autocorrelación parcial (PACF)** como

$$\begin{aligned} \phi_{1,1} &= \text{Corr}(X_1, X_0) = \rho_1, \\ \phi_{h,h} &= \text{Corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}) \text{ para } h \geq 2, \end{aligned}$$

donde

- $X_h^{h-1} = \beta_1 X_{h-1} + \beta_2 X_{h-2} + \dots + \beta_{h-1} X_1$ con los coeficientes $\beta_1, \beta_2, \dots, \beta_{h-1}$ obtenidos de una regresión lineal entre X_h y $\{X_1, X_2, \dots, X_{h-1}\}$.
- $X_0^{h-1} = \alpha_1 X_{h-1} + \alpha_2 X_{h-2} + \dots + \alpha_{h-1} X_1$ con los coeficientes $\alpha_1, \alpha_2, \dots, \alpha_{h-1}$ obtenidos de una regresión lineal entre X_0 y $\{X_1, X_2, \dots, X_{h-1}\}$.

Notas:

- La idea de tomar a $(X_h - X_h^{h-1})$ y $(X_0 - X_0^{h-1})$ es quitar la correlación lineal con el vector $(X_1, X_2, \dots, X_{h-1})$.
- Por estacionariedad, el PACF es la correlación entre X_t y X_{t-h} removiendo la dependencia lineal a $(X_1, X_2, \dots, X_{h-1})$.
- Existen métodos numéricos para calcular esta función, sin embargo, no se presentarán pues se puede obtener utilizando lenguajes de programación que tengan dicha función implementada.

Finalmente, el criterio de selección de modelos de una serie de tiempo ARMA se resume en la tabla presentada a continuación.

Tabla 2.1: Comportamiento de la ACF y PACF para un ARMA causal.

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Decrece asintótico a cero.	Se corta después de q .	Decrece asintótico a cero.
PACF	Se corta después de p .	Decrece asintótico a cero.	Decrece asintótico a cero.

2.5. Procesos GARCH

En la sección previa analizamos cierto tipo de modelos lineales, sin embargo, no todas las series de tiempo presentan un comportamiento lineal. A continuación, se presentarán ciertos tipos de modelos no lineales con varianza no constante a lo largo del tiempo. Previo a ello, se explicarán los hechos estilizados de las series de tiempo financieras, ya que muchas series de este tipo pueden ser descritas bajo dicho modelo.

2.5.1. Hechos estilizados

Los hechos estilizados de una serie de tiempo financiera son un conjunto de observaciones empíricas e inferencias sobre estas observaciones que parecen aplicar a la mayoría de series de tiempo diarias de cambios en factores de riesgo, como pueden ser los tipos de cambio y los log-retornos de índices. En muchas ocasiones, estos hechos se mantienen en series de tiempo con temporalidades distintas, aunque no siempre sucede. Una versión de estos hechos enunciada por McNeil et al. (2005) [10] es la siguiente:

- a) A pesar de que las series de retornos muestran una baja correlación serial, no son independientes ni idénticamente distribuidas.
- b) Las series de valores absolutos o cuadrados de los retornos muestran una alta correlación serial.
- c) La esperanza condicional de los retornos es cercana a cero.
- d) La volatilidad parece variar a lo largo del tiempo.
- e) Las series de retornos son leptocúrticas o de colas pesadas.
- f) Los valores extremos de las series de retornos se presentan en agrupaciones.

2.5.2. Proceso Autorregresivo de Heterocedasticidad Condicional (ARCH)

Definición 2.5.1. Sean $p \in \mathbb{N} \cup \{\infty\}$ y $\{Z_t\}_{t \in T}$ un $SWN(0, 1)$. Decimos que $\{X_t\}_{t \in T}$ es un **proceso autorregresivo de heterocedasticidad condicional de parámetro p** , denotado por $ARCH(p)$, si se cumple que

- 1) Es un proceso débilmente estacionario.
- 2) Para algún proceso $\{\sigma_t\}_{t \in T}$, $\alpha_0 > 0$ y $\alpha_i \geq 0$ para cada $i \in \{1, \dots, p\}$,

$$X_t = \sigma_t Z_t \quad y$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad .$$

Propiedades:

- El proceso $\{\sigma_t\}_{t \in T}$ es predecible con \mathcal{F}_t la información de X_t , es decir, $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$.
- $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t] = 0$.
- $\text{Var}[X_t | \mathcal{F}_{t-1}] = \sigma_t^2 \text{Var}[Z_t] = \sigma_t^2$.
- $\{X_t\}_{t \in T}$ es débilmente estacionario si y solo si $\sum_{i=1}^p \alpha_i < 1$.

La demostración de esta propiedad puede ser encontrada en el capítulo titulado “Stationarity, Mixing, Distributional Properties and Moments of GARCH (p,q)–Processes” de Andersen et al. (2009) [11].

Ejemplo 2.5.1. Un proceso $ARCH(1)$ se ve como

$$X_t = \sigma_t Z_t \text{ con } \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 .$$

2.5.3. Proceso Autorregresivo Generalizado de Heterocedasticidad Condicional (GARCH)

Definición 2.5.2. Sean $p, q \in \mathbb{N} \cup \{\infty\}$ y $\{Z_t\}_{t \in T}$ un $SWN(0, 1)$. Decimos que $\{X_t\}_{t \in T}$ es un **proceso autorregresivo generalizado de heterocedasticidad condicional de parámetros p, q** , denotado por $GARCH(p, q)$, si se cumple que

- 1) Es un proceso débilmente estacionario.
- 2) Para algún proceso $\{\sigma_t\}_{t \in T}$, $\alpha_0 > 0$, $\alpha_i \geq 0$ para cada $i \in \{1, \dots, p\}$ y $\beta_j \geq 0$ para cada $j \in \{1, \dots, q\}$,

$$X_t = \sigma_t Z_t \text{ y}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 .$$

Propiedades:

- El proceso $\{\sigma_t\}_{t \in T}$ es predecible con \mathcal{F}_t la información de X_t , es decir, $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$.
- $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \sigma_t \mathbb{E}[Z_t] = 0$.
- $\text{Var}[X_t | \mathcal{F}_{t-1}] = \sigma_t^2 \text{Var}[Z_t] = \sigma_t^2$.

- $\{X_t\}_{t \in T}$ es débilmente estacionario si y solo si $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.

La demostración de esta propiedad puede ser encontrada en el capítulo titulado “Stationarity, Mixing, Distributional Properties and Moments of GARCH (p,q)-Processes” de Andersen et al. (2009) [11].

Ejemplo 2.5.2. Un proceso $GARCH(1,1)$ se ve como

$$X_t = \sigma_t Z_t \text{ con } \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 .$$

2.5.4. Simulación de series de tiempo GARCH

En esta subsección se realizarán simulaciones de distintos procesos $GARCH(1,1)$ en R. A continuación, se definen dos funciones:

```

1 # install.packages("fGarch")
2 library("fGarch")
3
4 # Función auxiliar para generar simulaciones de un SWN:
5
6 simAleatoria <- function(l, n){
7   # Esta función auxiliar verifica las condiciones básicas y de
8   # formato que debe cumplir la especificación de distribución
9   # del ruido blanco, denotada por l. En caso de que se cumplan
10  # dichas condiciones, regresa una simulación de n elementos
11  # pertenecientes a dicha distribución. En su defecto, imprime
12  # los mensajes de error correspondientes y regresa FALSE.
13
14  sim <- FALSE
15
16  if (is.list(l) == FALSE){
17    cat("ERROR \nLa primera entrada del input debe ser una lista."
18      )
19  } else{
20
21    if (l[[1]] == "norm" & length(l) == 1){
22      sim <- rnorm(n, 0, 1)
23    } else{
24      if (l[[1]] == "tstudent" & length(l) == 2){
25        if (is.numeric(l[[2]]) & l[[2]] > 2){
26          sim <- rstd(n, 0, 1, nu = l[[2]]) }
27        } else{
28          if (l[[1]] == "sged" & length(l) == 3){
29            if (is.numeric( c(l[[2]], l[[3]]) )){
30              sim <- rsged(n, 0, 1, nu = l[[2]], xi = l[[3]]) }
31            }
32          }
33        }
34      }
35    }
36  }
37  }
38  }
39  }
40  }

```

```

31     }
32   }
33
34   if(sum(sim) == FALSE){
35     cat("ERROR \nIntroducir la distribución bajo el siguiente
36         formato:
37         \nDistribución normal: list('norm')
38         \nDistribución t de Student normalizada: list('tstudent',
39             nu), con nu > 2 los grados de libertad.
40         \nDistribución de Error Generalizada Sgegada: list('sged',
41             nu, xi), con nu el parámetro numérico de forma y xi el
42             parámetro numérico de asimetría.")
43   }
44 }
45
46 # Función principal para generar simulaciones de un GARCH(1,1):
47
48 simulacionGarch <- function(alpha0, alpha1, beta1, n, distr, x0,
49                             s0){
50   # Esta función simula una realización de n pasos de una serie de
51   # tiempo GARCH(1,1) con parámetros alpha0, alpha1 y beta1,
52   # utilizando un ruido blanco estricto con distribución distr
53   # y con los valores iniciales X0=x0 y sigma0=s0.
54
55   if(alpha1 + beta1 >= 1){
56     cat("ERROR \nLa suma de los parámetros alpha1 y beta1 debe ser menor
57         que 1.")
58   } else{
59     if(alpha0 <= 0 || alpha1 < 0 || beta1 < 0){
60       cat("ERROR \nEl parámetro alpha0 debe ser positivo, alpha1 debe ser
61           no negativo y beta1 debe ser no negativo.")
62     } else{
63       # Ya que se verificaron las condiciones básicas de los pará-
64       # metros, se realiza la simulación.
65       ruido <- simAleatoria(distr, n)
66       if(sum(ruido)){
67         s <- s0
68         X <- c(x0)
69         for(i in 1:n){
70           s2 <- alpha0 + (alpha1 * tail(X,1)^2) + (beta1 * s^2)
71           s <- sqrt(s2)
72           X <- append(X, s * ruido[i])
73         }
74         X <- X[2:(n+1)] # Quitamos el primer punto.

```

```

70     return(X)
71   }
72 }
73 }
74 }

```

En el código previo, *simAleatoria* es una función auxiliar que comprueba que la especificación de distribución ingresada sea correcta y, en caso de que lo sea, regresa una simulación de n elementos de dicha distribución. La segunda función *simulacionGarch* realiza una simulación de un proceso GARCH(1,1) con los parámetros especificados.

A continuación, utilizaremos la función *simulacionGarch* para simular distintos procesos GARCH(1,1).

Variando α_0 .

Para comenzar, consideraremos un ruido normal con distintos valores del parámetro α_0 . En este caso, simularemos las siguientes series de tiempo:

- $N1_t = \sigma_t Z_t$ con $\sigma_t^2 = 1 + 0.3 N1_{t-1}^2 + 0.5\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.
- $N2_t = \sigma_t Z_t$ con $\sigma_t^2 = 50 + 0.3 N2_{t-1}^2 + 0.5\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.
- $N3_t = \sigma_t Z_t$ con $\sigma_t^2 = 100 + 0.3 N3_{t-1}^2 + 0.5\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.

Para ello utilizaremos el siguiente código:

```

1 #install.packages("ggplot2")
2 #install.packages("gridExtra")
3 library("ggplot2")
4 library("gridExtra")
5
6 # Simulaciones normales con distintos parámetros
7 n <- 2000
8 t <- 1:n
9
10 # Distintos valores de alpha0
11 N1 <- simulacionGarch(1, 0.3, 0.5, n, list("norm"), 1, 0.5)
12 N1 <- data.frame(t, N1)
13 plotN1 <- ggplot(data = N1, aes(x = t, y = N1)) +
14   geom_line(color = 'deeppink4', linewidth = .5) +
15   labs(x = "t", y = "Xt",
16        title = "Simulación de un GARCH(1,1) con ruido normal",
17        subtitle = "Parámetros  $\alpha_0=1$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
18   theme_minimal()

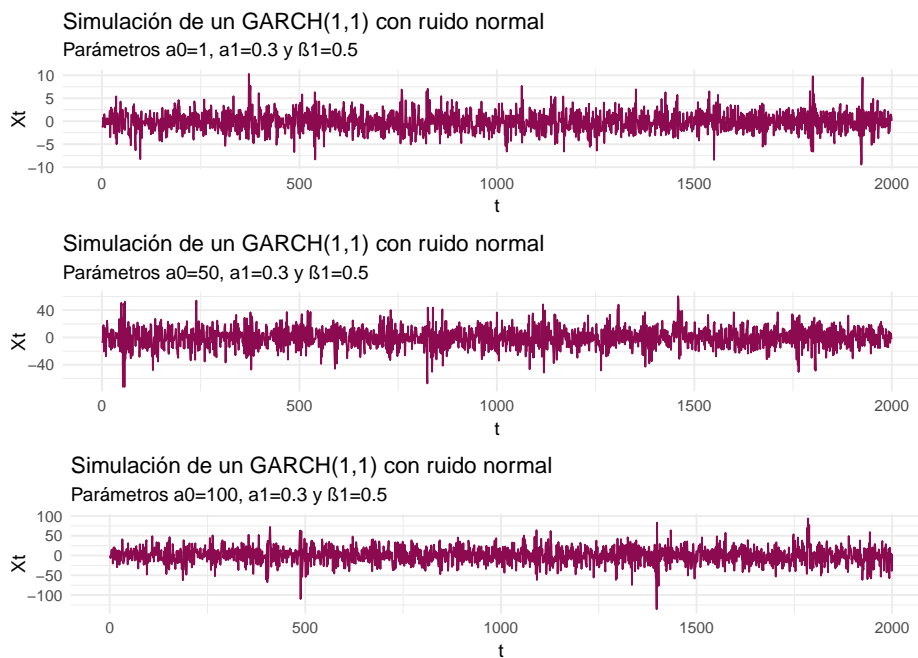
```

```

19
20 N2 <- simulacionGarch(50, 0.3, 0.5, n, list("norm"), 1, 0.5)
21 N2 <- data.frame(t, N2)
22 plotN2 <- ggplot(data = N2, aes(x = t, y = N2)) +
23   geom_line(color = 'deeppink4', linewidth = .5)+
24   labs(x = "t", y = "Xt",
25        title = "Simulación de un GARCH(1,1) con ruido normal",
26        subtitle = "Parámetros  $\alpha_0=50$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
27   theme_minimal()
28
29 N3 <- simulacionGarch(100, 0.3, 0.5, n, list("norm"), 1, 0.5)
30 N3 <- data.frame(t, N3)
31 plotN3 <- ggplot(data = N3, aes(x = t, y = N3)) +
32   geom_line(color = 'deeppink4', linewidth = .5)+
33   labs(x = "t", y = "Xt",
34        title = "Simulación de un GARCH(1,1) con ruido normal",
35        subtitle = "Parámetros  $\alpha_0=100$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
36   theme_minimal()
37
38 varianzas <- c(var(N1[2]), var(N2[2]), var(N3[2]))
39 grid.arrange(plotN1, plotN2, plotN3, ncol=1)

```

Con lo cual se obtuvo el vector de varianzas (4.741923, 226.234819, 436.508146) y la gráfica:



Podemos observar que mientras mayor es el valor del parámetro α_0 , la serie tiene una mayor variación, lo cual también es notorio en las varianzas numéricas obtenidas de cada serie. Esto tiene sentido analíticamente pues α_0 se suma a σ_t en cada paso. En realidad, al cambiar los valores de α_0 , las simulaciones se ven muy semejantes en distribución, únicamente con un cambio de escala. Esto tiene sentido pues este parámetro se puede interpretar como una especie de varianza mínima en la definición de la varianza de la serie.

Variando α_1 y β_1 .

Posteriormente, consideraremos un ruido normal con distintos valores de los parámetros α_1 y β_1 . En este caso, simularemos las siguientes series de tiempo:

- $N4_t = \sigma_t Z_t$ con $\sigma_t^2 = 1 + 0.1 N4_{t-1}^2 + 0.45\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.
- $N5_t = \sigma_t Z_t$ con $\sigma_t^2 = 1 + 0.45 N5_{t-1}^2 + 0.45\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.
- $N6_t = \sigma_t Z_t$ con $\sigma_t^2 = 1 + 0.45 N6_{t-1}^2 + 0.1\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.

Para ello utilizaremos el siguiente código:

```

1 # Distintos valores de alpha1 y beta1
2 N4 <- simulacionGarch(1, 0.1, 0.45, n, list("norm"), 1, 0.5)
3 N4 <- data.frame(t, N4)
4 plotN4 <- ggplot(data = N4, aes(x = t, y = N4)) +
5   geom_line(color = 'chartreuse4', linewidth = .5)+
6   labs(x = "t", y = "Xt",
7        title = "Simulación de un GARCH(1,1) con ruido normal",
8        subtitle = "Parámetros  $\alpha_0=1$ ,  $\alpha_1=0.1$  y  $\beta_1=0.45$ ") +
9   theme_minimal()+
10  ylim(-15, 15)
11
12 N5 <- simulacionGarch(1, 0.45, 0.45, n, list("norm"), 1, 0.5)
13 N5 <- data.frame(t, N5)
14 plotN5 <- ggplot(data = N5, aes(x = t, y = N5)) +
15   geom_line(color = 'chartreuse4', linewidth = .5)+
16   labs(x = "t", y = "Xt",
17        title = "Simulación de un GARCH(1,1) con ruido normal",
18        subtitle = "Parámetros  $\alpha_0=1$ ,  $\alpha_1=0.45$  y  $\beta_1=0.45$ ") +
19   theme_minimal()+
20   ylim(-15, 15)
21
22 N6 <- simulacionGarch(1, 0.45, 0.1, n, list("norm"), 1, 0.5)
23 N6 <- data.frame(t, N6)
24 plotN6 <- ggplot(data = N6, aes(x = t, y = N6)) +
25   geom_line(color = 'chartreuse4', linewidth = .5)+

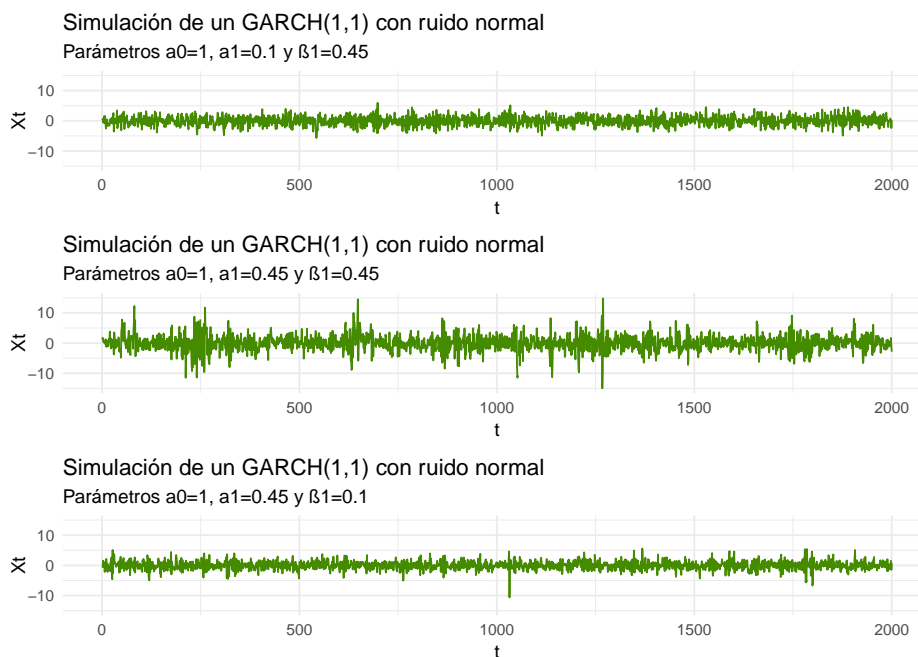
```

```

26 labs(x = "t", y = "Xt",
27       title = "Simulación de un GARCH(1,1) con ruido normal",
28       subtitle = "Parámetros  $\alpha_0=1$ ,  $\alpha_1=0.45$  y  $\beta_1=0.1$ ") +
29 theme_minimal()+
30 ylim(-15, 15)
31
32 varianzas <- c(var(N4[2]), var(N5[2]), var(N6[2]))
33 grid.arrange(plotN4, plotN5, plotN6, ncol=1)

```

Con lo cual se obtuvo el vector de varianzas (2.375247, 8.017753, 2.147776) y la gráfica:



Podemos observar que mientras mayores sean los valores de α_1 y β_1 , la serie tiene una mayor variación, lo cual también es notorio en las varianzas numéricas obtenidas de cada serie. Esto tiene sentido analíticamente pues α_1 y β_1 multiplican a valores positivos y se suman a σ_t^2 en cada paso. En este caso, a diferencia del previo, al variar α_1 y β_1 las simulaciones no parecen tener la misma distribución.

Variando la distribución del ruido blanco estricto.

Finalmente, utilizando valores fijos de todos los parámetros, consideraremos ruidos con distintas distribuciones. Además, graficaremos tanto la serie de tiempo simulada X_t , como un histograma de los datos. En este caso, simu-

laremos las siguientes series de tiempo:

- $X1_t = \sigma_t Z_t$ con $\sigma_t^2 = 2 + 0.3 X1_{t-1}^2 + 0.5\sigma_{t-1}^2$ y $Z_t \sim N(0,1)$.
- $X2_t = \sigma_t Z_t$ con $\sigma_t^2 = 2 + 0.3 X2_{t-1}^2 + 0.5\sigma_{t-1}^2$ y Z_t con distribución t de Student estandarizada con 5 grados de libertad.
- $X3_t = \sigma_t Z_t$ con $\sigma_t^2 = 2 + 0.3 X3_{t-1}^2 + 0.5\sigma_{t-1}^2$ y Z_t con distribución de Error Generalizada Ssgada (SGED)¹ con parámetro de forma 2 y parámetro de asimetría 1.5.

Para ello utilizaremos el siguiente código:

```

1 # Simulaciones con distintas distribuciones
2 n <- 5000
3 t <- 1:n
4
5 X1 <- simulacionGarch(2, 0.3, 0.5, n, list("norm"), 1, 0.5)
6 X1 <- data.frame(t, X1)
7 plotX1 <- ggplot(data = X1, aes(x = t, y = X1)) +
8   geom_line(color = 'cyan4', linewidth = .5)+
9   labs(x = "t", y = "Xt",
10        title = "Simulación de un GARCH(1,1) con ruido normal",
11        subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
12   theme_minimal()+
13   ylim(-20, 20)
14 plotX1h <- ggplot(data = X1, aes(x = X1)) +
15   geom_histogram(color = 'cyan4', fill = 'cyan3', linewidth = 0.5)
16   +
17   labs(x = "Xt", y = "Frecuencia",
18        title = "Simulación de un GARCH(1,1) con ruido normal",
19        subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
20   theme_minimal()
21 X2 <- simulacionGarch(2, 0.3, 0.5, n, list("tstudent",5), 1, 0.5)
22 X2 <- data.frame(t, X2)
23 plotX2 <- ggplot(data = X2, aes(x = t, y = X2)) +
24   geom_line(color = 'coral4', linewidth = .5)+
25   labs(x = "t", y = "Xt",
26        title = "Simulación de un GARCH(1,1) con ruido t de Student",
27        subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
28   theme_minimal()+
29   ylim(-20, 20)
30 plotX2h <- ggplot(data = X2, aes(x = X2)) +

```

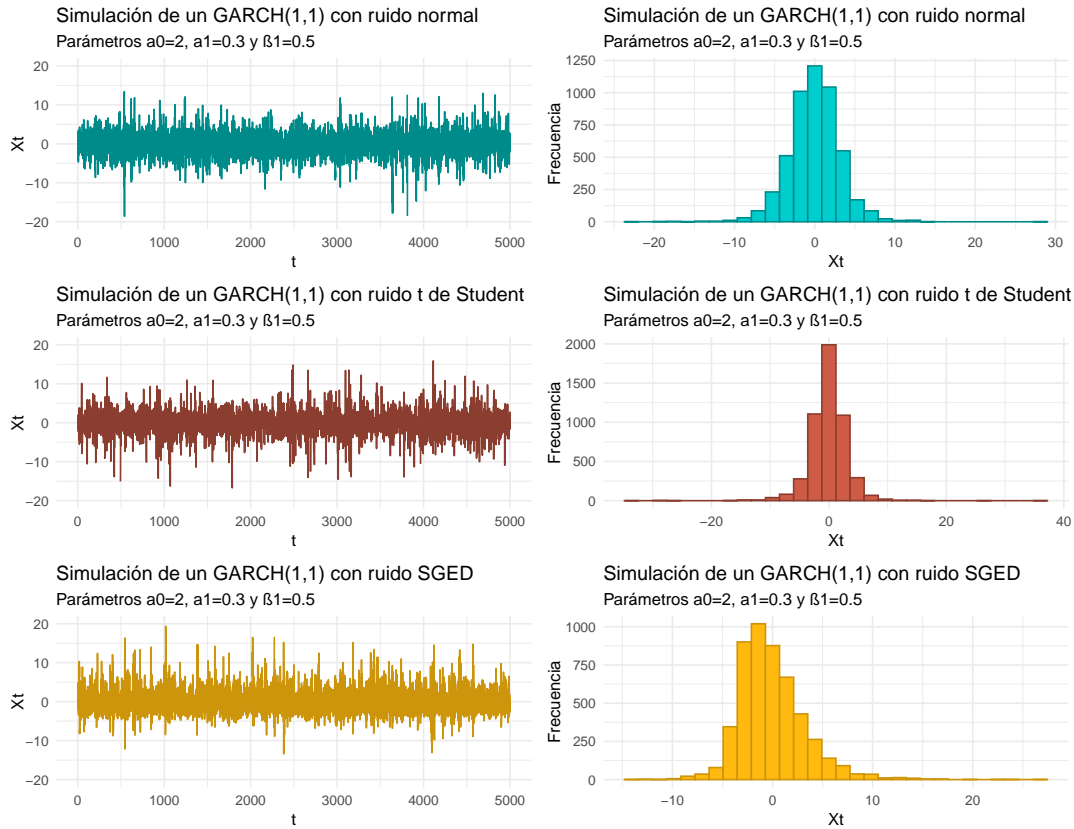
¹Consultar el Apéndice A para detalles acerca de esta distribución.

```

31 geom_histogram(color = 'coral4', fill = 'coral3', linewidth =
    0.5)+
32 labs(x = "Xt", y = "Frecuencia",
33       title = "Simulación de un GARCH(1,1) con ruido t de Student",
34       subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
35 theme_minimal()
36
37 X3 <- simulacionGarch(2, 0.3, 0.5, n, list("sged", 2, 1.5), 1, 0.5)
38 X3 <- data.frame(t, X3)
39 plotX3 <- ggplot(data = X3, aes(x = t, y = X3)) +
40   geom_line(color = 'darkgoldenrod3', linewidth = .5)+
41   labs(x = "t", y = "Xt",
42        title = "Simulación de un GARCH(1,1) con ruido SGED",
43        subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
44   theme_minimal()+
45   ylim(-20, 20)
46 plotX3h <- ggplot(data = X3, aes(x = X3)) +
47   geom_histogram(color = 'darkgoldenrod3', fill = 'darkgoldenrod1',
48                 , linewidth = 0.5)+
49   labs(x = "Xt", y = "Frecuencia",
50        title = "Simulación de un GARCH(1,1) con ruido SGED",
51        subtitle = "Parámetros  $\alpha_0=2$ ,  $\alpha_1=0.3$  y  $\beta_1=0.5$ ") +
52   theme_minimal()
53 varianzas <- c(var(X1[2]), var(X2[2]), var(X3[2]))
54 grid.arrange(plotX1, plotX1h, plotX2, plotX2h,
55              plotX3, plotX3h, ncol=2)

```

Con lo cual se obtuvo el vector de varianzas (10.105654, 9.681584, 11.037851) y la gráfica:



Podemos observar que las tres series no presentan diferencias significativas en la varianza, lo cual tiene sentido ya que los parámetros α_0 , α_1 y β_1 de la serie GARCH utilizados son los mismos en todos los casos, mientras que el ruido blanco tiene la misma media y varianza. Por otro lado, en los histogramas se pueden percibir más claramente las distintas tendencias de los datos:

- En la gráfica con ruido normal, los datos presentan una tendencia simétrica con respecto a la media cero. Es semejante a un histograma normal, aunque con colas más pesadas.
- En la gráfica con ruido t de Student, los datos presentan una tendencia de leptocurtosis, es decir, colas pesadas y concentración en el centro.
- En la gráfica con ruido SGED, los datos se encuentran cargados hacia la derecha, de manera que la moda y mediana no coinciden.

2.6. Estimación de parámetros

Ya que se desarrollaron los modelos teóricos, es necesario relacionarlos con los datos muestrales que se tengan de algún fenómeno. Para ello, el objetivo será asociarlos con cierta distribución de probabilidad conocida que cuenta con parámetros específicos, de manera que se puedan realizar predicciones mediante dicho modelo. Cualquier función de una muestra aleatoria cuyo objetivo es aproximar un parámetro se denomina **estimador**.

2.6.1. Métodos de estimación

Para comenzar, se presentarán los métodos de **estimación puntual**, los cuales se refieren a hallar una aproximación numérica puntual de los parámetros con base en los datos que se tienen de la muestra. Para este fin, se utilizan principalmente dos procedimientos.

Método de máxima verosimilitud

Definición 2.6.1. Sea w_1, w_2, \dots, w_n una muestra aleatoria de $f_W(w; \theta)$, donde θ es el parámetro desconocido (también puede ser un vector de parámetros).

La **función de verosimilitud**, denotada por $\mathcal{L}(\theta)$, se define como el producto de la función $f_W(w; \theta)$ evaluada en los n datos muestrales, es decir,

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_W(w_i; \theta) .$$

Si $\mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta)$ para todos los posibles valores de θ , entonces $\hat{\theta}$ se denomina como el **estimador de máxima verosimilitud (MLE)** para θ . La idea de esta definición es hallar el valor del parámetro que maximiza la probabilidad de obtener la muestra aleatoria observada.

Para obtener este estimador, basta con obtener el máximo de la función $\mathcal{L}(\theta)$ mediante los métodos de cálculo usuales. Un método común consiste en obtener el máximo del logaritmo natural de $\mathcal{L}(\theta)$, ya que preserva el máximo y puede simplificar los cálculos.

Método de momentos

Sea w_1, w_2, \dots, w_n una muestra aleatoria de $f_W(w; \theta_1, \theta_2, \dots, \theta_k)$. El **método de momentos** propone los estimadores $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ como las soluciones

del siguiente sistema de k ecuaciones

$$\int_{-\infty}^{\infty} w^j f_W(w; \theta_1, \theta_2, \dots, \theta_k) dw = \mathbb{E}[W^j] = \frac{1}{n} \sum_{i=1}^n w_i^j, \text{ para todo } j \in \{1, \dots, k\}.$$

Ya que se cuenta con una propuesta de estimadores, es necesario considerar qué tan adecuados resultan para describir a la muestra aleatoria. Para ello, se puede considerar su variabilidad mediante los **intervalos de confianza**, los cuales se refieren a rangos de números que tienen cierta probabilidad dada $100(1 - \alpha)\%$ de contener al parámetro, donde α generalmente toma valores pequeños como son 0.1, 0.05 o 0.01.

Para obtener intervalos de confianza y estimaciones adecuadas, es necesario considerar el **tamaño de la muestra**, ya que una muestra muy pequeña aporta menos información, mientras que una muestra muy grande resulta más difícil de obtener prácticamente.

2.6.2. Propiedades de los estimadores

Otra manera de evaluar qué tan adecuados son los estimadores es mediante sus propiedades.

Insesgadez.

Definición 2.6.2. Sea W_1, W_2, \dots, W_n una muestra aleatoria de $f_W(w; \theta)$. Un estimador $\hat{\theta} = h(W_1, W_2, \dots, W_n)$ se denomina **insesgado para θ** si $\mathbb{E}[\hat{\theta}] = \theta$ para todo θ .

Eficiencia.

Definición 2.6.3. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados para el parámetro θ . Si se cumple que $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$ entonces decimos que $\hat{\theta}_1$ es **más eficiente** que $\hat{\theta}_2$. Además, la eficiencia relativa de $\hat{\theta}_1$ con respecto a $\hat{\theta}_2$ está dada por $\frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$.

Así, también podemos comparar qué estimador es mejor en caso de que tengamos más de una propuesta.

Adicionalmente, es razonable buscar al estimador insesgado más eficiente, para lo cual se utiliza un resultado conocido como la Cota de Cramér-Rao, la cual representa la varianza mínima que un estimador insesgado puede tener.

De esta manera, si la varianza de cierto estimador insesgado coincide con la cota, podemos afirmar que es el estimador más eficiente. Este resultado se puede consultar en Larsen y Marx (2001) [7].

Suficiencia.

Definición 2.6.4. Sea W_1, W_2, \dots, W_n una muestra aleatoria de $f_W(w; \theta)$. Un estimador $\hat{\theta} = h(W_1, W_2, \dots, W_n)$ se denomina **suficiente para θ** si para todo θ , la función de distribución condicional de W_1, W_2, \dots, W_n dado $\hat{\theta}$ no depende de θ .

Para probar la suficiencia, es común utilizar el **Teorema de Fisher-Neyman**, el cual se puede consultar en Larsen y Marx (2001) [7].

Consistencia.

Definición 2.6.5. Un estimador $\hat{\theta}_n = h(W_1, W_2, \dots, W_n)$ se denomina **consistente para θ** si converge en probabilidad a θ , es decir,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[|\hat{\theta}_n - \theta| < \varepsilon \right] = 1 \text{ para todo } \varepsilon > 0 .$$

En el capítulo 5 de Larsen y Marx (2001) [7] se pueden consultar la construcción, más propiedades y ejemplos de las definiciones contenidas en esta sección.

Capítulo 3

Modelos markovianos

En general, un modelo markoviano de cambio de régimen se construye combinando dos o más modelos dinámicos por medio de un switch markoviano. En el presente capítulo, se describirán dos modelos de series de tiempo con switch markoviano. Para comenzar, se describirá el modelo autorregresivo con cambio de régimen markoviano, considerando el modelo básico, algunas extensiones y dos métodos para su estimación estadística. Posteriormente, se describirá y simulará un modelo GARCH con switch markoviano.

El contenido de este capítulo está basado en las ideas de Kuan (2002) [12]; y Ardia et al. (2019) [13].

3.1. Modelo autorregresivo de media condicional

Para comenzar, se explicará detalladamente el modelo autorregresivo markoviano con cambio de régimen en la media presentado por Kuan (2002) [12].

En todo el desarrollo de este modelo, asumiremos que los datos son consistentes con una serie de tiempo no-estacional. En caso de que este supuesto no se cumpla, si d es el periodo de la serie bastará con aplicarle el operador de diferencia estacional de d periodos, tal como se explicó en el capítulo anterior.

3.1.1. Modelo básico

El modelo más sencillo presentado es un AR(1) con cambio de régimen markoviano definido como a continuación.

Definición 3.1.1. (*Modelo de switch markoviano para un AR(1)*) Sea $\{S_t\}_{t \in T}$ una sucesión de variables aleatorias con soporte $\{0, 1\}$, las cuales se conocerán como **variables de estado**. Entonces definimos a la serie de tiempo $\{Z_t\}_{t \in T}$ como

$$Z_t = \alpha_0 + \alpha_1 S_t + \beta Z_{t-1} + \epsilon_t, \text{ para todo } t \in T,$$

donde $\alpha_0, \alpha_1, \beta \in \mathbb{R}$, $|\beta| < 1$ y $\{\epsilon_t\}_{t \in T}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_ϵ^2 .

Observación. En este modelo las características aleatorias de ϵ_t y la variable de estado S_t determinan el valor de la variable Z_t , la cual tiene dos estructuras distintas dependiendo del valor de S_t en cada punto del tiempo:

- Si $S_t = 0$, la serie de tiempo toma el valor $Z_t = \alpha_0 + \beta Z_{t-1} + \epsilon_t$.

Sea $\mu = \mathbb{E}[Z_l]$ para todo $l \in T$, entonces se puede calcular el primer momento como

$$\begin{aligned} \mu &= \mathbb{E}[Z_t] = \mathbb{E}[\alpha_0 + \beta Z_{t-1} + \epsilon_t] = \mathbb{E}[\alpha_0] + \mathbb{E}[\beta Z_{t-1}] + \mathbb{E}[\epsilon_t] \\ &= \alpha_0 + \beta \mathbb{E}[Z_{t-1}] + 0 = \alpha_0 + \beta \mu. \end{aligned}$$

Entonces $\mu = \frac{\alpha_0}{1-\beta}$, siempre que $\beta \neq 1$. Es decir, cuando $S_t = 0$ este es un proceso AR(1) con media $\frac{\alpha_0}{1-\beta}$.

- Si $S_t = 1$, la serie de tiempo toma el valor $Z_t = \alpha_0 + \alpha_1 + \beta Z_{t-1} + \epsilon_t$.

Sea $\mu = \mathbb{E}[Z_l]$ para todo $l \in T$, entonces se puede calcular el primer momento como

$$\begin{aligned} \mu &= \mathbb{E}[Z_t] = \mathbb{E}[\alpha_0 + \alpha_0 + \beta Z_{t-1} + \epsilon_t] = \mathbb{E}[\alpha_0 + \alpha_1] + \mathbb{E}[\beta Z_{t-1}] + \mathbb{E}[\epsilon_t] \\ &= \alpha_0 + \alpha_1 + \beta \mathbb{E}[Z_{t-1}] + 0 = \alpha_0 + \alpha_1 + \beta \mu. \end{aligned}$$

Entonces $\mu = \frac{\alpha_0 + \alpha_1}{1-\beta}$, siempre que $\beta \neq 1$. Es decir, cuando $S_t = 1$ este es un proceso AR(1) con media $\frac{\alpha_0 + \alpha_1}{1-\beta}$.

Así, siempre que $\alpha_1 \neq 0$, la serie de tiempo alterna entre un proceso AR(1) con media $\frac{\alpha_0}{1-\beta}$ y un proceso AR(1) con media $\frac{\alpha_0 + \alpha_1}{1-\beta}$. Esto se conoce como un

cambio de régimen, ya que permite dos estructuras distintas que dependen del valor que tome la variable de estado S_t .

La variable de estado puede definirse de diversas maneras, a continuación se expondrán cuatro propuestas posibles.

Modelo con un único cambio estructural.

Es el modelo resultante de la definición **3.1.1** con la sucesión de variables aleatorias independientes e idénticamente distribuidas $\{S_t\}_{t \in T}$ dadas por

$$S_t = \begin{cases} 0 & \text{si } t \leq t_0 \\ 1 & \text{si } t > t_0 \end{cases}, \text{ para algún } t_0 \in T.$$

En este caso, el modelo presenta únicamente un cambio de régimen después del tiempo t_0 .

Ventajas:

- La definición es sencilla de analizar y directa.

Desventajas:

- El modelo es muy restrictivo debido a que únicamente permite un cambio estructural.
- Aunque el modelo se puede extender directamente para permitir más cambios en el tiempo, la estimación estadística se complica.
- El cambio de régimen depende únicamente del tiempo, el cual es externo al modelo. De esta manera, depende únicamente de una variable que no se ve influida por el modelo y no es posible definir el cambio utilizando la información relativa a lo que ya ocurrió en tiempos previos.

Modelo de cambio aleatorio de Quandt.

Es el modelo resultante de la definición **3.1.1** con la sucesión $\{S_t\}_{t \in T}$ compuesta por variables aleatorias Bernulli independientes, es decir,

$$f_{S_t}(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases}, \text{ con } p \in (0, 1).$$

En este caso, el modelo puede presentar múltiples cambios de régimen en el tiempo dependiendo del valor de p .

Ventajas:

- El modelo permite múltiples cambios.

Desventajas:

- El cambio de régimen depende únicamente de una variable externa al modelo. La realización de cada S_t es independiente al pasado y al futuro, de manera que el proceso Z_t puede alternar entre distintos estados muy frecuentemente.
- Debido a que las variables de estado son independientes, puede que no sean aplicables a los datos de una serie de tiempo.

Modelo de umbral.

Es el modelo resultante de la definición **3.1.1** con la sucesión de variables aleatorias $\{S_t\}_{t \in T}$ dadas por

$$S_t = \mathbb{I}_{\{\lambda_t \leq c\}} = \begin{cases} 1 & \text{si } \lambda_t \leq c \\ 0 & \text{si } \lambda_t > c \end{cases},$$

para alguna sucesión de variables aleatorias $\{\lambda_t\}_{t \in T}$.

En este caso, el modelo presenta cambios de régimen dependientes de si el valor de λ_t de cada tiempo es mayor que el valor de umbral $c \in \mathbb{R}$. Es común escoger una variable aleatoria desfasada, es decir, $\lambda_t = Z_{t-d}$.

Ventajas:

- El modelo permite múltiples cambios.
- Las variables de estado pueden depender del modelo en lugar de ser externas.

Desventajas:

- Escoger variables aleatorias λ_t y el valor de umbral c para este modelo generalmente es complicado.

Aunque todas estos modelos pueden ser útiles para describir diferentes conjuntos de datos con diversas ventajas y desventajas, para fines de este escrito nos limitaremos al modelo de switch markoviano definido como sigue.

Modelo de switch markoviano.

Definición 3.1.2. (Modelo de switch markoviano) Es el modelo resultante de la definición 3.1.1 con la sucesión $\{S_t\}_{t \in T}$ definida como una cadena de Markov discreta con espacio de estados $\varepsilon = \{0, 1\}$ y matriz de probabilidades de transición \mathbb{P} , la cual se puede expresar como

$$\mathbb{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} \mathbb{P}[S_t = 0 \mid S_{t-1} = 0] & \mathbb{P}[S_t = 1 \mid S_{t-1} = 0] \\ \mathbb{P}[S_t = 0 \mid S_{t-1} = 1] & \mathbb{P}[S_t = 1 \mid S_{t-1} = 1] \end{pmatrix},$$

donde para todos los $i, j \in \varepsilon$, p_{ij} representa la probabilidad de transición del estado i al estado j , es decir, la probabilidad de que $S_t = j$ dado que ya ocurrió que $S_{t-1} = i$.

En este caso, el cambio está dado por una cadena de Markov de dos estados, de manera que depende únicamente de lo que sucedió en el tiempo inmediatamente anterior. Así, el modelo presenta cambios de régimen en el tiempo frecuentes y aleatorios, cuya persistencia depende de los valores de las probabilidades de transición.

Ventajas:

- El modelo es sencillo de implementar, pues la clasificación de régimen es probabilística y queda determinada por los datos.

Desventajas:

- Este modelo no es tan fácil de interpretar, ya que las variables de estado no son observables.

Observación. Dado que $p_{00} + p_{01} = 1$ y $p_{10} + p_{11} = 1$, la matriz únicamente tiene dos parámetros libres, de manera que se puede expresar como

$$\mathbb{P} = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}, \text{ con } p_{00}, p_{11} \in [0, 1].$$

3.1.2. Extensiones del modelo

El modelo de switch markoviano presentado anteriormente se puede extender para permitir una mayor variedad de estructuras dinámicas. A continuación, presentaremos brevemente cuatro posibles extensiones del modelo.

Definición 3.1.3. (Modelo de switch markoviano para un AR(p)) Sea $\{S_t\}_{t \in T}$ una cadena de Markov discreta con espacio de estados $\varepsilon = \{0, 1\}$

y matriz de probabilidades de transición \mathbb{P} . Entonces definimos a la serie de tiempo $\{Z_t\}_{t \in T}$ como

$$\begin{aligned} Z_t &= \alpha_0 + \alpha_1 S_t + \sum_{k=1}^p \beta_k Z_{t-k} + \epsilon_t \\ &= \alpha_0 + \alpha_1 S_t + \beta_1 Z_{t-1} + \cdots + \beta_p Z_{t-p} + \epsilon_t, \text{ para todo } t \in T, \end{aligned}$$

donde $\alpha_0, \alpha_1 \in \mathbb{R}$, $\beta_k \in \mathbb{R}$ para cada $k \in \{1, \dots, p\}$ y $\{\epsilon_t\}_{t \in T}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_ϵ^2 .

En lo que resta de esta subsección, nos limitaremos al estudio de esta extensión del modelo, con el objetivo de permitir una estructura más compleja que un AR(1).

Definición 3.1.4. (Modelo de switch markoviano para un AR(p) con m regímenes) Sea $\{S_t\}_{t \in T}$ una cadena de Markov discreta con espacio de estados ε de cardinalidad m y matriz de probabilidades de transición \mathbb{P} . Entonces definimos a la serie de tiempo $\{Z_t\}_{t \in T}$ como

$$Z_t = \alpha_0 + \alpha_1 S_t + \sum_{k=1}^p \beta_k Z_{t-k} + \epsilon_t,$$

donde $\alpha_0, \alpha_1 \in \mathbb{R}$, $\beta_k \in \mathbb{R}$ para cada $k \in \{1, \dots, p\}$ y $\{\epsilon_t\}_{t \in T}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_ϵ^2 .

Este modelo es muy semejante al previo, con la diferencia de que la cadena puede tener un espacio de estados distinto a $\{0, 1\}$, de manera que la serie de tiempo $\{Z_t\}_{t \in T}$ puede alternar entre m regímenes distintos (cada uno con su respectiva media). En este caso, la matriz de probabilidades de transición \mathbb{P} es una matriz m -dimensional y cuenta con una mayor cantidad de parámetros libres.

Definición 3.1.5. (Modelo de switch markoviano d-dimensional para un AR(p)) Sea $\{S_t\}_{t \in T}$ una cadena de Markov discreta con espacio de estados $\varepsilon = \{0, 1\}$ y matriz de probabilidades de transición \mathbb{P} . Entonces definimos a la serie de tiempo d -dimensional $\{Z_t\}_{t \in T}$ como

$$Z_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 S_t + \sum_{k=1}^p \mathbf{B}_k Z_{t-k} + \epsilon_t,$$

donde $\alpha_0, \alpha_1 \in \mathbb{R}^d$ son vectores d -dimensionales, para cada $k \in \{1, \dots, p\}$ B_k son matrices cuadradas d -dimensionales de parámetros y $\{\epsilon_t\}_{t \in T}$ es una sucesión de vectores aleatorios independientes e idénticamente distribuidas con media cero y matriz de varianzas-covarianzas Σ_0 .

En este caso, Z_t es un modelo de vector aleatorio autorregresivo con switch markoviano.

Definición 3.1.6. (*Modelo de switch markoviano AR(p) con dependencia a las variables de estado previas*) Sea $\{S_t\}_{t \in T}$ una cadena de Markov discreta con espacio de estados $\varepsilon = \{0, 1\}$ y matriz de probabilidades de transición \mathbb{P} . Entonces definimos a la serie de tiempo auxiliar $\{\tilde{Z}_t\}_{t \in T}$ como

$$\tilde{Z}_t = \sum_{k=1}^p \beta_k \tilde{Z}_{t-k} + \epsilon_t ,$$

donde $\beta_k \in \mathbb{R}$ para cada $k \in \{1, \dots, p\}$ y $\{\epsilon_t\}_{t \in T}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_ϵ^2 . Posteriormente, definimos a la serie de tiempo $\{Z_t\}_{t \in T}$ como

$$Z_t = \tilde{Z}_t + \alpha_0 + \alpha_1 S_t ,$$

donde $\alpha_0, \alpha_1 \in \mathbb{R}$.

Observación. En este modelo, la serie de tiempo $\{Z_t\}_{t \in T}$ tiene dependencia a las variables de estado previas.

Tenemos que $\tilde{Z}_l = Z_l - \alpha_0 - \alpha_1 S_l$ para todo $l \in T$, entonces, para todo $t \in T$,

$$\begin{aligned} Z_t &= \tilde{Z}_t + \alpha_0 + \alpha_1 S_t = \left(\sum_{k=1}^p \beta_k \tilde{Z}_{t-k} + \epsilon_t \right) + \alpha_0 + \alpha_1 S_t \\ &= \sum_{k=1}^p \beta_k (Z_{t-k} - \alpha_0 - \alpha_1 S_{t-k}) + \alpha_0 + \alpha_1 S_t + \epsilon_t \\ &= \sum_{k=1}^p \beta_k Z_{t-k} - \alpha_0 \sum_{k=1}^p \beta_k - \alpha_1 \sum_{k=1}^p \beta_k S_{t-k} + \alpha_0 + \alpha_1 S_t + \epsilon_t \\ &= \sum_{k=1}^p \beta_k Z_{t-k} - \alpha_0 \left(\sum_{k=1}^p \beta_k - 1 \right) - \alpha_1 \left(\sum_{k=1}^p \beta_k S_{t-k} - S_t \right) + \epsilon_t \\ &= \sum_{k=1}^p \beta_k Z_{t-k} - \alpha_0 \sum_{k=0}^p \beta_k - \alpha_1 \sum_{k=0}^p \beta_k S_{t-k} , \text{ con } \beta_0 = -1. \end{aligned}$$

Es decir, en este modelo los valores de Z_t dependen tanto de S_t como de las variables de estado en tiempos anteriores S_{t-1}, \dots, S_{t-p} .

3.1.3. Estimación estadística

Supongamos que $\{Z_t\}_{t \in \{1, \dots, T_0\}}$ es una serie de tiempo descrita como el modelo de switch markoviano para un AR(p) presentado en la definición 3.1.3, es decir,

$$Z_t = \alpha_0 + \alpha_1 S_t + \sum_{k=1}^p \beta_k Z_{t-k} + \epsilon_t.$$

A continuación, desarrollaremos dos métodos de estimación estadística para dicho modelo.

Estimador de máxima quasi-verosimilitud (QMLE)

Sean $\{z_t\}_{t \in \{1, \dots, T_0\}}$ los datos observados de la serie de tiempo. A partir de dichos datos, el objetivo es estimar al vector de parámetros

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \dots, \beta_p, \sigma_\epsilon^2, p_{00}, p_{11}).$$

Sea $\mathcal{Z}^t = \{z_1, \dots, z_t\}$ el conjunto de todas las variables observadas a tiempo t . De tal manera, $\mathcal{Z}^{T_0} = \{z_1, \dots, z_{T_0}\}$ representa la información observada en toda la muestra.

Para obtener la verosimilitud de la variable de estado S_t , evaluaremos las esperanzas condicionales de S_t dado que se tienen diferentes conjuntos de información.

Definición 3.1.7. (Probabilidades a estimar)

1. Probabilidades de predicción.

$$\mathbb{P}[S_t = i \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}], \text{ para } S_t = 0, 1.$$

Estas utilizan la información del tiempo previo para obtener una predicción de lo que sucedió con la variable S_t .

2. Probabilidades de filtrado.

$$\mathbb{P}[S_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta}], \text{ para } S_t = 0, 1.$$

En este caso, el filtrado se refiere a que se obtiene una estimación del estado S_t utilizando la información del tiempo actual y tiempos previos.

3. Probabilidades de suavizado.

$$\mathbb{P}[S_t = i \mid \mathcal{Z}^{T_0}; \boldsymbol{\theta}], \text{ para } S_t = 0, 1.$$

Estas utilizan la información de toda la muestra para obtener una predicción de lo que sucedió con la variable S_t .

A continuación, obtendremos una estimación de dichas probabilidades basadas en la muestra $\{z_t\}_{t \in \{1, \dots, T_0\}}$ mediante el siguiente procedimiento:

Paso 1: Obtener la distribución de Z_t dado que conocemos los valores de S_t , \mathcal{Z}^{t-1} y $\boldsymbol{\theta}$.

Supongamos que ϵ_t son variables aleatorias normales independientes, entonces la densidad de la variable aleatoria $(Z_t \mid S_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ está dada por

$$\begin{aligned} F_{Z_t \mid S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) &= \mathbb{P} \left[Z_t \leq z_t \mid S_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta} \right] \\ &= \mathbb{P} \left[\alpha_0 + \alpha_1 S_t + \sum_{k=1}^p \beta_k Z_{t-k} + \epsilon_t \leq z_t \mid S_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta} \right] \\ &= \mathbb{P} \left[\epsilon_t \leq z_t - \alpha_0 - \alpha_1 S_t - \sum_{k=1}^p \beta_k Z_{t-k} \mid S_t = i, \mathcal{Z}^{t-1} = \{z_1, \dots, z_{t-1}\}; \boldsymbol{\theta} = \boldsymbol{\theta}' \right] \\ &= \mathbb{P} \left[\epsilon_t \leq z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \mid S_t = i, \mathcal{Z}^{t-1} = \{z_1, \dots, z_{t-1}\}; \boldsymbol{\theta} = \boldsymbol{\theta}' \right] \\ &= \mathbb{P} \left[\epsilon_t \leq z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right]. \end{aligned}$$

El último paso se cumple puesto que ϵ_t es independiente y $z_t, \alpha'_0, \alpha'_1, i, \beta'_k, z_{t-k}$ para cada $k \in \{1, \dots, p\}$ toman valores fijos conocidos, de manera que ya no dependen de la información contenida en la condicional. Entonces,

$$\begin{aligned} F_{Z_t \mid S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) &= \mathbb{P} \left[\epsilon_t \leq z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right] \\ &= F_{\epsilon_t} \left(z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right). \end{aligned}$$

Derivando para obtener la función de masa,

$$\begin{aligned}
f_{Z_t | S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) &= \frac{d}{dz_t} F_{Z_t | S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) \\
&= \frac{d}{dz_t} F_{\epsilon_t} \left(z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right) \\
&= f_{\epsilon_t} \left(z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right) \\
&\quad \times \frac{d}{dz_t} \left[z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right] \\
&= f_{\epsilon_t} \left(z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k} \right).
\end{aligned}$$

Por lo tanto,

$$f_{Z_t | S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp \left\{ -\frac{(z_t - \alpha'_0 - \alpha'_1 i - \sum_{k=1}^p \beta'_k z_{t-k})^2}{2\sigma_\epsilon^2} \right\}.$$

Con esto, la distribución de la variable aleatoria $(Z_t | S_t = i, \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ es conocida y se puede utilizar para el siguiente paso.

Paso 2: Obtener las probabilidades de filtrado, las probabilidades de predicción y las densidades condicionales $f_{Z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)$.

Observamos que se cumplen las siguientes ecuaciones:

- (i) *Distribución de $(Z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ a partir de las probabilidades de predicción.* Si las probabilidades de predicción $\mathbb{P}[S_t = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta}]$ son conocidas para todo $i \in \{0, 1\}$, utilizando el teorema de la probabilidad total,

$$\begin{aligned}
f_{Z_t | \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) &= f_{Z_t | S_t=0, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) \mathbb{P}[S_t = 0 | \mathcal{Z}^{t-1}; \boldsymbol{\theta}] \\
&\quad + f_{Z_t | S_t=1, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t) \mathbb{P}[S_t = 1 | \mathcal{Z}^{t-1}; \boldsymbol{\theta}].
\end{aligned}$$

Esta es una distribución conocida, ya que para $i = 0, 1$ conocemos el valor de $f_{Z_t | S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)$ por el **Paso 1**.

- (ii) *Probabilidades de filtrado a partir de las probabilidades de predicción.* Si las probabilidades de predicción $\mathbb{P}[S_t = i | \mathcal{Z}^{t-1}; \boldsymbol{\theta}]$ son conocidas para

todo $i \in \{0, 1\}$, utilizando el teorema de Bayes,

$$\mathbb{P}[S_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta}] = \frac{\mathbb{P}[S_t = i \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}] f_{Z_t \mid S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)}{f_{Z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)}.$$

Este valor es conocido, ya que para $i = 0, 1$ conocemos el valor de $f_{Z_t \mid S_t=i, \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)$ por el **Paso 1** y la distribución de $(Z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta})$ es conocida a partir de las probabilidades de predicción como se describe en (i).

(iii) *Probabilidades de predicción a partir de las probabilidades de filtrado.* Si las probabilidades de filtrado $\mathbb{P}[S_t = i \mid \mathcal{Z}^t; \boldsymbol{\theta}]$ son conocidas para todo $i \in \{0, 1\}$, utilizando el teorema de la probabilidad total,

$$\begin{aligned} \mathbb{P}[S_{t+1} = i \mid \mathcal{Z}^t; \boldsymbol{\theta}] &= \mathbb{P}[S_{t+1} = i \mid S_t = 0, \mathcal{Z}^t; \boldsymbol{\theta}] \mathbb{P}[S_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}] \\ &\quad + \mathbb{P}[S_{t+1} = i \mid S_t = 1, \mathcal{Z}^t; \boldsymbol{\theta}] \mathbb{P}[S_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta}] \\ &= p_{0i} \mathbb{P}[S_t = 0 \mid \mathcal{Z}^t; \boldsymbol{\theta}] + p_{1i} \mathbb{P}[S_t = 1 \mid \mathcal{Z}^t; \boldsymbol{\theta}]. \end{aligned}$$

Observemos que las ecuaciones (ii) y (iii) forman un sistema recursivo para $t = k, \dots, T_0$. Entonces debemos tomar un valor inicial de la primera probabilidad de predicción $\mathbb{P}[S_k = i \mid \mathcal{Z}^{k-1}; \boldsymbol{\theta}]$ y sustituir recursivamente en las ecuaciones hasta obtener las probabilidades de filtrado, las probabilidades de predicción y las densidades condicionales $f_{Z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)$.

Este valor inicial se puede definir de distintas maneras, por ejemplo, en Kuan (2002) [12] se propone iniciar con su contraparte no condicional dada por la tercera columna de la matriz $(A'A)^{-1}A'$, donde $A = \begin{pmatrix} I - P \\ 1' \end{pmatrix}$ con I la matriz identidad y 1 la matriz de unos de dos dimensiones.

Paso 3: Obtener el estimador de máxima quasi-verosimilitud (QM-LE).

Ya que contamos con las densidades condicionales $f_{Z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t)$, definimos la función de quasi-log-verosimilitud como

$$\mathcal{L}_{T_0}(\boldsymbol{\theta}) = \frac{1}{T_0} \sum_{i=1}^{T_0} \ln f_{Z_t \mid \mathcal{Z}^{t-1}; \boldsymbol{\theta}}(z_t).$$

Esto tiene sentido puesto que T_0 es un valor fijo no negativo y la función

de log-verosimilitud bajo supuestos de normalidad estaría dada por

$$\ln(\mathcal{L}_{T_0}(\boldsymbol{\theta})) = \ln \left(\prod_{i=1}^{T_0} f_{Z_t | Z^{t-1}; \boldsymbol{\theta}}(z_t) \right) = \sum_{i=1}^{T_0} \ln f_{Z_t | Z^{t-1}; \boldsymbol{\theta}}(z_t).$$

Para obtener el QMLE $\widehat{\boldsymbol{\theta}}_{T_0}$, únicamente debemos maximizar la función $\mathcal{L}_{T_0}(\boldsymbol{\theta})$ mediante cualquier método numérico, para lo cual es común utilizar algoritmos computacionales de búsqueda.

Ya que se cuenta con este estimador, se pueden obtener estimadores de las probabilidades de filtrado y predicción sustituyendo los valores de $\widehat{\boldsymbol{\theta}}_{T_0}$ en las ecuaciones correspondientes.

Paso 4: Obtener las probabilidades de suavizado.

Ya que se realizaron los pasos previos, es posible obtener las probabilidades de suavizado para $i, j \in \{0, 1\}$. Para comenzar,

$$\mathbb{P}[S_t = i | S_{t+1} = j, \mathcal{Z}^{T_0}; \boldsymbol{\theta}] = \mathbb{P}[S_t = i | S_{t+1} = j, \mathcal{Z}^t; \boldsymbol{\theta}] = \frac{p_{ij} \mathbb{P}[S_t = i | \mathcal{Z}^t; \boldsymbol{\theta}]}{\mathbb{P}[S_{t+1} = j | \mathcal{Z}^t; \boldsymbol{\theta}]}.$$

Entonces las probabilidades de suavizado están dadas por

$$\begin{aligned} \mathbb{P}[S_t = i | \mathcal{Z}^{T_0}; \boldsymbol{\theta}] &= \mathbb{P}[S_t = i | S_{t+1} = 0, \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \mathbb{P}[S_{t+1} = 0 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \\ &\quad + \mathbb{P}[S_t = i | S_{t+1} = 1, \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \mathbb{P}[S_{t+1} = 1 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \\ &= \frac{p_{i0} \mathbb{P}[S_t = i | \mathcal{Z}^t; \boldsymbol{\theta}]}{\mathbb{P}[S_{t+1} = 0 | \mathcal{Z}^t; \boldsymbol{\theta}]} \mathbb{P}[S_{t+1} = 0 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \\ &\quad + \frac{p_{i1} \mathbb{P}[S_t = i | \mathcal{Z}^t; \boldsymbol{\theta}]}{\mathbb{P}[S_{t+1} = 1 | \mathcal{Z}^t; \boldsymbol{\theta}]} \mathbb{P}[S_{t+1} = 1 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}] \\ &= \mathbb{P}[S_t = i | \mathcal{Z}^t; \boldsymbol{\theta}] \times \\ &\quad \left(\frac{p_{i0} \mathbb{P}[S_{t+1} = 0 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}]}{\mathbb{P}[S_{t+1} = 0 | \mathcal{Z}^t; \boldsymbol{\theta}]} + \frac{p_{i1} \mathbb{P}[S_{t+1} = 1 | \mathcal{Z}^{T_0}; \boldsymbol{\theta}]}{\mathbb{P}[S_{t+1} = 1 | \mathcal{Z}^t; \boldsymbol{\theta}]} \right). \end{aligned}$$

Observemos que ya conocemos la probabilidad de filtrado $\mathbb{P}[S_{T_0} = i | \mathcal{Z}^{T_0}; \boldsymbol{\theta}]$ que también es la última probabilidad de suavizado, de manera que podemos tomarla como valor inicial y obtener las probabilidades de filtrado correspondientes a $S_{T_0-1}, S_{T_0-2}, \dots, S_{k+1}$ de manera recursiva sustituyendo en las ecuaciones. Adicionalmente, estas probabilidades también son funciones de $\boldsymbol{\theta}$, de manera que los estimadores de las probabilidades de suavizado también se pueden obtener sustituyendo los valores de $\widehat{\boldsymbol{\theta}}_{T_0}$ en las ecuaciones correspondientes.

Muestreo de Gibbs

Los métodos de Montecarlo basados en cadenas de Markov (MCMC) son métodos de simulación para el muestreo de una distribución de probabilidad, en los cuales dada dicha distribución se puede construir una cadena de Markov que la aproxime. Uno de estos métodos es el muestreo de Gibbs, el cual es un algoritmo que se utiliza para tomar una muestra de una probabilidad multivariada específica cuando la distribución conjunta es complicada, pero ciertas distribuciones condicionales son más fáciles de calcular. Este método consiste en obtener simulaciones aleatorias de parámetros como se expone a continuación.

Sean $\{z_t\}_{t \in \{1, \dots, T_0\}}$ los datos observados de la serie de tiempo. Para este método, consideraremos a los parámetros de $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \beta_1, \dots, \beta_p, \sigma_\epsilon^2, p_{00}, p_{11})$ como variables aleatorias. Además, consideraremos las variables de estado no observadas S_1, S_2, \dots, S_{T_0} como parámetros.

Este vector puede ser clasificado en cuatro grupos:

- Variables de estado: $\boldsymbol{\theta}_1 = \{S_t\}_{t \in \{1, \dots, T_0\}}$.
- Probabilidades de transición: $\boldsymbol{\theta}_2 = \{p_{00}, p_{11}\}$.
- Parámetros del modelo AR(p): $\boldsymbol{\theta}_3 = \{\alpha_0, \alpha_1, \beta_1, \dots, \beta_p\}$.
- Varianza: $\boldsymbol{\theta}_4 = \{\sigma_\epsilon^2\}$.

Entonces el vector aumentado de parámetros, denotado por $\boldsymbol{\theta}'$, está dado por la unión de $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_3$ y $\boldsymbol{\theta}_4$, es decir,

$$\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4).$$

Además, para este método consideramos que contamos con las **distribuciones condicionales completas** definidas como

$$\pi(\boldsymbol{\theta}_i \mid \boldsymbol{z}^T, \{\boldsymbol{\theta}_j, j \neq i\}), \text{ para } i = 1, \dots, 4.$$

Estas distribuciones se pueden obtener a partir de distribuciones previas dadas de los parámetros y las funciones de verosimilitud.

Paso 1: Obtener aleatoriamente valores iniciales para el algoritmo.

El muestreo de Gibbs comienza generando aleatoriamente valores iniciales

$$\boldsymbol{\theta}^{(0)'} = (\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \boldsymbol{\theta}_3^{(0)}, \boldsymbol{\theta}_4^{(0)}).$$

Paso 2: Obtener N realizaciones de los parámetros.

Posteriormente, se obtiene la i -ésima realización de $\boldsymbol{\theta}'$.

1. Obtener una realización de $\boldsymbol{\theta}_1$, la cual denotaremos como $\boldsymbol{\theta}_1^{(i)}$, de la distribución condicional completa como

$$\pi \left(\boldsymbol{\theta}_1 \mid \mathcal{Z}^T, \boldsymbol{\theta}_2^{(i-1)}, \boldsymbol{\theta}_3^{(i-1)}, \boldsymbol{\theta}_4^{(i-1)} \right).$$

2. Obtener una realización de $\boldsymbol{\theta}_2$, la cual denotaremos como $\boldsymbol{\theta}_2^{(i)}$, de la distribución condicional completa como

$$\pi \left(\boldsymbol{\theta}_2 \mid \mathcal{Z}^T, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_3^{(i-1)}, \boldsymbol{\theta}_4^{(i-1)} \right).$$

3. Obtener una realización de $\boldsymbol{\theta}_3$, la cual denotaremos como $\boldsymbol{\theta}_3^{(i)}$, de la distribución condicional completa como

$$\pi \left(\boldsymbol{\theta}_3 \mid \mathcal{Z}^T, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_4^{(i-1)} \right).$$

4. Obtener una realización de $\boldsymbol{\theta}_4$, la cual denotaremos como $\boldsymbol{\theta}_4^{(i)}$, de la distribución condicional completa como

$$\pi \left(\boldsymbol{\theta}_4 \mid \mathcal{Z}^T, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3^{(i)} \right).$$

De esta manera, la i -ésima realización de $\boldsymbol{\theta}'$ está dada por

$$\boldsymbol{\theta}^{(i)'} = \left(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3^{(i)}, \boldsymbol{\theta}_4^{(i)} \right).$$

Este procedimiento se debe repetir N veces, con lo cual se obtiene la sucesión de Gibbs

$$\{\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'}, \dots, \boldsymbol{\theta}^{(N)'}\}.$$

Sean $N_1 + N_2 = N$. Es común considerar la subsucesión resultante de eliminar los primeros N_1 estimadores y conservar los N_2 estimadores restantes, con el fin de evitar efectos derivados de la selección de valores iniciales.

Paso 3: Calcular los estimadores.

Finalmente, los estimadores buscados se definen como el promedio simple de los elementos de la sucesión de Gibbs, es decir,

Sea $\boldsymbol{\theta}^{(i)'} = (\alpha_0^{(i)}, \alpha_1^{(i)}, \beta_1^{(i)}, \dots, \beta_p^{(i)}, \sigma_\epsilon^2{}^{(i)}, p_{00}^{(i)}, p_{11}^{(i)})$, para todo $i \in \{1, \dots, N\}$.

Entonces los estimadores están dados por

$$\begin{aligned} \alpha_0 &= \frac{1}{N} \sum_{i=1}^N \alpha_0^{(i)} \\ \alpha_1 &= \frac{1}{N} \sum_{i=1}^N \alpha_1^{(i)} & \sigma_\epsilon^2 &= \frac{1}{N} \sum_{i=1}^N \sigma_\epsilon^2{}^{(i)} \\ \beta_1 &= \frac{1}{N} \sum_{i=1}^N \beta_1^{(i)} & p_{00} &= \frac{1}{N} \sum_{i=1}^N p_{00}^{(i)} \\ & \vdots & p_{11} &= \frac{1}{N} \sum_{i=1}^N p_{11}^{(i)} \\ \beta_p &= \frac{1}{N} \sum_{i=1}^N \beta_p^{(i)} \end{aligned}$$

Propiedades de convergencia.

De acuerdo con Kuan (2002) [12], esta sucesión converge exponencialmente en distribución a la verdadera distribución de $\boldsymbol{\theta}'$, es decir,

$$\boldsymbol{\theta}^{(N)'} \xrightarrow{D} \boldsymbol{\theta}', \text{ cuando } N \rightarrow \infty .$$

También se cumple que para todo $k \in \{1, 2, 3, 4\}$, cada subvector de parámetros $\boldsymbol{\theta}_k^{(N)}$ converge exponencialmente en distribución a la verdadera densidad de $\boldsymbol{\theta}_k$.

Más aún, para toda función medible g y para todo $k \in \{1, 2, 3, 4\}$, se cumple que

$$\frac{1}{N} \sum_{j=1}^N g(\boldsymbol{\theta}^{(j)'}) \xrightarrow{c.s.} \mathbb{E}[g(\boldsymbol{\theta}')], \text{ cuando } N \rightarrow \infty .$$

3.2. Modelo GARCH con switch markoviano

3.2.1. Descripción del modelo

Los modelos GARCH son muy utilizados para el estudio de fenómenos financieros, ya que la varianza no constante es una característica presente en diversas series de datos. Sin embargo, existen algunos fenómenos en los cuales dicha definición de la varianza no basta para describir el comportamiento de la serie. Para esto se han desarrollado distintas generalizaciones en las cuales se realizan cambios de régimen dentro de un proceso GARCH, las cuales se conocen como modelos GARCH de cambio de régimen markoviano.

Una de estas generalizaciones se utiliza para ciertos fenómenos en los cuales el comportamiento de la varianza no puede ser descrito mediante un único conjunto de parámetros, sino que adopta distintos conjuntos de parámetros a lo largo del tiempo. Una posible solución a este problema es permitir que los parámetros de la varianza se modifiquen de acuerdo con una cadena de Markov discreta. De esta manera, cada régimen captura una dinámica de varianza distinta perteneciente a un modelo GARCH, así que se pueden modelar periodos de alta y baja volatilidad.

Este enfoque se conoce como el modelo GARCH con switch markoviano, lo cual se abrevia como MSGARCH por sus siglas en inglés. La idea de este modelo es definir K procesos GARCH distintos y permitir que el proceso alterne entre dichos procesos de acuerdo con la matriz de probabilidades de transición de una cadena de Markov de K estados, ocasionando que el proceso presente cambios entre estos K regímenes distintos. A continuación, estudiaremos este modelo para procesos GARCH(1,1).

Definición 3.2.1. (Modelo GARCH(1,1) con switch markoviano)
Sean $\{Z_t\}_{t \in \mathbb{N}}$ un $SWN(0,1)$ y $K \in \mathbb{N}$ el número de regímenes del modelo. Para toda $k \in \{1, 2, \dots, K\}$, definimos a los conjuntos de parámetros como $\theta_k = (\alpha_{0,k}, \alpha_{1,k}, \beta_{1,k})$ con $\alpha_{0,k} > 0$, $\alpha_{1,k} \geq 0$ y $\beta_{1,k} \geq 0$ tales que $\alpha_{1,k} + \beta_{1,k} < 1$. Sea $\{S_t\}_{t \in \mathbb{N}}$ una cadena de Markov con espacio de estados $\varepsilon = \{1, 2, \dots, K\}$ y matriz de probabilidades de transición \mathbb{P} .

Decimos que $\{Y_t\}_{t \in \mathbb{N}}$ es un **modelo GARCH(1,1) con switch markoviano**, denotado por $MSGARCH(1,1)$, si se cumple que

$$Y_t = \sqrt{h_t} Z_t \text{ y}$$

$$h_t = \alpha_{0,S_t} + \alpha_{1,S_t} Y_{t-1}^2 + \beta_{1,S_t} h_{t-1}.$$

En esta definición, la cadena de Markov $\{S_t\}_{t \in \mathbb{N}}$ rige los cambios de régimen entre los K conjuntos de parámetros θ_k , los cuales definen los K -regímenes GARCH(1,1) del modelo. En cada tiempo t , el conjunto de parámetros es seleccionado de acuerdo con el valor que tomó la cadena de Markov a tiempo t , lo cual depende únicamente de lo que sucedió en el tiempo $t - 1$ por la propiedad de Markov.

Ejemplo 3.2.1. Forma general del modelo GARCH(1,1) con switch markoviano de 3 regímenes.

Consideramos un modelo con las siguientes especificaciones:

■ **Ruido blanco estricto.**

Sea $\{Z_t\}_{t \in \mathbb{N}}$ un $SWN(0, 1)$ cualquiera.

■ **Parámetros del proceso GARCH.**

Los parámetros de cada uno de los regímenes están dados por:

	Estado 1	Estado 2	Estado 3
α_0	$\alpha_{0,1}$	$\alpha_{0,2}$	$\alpha_{0,3}$
α_1	$\alpha_{1,1}$	$\alpha_{1,2}$	$\alpha_{1,3}$
β_1	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{1,3}$

Donde para todo $k \in \{1, 2, 3\}$ se cumple que $\alpha_{0,k} > 0$, $\alpha_{1,k} \geq 0$, $\beta_{1,k} \geq 0$ y $\alpha_{1,k} + \beta_{1,k} < 1$.

Es decir, tenemos a los conjuntos de parámetros $\theta_1 = (\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1})$, $\theta_2 = (\alpha_{0,2}, \alpha_{1,2}, \beta_{1,2})$ y $\theta_3 = (\alpha_{0,3}, \alpha_{1,3}, \beta_{1,3})$.

■ **Variables de estado.**

Sea $\{S_t\}_{t \in \mathbb{N}}$ una cadena de Markov con espacio de estados $\varepsilon = \{1, 2, 3\}$ y matriz de probabilidades de transición

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

donde \mathbb{P} es una matriz estocástica.

Con esto, podemos definir a $\{Y_t\}_{t \in \mathbb{N}}$ como el proceso dado por

$$Y_t = \sqrt{h_t} Z_t \text{ y}$$

$$h_t = \alpha_{0,S_t} + \alpha_{1,S_t} Y_{t-1}^2 + \beta_{1,S_t} h_{t-1}.$$

Este es un modelo GARCH(1,1) con switch markoviano que alterna entre 3 regímenes. Es decir, a lo largo del tiempo el proceso cambia entre las especificaciones de tres procesos GARCH(1,1) de diferentes parámetros, de acuerdo con una cadena de Markov.

3.2.2. Simulación del modelo

En la presente sección, desarrollaremos dos simulaciones del modelo GARCH con switch markoviano descrito en el **Ejemplo 3.2.1**.

A continuación, se define la siguiente función:

```

1 # install.packages("fGarch")
2 library("fGarch")
3 # PREVIO – Cargar la función simAleatoria
4
5 simulacionGarchMarkoviano <- function(estado1, estado2, estado3,
6   P, n, distr, x0, s0, edo0){
7   # Esta función simula una realización de n pasos de una serie de
8   # tiempo GARCH(1,1) con switch markoviano que alterna entre los
9   # parámetros de estado1, estado2 y estado3 de acuerdo con la
10  # matriz de probabilidades de transición P, utilizando un ruido
11  # blanco estricto con distribución distr y con los valores
12  # iniciales X0=x0, σ0=s0, y comenzando en los parámetros del
13  # estado inicial edo0.
14
15  if(sum(rowSums(P)!=1) != 0 || sum(P<0) != 0){
16    cat("ERROR \nLa matriz no es una matriz estocástica")
17  } else{
18    if(sum(dim(P) == 3) != 2 || sum( c(length(estado1), length(
19      estado2), length(estado3)) == 3 ) != 3){
20      cat("ERROR \nLa matriz debe tener dimensiones 3x3 y los
21        vectores de estado dimensión 3.")
22    } else{
23      p <- data.frame(estado1, estado2, estado3)
24      if(sum(p[2,] + p[3,] >= 1) != 0 || sum(p[1,] <= 0) != 0 ||
25        sum(p[2,] < 0) != 0 || sum(p[3,] < 0) != 0){
26        cat("ERROR \nPara todos los vectores de estados:
27          \nLa suma de los parámetros α1 y β1 debe ser menor que 1.

```

```

24     \nEl parámetro  $\alpha_0$  debe ser positivo,  $\alpha_1$  debe ser no
      negativo y  $\beta_1$  debe ser no negativo.")
25   } else{
26     # Ya que se verificaron las condiciones básicas de los
      estados, la matriz y los parámetros, se realiza la
      simulación.
27     parametros <- data.frame(estado1, estado2, estado3,
28                               row.names = c("alpha0", "alpha1",
                                              "beta1"))
29     ruido <- simAleatoria(distr, n)
30     if(sum(ruido)){
31       s <- s0
32       X <- c(x0)
33       edo <- edo0
34       historEdo <- c()
35       for(i in 1:n){
36         p <- paste("estado", edo, sep = "")
37         s2 <- parametros[1,p] + ( parametros[2,p] * tail(X,1)
38                               ^2 ) + ( parametros[3,p] * s^2 )
39         s <- sqrt(s2)
40         X <- append(X, s * ruido[i])
41         historEdo <- append(historEdo, edo)
42         edo <- sample(c(1, 2, 3), 1, prob = P[edo,] )
43       }
44       X <- X[2:(n+1)] # Quitamos el primer punto.
45       return(data.frame(X, historEdo))
46     }
47   }
48 }
49 }
50 }

```

En el código previo, *simulacionGarchMarkoviano* realiza una simulación de un proceso GARCH(1,1) con switch markoviano utilizando los parámetros especificados. Para ejecutar esta función, es necesario tener previamente cargada la función *simAleatoria* que fue presentada en la sección **2.5.4. Simulación de series de tiempo GARCH** del capítulo previo.

A continuación, utilizaremos la función *simulacionGarchMarkoviano* para simular dos procesos GARCH(1,1) distintos. Para ambos ejemplos, consideraremos una simulación de trayectoria de tamaño 1000 con ruido normal, valor inicial de la serie 1, varianza inicial 0.5 y que comienzan en el estado 1. Además, consideraremos tres estados con los parámetros:

k	1	2	3
$\alpha_{0,k}$	0.001	30	100
$\alpha_{1,k}$	0.3	0.3	0.3
$\beta_{1,k}$	0.5	0.5	0.5

En este caso solo varía el parámetro α_0 , lo que implica que la distribución de las series de tiempo es similar, pero con un cambio de escala. Seleccionamos estos valores debido a que el cambio en la varianza nos permitirá distinguir gráficamente el comportamiento del proceso y en qué estado se encuentra.

Ejemplo 3.2.2. Como primer ejemplo, consideramos la matriz de probabilidades de transición

$$\mathbb{P} = \begin{pmatrix} 0.95 & 0.03 & 0.02 \\ 0.25 & 0.7 & 0.05 \\ 0.1 & 0.05 & 0.85 \end{pmatrix}.$$

Esta matriz presenta pocos cambios de estado, ya que las probabilidades en la diagonal son altas.

Para ello utilizaremos el siguiente código:

```

1 #install.packages("ggplot2")
2 #install.packages("gridExtra")
3 library("ggplot2")
4 library("gridExtra")
5
6 # Ejemplo 1 : Matriz con pocos cambios (probabilidad alta en la
7   diagonal) y estados variando únicamente el parámetro alpha0.
8 # Simulación:
9 estado1 <- c(0.001, 0.3, 0.5)
10 estado2 <- c(30, 0.3, 0.5)
11 estado3 <- c(100, 0.3, 0.5)
12 P <- matrix(c(0.95, 0.03, 0.02,
13              0.25, 0.7, 0.05,
14              0.1, 0.05, 0.85),
15            nrow=3,byrow=T)
16 n <- 1000
17 distr <- list("norm")
18 x0 <- 1
19 s0 <- 0.5
20 edo0 <- 1
21
22 sim1 <- simulacionGarchMarkoviano(estado1, estado2, estado3, P, n,
   distr, x0, s0, edo0)

```

```

23
24 # Gráficas:
25 t <- 1:n
26 X1 <- data.frame(t, X1 = sim1$X)
27 historEdo1 <- data.frame(t, historEdo1 = sim1$historEdo)
28
29 plot1 <- ggplot(data = X1, aes(x = t, y = X1)) +
30   geom_line(color = 'firebrick', linewidth = .5) +
31   labs(x = "t", y = "Xt",
32        title = "Simulación de un GARCH(1,1) con switch markoviano"
33          ,
34        subtitle = "Serie de tiempo") +
35   scale_x_continuous(minor_breaks = t[1:(length(t)/20)==(length(t)
36     /20)]) # Esta sintaxis regresa cada (length(t)/20) elementos
37     de la secuencia
38 plot2 <- ggplot(data = historEdo1, aes(x = t, y = historEdo1)) +
39   geom_line(color = 'firebrick', linewidth = .5)+
40   geom_point(color = 'firebrick') +
41   labs(x = "t", y = "Estado",
42        title = "Simulación de un GARCH(1,1) con switch markoviano"
43          ,
44        subtitle = "Estados") +
45   scale_x_continuous(minor_breaks = t[1:(length(t)/20)==(length(t)
46     /20)])
47 grid.arrange(plot1, plot2, ncol=1)

```

Con lo cual se obtuvieron las gráficas:



Podemos observar que en efecto la gráfica de la serie de tiempo presenta distintos niveles de volatilidad dependiendo del estado en el que se encuentre: cuando se encuentra en el estado 1 presenta una volatilidad muy baja, cuando se encuentra en el estado 2 presenta una volatilidad media y cuando se encuentra en el estado 3 presenta una volatilidad alta. Por la selección de los parámetros de los estados, es fácil distinguir cuándo la serie de tiempo se encuentra en cada estado, lo cual también se puede corroborar con la gráfica de estados. Además, no existen cambios de régimen frecuentes, de manera que cada régimen se mantiene por un periodo relativamente largo de tiempo y la serie se puede identificar por pedazos.

A comparación de un GARCH(1,1) sin switch markoviano, en este caso podemos notar que el cambio de régimen ocasiona que el nivel de volatilidad sea diferente a lo largo del tiempo. Además, si dividimos la serie de acuerdo con el estado en el que se encuentra, cada pedazo de la gráfica se ve como un proceso GARCH(1,1) sin cambio de régimen.

Ejemplo 3.2.3. Como segundo ejemplo, consideramos la matriz de probabilidades de transición

$$\mathbb{P} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Esta matriz presenta cambios de estado frecuentes, ya que todos los estados son equiprobables partiendo de cualquier estado.

Para ello utilizaremos el siguiente código:

```

1 # Ejemplo 2 : Matriz con cambios frecuentes y estados variando ú
   nicamente el parámetro alpha0.
2
3 # Simulación:
4 estado1 <- c(0.001, 0.3, 0.5)
5 estado2 <- c(30, 0.3, 0.5)
6 estado3 <- c(100, 0.3, 0.5)
7 P <- matrix(c(1/3, 1/3, 1/3,
8               1/3, 1/3, 1/3,
9               1/3, 1/3, 1/3),
10            nrow=3,byrow=T)
11 n <- 1000
12 distr <- list("norm")
13 x0 <- 1
14 s0 <- 0.5
15 edo0 <- 1

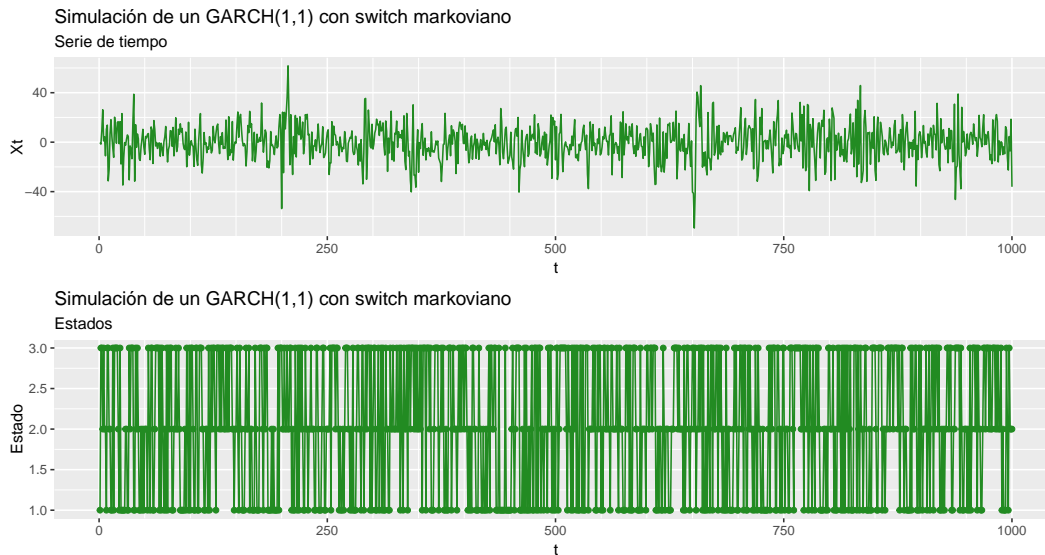
```

```

16
17 sim2 <- simulacionGarchMarkoviano(estado1, estado2, estado3, P, n,
    distr, x0, s0, edo0)
18
19 # Gráficas:
20 t <- 1:n
21 X2 <- data.frame(t, X2 = sim2$X)
22 historEdo2 <- data.frame(t, historEdo2 = sim2$historEdo)
23
24 plot3 <- ggplot(data = X2, aes(x = t, y = X2)) +
25   geom_line(color = 'forestgreen', linewidth = .5) +
26   labs(x = "t", y = "Xt",
27        title = "Simulación de un GARCH(1,1) con switch markoviano"
28        ,
29        subtitle = "Serie de tiempo") +
30   scale_x_continuous(minor_breaks = t[1:(length(t)/20)==(length(t)
31   /20)])
32 plot4 <- ggplot(data = historEdo2, aes(x = t, y = historEdo2)) +
33   geom_line(color = 'forestgreen', linewidth = .5)+
34   geom_point(color = 'forestgreen') +
35   labs(x = "t", y = "Estado",
36        title = "Simulación de un GARCH(1,1) con switch markoviano"
37        ,
38        subtitle = "Estados") +
39   scale_x_continuous(minor_breaks = t[1:(length(t)/20)==(length(t)
40   /20)])
41 grid.arrange(plot3, plot4, ncol=1)

```

Con lo cual se obtuvieron las gráficas:



Podemos observar que en efecto la gráfica de la serie de tiempo presenta diferentes niveles de volatilidad dependiendo del estado en el que se encuentre; cuando se encuentra en el estado 1 presenta una volatilidad muy baja, cuando se encuentra en el estado 2 presenta una volatilidad media y cuando se encuentra en el estado 3 presenta una volatilidad alta. Por la selección de los parámetros de los estados, es fácil distinguir cuándo la serie de tiempo se encuentra en cada estado, lo cual también se puede corroborar con la gráfica de estados. Además, existen cambios de régimen frecuentes, de manera que cada régimen se mantiene por un periodo corto de tiempo y la serie no se puede identificar por pedazos.

Este proceso se ve mucho más semejante a un GARCH(1,1) sin switch markoviano, pues los cambios de régimen ya no son tan fácilmente identificables a partir de la gráfica de la serie de tiempo.

Capítulo 4

Aplicación al S&P/BMV IPC

En el presente capítulo, se aplicará el modelo markoviano de cambio de régimen MSGARCH a la serie de datos del Índice de Precios y Cotizaciones evaluado por S&P Dow Jones Indices y la Bolsa Mexicana de Valores. Para comenzar, se realizará un análisis descriptivo de los datos, que incluirá una descripción básica y un análisis de estadística descriptiva. Posteriormente, se realizará estimación estadística sobre los datos incluyendo un ajuste del modelo y simulaciones.

Los contenidos de este capítulo están basados en el análisis hecho por Ferrer (2022) [14] del índice S&P 500 utilizado en Estados Unidos, mientras que la serie de datos analizada consta de la información histórica del índice equivalente en México, S&P/BMV IPC, obtenida de la página de S&P Dow Jones Indices [15]. Por otro lado, la información teórica del índice fue consultada en la página de S&P Dow Jones Indices [15] y en Castillo (2024) [16]. El código mediante el cual se obtuvieron los resultados de esta sección se encuentra contenido en el Apéndice B.

4.1. Análisis descriptivo de los datos

4.1.1. Descripción de los datos

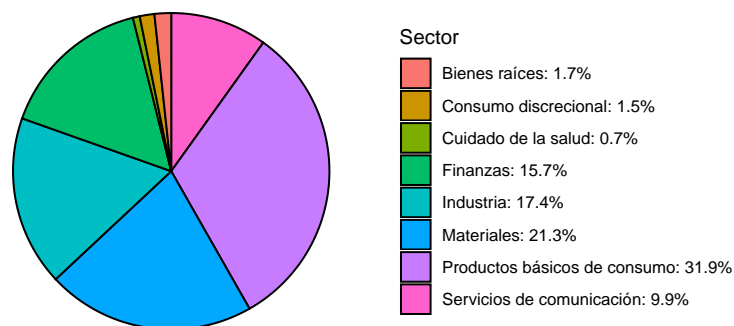
El S&P/BMV Índice de Precios y Cotizaciones (S&P/BMV IPC) es el principal indicador de la tendencia y comportamiento del mercado accionario mexicano. Este índice mide el desempeño de las 35 acciones más representa-

tivas listadas en la Bolsa Mexicana de Valores (BMV), es decir, aquellas con mayor volumen y liquidez. Se consideran 8 sectores económicos:

- Productos básicos de consumo
- Finanzas
- Materiales
- Industria
- Servicios de comunicación
- Consumo discrecional
- Bienes raíces
- Cuidado de la salud

Los cuales presentan el siguiente porcentaje de participación en el índice:

Composición por Sector Industrial



En particular, de acuerdo con S&P Dow Jones Indices [15], las empresas consideradas al 29 de noviembre de 2024 son las siguientes:

1. Grupo Financiero Banorte O
2. Grupo México SAB de CV B
3. Cemex SA CPO
4. Fomento Económico Mexicano S.A.B. de C.V.
5. América Movil SAB de CV B
6. Walmart de México SAB de CV
7. Grupo Aeroportuario del Pacífico, S.A.B. de C.V.
8. Grupo Aeroportuario del Sureste SAB de CV B

9. Arca Continental, SAB de CV
10. Grupo Bimbo S.A.B.
11. Alfa SA A
12. Alsea SA
13. Banco del Bajío, S.A.
14. Becele, S.A. De C.V.
15. Bolsa Mexicana de Valores SA de CV
16. Coca-Cola Femsa SAB de CV UBL
17. Corporación Inmobiliaria Vesta, S.A.B. DE C.V.
18. El Puerto de Liverpool SAB de CV
19. Genomma Lab Internacional SA de CV
20. Gentera SAB de CV
21. Gruma SAB B
22. Grupo Aeroportuario del Centro Norte, S.A.B. de C.V.
23. Grupo Carso SAB de CV
24. Grupo Cementos de Chihuahua SAB de CV
25. Grupo Comercial Chedraui SA de CV
26. Grupo Financiero Inbursa O
27. Grupo Televisa SAB CPO
28. Industrias Penoles
29. Kimberly Clark de Mexico S.A.B. de C.V. A
30. La Comer S.A.B. de C.V. UBC
31. Megacable Holdings SAB de CV

32. ORBIA ADVANCE CORPORATION SAB de CV
33. Promotora y Operadora de Infraestructura SAB de CV
34. Qualitas Controladora S.A.B de C.V.
35. Regional, S.A. de C.V.

Las primeras 10 empresas mencionadas en dicha lista son aquellas con mayor peso en el mercado, mientras que el resto se presentan en orden alfabético.

El índice S&P/BMV IPC ha existido desde 1978 en la forma del BMV Índice de Precios y Cotizaciones (BMV IPC), el cual era calculado únicamente por la Bolsa Mexicana de Valores. Sin embargo, en 2015 se firmó un acuerdo de licenciamiento con la empresa S&P Dow Jones Indices con el objetivo de alinear el IPC con los estándares internacionales y globalizar el mercado de acciones mexicano. De tal manera, a partir de 2017 este índice se calcula en conjunto por la Bolsa Mexicana de Valores (BMV) y el organismo internacional S&P Dow Jones Indices.

En cuanto a la metodología de cálculo, el S&P/BMV IPC se obtiene con base en las variaciones de precios de las 35 acciones seleccionadas, las cuales deben ser una muestra balanceada y representativa de todas las acciones cotizadas en la Bolsa Mexicana de Valores. Aunque este índice se calcula a tiempo real, la selección de las empresas a considerar se realiza semestralmente en marzo y septiembre mediante un proceso llamado rebalanceo, el cual toma en cuenta distintas variables relativas a su operación y volumen. Adicionalmente, se pondera la participación de cada una de las empresas que comprende la muestra con base en el valor de mercado de sus acciones en circulación. Es importante mencionar que en esta ponderación ningún componente deberá rebasar el 25 % y la ponderación total de las 5 acciones de mayor tamaño no deberá rebasar el 60 %, con la finalidad de obtener suficiente diversificación en la muestra.

De acuerdo con S&P Dow Jones Indices y la Bolsa Mexicana de Valores [17], el cálculo de los dividendos de un índice dado se realiza determinando los dividendos totales pagados en cada día de negociación en bolsa como

$$DividendosTotalesPagados = \sum_i Dividendo_i \times Número\ de\ acciones_i ,$$

donde $Dividendo_i$ es el dividendo pagado por la compañía i y $Número\ de$

$acciones_i$ es el número de acciones de la compañía i . El valor de $Dividendo_i$ es cero en la mayoría de los casos, de manera que esta suma se reduce.

Posteriormente, esta variable se expresa en términos de puntos del índice, es decir, se realiza un tipo de cambio de escala como

$$Dividendos\ del\ indice = \frac{Dividendos\ Totales\ Pagados}{Divisor}.$$

Este divisor es un valor establecido desde el comienzo del índice, el cual se selecciona de tal manera que la suma se vuelva un número más interpretable. De esta manera, el índice está medido en puntos y no en una unidad financiera como son los pesos mexicanos.

Debido a esta selección de divisor, resulta más importante analizar los rendimientos sobre este índice que el valor numérico que presenta, por lo cual calcularemos la serie de tiempo de rendimientos. Sea $\{P_1, P_2, \dots, P_T\}$ el conjunto de datos que se tienen de la serie financiera del índice S&P/BMV IPC, entonces la serie financiera de retornos sobre la inversión (ROI) está dada por

$$R_t = \frac{P_{t+1} - P_t}{P_t} * 100, \text{ para cada } t \in \{1, 2, \dots, T - 1\}.$$

Esta serie representa los rendimientos que se obtienen sobre la inversión como un porcentaje y será la serie de tiempo sobre la cual realizaremos estimación estadística.

Para consultar más información con respecto a la selección del divisor de un índice, el cálculo de los índices y el proceso específico para el índice S&P/BMV IPC, se pueden consultar los manuales Index Mathematics Methodology [18] y S&P/BMV Indices Methodology [19].

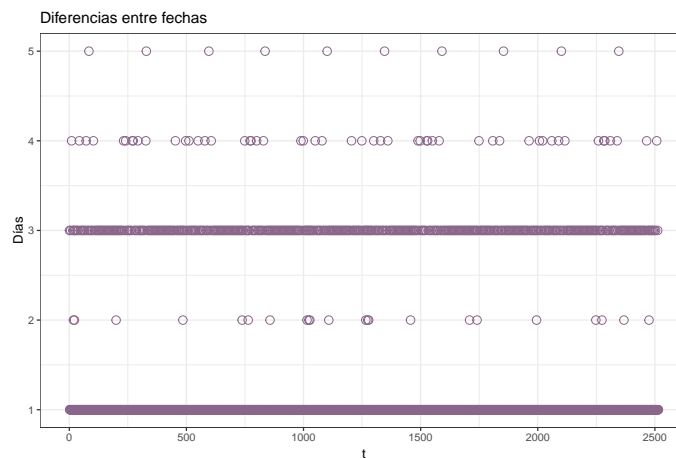
4.1.2. Estadística descriptiva

Para el presente análisis se utilizará la serie de datos del índice S&P/BMV IPC. En total, se cuenta con 2,516 observaciones desde el 28 de noviembre de 2014 hasta el 28 de noviembre de 2024, es decir, durante 10 años.

La serie de tiempo se comporta de la siguiente manera:



Los datos de esta serie de tiempo han sido reportados diariamente considerando los días hábiles, es decir, los días naturales excluyendo sábados, domingos y días festivos. Las diferencias entre las fechas reportadas se comportan como:

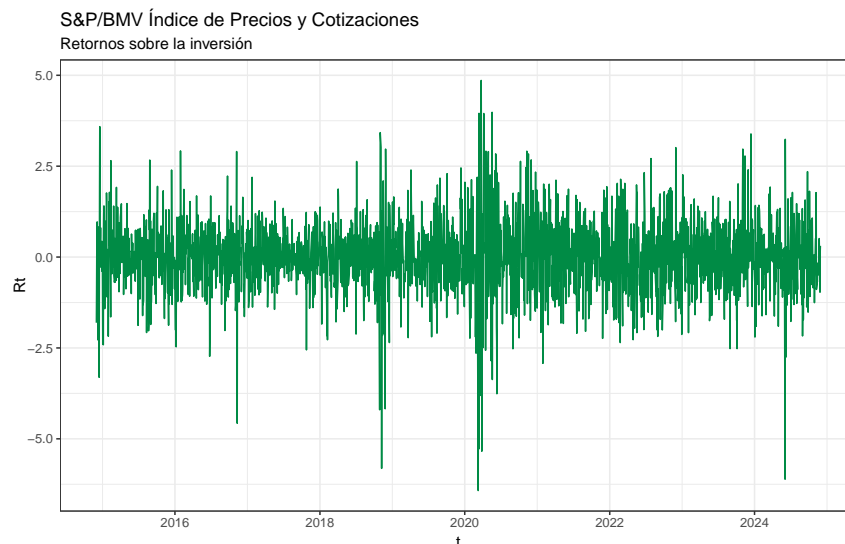


Lo cual corresponde a la tabla de frecuencias:

Diferencia	Frecuencia
1	1971
2	22
3	460
4	52
5	10

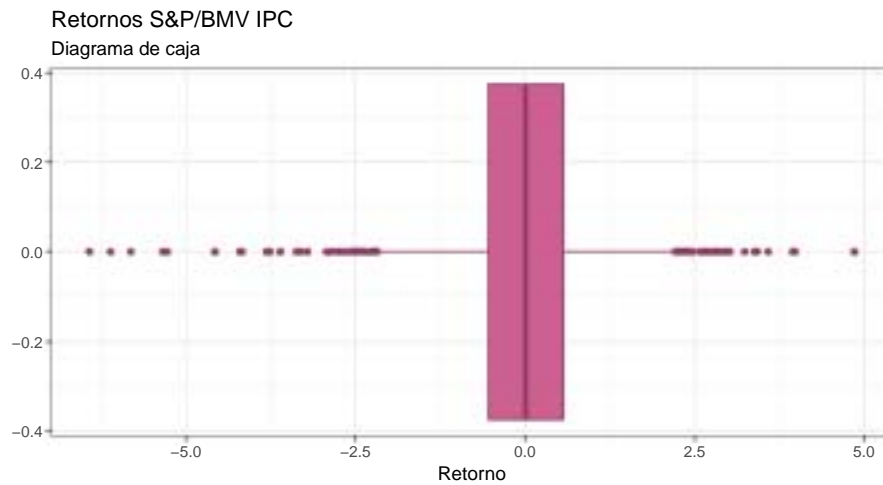
Podemos notar que casi todas las observaciones fueron reportadas con un día de diferencia, lo cual corresponde a las observaciones diarias de lunes a viernes. La siguiente diferencia más frecuente fue de tres días, lo cual corresponde a los sábados y domingos, que no son reportados. También se presentan diferencias de dos, cuatro y cinco días, las cuales corresponden a días festivos y otros datos faltantes. Finalmente, la media de la diferencia fue de 1.452 días y su desviación estándar fue de 0.8863573.

Ya que analizamos la frecuencia en la cual se presentan los datos, la serie de tiempo de los rendimientos del S&P/BMV se comporta de la siguiente manera:



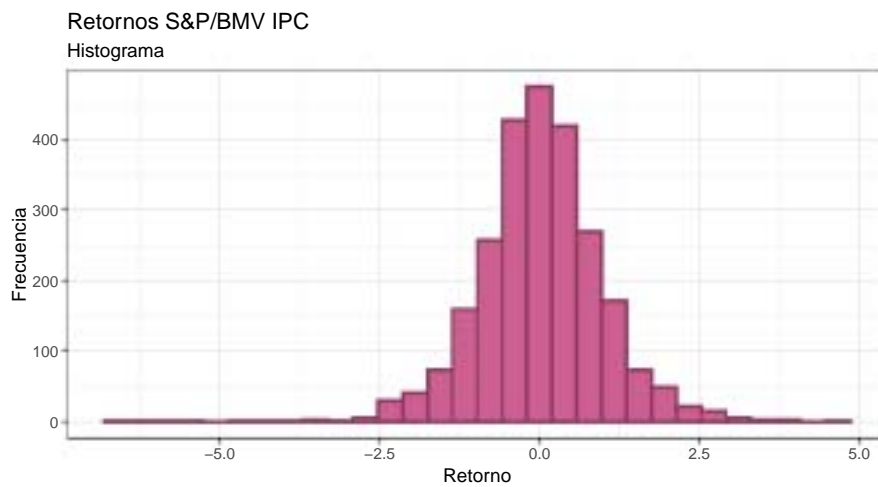
Observemos que la volatilidad varía a lo largo del tiempo y los valores extremos ocurren en agrupaciones, lo cual concuerda con los hechos estilizados de una serie de tiempo financiera. Además, notamos que los retornos absolutos grandes son más comunes alrededor de ciertas fechas específicas en 2017, 2019, 2020 y 2024, a comparación de otros periodos. Esto sugiere que la varianza condicional varía con el tiempo de acuerdo con una especificación de cambio de régimen.

La serie de tiempo de rendimientos presenta un promedio cercano a cero y desviación estándar cercana a 1, específicamente 0.009785 % y 0.9909876 %, respectivamente. Se tiene un valor mínimo de -6.422562 % y un valor máximo de 4.858268 %. Podemos observar el resumen de este comportamiento mediante la gráfica de caja:



Además, la curtosis de esta serie es de 6.498854, por lo cual la serie es leptocúrtica como se indica en los hechos estilizados.

El histograma de los datos es:

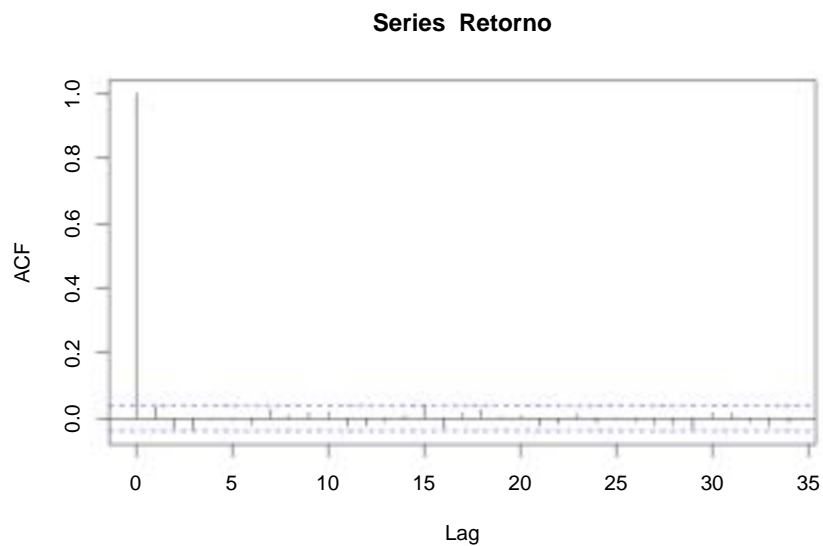


Podemos notar que los datos se concentran en el cero y su concentración disminuye conforme se alejan. Sin embargo, no presenta un comportamiento normal, como se puede observar en el siguiente gráfico Q-Q :



En el centro, los puntos se ajustan a una curva normal; sin embargo, la serie de retornos presenta colas más ligeras.

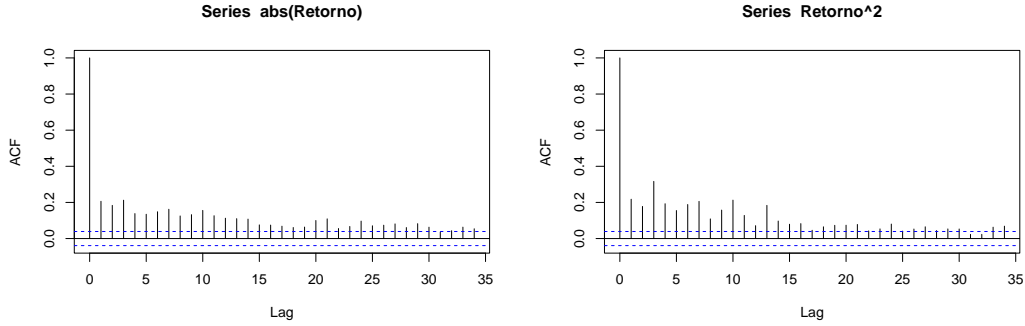
Por otra lado, calcularemos las ACF y PACF muestrales de la serie.



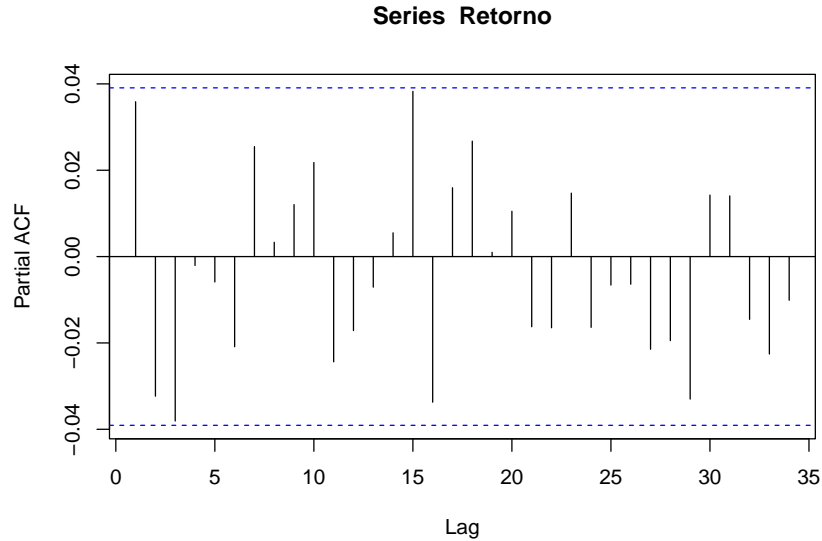
Es decir, la función de autocorrelación es significativamente cero desde un rezago de 1. Además, debido a que la función no decrece asintóticamente a partir de cierto punto, no contamos con evidencia de que la serie sea un modelo MA.

Sin embargo, al calcular la función de autocorrelación del valor absoluto o

el cuadrado, sí se obtiene una correlación significativa, como se puede observar en las siguientes gráficas:



Ambas observaciones concuerdan con los hechos estilizados. En cuanto a la función de autocorrelaciones parciales:



En este caso, debido a que la función no decrece asintóticamente a partir de cierto punto, no contamos con evidencia de que la serie sea un modelo AR.

4.2. Estimación estadística

La estimación estadística de este modelo se realiza mediante la librería “MSGARCH” en R, cuyo manual puede ser consultado en Ardia et al. (2019)

[13].

4.2.1. Selección y ajuste de modelo

Lo primero que se debe realizar para utilizar esta librería es crear la especificación del modelo mediante la función `CREATE_SPEC()`, la cual tiene la siguiente sintaxis:

```
CREATE_SPEC(
    VARIANCE.SPEC = LIST(MODEL = C("SGARCH", "SGARCH")),
    DISTRIBUTION.SPEC = LIST(DISTRIBUTION = C("NORM", "NORM")),
    SWITCH.SPEC = LIST(DO.MIX = FALSE, K = NULL),
    CONSTRAINT.SPEC = LIST(FIXED = LIST(), REGIME.CONST = NULL)
)
```

Donde se tienen los siguientes argumentos principales:

- **VARIANCE.SPEC:** Es la especificación del modelo que presenta cada régimen, en donde "SGARCH" es el modelo MSGARCH que hemos descrito en el capítulo previo. Este argumento debe ser una lista con número de elementos igual al número de regímenes. En caso de que se desee que todos los regímenes consideren el mismo modelo, basta con incluir un vector con una entrada y fijar el valor deseado de K en el tercer argumento.
- **DISTRIBUTION.SPEC:** Es la especificación de la distribución del ruido blanco que presenta cada régimen, en donde "NORM", "STD" y "SGED" corresponden a la distribución normal, t de Student y SGED, respectivamente. Este argumento debe ser una lista con número de elementos igual al número de regímenes. En caso de que se desee que todos los regímenes consideren el mismo modelo, basta con incluir un vector con una entrada y fijar el valor deseado de K en el tercer argumento.
- **SWITCH.SPEC:** Una lista en la cual el primer elemento `DO.MIX = FALSE` representa la especificación de un modelo MSGARCH y K representa el número de regímenes. Si `K = NULL` se toma la misma longitud de las listas `MODEL` y `DISTRIBUTION`.
- **CONSTRAINT.SPEC:** Una lista donde el primer elemento es una lista numérica de entradas numéricas a fijar en la estimación y el segundo elemento controla los parámetros que deben tomar el mismo valor den-

tro de la estimación.

Posteriormente se obtiene estimación por Máxima Verosimilitud mediante la función `FITML()`, la cual tiene la siguiente sintaxis:

`FITML(SPEC, DATA)`

Donde se tienen los siguientes argumentos principales:

- `SPEC`: La especificación del modelo creada con `CreateSpec`.
- `DATA`: Vector de observaciones a ajustar.

La salida principal que nos regresa esta función es una tabla con las estimaciones estadísticas de los parámetros α y β de cada uno de los regímenes y las probabilidades de transición, en la cual se tienen las siguientes columnas:

- *Estimate*: El estimador del parámetro.
- *Std. Error*: La desviación estándar que presenta el parámetro.
- *t value*: Es el cociente entre el estimador y la desviación estándar. Es deseable que los coeficientes tengan un *t value* grande pues esto indicaría que el error estándar es pequeño en comparación con el valor del coeficiente.
- *Pr(> |t|)*: Es el *p value* que se calcula utilizando el estadístico *t* de la distribución T. En la práctica, si el *p value* es inferior a 0.05 usualmente es considerado como significativo, es decir, el coeficiente no es cero con un nivel de confianza del 95 % y en efecto aporta valor a nuestro modelo. Es deseable tener *p value* pequeños.

Adicionalmente, nos regresa los índices AIC y BIC del modelo ¹, los cuales sirven para determinar el modelo más adecuado para describir los datos.

A continuación, realizaremos la estimación con cuatro modelos distintos y definiremos cuál utilizar bajo el criterio del menor promedio ponderado del AIC y BIC.

1. Modelo de 2 regímenes y ruido blanco normal.

¹Consultar el Apéndice C para detalles acerca de estos criterios.

```

-----
Fitted parameters:
      Estimate Std. Error  t value  Pr(>|t|)
alpha0_1  0.0041    0.0024   1.6902 4.549e-02
alpha1_1  0.0250    0.0146   1.7136 4.330e-02
beta_1    0.9548    0.0062 153.0132 <1e-16
alpha0_2  0.1281    0.0642   1.9958 2.298e-02
alpha1_2  0.1048    0.1186   0.8832 1.886e-01
beta_2    0.8719    0.0263  33.1048 <1e-16
P_1_1    0.9006    0.0838  10.7429 <1e-16
P_2_1    0.2868    0.0340   8.4318 <1e-16
-----

```

El promedio del *t value* es de 26.44695 y el promedio del *p value* es de 0.03754054.

2. Modelo de 3 regímenes y ruido blanco normal.

```

-----
Fitted parameters:
      Estimate Std. Error  t value  Pr(>|t|)
alpha0_1  0.0010    0.0021   0.4565 3.240e-01
alpha1_1  0.0312    0.0185   1.6813 4.635e-02
beta_1    0.9532    0.0053 181.2583 <1e-16
alpha0_2  0.5691    0.3685   1.5444 6.124e-02
alpha1_2  0.0560    0.0854   0.6558 2.560e-01
beta_2    0.4428    0.3281   1.3498 8.854e-02
alpha0_3  7.2903   22.1576   0.3290 3.711e-01
alpha1_3  0.0000    0.0000   0.0434 4.827e-01
beta_3    0.0399    2.9095   0.0137 4.945e-01
P_1_1    0.8752    2.6253   0.3334 3.694e-01
P_1_2    0.1244    0.7540   0.1650 4.345e-01
P_2_1    0.2084    0.1036   2.0127 2.207e-02
P_2_2    0.7849    0.0008 935.0822 <1e-16
P_3_1    0.1105    0.0002 479.4993 <1e-16
P_3_2    0.0005    0.0838   0.0060 4.976e-01
-----

```

El promedio del *t value* es de 106.9621 y el promedio del *p value* es de 0.2298654.

3. Modelo de 2 regímenes y ruido blanco t de Student.

```

-----
Fitted parameters:
      Estimate Std. Error  t value  Pr(>|t|)
alpha0_1  0.0160    0.0065   2.4656 6.839e-03
alpha1_1  0.0459    0.0129   3.5693 1.789e-04
beta_1    0.9316    0.0179  51.9717 <1e-16
nu_1     11.5606    2.8443   4.0645 2.407e-05
alpha0_2  0.8476    1.1756   0.7210 2.355e-01
alpha1_2  0.0730    0.1143   0.6386 2.615e-01
beta_2    0.8446    0.1992   4.2406 1.115e-05
P_1_1    0.9972    0.0021 469.6027 <1e-16
P_2_1    0.1263    0.0692   1.8255 3.397e-02
-----

```

El promedio del *t value* es de 59.89995 y el promedio del *p value* es de 0.05977904.

4. Modelo de 3 regímenes y ruido blanco *t* de Student.

```

-----
Fitted parameters:
      Estimate Std. Error  t value  Pr(>|t|)
alpha0_1  0.0026    0.0028   0.9211 1.785e-01
alpha1_1  0.0182    0.0092   1.9832 2.367e-02
beta_1    0.9617    0.0165  58.2875 <1e-16
nu_1     98.4450   15.8646   6.2053 2.729e-10
alpha0_2  0.0256    0.0223   1.1462 1.259e-01
alpha1_2  0.0290    0.0230   1.2648 1.030e-01
beta_2    0.9565    0.0338  28.2915 <1e-16
alpha0_3  0.7633    2.3875   0.3197 3.746e-01
alpha1_3  0.0192    0.0727   0.2639 3.959e-01
beta_3    0.8943    0.3345   2.6739 3.749e-03
P_1_1    0.8555    0.0629  13.5971 <1e-16
P_1_2    0.1435    0.0631   2.2719 1.155e-02
P_2_1    0.1722    0.0777   2.2160 1.335e-02
P_2_2    0.8242    0.0784  10.5166 <1e-16
P_3_1    0.0986    0.0585   1.6850 4.600e-02
P_3_2    0.0023    0.0293   0.0770 4.693e-01
-----

```

El promedio del *t value* es de 8.23254 y el promedio del *p value* es de 0.1090912.

Observamos que los modelos 2 y 4 presentan un *p value* alto, mientras que los otros dos modelos presentan un *p value* significativamente cero. Por lo tanto, podemos seleccionar cualquiera de estos dos modelos para la estimación. En cuanto al *t value*, el modelo 2 presenta el valor más alto, seguido por los modelos 3, 1 y 4.

En cuanto a los criterios de selección de modelos, se obtuvieron los siguientes resultados:

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
AIC	6,650.1845	6,639.5277	6,647.6541	6,640.8603
BIC	6,696.8248	6,726.9781	6,700.1244	6,739.9708
Prom. ponderado	6,668.8406	6,674.5079	6,668.6422	6,680.5045

Donde el promedio ponderado se obtuvo dándole un peso del 60 % al AIC, debido a que nos interesa una penalización más leve por exceso de parámetros. Debido a que el modelo 3 tiene un promedio ponderado más bajo y es adecuado de acuerdo con sus *p value* y *t value*, proseguiremos con la estimación considerando este modelo. A continuación, presentamos el modelo MSGARCH ajustado.

Sean $\{Z_t\}_{t \in \mathbb{N}}$ variables aleatorias independientes e idénticamente distribuidas *t* de Student de parámetro 11.5606 y $\{S_t\}_{t \in \mathbb{N}}$ una cadena de Markov con espacio de estados $\varepsilon = \{1, 2\}$ y matriz de probabilidades de transición

$$\mathbb{P} = \begin{pmatrix} 0.9972 & 0.0028 \\ 0.1263 & 0.8737 \end{pmatrix}.$$

Sean los conjuntos de parámetros

$$\theta_1 = (\alpha_{0,1} = 0.01596231 , \alpha_{1,1} = 0.04593637 , \beta_{1,1} = 0.93162042),$$

$$\theta_2 = (\alpha_{0,2} = 0.84764282 , \alpha_{1,2} = 0.07299234 , \beta_{1,2} = 0.84456307).$$

Entonces la serie de tiempo de los rendimientos del S&P/BMV IPC se ajusta al modelo

$$Y_t = \sqrt{h_t} Z_t y$$

$$h_t = \alpha_{0,S_t} + \alpha_{1,S_t} Y_{t-1}^2 + \beta_{1,S_t} h_{t-1} .$$

El primer régimen presenta una volatilidad no condicional de 0.8434259 y probabilidad de permanecer en este régimen de 0.9972, es decir, presenta baja volatilidad y alta persistencia. El segundo régimen presenta una volatilidad no condicional de 3.2010593 y probabilidad de permanecer en este régimen de 0.8737, es decir, presenta alta volatilidad y alta persistencia.

4.2.2. Simulaciones con el modelo ajustado

A continuación realizaremos simulaciones del modelo ajustado MSGARCH. Para comenzar, se define la función equivalente a la función *simulacionGarch-Markoviano* de la subsección 4.2.2. **Simulación del modelo** del capítulo previo, modificada para permitir dos regímenes en lugar de tres:

```

1 # install.packages("fGarch")
2 library("fGarch")
3 # PREVIO – Cargar la función simAleatoria
4
5 simulacionGarchMarkoviano2 <- function(estado1, estado2, P, n,
6   distr, x0, s0, edo0){
7   # Esta función simula una realización de n pasos de una serie de
8   # tiempo GARCH(1,1) con switch markoviano que alterna entre los
9   # parámetros de estado1 y estado2 de acuerdo con la
10  # matriz de probabilidades de transición P,
11  # utilizando un ruido blanco estricto con distribución distr
12  # y con los valores iniciales X0=x0,  $\sigma_0=s_0$  y comenzando en
13  # los parámetros del estado inicial edo0.
14
15  if(sum(rowSums(P)!=1) != 0 || sum(P<0) != 0){
16    cat("ERROR \nLa matriz no es una matriz estocástica")
17  } else{
18    if(sum(dim(P) == 2) != 2 || sum( c(length(estado1), length(
19      estado2)) == 3 ) != 2){
20      cat("ERROR \nLa matriz debe tener dimensiones 2x2 y los
21        vectores de estado dimensión 3.")
22    } else{
23      p <- data.frame(estado1, estado2)
24      if(sum(p[2,] + p[3,] >= 1) != 0 || sum(p[1,] <= 0) != 0 ||
25        sum(p[2,] < 0) != 0 || sum(p[3,] < 0) != 0 ){
26        cat("ERROR \nPara todos los vectores de estados:
27          \nLa suma de los parámetros  $\alpha_1$  y  $\beta_1$  debe ser menor que 1.
28          \nEl parámetro  $\alpha_0$  debe ser positivo,  $\alpha_1$  debe ser no
29          negativo y  $\beta_1$  debe ser no negativo.")
30      } else{
31        # Ya que se verificaron las condiciones básicas de los
32        # estados, la matriz y los parámetros, se realiza la
33        # simulación.
34        parametros <- data.frame(estado1, estado2,
35          row.names = c("alpha0", "alpha1",
36            "beta1"))
37        ruido <- simAleatoria(distr, n)
38        if(sum(ruido)){
39          s <- s0
40          X <- c(x0)
41          edo <- edo0
42          historEdo <- c()
43          for(i in 1:n){
44            p <- paste("estado", edo, sep = "")
45            s2 <- parametros[1,p] + ( parametros[2,p] * tail(X,1)
46              ^2 ) + ( parametros[3,p] * s^2 )

```

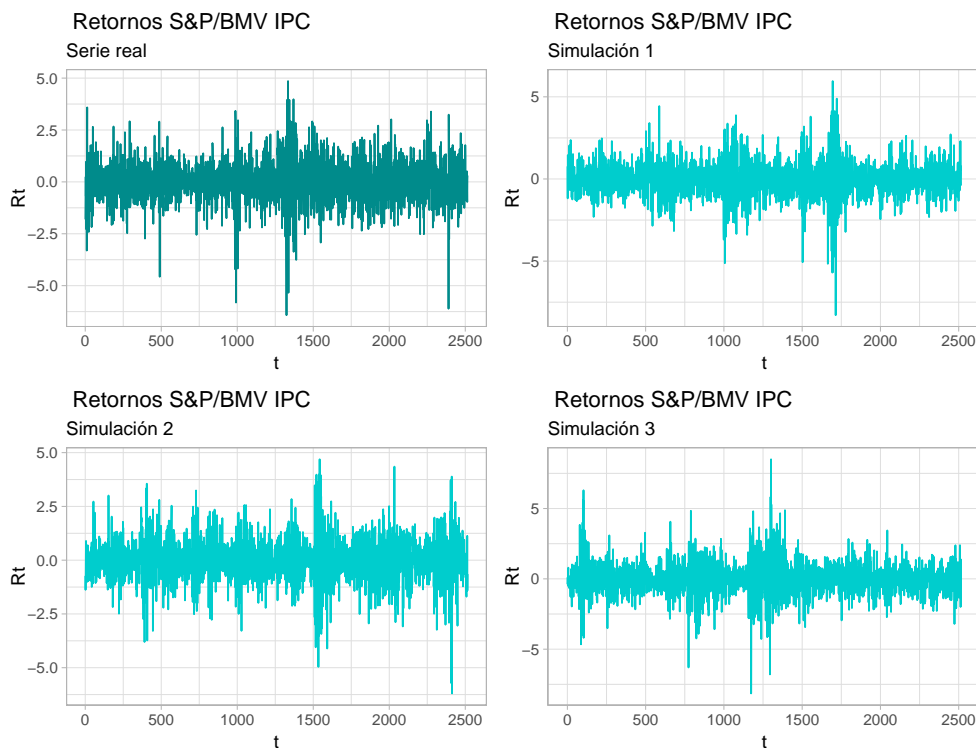
```

38     s <- sqrt(s2)
39     X <- append(X, s * ruido[i])
40     historEdo <- append(historEdo, edo)
41     edo <- sample(c(1, 2), 1, prob = P[edo,] )
42   }
43   X <- X[2:(n+1)] # Quitamos el primer punto.
44
45   return(data.frame(X, historEdo))
46 }
47 }
48 }
49 }
50 }

```

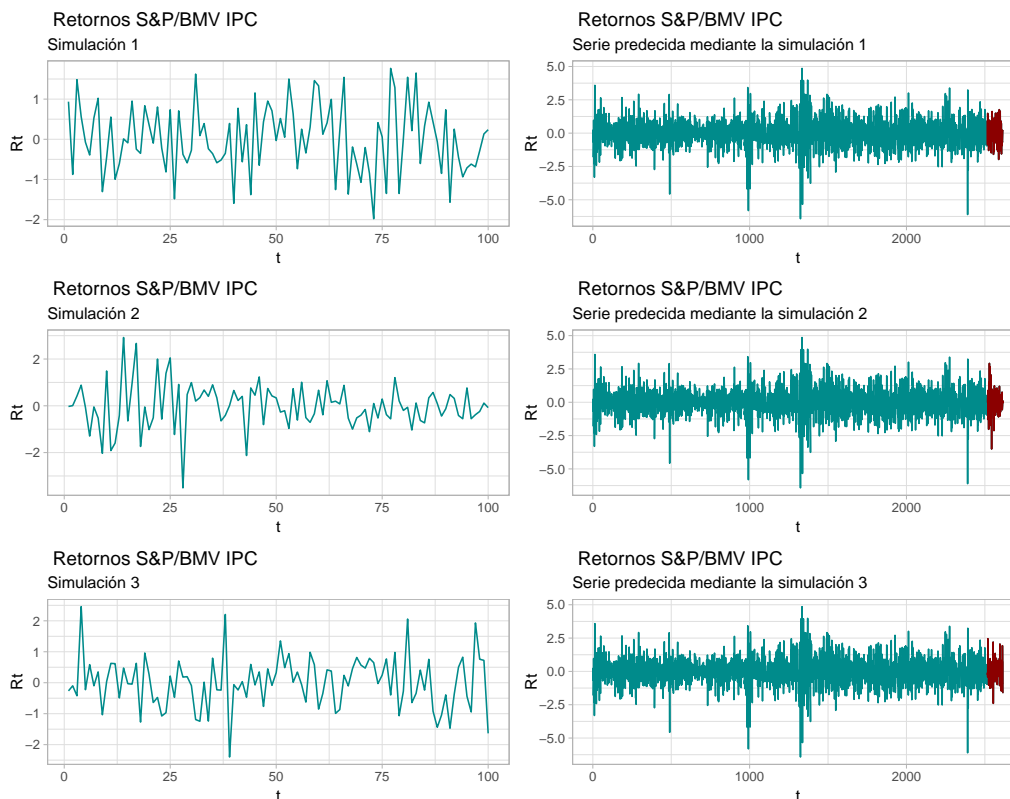
En el código previo, *simulacionGarchMarkoviano2* realiza una simulación de un proceso GARCH(1,1) con switch markoviano utilizando los parámetros especificados. Para ejecutar esta función, es necesario tener previamente cargada la función *simAleatoria* que fue presentada en la sección **3.5.4. Simulación de series de tiempo GARCH** del capítulo 3.

A continuación, utilizando dicha función se realizaron tres simulaciones distintas durante el mismo intervalo de tiempo de la serie original.



Observamos que estas simulaciones presentan un comportamiento semejante a la serie de tiempo real, de manera que el modelo parece ser una buena descripción de la serie.

Por otro lado, se realizaron tres simulaciones para 100 puntos futuros de la serie.



En muchas ocasiones, ya que se tiene el modelo teórico de una serie de tiempo, basta con utilizar los estimadores para predecir los valores futuros. Sin embargo, en este caso existe el componente aleatorio dado por el cambio de régimen markoviano, de manera que dicho método no representa la división de la serie de tiempo por trozos de acuerdo con el régimen en el que se encuentra. Por lo tanto, para la predicción de este modelo se propone aplicar los métodos de Montecarlo partiendo del último punto conocido.

Los métodos de Montecarlo son procedimientos en los cuales se generan puntos aleatorios que simulan casos, con los cuales se puede calcular la distribución de la variable de interés vista como una variable aleatoria. Si el número

de simulaciones es suficientemente grande, se obtendrá una buena aproximación. En el caso de las series de tiempo MSGARCH, se debe considerar a cada punto en el futuro como una variable de interés y realizar un número grande de simulaciones, tal como se realizaron las tres simulaciones mostradas en la gráfica previa. Posteriormente, se deben utilizar los resultados para expresar dicho punto como una variable aleatoria con momentos conocidos. Con ello, se obtiene la predicción de la serie describiendo una variable aleatoria por cada uno de los puntos a predecir.

4.3. Conclusiones

A continuación, presentaremos algunas conclusiones acerca del ajuste de datos realizado en el presente capítulo.

Para comenzar, recordemos que en el análisis descriptivo se evidenció que la serie de tiempo trabajada cumple con los hechos estilizados de una serie de tiempo financiera: la función de autocovarianza es cero desde un rezago de 1 aunque la serie no es independiente ni idénticamente distribuida, las funciones de autocovarianza del valor absoluto y cuadrado sí tienen correlación significativa, la volatilidad varía a lo largo del tiempo, la serie es leptocúrtica y los valores extremos ocurren en agrupaciones. Únicamente faltaría comprobar que la esperanza condicional de los retornos es cercana a cero. Es decir, podemos observar que los datos se comportan como es esperado de una serie financiera.

El análisis exploratorio resultó muy importante para la selección de modelo, ya que pudimos observar que los datos se comportan como modelos GARCH distintos por intervalos de tiempo, es decir, observando a la serie por pedazos se presentan comportamientos correspondientes a modelos GARCH distintos.

En cuanto al ajuste de modelos, se prefiere un modelo más sencillo con alto poder descriptivo, así que tiene sentido haber seleccionado un modelo con menos parámetros. En particular, el modelo MSGARCH de dos regímenes con ruido t de Student resultó adecuado para describir a nuestros datos, ya que tuvo un promedio ponderado de los criterios AIC y BIC más bajo que las otras opciones y tiene un p value bajo en todos los coeficientes, lo cual indica que ningún coeficiente es significativamente cero a un nivel de confianza del 94.022096 % y nuestro modelo no tiene parámetros de más. Además, en general cuenta con t value altos, lo cual indica que los errores estándar son pequeños en comparación con los valores de los coeficientes.

Capítulo 5

Conclusiones

Debido a la importancia que presentan los datos financieros en la sociedad actual, es necesario continuar desarrollando modelos con mayor capacidad descriptiva que se acoplen a la creciente complejidad del mundo. Esto es y seguirá siendo una etapa particularmente importante para la predicción de ciertos fenómenos que impactan la vida cotidiana de las personas. Por ejemplo, es indispensable estudiar la inflación para que la población mantenga una vida digna en la vejez mediante la creación de fondos robustos de ahorro para el retiro.

A lo largo de este trabajo, hemos logrado describir la serie financiera del Índice de Precios y Cotizaciones mediante un modelo GARCH con switch markoviano, a partir de lo cual se han realizado simulaciones del proceso. Para esto, resultó necesario presentar conceptos básicos de cadenas de Markov y de series de tiempo, así como dedicar un capítulo a la presentación y estudio de algunos modelos markovianos de cambio de régimen. En particular, comenzamos introduciendo el modelo autorregresivo de media condicional con cambio de régimen markoviano, el cual sirve como primer ejemplo sencillo para obtener la intuición necesaria acerca de esta clase de modelos. Posteriormente, describimos detalladamente el modelo de interés y lo simulamos en el lenguaje de programación R. Finalmente, en el último capítulo realizamos la estimación estadística de este modelo mediante una librería específica de R denominada “MSGARCH” y utilizamos dicho ajuste para realizar simulaciones con el objetivo de presentar una breve idea respecto a la predicción de valores futuros.

Dentro de los resultados que se obtuvieron, notamos que bastó con con-

siderar dos regímenes para obtener un buen ajuste de los datos, lo cual está relacionado con la preferencia que tenemos por modelos sencillos con alto poder predictivo. Sin embargo, mediante el estudio de modelos con mayor número de regímenes también se puede lograr un buen ajuste de datos, además de que es factible desarrollarlos utilizando métodos computacionales diversos. También es importante notar que este modelo específico es una de las muchas posibilidades de extender y adaptar series de tiempo para incluir cambios de régimen, los cuales además pueden estar regidos por diversas distribuciones distintas a la markoviana.

Finalmente, las simulaciones de valores futuros como método para obtener predicciones estadísticas es un punto que puede ser desarrollado en trabajos posteriores. En este escrito, se introdujo brevemente la idea de obtener predicciones mediante métodos de Montecarlo, para lo cual se utilizó un ejemplo de tres simulaciones que parece presentar resultados satisfactorios. No se profundizó en la idea ni en el procedimiento puesto que está fuera del alcance del trabajo.

En conclusión, este es un trabajo introductorio a cierto tipo de modelos de serie de tiempo con cambio de régimen, en donde se resaltó su importancia mediante su aplicación para ajustar datos sobre el Índice de Precios y Cotizaciones. Sin embargo, con el objetivo de lograr un mejor análisis y ajuste de datos reales provenientes de fenómenos aleatorios, queda mucho por desarrollar acerca de estos modelos y, en general, dentro del amplio campo de conocimientos que es la estadística.

Apéndice A

Otras distribuciones

A.1. Distribución de Error Generalizada (GED)

Una distribución utilizada frecuentemente para modelar errores en fenómenos financieros es la distribución de error generalizada, la cual fue propuesta en 1991 por Daniel B. Nelson. A continuación, presentamos su definición.

Sea X una variable aleatoria perteneciente a la familia de la Distribución de Error Generalizada de parámetro ν , donde $0 < \nu \leq \infty$ es el parámetro de forma. Entonces su función de densidad es

$$f_X(x) = \frac{\nu}{\lambda_\nu 2^{1+\frac{1}{\nu}} \Gamma\left(\frac{1}{\nu}\right)} \exp\left\{-\frac{1}{2} \left|\frac{x}{\lambda_\nu}\right|^\nu\right\}, \text{ con } \lambda_\nu = \left(\frac{2^{-\frac{2}{\nu}} \Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)}\right)^{\frac{1}{2}}.$$

Esta densidad está estandarizada, de manera que tiene media cero y varianza 1. También es posible obtener la distribución con media μ y varianza σ^2 mediante la transformación $(\sigma X + \mu)$.

De esta definición podemos observar que es una densidad simétrica, de manera que los momentos centrales impares son cero y los pares pueden ser calculados como

$$\mu_{2r} = \sigma^{2r} \frac{\left(2^{\frac{1}{\nu}} \lambda_\nu\right)^{2r}}{\Gamma\left(\frac{1}{\nu}\right)} \Gamma\left(\frac{2r+1}{\nu}\right).$$

Finalmente, los coeficientes de asimetría γ_1 y curtosis γ_2 están dados por

$$\gamma_1 = 0 \quad y \quad \gamma_2 = \frac{\Gamma\left(\frac{1}{\nu}\right) \Gamma\left(\frac{5}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)^2} - 3.$$

El contenido presentado previamente está basado en el capítulo 4 de Würtz et al. (2006) [20].

A.2. Distribuciones Sesgadas Estandarizadas

Sea X una distribución estandarizada simétrica y unimodal con función de densidad f_X . Entonces se puede introducir un sesgo mediante el cambio de escala

$$f_X^*(x) = \frac{2}{\xi + \frac{1}{\xi}} f(x_{\mu_\xi \sigma_\xi}); \quad x_{\mu_\xi \sigma_\xi} = \xi^{\text{sign}(\sigma_\xi x + \mu_\xi)} (\sigma_\xi x + \mu_\xi), \text{ donde}$$

$$\mu_\xi = M_1 \left(\xi - \frac{1}{\xi} \right) \quad y \quad \sigma_\xi^2 = (M_2 - M_1^2) \left(\xi^2 - \frac{1}{\xi^2} \right) + 2M_1^2 - M_2, \text{ con}$$

$$M_r = 2 \int_0^\infty x^r f_X(x) dx .$$

Donde $0 < \xi < \infty$ es el parámetro de asimetría.

En particular, al aplicar esta transformación a la Distribución de Error Generalizada, se obtiene la Distribución de Error Generalizada Sesgada (SGED).

El contenido presentado previamente está basado en el capítulo 4 de Würtz et al. (2006) [20].

Apéndice B

Código en R

Para obtener los resultados presentados en el capítulo 5 se utilizó el siguiente código en R:

B.1. Serie de tiempo del S&P/BMV IPC

```
1 #install.packages("readxl")
2 library("readxl")
3 #install.packages("ggplot2")
4 library("ggplot2")
5
6 d1 <- read_excel("C:/Users/Itzel Rmz/OneDrive/Escritorio/Datos 1 -
7     BMV IPC.xls",
8     range = "A7:B2523")
9 df1 <- data.frame(Fecha = as.Date(d1$Fecha), IPC = d1$'S&P/BMV IPC
10     ')
11 # Gráfica
12 ggplot(data = df1, aes(x = Fecha, y = IPC)) +
13     geom_line(color = 'firebrick', linewidth = .5) +
14     labs(x = "Fecha", y = "IPC",
15         title = "S&P/BMV Índice de Precios y Cotizaciones",
16         subtitle = "Valor del índice")+
17     theme_bw()
```

B.2. Análisis de las diferencias entre las fechas reportadas

```

1 # Cálculo de las diferencias
2 # install.packages("dplyr")
3 library("dplyr")
4
5 dif <- c()
6 for (i in 2:length(df1$Fecha)){
7   dif <- append(dif, as.double(difftime(d$Fecha[i],d$Fecha[i-1],
8     units = "days")))
9 }
10 dif <- data.frame(dif)
11
12 # Gráfica:
13 t <- 1:length(dif$dif)
14 X <- data.frame(t, dif = dif$dif)
15 ggplot(data = X, aes(x = t, y = dif)) +
16   geom_point(shape = 1, color = 'plum4', size = 3)+
17   labs(x = "t", y = "Días",
18     title = "Diferencias entre fechas")+
19   theme_bw()
20
21 # Tabla de frecuencias:
22 dif %>%
23   group_by(dif) %>%
24   summarize(Freq=n())

```

B.3. Cálculo del retorno sobre la inversión y sus gráficas

```

1 # Retorno sobre la inversión
2 Retorno <- c()
3 for (i in 2:length(df1$IPC)){
4   Retorno <- append(Retorno, (df1$IPC[i]-df1$IPC[i-1])/df1$IPC[i-1]
5     *100)
6 }
7
8 ret <- data.frame(Fecha = df1$Fecha[2:length(df1$Fecha)], Retorno
9   = Retorno)
10
11 # Gráficas:
12 ggplot(data = ret, aes(x = Fecha, y = Retorno)) +
13   geom_line(color = 'springgreen4', linewidth = .5) +
14   labs(x = "t", y = "Rt",

```

```

12     title = "S&P/BMV Índice de Precios y Cotizaciones",
13     subtitle = "Retornos sobre la inversión")+
14   theme_bw()
15
16 summary(Retorno) # Media cercana a cero
17 sd(Retorno) # Desviación estándar cercana a uno
18 # install.packages("moments")
19 library("moments")
20 kurtosis(Retorno)
21
22 ggplot(data = ret, aes(x = Retorno)) +
23   geom_boxplot(color = 'hotpink4', fill='hotpink3') +
24   labs(x = "Retorno",
25        title = "Retornos S&P/BMV IPC",
26        subtitle = "Diagrama de caja")+
27   theme_bw()
28
29 ggplot(data = ret, aes(x = Retorno)) +
30   geom_histogram(color = 'hotpink4', fill='hotpink3', linewidth =
31     0.7) +
32   labs(x = "Retorno", y = "Frecuencia",
33        title = "Retornos S&P/BMV IPC",
34        subtitle = "Histograma")+
35   theme_bw()
36
37 ggplot(data = ret, aes(sample = Retorno)) +
38   geom_qq(color = 'hotpink3') +
39   labs(x = "Teórico", y = "Muestra",
40        title = "Retornos S&P/BMV IPC",
41        subtitle = "Gráfico Q-Q")+
42   theme_bw()
43 acf(Retorno)
44
45 pacf(Retorno)

```

B.4. Ajuste de los datos a modelo MSGARCH y comparación de parámetros

```

1 #install.packages("MSGARCH")
2 library("MSGARCH")
3
4 # Ajuste a MSGARCH y comparación de modelos
5 spec1a <- CreateSpec(
6   variance.spec = list(model = c("sGARCH")),

```

```

7   distribution.spec = list(distribution = c("norm")),
8   switch.spec = list(K = 2)
9 ) # 2 regímenes normal
10 flaML <- FitML(spec = spec1a, data = ret$Retorno)
11 m1t <- mean(flaML$Inference$MatCoef[,3])
12 m1p <- mean(flaML$Inference$MatCoef[,4])
13
14 spec1b <- CreateSpec(
15   variance.spec = list(model = c("sGARCH")),
16   distribution.spec = list(distribution = c("norm")),
17   switch.spec = list(K = 3)
18 ) # 3 regímenes normal
19 flbML <- FitML(spec = spec1b, data = ret$Retorno)
20 m2t <- mean(flbML$Inference$MatCoef[,3])
21 m2p <- mean(flbML$Inference$MatCoef[,4])
22
23 spec1c <- CreateSpec(
24   variance.spec = list(model = c("sGARCH")),
25   distribution.spec = list(distribution = c("std")),
26   switch.spec = list(K = 2),
27   constraint.spec = list(regime.const = "nu")
28 ) # 2 regímenes t de Student
29 flcML <- FitML(spec = spec1c, data = ret$Retorno)
30 m3t <- mean(flcML$Inference$MatCoef[,3])
31 m3p <- mean(flcML$Inference$MatCoef[,4])
32
33 spec1d <- CreateSpec(
34   variance.spec = list(model = c("sGARCH")),
35   distribution.spec = list(distribution = c("std")),
36   switch.spec = list(K = 3),
37   constraint.spec = list(regime.const = "nu")
38 ) # 3 regímenes t de Student
39 fldML <- FitML(spec = spec1d, data = ret$Retorno)
40 m4t <- mean(fldML$Inference$MatCoef[,3])
41 m4p <- mean(fldML$Inference$MatCoef[,4])

```

B.5. Descripción y estimación del modelo MS-GARCH

```

1 fit1 <- flcML
2 vol <- sapply(ExtractStateFit(fit1), UncVol) #UncVol regresa la
   volatilidad no condicional del proceso
3
4 # Parámetros estimados:
5 fit1$par

```

```

6 estado1 <- c(fit1$par["alpha0_1"],
7             fit1$par["alpha1_1"],
8             fit1$par["beta_1"])
9 estado2 <- c(fit1$par["alpha0_2"],
10            fit1$par["alpha1_2"],
11            fit1$par["beta_2"])
12 P1 <- fit1$par["P_1_1"]
13 P2 <- fit1$par["P_2_1"]
14 P <- matrix(c(P1, 1-P1,
15             P2, 1-P2),
16            nrow=2,byrow=T)
17 distr <- list("tstudent", fit1$par["nu_1"])

```

B.6. Simulaciones del modelo ajustado MS-GARCH

```

1 # Simulaciones en el mismo tiempo que la serie de retornos
2 #install.packages("gridExtra")
3 library("gridExtra")
4 edo0 <- 1
5 s0 <- vol[1]
6
7 x0 <- 0
8 n <- length(ret$Retorno)
9
10 X0 <- data.frame(t = 1:n, Rt = ret$Retorno)
11 p0 <- ggplot(data = X0, aes(x = t, y = Rt)) +
12   geom_line(color = 'cyan4', linewidth = .5) +
13   labs(x = "t", y = "Rt",
14        title = " Retornos S&P/BMV IPC",
15        subtitle = " Serie real")+
16   theme_light()
17 X1 <- simulacionGarchMarkoviano2(estado1, estado2, P, n, distr, x0
18   , s0, edo0)
19 X1 <- data.frame(t = 1:n, Rt = X1$X)
20 p1 <- ggplot(data = X1, aes(x = t, y = Rt)) +
21   geom_line(color = 'cyan3', linewidth = .5) +
22   labs(x = "t", y = "Rt",
23        title = " Retornos S&P/BMV IPC",
24        subtitle = " Simulación 1")+
25   theme_light()
26 X2 <- simulacionGarchMarkoviano2(estado1, estado2, P, n, distr, x0
27   , s0, edo0)
28 X2 <- data.frame(t = 1:n, Rt = X2$X)

```

```

28 p2 <- ggplot(data = X2, aes(x = t, y = Rt)) +
29   geom_line(color = 'cyan3', linewidth = .5) +
30   labs(x = "t", y = "Rt",
31        title = " Retornos S&P/BMV IPC",
32        subtitle = " Simulación 2")+
33   theme_light()
34
35 X3 <- simulacionGarchMarkoviano2(estado1, estado2, P, n, distr, x0
   , s0, edo0)
36 X3 <- data.frame(t = 1:n, Rt = X3$X)
37 p3 <- ggplot(data = X3, aes(x = t, y = Rt)) +
38   geom_line(color = 'cyan3', linewidth = .5) +
39   labs(x = "t", y = "Rt",
40        title = " Retornos S&P/BMV IPC",
41        subtitle = " Simulación 3")+
42   theme_light()
43
44 grid.arrange(p0, p1, p2, p3, ncol=2)
45
46 # Predicción mediante simulación
47 edo0 <- 1
48 s0 <- vol[1]
49 x0 <- tail(ret$Retorno,1)
50 s <- 100 # Número de puntos a estimar
51
52 sim1 <- simulacionGarchMarkoviano2(estado1, estado2, P, s, distr,
   x0, s0, edo0)
53 sim1 <- data.frame(t = 1:s, sim = sim1$X)
54 pred <- data.frame(t = (length(ret$Retorno)+1):(length(ret$Retorno
   )+s),
55                   Rt = sim1$sim)
56 pla <- ggplot(data = sim1, aes(x = t, y = sim)) +
57   geom_line(color = 'cyan4', linewidth = .5) +
58   labs(x = "t", y = "Rt",
59        title = " Retornos S&P/BMV IPC",
60        subtitle = " Simulación 1")+
61   theme_light()
62 X <- data.frame( t = 1:(length(ret$Retorno)+s), Rt = c(ret$Retorno
   , sim1$sim) )
63 plb <- ggplot(data = X, aes(x = t, y = Rt)) +
64   geom_line(color = 'cyan4', linewidth = .5) +
65   labs(x = "t", y = "Rt",
66        title = " Retornos S&P/BMV IPC",
67        subtitle = " Serie predecida mediante la simulación 1")+
68   theme_light()
69 plb <- plb + geom_line(data=pred, color="red4", linewidth = .5)
70

```

```

71 sim2 <- simulacionGarchMarkoviano2(estado1, estado2, P, s, distr,
   x0, s0, edo0)
72 sim2 <- data.frame(t = 1:s, sim = sim2$X)
73 pred <- data.frame(t = (length(ret$Retorno)+1):(length(ret$Retorno)
   )+s),
74                   Rt = sim2$sim)
75 p2a <- ggplot(data = sim2, aes(x = t, y = sim)) +
76   geom_line(color = 'cyan4', linewidth = .5) +
77   labs(x = "t", y = "Rt",
78        title = " Retornos S&P/BMV IPC",
79        subtitle = " Simulación 2")+
80   theme_light()
81 X <- data.frame( t = 1:(length(ret$Retorno)+s), Rt = c(ret$Retorno
   , sim2$sim) )
82 p2b <- ggplot(data = X, aes(x = t, y = Rt)) +
83   geom_line(color = 'cyan4', linewidth = .5) +
84   labs(x = "t", y = "Rt",
85        title = " Retornos S&P/BMV IPC",
86        subtitle = " Serie predecida mediante la simulación 2")+
87   theme_light()
88 p2b <- p2b + geom_line(data=pred, color="red4", linewidth = .5)
89
90 sim3 <- simulacionGarchMarkoviano2(estado1, estado2, P, s, distr,
   x0, s0, edo0)
91 sim3 <- data.frame(t = 1:s, sim = sim3$X)
92 pred <- data.frame(t = (length(ret$Retorno)+1):(length(ret$Retorno)
   )+s),
93                   Rt = sim3$sim)
94 p3a <- ggplot(data = sim3, aes(x = t, y = sim)) +
95   geom_line(color = 'cyan4', linewidth = .5) +
96   labs(x = "t", y = "Rt",
97        title = " Retornos S&P/BMV IPC",
98        subtitle = " Simulación 3")+
99   theme_light()
100 X <- data.frame( t = 1:(length(ret$Retorno)+s), Rt = c(ret$Retorno
   , sim3$sim) )
101 p3b <- ggplot(data = X, aes(x = t, y = Rt)) +
102   geom_line(color = 'cyan4', linewidth = .5) +
103   labs(x = "t", y = "Rt",
104        title = " Retornos S&P/BMV IPC",
105        subtitle = " Serie predecida mediante la simulación 3")+
106   theme_light()
107 p3b <- p3b + geom_line(data=pred, color="red4", linewidth = .5)
108
109 grid.arrange(p1a, p1b, p2a, p2b, p3a, p3b, ncol=2)

```

Apéndice C

Índices AIC y BIC

Los índices AIC y BIC proporcionan criterios para la selección de modelos, es decir, nos permiten determinar el modelo más adecuado para describir a los datos dado que se tienen distintas propuestas de modelos.

El **criterio de información de Akaike (AIC)** fue propuesto por Hirotugu Akaike en 1974 y tiene como objetivo estimar la calidad de distintos modelos penalizando la complejidad. Para esto, el AIC intenta encontrar un modelo que maximice la verosimilitud de los datos tomando en cuenta el número de parámetros utilizados. Su fórmula es

$$AIC = 2k - 2 \ln(\mathcal{L}(\hat{\theta})),$$

donde $\mathcal{L}(\hat{\theta})$ es la función de verosimilitud maximizada y k es el número de parámetros en el modelo. Cuando la función de verosimilitud resulta muy difícil de calcular, en muchas ocasiones se utiliza la suma de errores cuadráticos en lugar de la verosimilitud maximizada.

Al incorporar a la verosimilitud y al número de parámetros, el AIC trata de encontrar un balance entre el ajuste del modelo y su complejidad. Cuando se desea seleccionar un modelo bajo este criterio, el mejor es aquel con un menor AIC.

El **criterio de información Bayesiano (BIC)** fue propuesto por Gideon E. Schwarz en 1978 e igualmente tiene como objetivo estimar la calidad de distintos modelos penalizando la complejidad. Su fórmula es

$$BIC = k \ln(n) - 2 \ln(\mathcal{L}(\hat{\theta})),$$

donde $\mathcal{L}(\hat{\theta})$ es la función de verosimilitud maximizada, k es el número de parámetros en el modelo y n es el tamaño de la muestra. Cuando se desea seleccionar un modelo bajo este criterio, el mejor es aquel con un menor BIC.

La diferencia entre estos dos criterios es únicamente el término de penalización por complejidad del modelo, ya que en el BIC la penalización aumenta logarítmicamente con el tamaño de la muestra. De tal manera, el BIC tiende a favorecer modelos más simples.

El contenido presentado previamente está basado en los contenidos de Montesinos (2011) [21].

Bibliografía

- [1] Ross S. *Introduction to Probability Models*. 10.^a ed. Academic Press, 2010.
- [2] Suhov Y. y Kelbert M. “Probability and Statistics by Example. Volume 2: Markov chains: A primer in random processes and their applications”. En: *Cambridge University Press* (nov. de 2008). DOI: 10.1017/CB09780511813641.
- [3] Craig B. y P. Sendi. “Estimation of the transition matrix of a discrete-time Markov chain”. En: *Health economics* 11 (2002), págs. 33-42. DOI: 10.1002/hec.654.
- [4] Spedicato G. et al. *The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R*. 2014.
- [5] Wei W. *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley Publishing Company, Inc., 1990. ISBN: 0-201-15911-2.
- [6] Chatfield C. *The Analysis of Time Series: An Introduction*. 6.^a ed. Texts in Statistical Science. Chapman Hall/CRC, 2004. ISBN: 1-58488-317-0.
- [7] Larsen R. y Marx M. *An Introduction to Mathematical Statistics and Its Applications*. 3.^a ed. Prentice Hall, 2001. ISBN: 0-13-922303-7.
- [8] Nosedal A. *The Moving Average Models MA(1) and MA(2)*. Feb. de 2019.
- [9] Shumway R. y Stoffer D. *Time Series Analysis and Its Applications: With R Examples*. 2.^a ed. Springer, 2006. ISBN: 978-0-387-29317-2. DOI: 10.1007/0-387-36276-2.
- [10] McNeil A.J., Frey R. y Embrechts P. *Quantitative Risk Management. Concepts, Techniques and Tools*. 1.^a ed. Princeton University Press, 2005. ISBN: 0-691-12255-5.
- [11] Andersen T. et al., eds. *Handbook of Financial Time Series*. Springer-Verlag, 2009. ISBN: 978-3-540-71296-1.
- [12] Kuan C.-M. “Lecture on the markov switching model”. En: *Institute of Economics, Academia Sinica* (abr. de 2002).

- [13] Ardia D. et al. “Markov-Switching GARCH Models in R: The MS-GARCH Package”. En: *Journal of statistical software* 91 (oct. de 2019). DOI: 10.18637/jss.v091.i04.
- [14] Ferrer M. *Modelling Volatility with Markov-Switching GARCH Models*. University of Liverpool, ago. de 2022.
- [15] S&P Dow Jones Indices. *S&P/BMV IPC*. 2024. URL: <https://www.spglobal.com/spdji/es/indices/equity/sp-bmv-ipc/#overview> (visitado 21-12-2024).
- [16] Castillo J. *¿Qué es el índice SP BMV IPC y cómo se calcula?* Jun. de 2024. URL: <https://www.rankia.mx/acciones/indice-precios-cotizaciones-ipc/blog/6422842-que-como-calcular-indice-sp-bmv-ipc> (visitado 21-12-2024).
- [17] *Consulta sobre la Metodología de Cálculo de los Índices de Retorno Total de la BMV*. S&P Dow Jones Indices y Bolsa Mexicana de Valores.
- [18] *Index Mathematics Methodology*. S&P Dow Jones Indices. Mar. de 2024.
- [19] *S&P/BMV Indices Methodology*. S&P Dow Jones Indices. Oct. de 2024.
- [20] Würtz D., Chalabi Y. y Luksan L. “Parameter Estimation of ARMA Models with GARCH/APARCH Errors An R and SPlus Software Implementation”. En: *JSS Journal of Statistical Software* (2006).
- [21] Montesinos A. “Estudio del AIC y del BIC en la selección de modelos de vida con datos censurados”. Tesis de maestría. Guanajuato, Guanajuato: CIMAT, 2011.