



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Dos métodos para la detección de cambios
abruptos en series de tiempo

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Eduardo Nava Rosales

TUTORA

Dra. Ana Meda Guardiola



Ciudad Universitaria, Ciudad de México

2018



Universidad Nacional
Autónoma de México



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.

1.Datos del alumno

Nava
Rosales
Eduardo
55 43634917
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
412055988

2.Datos del tutor

Dra.
Ana
Meda
Guardiola

3.Datos del sinodal 1

Dra.
Guillermina
Eslava
Gómez

4.Datos del sinodal 2

Dr.
Sergio Iván
López
Ortega

5.Datos del sinodal 3

Act.
Miguel Ángel
Chong
Rodríguez

6.Datos del sinodal 4

Act.
Elsa Lorena
García
Reyes

7.Datos del trabajo escrito

Dos métodos para la detección de cambios abruptos en series de tiempo
Nn 90
2018

¡México, Pumas, Universidad!

Agradecimientos

Quiero agradecer de todo corazón a todas las personas con las que he compartido este camino, es imposible nombrar a todos. A todos mis compañeros y amigos que estuvieron desde los primeros intentos de demostraciones, los que conocí a la mitad del camino y con los que compartí los últimos días en la facultad. Diego, Glo, Daniel, Diana, Lor, Maf, Sonny, Mau e Isaas, gracias por el apoyo brindado. Muchas gracias a cada uno de mis profesores y ayudantes quienes muchas veces dedicaron más tiempo del que estaba programado para aclarar mis múltiples dudas, especialmente a la profesora Ana Meda quien me tuvo la paciencia, consideración y confianza para poder llevar este trabajo a buen puerto. Por último, quiero agradecer a mi familia, que con su apoyo y enseñanzas hicieron que este camino fuera menos complicado, a mi papá Fernando. A mis hermanos: Fer, Kari, Ari y Nan. Y a mí mamá Lorena quien ha estado todo el tiempo conmigo.

Índice general

1. Conceptos Básicos	9
1.1. Pruebas de Hipótesis	10
1.2. Análisis Secuencial	11
1.3. Detección de cambios abruptos	12
1.3.1. Detección en directo	13
1.3.2. Detección con una muestra fija	13
2. El método de CUSUM	17
2.1. CUSUM	17
2.2. Teoría ergódica	20
2.2.1. Teorema Ergódico de Birkhoff-Khinchin	22
2.3. Optimalidad del algoritmo CUSUM	28

3. Algoritmo CSSM	35
3.1. Procesos fuertemente mezclantes (α – <i>mixing</i>)	36
3.1.1. Teorema Funcional del Límite Central	37
3.2. Detección de cambios en procesos α - mezclantes	48
3.2.1. CSSM	53
4. Simulaciones y Resultados	61
4.1. Comparación de los algoritmos <i>CSSM</i> y <i>CUSUM</i> en un proceso <i>ARMA</i> (1,1)	62
4.1.1. Resultados <i>CUSUM</i> ante un <i>ARMA</i> (1,1)	63
4.1.2. Resultado <i>CSSM</i> ante un <i>ARMA</i> (1,1)	69
4.1.3. El <i>CUSUM</i> y el <i>CSSM</i> ante un <i>ARMA</i> (2,3)	73
4.1.4. El <i>CSSM</i> y un proceso <i>GARCH</i> (1,1)	77
A. Más teoremas	85
B. Códigos R	87
B.1. Código <i>CUSUM</i>	87
B.2. Código <i>CSSM</i>	88

Notación

Notación	Significado
\mathbb{N}	Conjunto de los números naturales.
\mathbb{R}	Conjunto de los números reales.
$(\Omega, \mathcal{F}, \mathbb{P})$	Espacio de probabilidad.
$\mathbb{P}(A)$	Probabilidad del evento A .
$F(x)$	Función de distribución de X .
$f(x)$	Función de densidad de X .
$\mathbf{E}(X)$	Esperanza de X .
$\mathbf{Var}(X)$	Varianza de X .
$\mathbf{Cov}(X, Y)$	Covarianza entre X y Y .
<i>v.a.</i>	Variable(s) aleatoria(s).
<i>v.a.i.i.d.</i>	Variables aleatorias independientes idénticamente distribuidas.
θ	Parámetro o vector de parámetros.
$F_\theta(x)$	Función de distribución X bajo θ .
$f_\theta(x)$	Función de densidad X bajo θ .
$\mathbf{E}_\theta(X)$	Esperanza de X bajo θ .
$\{X_n\}$	Sucesión de <i>v.a.</i> con subíndices en \mathbb{N} .
$\{a_n\}$	Sucesión de números reales con subíndices en \mathbb{N} .
\mathcal{F}_m^n	σ - álgebra generada por las <i>v. a.</i> X_m, X_{m+1}, \dots, X_n .
$\int f d\mu$	Integral la función f con respecto a la medida μ .
$\int_A x dF(x)$	Esperanza X sobre A .
$\mathbb{1}_A$	Función indicadora del evento A .
$\stackrel{\mathcal{D}}{=}$	Igualdad en distribución.
$\ X\ _\beta$	$(\mathbf{E}(X^\beta))^{1/\beta}$.

Resumen

Las series de tiempo llegan a presentar cambios repentinos en su estructura. En este trabajo se presentan dos algoritmos que sirven para detectar estos cambios y en su caso estimar el punto donde ocurrió el cambio. El primer algoritmo, conocido como CUSUM, para el cual existe una prueba que muestra que es el mejor en el sentido que minimiza la probabilidad de detectar un cambio cuando este no existe. El segundo algoritmo conocido como CSSM sirve para detectar cambios en modelos de series de tiempo, que cumplen una condición conocida como *alpha – mezclante*. Existe una gran diferencia en la implementación de cada algoritmo, mientras que el CUSUM necesita conocer los parámetros de la serie tanto antes como después del cambio, el CSSM no necesita otra cosa más que los datos. En el trabajo se compararon ambos algoritmos a través de simulaciones de procesos $ARMA(1, 1)$ y $ARMA(3, 2)$. Los resultados muestran que, en ambos casos, el CUSUM tiene casi el 100% de efectividad en la detección y menos del 4% de falsos positivos, mientras que el CSSM tuvo una efectividad mayor al 80% cuando el cambio en ambos parámetros del $ARMA(1, 1)$ fue mayor a .1, sin embargo el porcentaje de falsos positivos fue de 21%, para el caso del $ARMA(2, 3)$ el porcentaje de efectividad en la detección fue de 95% y el de falsas alarmas fue de 4.1%. Se pudo observar que ambos algoritmos aumentan su efectividad conforme aumenta el tamaño del cambio y también aumenta la precisión en la estimación del punto de cambio, siendo el CUSUM el más preciso de los dos. También se presentan los teoremas en se basan ambos algoritmos, entre estos, se encuentran el Teorema Ergódico de Birkhoff-Khinchin y el Teorema Funcional del Límite Central de Herrndorf.

Introducción

Las series de tiempo son un conjunto de observaciones x_t , cada una registrada en un tiempo específico $t \leq 0$, se pueden considerar como la realización de una sucesión de variables aleatorias y en ocasiones pueden presentar cambios repentinos en su comportamiento, lo que se puede interpretar como un cambio en los parámetros de las variables o incluso como un cambio en la distribución de las mismas.

La motivación del presente trabajo fue conocer el sustento teórico de dos algoritmos para la detección de cambios abruptos en series de tiempo y comparar su funcionamiento. La detección tiene dos enfoques; el primero, teniendo una muestra fija inferir si ha ocurrido algún cambio de parámetros en la serie y el segundo, detectar si ocurre algún cambio en un proceso del que se sigue observando.

En la detección con una muestra fija se tiene que dar respuesta a dos problemas, decidir si hubo o no un cambio en la estructura de la serie y en caso de haberlo estimar lo mejor posible el punto en el tiempo en que ocurrió. En el segundo caso el principal problema es detectar lo más rápido posible si es que hubo un cambio y tener una probabilidad pequeña de no detectar un cambio cuando lo hubo, minimizando la probabilidad de detectarlo sin que haya ocurrido. Los algoritmos que se presentarán sirven para detectar cambios en ambos casos.

El primer algoritmo es llamado CUSUM (Cumulative Sums) y está basado en la razón de verosimilitudes y en el *Teorema Ergódico de Birkhoff-Khinchin*. Este algoritmo requiere que la serie siga un modelo paramétrico y conocer los parámetros de antes y después del cambio. Esto supone varios problemas. En primer lugar, en la práctica se pueden estimar los parámetros que sigue una serie, pero no tener su valor real, por otra parte, si no se sabe si hay o no cambio, ¿cómo se podrían estimar los parámetros a los que se cambia? Este algoritmo ha servido como base para la creación de nuevos algoritmos; además se ha demostrado que es el algoritmo más eficiente en

varios sentidos, por lo que sirve como referencia para comparar la eficiencia de otros. Una gran parte de este trabajo está dedicada a su demostración,

El segundo algoritmo es una solución al problema de no conocer los parámetros. Este algoritmo utiliza un estimador que tiende al máximo de la suma de Puentes Brownianos y otra gran parte de esta tesis está dedicada a su demostración. Este algoritmo, conocido como CSSM (Cumulative Sums Strong Mixing) sirve para aquellas series que cumplen con una condición conocida como α -mixing. La condición α -mixing no es demasiado restrictiva dado que la cumplen diversos tipos de modelos como lo son: ARIMA, GARCH, que son de los más utilizados para modelar series de tiempo, así como Procesos de Markov Ergódicos o procesos m-dependientes. Mediante la comparación con el CUSUM se discutirá su eficiencia.

Para las simulaciones se utilizó el software R-studio que es un entorno para programar en el lenguaje R. Se decidió usar este lenguaje ya que es un software libre y bastante utilizado para análisis estadístico. Se programaron ambos algoritmos (ver <https://gist.github.com/Eduardo-NR90>), se realizaron múltiples simulaciones de procesos tipo ARMA y GARCH y se les aplicaron las pruebas. Cabe destacar que no se utilizó ningún paquete base en particular.

El trabajo está organizado de la siguiente manera:

En el primer capítulo se dan todas las definiciones y conceptos necesarios para abordar el tema, la principal fuente bibliográfica es el libro *Detection of Abrupt Changes: Theory and Application* ([2] capítulos 2, 5 y 8). En los capítulos subsecuentes se presentan los algoritmos, así como los teoremas que justifican su uso. Se dedica un capítulo para cada algoritmo.

En el segundo capítulo se presenta el algoritmo CUSUM, que se puede encontrar en *Detection of Abrupt Changes: Theory and Application* (ver [2] Sección 8.2.1.2, pág. 306), posteriormente, para los teoremas y demostraciones las principales referencias son: el artículo *Procedures for Reacting to a Change in Distribution* [11] y el libro *Real Analysis and Probability* ([7] Cap. 3 y Teo. 8.4.1, pág. 268).

En el tercer capítulo se presentan los teoremas y sus respectivas demostraciones bajo los cuales está basado el CSSM y posteriormente se muestra el algoritmo. Las principales referencias para esta parte son: los artículos *The CUSUM Test for Detecting Structural Changes in Strong Mixing Processes* [1] y *A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables* [9].

Las demostraciones que se incluyen tanto en el segundo como en el tercer capítulo son las de los resultados más importantes y que en los libros y artículos consultados donde aparecen, citan a su vez otros artículos para su demostración.

En el último capítulo se presenta la aplicación de estos algoritmos sobre datos simulados y se realiza la comparación de su eficiencia. Una sección de conclusiones está incluida al final del trabajo.

Capítulo 1

Conceptos Básicos

El propósito de este capítulo es dejar claros los conceptos y definiciones que se usarán a lo largo del trabajo, así como presentar algunos resultados preliminares que serán de utilidad para los capítulos subsecuentes.

Un *Cambio Abrupto* hace referencia a un cambio en alguna característica del proceso, que ocurre manera muy rápida (el cambio no tiene que ser de gran magnitud). Existen diversos tipos de cambios que pueden sufrir los parámetros:

Cambios Aditivos se refieren a cualquier cambio en el sistema proveniente de cambios en la media de la secuencia.

Cambios no Aditivos son los cambios que ocurren en la varianza, correlación, características espectrales o dinámica del proceso. Estos cambios suponen un problema más grande que el de los cambios aditivos.

Los algoritmos que se presentan en este trabajo están diseñados para detectar los cambios aditivos.

Definición 1.0.1. Una *Serie de Tiempo* es un conjunto de observaciones y_t , cada una registrada en un tiempo específico t , tal que cada observación está separada por una unidad equidistante de tiempo.

Definición 1.0.2. Un *Tiempo de Paro* respecto a la sucesión de variables aleatorias X_1, X_2, \dots es una función $\tau : \Omega \rightarrow \mathbb{N}$ tal que:

$$\begin{aligned}\tau \text{ es } \mathcal{F}_1^\infty \text{ -medible,} \\ \{\tau = k\} \in \mathcal{F}_1^k,\end{aligned}$$

donde \mathcal{F}_n^m es la sigma álgebra generada por las variables X_n, X_{n+1}, \dots, X_m .

1.1. Pruebas de Hipótesis

Una *Hipótesis Estadística* \mathbf{H} es una afirmación o conjetura acerca de la distribución de una o mas variables aleatorias ([13] Def. 1, pág. 402).

Una *Prueba de Hipótesis Estadística* R es una regla o procedimiento mediante el cual, basándose en las observaciones, se decide si se rechaza la hipótesis \mathbf{H} ([13] , Def. 2, pág. 403). Los algoritmos que se describirán pretenden identificar si hubo algún cambio en el comportamiento de la serie, lo que finalmente es una prueba de hipótesis.

Sea R una Prueba de Hipótesis para probar \mathbf{H} definido como: Rechazar \mathbf{H}_0 si y solo si $(y_1, \dots, y_n) \in C_R$, donde $C_R \subset \mathbb{R}^n$ entonces C_R es llamada la *Región Crítica* de la Prueba R .

En las pruebas de hipótesis comúnmente se contrastan dos hipótesis: la hipótesis nula H_0 y la hipótesis alternativa H_1 .

Toda prueba de hipótesis esta sujeta a errores.

- Error tipo I: Rechazar \mathbf{H}_0 cuando \mathbf{H}_0 es cierta.
- Error tipo II: No rechazar \mathbf{H}_0 cuando \mathbf{H}_0 es falsa.

Estos errores tendrán cierta probabilidad de ocurrencia llamada *Tamaño de Error* (ver [13] pág. 405, Def. 5).

$$\alpha = \mathbb{P}(\text{error tipo I}) = \mathbb{P}((y_1, \dots, y_n) \in C_R | \mathbf{H}_0 \text{ cierta}). \quad (1.1)$$

$$\beta = \mathbb{P}(\text{error tipo II}) = \mathbb{P}((y_1, \dots, y_n) \notin C_R | \mathbf{H}_0 \text{ falsa}). \quad (1.2)$$

Es de esperar que si se intenta disminuir alguno de los tamaños de error, el otro aumente. La única forma para disminuir el tamaño de ambos errores simultáneamente es aumentar el tamaño de la muestra. Cuando se tienen solo dos hipótesis, por lo general se deja fijo el tamaño de error tipo I, α , y se intenta minimizar el tamaño de error tipo II, β .

1.2. Análisis Secuencial

El análisis secuencial es la teoría de resolver problemas de pruebas de hipótesis cuando el tamaño de la muestra no es fijo desde un principio, pero depende de los datos que ya han sido observados (ver [2] pág. 131), es decir, se realizan las pruebas cada vez que se tiene una nueva observación del proceso, lo que ocasiona que la muestra aumente.

Definición 1.2.1. Un *Test Estadístico Secuencial* para probar entre las hipótesis \mathbf{H}_0 y \mathbf{H}_1 se define como una pareja (R, T) , donde T es un tiempo de paro y $R(y_1, \dots, y_T)$ es una Prueba de Hipótesis.

Para saber el rendimiento de este tipo de test comúnmente se usa el siguiente índice como criterio.

Definición 1.2.2. El *ASN* (Average Sample Number) es la cantidad promedio de observaciones $E_\theta(T)$ que se necesitan para hacer una prueba de hipótesis con probabilidades aceptables de error de tipo I y II.

Prueba Secuencial entre dos hipótesis simples

En el caso de pruebas secuenciales entre dos hipótesis simples existen diversos resultados, entre ellos una prueba óptimo el cuál se muestra más adelante. Sin embargo, varios de estos resultados suponen hipótesis que en la práctica no se cumplen y los test que se adaptan a la realidad dejan de ser óptimos.

Definición 1.2.3. ([2] Def. 4.3.6, pág. 131) Sea (Y_n) una sucesión de variables aleatorias y sea $f_\theta(y_i)$ la función de densidad de Y_i , $i = 1, 2, \dots$, con θ el parámetro o vector de parámetros asociado a la distribución. El test (R, T) es un *Test de Razón de Probabilidades Secuencial (SPRT)* (Sequential Probability Ratio Test) para probar entre dos hipótesis simples $\mathbf{H}_0 = \{\theta : \theta = \theta_0\}$ vs $\mathbf{H}_1 = \{\theta : \theta = \theta_1\}$ si los datos $(y_n)_{n \geq 1}$ son observados periódicamente y si, a tiempo n , se puede tomar una de las siguientes decisiones:

- Rechazar \mathbf{H}_0 cuando $S_n \geq h$,
- Rechazar \mathbf{H}_1 cuando $S_n \leq -a$,
- Continuar el test y seguir observando cuando $-a < S_n < h$,

donde:

$$S_n = \ln \frac{f_{\theta_1}(y_1, \dots, y_n)}{f_{\theta_0}(y_1, \dots, y_n)},$$

$-a, h$ son barreras tales que $-\infty < -a < h < \infty$.

Esta definición puede ser reescrita como:

$$R(y_1, \dots, y_T) = \begin{cases} \mathbf{H}_1 & \text{cuando } S_T \geq h \\ \mathbf{H}_0 & \text{cuando } S_T \leq -a \end{cases}$$

donde T es el *tiempo de salida* definido como:

$$T = T_{-a,h} = \min\{n \geq 1 : (S_n \geq h) \cup (S_n \leq -a)\}.$$

1.3. Detección de cambios abruptos

A lo largo del trabajo el único supuesto general que se tendrá es que los datos provienen de la realización de una variable aleatoria (*v.a.*) cuya función de densidad f_θ depende de un número finito de parámetros θ , es decir, de una familia paramétrica. El problema que se desea resolver es detectar si hubo algún cambio en θ , que puede ser un parámetro o un vector de parámetros.

Definición 1.3.1. Sea (Y_n) una sucesión de variables aleatorias con función de densidad $f_{\theta_0}(y_i)$ para $1 \leq i \leq t_0 - 1$ y $f_{\theta_1}(y_i)$ para $t_0 \leq i$, $f_{\theta_0} \neq f_{\theta_1}$. El *Tiempo de cambio* t_0 es el momento donde la serie cambió de densidad o de parámetro.

Dependiendo qué interese, se tendrá que dar respuesta a alguno de los siguientes tres problemas:

- Detección en directo de un cambio.
- Detección con una muestra fija de algún cambio.
- Estimación con una muestra fija del tiempo de cambio.

Dada una serie de tiempo se considerará que antes del Tiempo de Cambio t_0 , el parámetro θ es igual a θ_0 . Después del cambio, $\theta = \theta_1$. El tiempo t_0 es, por supuesto, desconocido.

1.3.1. Detección en directo

El problema en la detección en directo radica en detectar el cambio lo más rápido posible conforme se tienen más observaciones de cierto proceso, teniendo un tamaño de error tipo I fijo, también llamado tasa de falsas alarmas.

En general, para este problema la detección se realiza mediante alguna regla considerando un tiempo de paro, la cuál comúnmente tiene la siguiente forma:

$$t_a = \inf\{n : g_n(y_1, \dots, y_n) \geq \lambda\}, \quad \lambda \in \mathbb{R}.$$

A t_a se le conoce como *Tiempo de Alarma*, es el instante en el que se detecta el cambio. Claramente g es una función dependiente de las observaciones que se tienen.

1.3.2. Detección con una muestra fija

Pruebas de Hipótesis

Dada una muestra finita y_1, y_2, \dots, y_N , $N \in \mathbb{N}$ de algún proceso interesa saber si dentro de esta muestra existe algún cambio. Para ello se puede realizar el siguiente tipo de prueba de hipótesis.

$$\mathbf{H}_0 : \text{ Para } 1 \leq k \leq N; f_{\theta}(y_k|y_1, \dots, y_N) = f_{\theta_0}(y_k|y_1, \dots, y_N).$$

vs

$$\mathbf{H}_1 : \text{ Existe } 1 \leq t_0 \leq N, \text{ desconocido tal que:}$$

$$\text{ Para } 1 \leq k \leq t_0 - 1; f_{\theta}(y_k|y_1, \dots, y_{k-1}) = f_{\theta_0}(y_k|y_1, \dots, y_{k-1}).$$

$$\text{ Para } t_0 \leq k \leq N; f_{\theta}(y_k|y_1, \dots, y_{k-1}) = f_{\theta_1}(y_k|y_1, \dots, y_{k-1}).$$

\mathbf{H}_0 implica que no hubo cambios, mientras que \mathbf{H}_1 quiere decir que existe un instante t_0 en donde ocurrió un cambio. Además, se requiere maximizar la probabilidad de decidir \mathbf{H}_1 cuando esta es realmente cierta, sujeto a una probabilidad fija de decidir \mathbf{H}_1 cuando \mathbf{H}_0 es verdadera.

Estimación del tiempo de cambio

Si se considera la misma prueba de hipótesis anterior, y se supone que se elige H_1 , es decir, que en la secuencia de observaciones $(y_k)_{1 \leq k \leq N}$, con densidad $f_\theta(y_k|y_1, \dots, y_{k-1})$ ocurre un cambio a tiempo desconocido t_0 , $1 \leq t_0 \leq N$, entonces se tiene que $\theta = \theta_0$ antes de t_0 y $\theta = \theta_1$ después del cambio. El tiempo t_0 tiene que ser estimado con la mayor precisión posible. Una forma de hacerlo es por máxima verosimilitud:

1. Si se supone que después del cambio la serie aún depende de las observaciones anteriores, se tiene:

$$(\hat{t}_0; \theta_0; \theta_1) = \arg \max_{2 \leq k \leq N} \{ \ln f_{\theta_0}(y_1, \dots, y_{k-1}) + \ln f_{\theta_1}(y_k, \dots, y_N | y_1, \dots, y_{k-1}) \}. \quad (1.3)$$

2. En el caso que las observaciones del cambio sean independientes de los datos pasados:

$$(\hat{t}_0; \theta_0; \theta_1) = \arg \max_{2 \leq k \leq N} \{ \ln f_{\theta_0}(y_1, \dots, y_{k-1}) + \ln f_{\theta_1}(y_k, \dots, y_N) \}. \quad (1.4)$$

Índices de eficiencia

Para poder evaluar el desempeño de los diferentes algoritmos se cuenta con los siguientes índices que también pueden ayudar para el diseño de los mismos (ver [2] Sección 4.4, pág. 151-153).

1. Tiempo medio entre falsas alarmas.
2. Probabilidad de detección falsa, α (1.1).
3. Tiempo medio de retraso para la detección.
4. Probabilidad de no detección del cambio, β (1.2).

5. Exactitud del tiempo de cambio.

Definición 1.3.2. El *Tiempo Medio entre Falsas Alarmas* se define como:

$$\bar{T} = \mathbf{E}_{\theta_0}(t_a),$$

que es la esperanza bajo θ_0 del tiempo que tarda el algoritmo en detectar un cambio.

Se desea que este tiempo tienda a infinito, es decir, que el número esperado de falsas alarmas se aproxime a 0.

Definición 1.3.3. *Retraso Medio Condicionado*, es el tiempo promedio que tarda el algoritmo en detectar un cambio real en θ . Se define como:

$$\mathbf{E}_{\theta_1}(t_a - t_0 + 1 | t_a \geq t_0, y_1 \dots y_{t_0-1})$$

Lo deseable es que este sea lo más pequeño posible.

Definición 1.3.4. *Peor Retraso Medio*, $\bar{\tau}$, es el tiempo promedio que tarda el algoritmo en detectar un cambio real en el parámetro θ cuando se supone que el conjunto de variables aleatorias hasta el momento del cambio es el que hace que el algoritmo se tarde más tiempo en detectar el cambio.

$$\bar{\tau} = \sup_{t_0 \geq 1} \text{ess sup } \mathbf{E}_{\theta_1}(t_a - t_0 + 1 | t_a \geq t_0, y_1 \dots y_{t_0-1}).$$

Se desea que el índice sea lo más pequeño posible.

El algoritmo CUSUM es el mejor algoritmo tomando en cuenta los índices recién descritos, aunque, como se muestra más adelante, es difícil de implementar en la práctica. Por esta razón el CUSUM es utilizado como referencia para estudiar la eficiencia de otros algoritmos.

Para conocer la eficiencia del algoritmo CSSM se compararon ciertos índices con los del CUSUM, en ambos casos los índices se calcularon mediante simulaciones.

Capítulo 2

El método de CUSUM

En este capítulo primero se describirá el algoritmo CUSUM y posteriormente se probarán algunos teoremas que demuestran porqué este algoritmo es el que tarda menos en la detección en directo. Las principales fuentes para este capítulo fueron: el libro *Detection of Abrupt Changes: Theory and Application* [2], donde se encuentra descrito el CUSUM, el libro *Real Analysis and Probability*, [7], en el cual se incluye un Teorema ergódico de suma importancia para poder demostrar que el CUSUM es el mejor algoritmo respecto al peor retraso medio, esta última prueba se localiza en el artículo *Procedures for Reacting to a Change in Distribution*, [11].

2.1. CUSUM

El algoritmo CUSUM (ver [2] Sección 8.2.1.2, pág. 306) está basado en la prueba del logaritmo de la razón de verosimilitudes y toma la idea de los SRPT, Definición 1.2.3 del capítulo anterior, tomando como barrera inferior $a = -\infty$ y como barrera superior $0 < h$. Esto es; sea

$$S_j^k = \ln \frac{f_{\theta_1}(y_j, \dots, y_k | y_1, \dots, y_{j-1})}{f_{\theta_0}(y_j, \dots, y_k | y_1, \dots, y_{j-1})}, \quad (2.1)$$

sea $g_0 = 0$,

$$g_k = (S_{k-N_k+1}^k)^+ = \max\{S_{k-N_k+1}^k, 0\}, \quad (2.2)$$

donde:

$$N_k = N_{k-1} \mathbb{1}_{\{g_{k-1} > 0\}} + 1.$$

Lo anterior implica que si el logaritmo de la razón de verosimilitudes S_j^k es negativo, se van descartando las observaciones, como si g_k se reiniciara cada vez que se vuelve negativa, es decir, g_k va acumulando solo cuando $0 < S_j^k$.

Dada $h > 0$, el tiempo de alarma para este algoritmo se define como:

$$t_a = \text{mín} \{k : g_k \geq h\}. \quad (2.3)$$

Si g_k supera cierta barrera $h > 0$, quiere decir que es más probable que las observaciones provengan de un proceso con parámetro θ_1 , por lo que se decide que hubo cambio. Para estimar el punto de cambio \hat{t}_0 se toma el tiempo en el que g_k fue igual a 0 por última vez. Esto, claro, para la detección en directo (cuando siguen “llegando observaciones”).

Para saber si hubo cambio cuando se cuenta con una muestra fija $y_1, y_2 \dots y_K$ se utiliza el siguiente estadístico:

$$g_K = \text{máx}_{1 \leq j \leq K} S_j^K,$$

si este rebasa la barrera h entonces se decide que hubo cambio y para saber donde ocurrió se cuenta con el estimador presentado en el capítulo anterior, Ecuación (1.3):

$$\hat{t}_0 = \arg \text{máx}_{2 \leq k \leq N-1} \{\ln f_{\theta_0}(y_1, \dots, y_{k-1}) + \ln f_{\theta_1}(y_k, \dots, y_N)\}.$$

Observación 2.1.1. Si se supone que las observaciones son independientes entonces el logaritmo de la razón de verosimilitudes se puede escribir como:

$$S_j^k = \ln \frac{f_{\theta_1}(y_j, \dots, y_k)}{f_{\theta_0}(y_j, \dots, y_k)} = \sum_{i=j}^k \ln \frac{f_{\theta_1}(y_i)}{f_{\theta_0}(y_i)},$$

y el tiempo estimado de cambio como:

$$\hat{t}_0 = \arg \text{máx}_{1 \leq j \leq K} \ln \left[\prod_{i=1}^{j-1} f_{\theta_0}(y_i) \prod_{i=j}^K f_{\theta_1}(y_i) \right].$$

Note que cuando se detecta cambio con una muestra fija, el punto estimado de cambio, \hat{t}_0 , es donde S_0^K alcanza su mínimo. En ambos casos el punto estimado de cambio es el mismo, esto se

puede observar en las siguientes gráficas, donde también se puede apreciar que el comportamiento de ambas gráficas es idéntico a partir de la última vez que g_k fue cero.

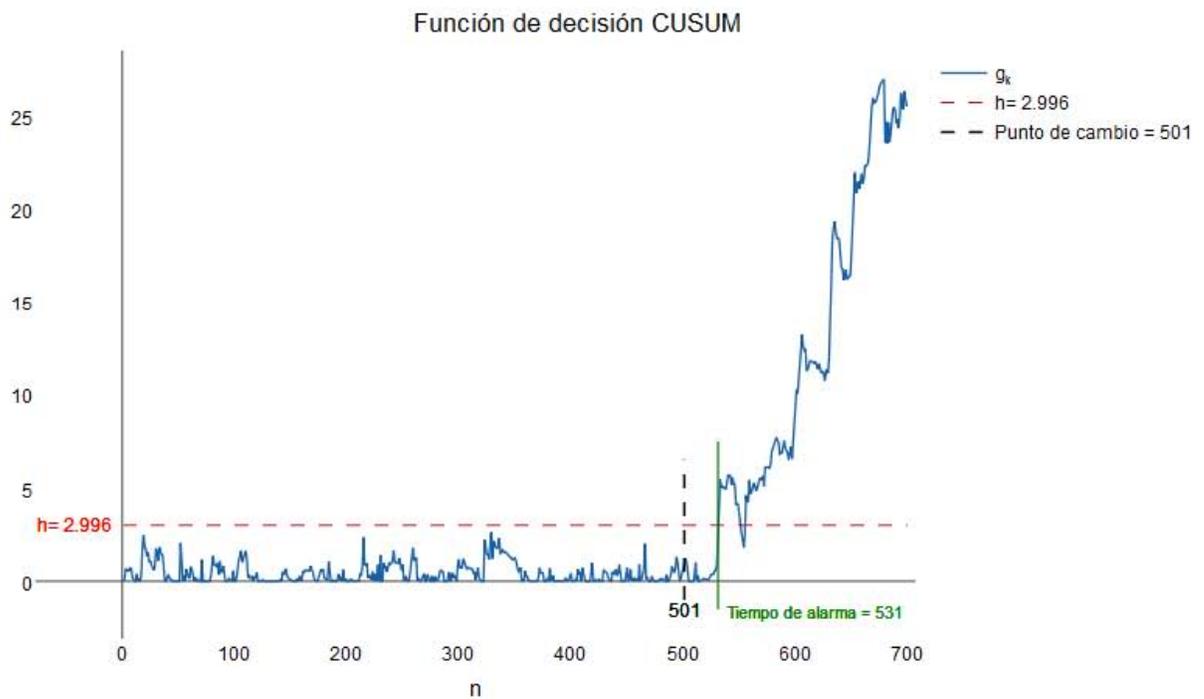


Figura 2.1: Comportamiento típico de la función de decisión g_k del CUSUM, tiempo de cambio 501.

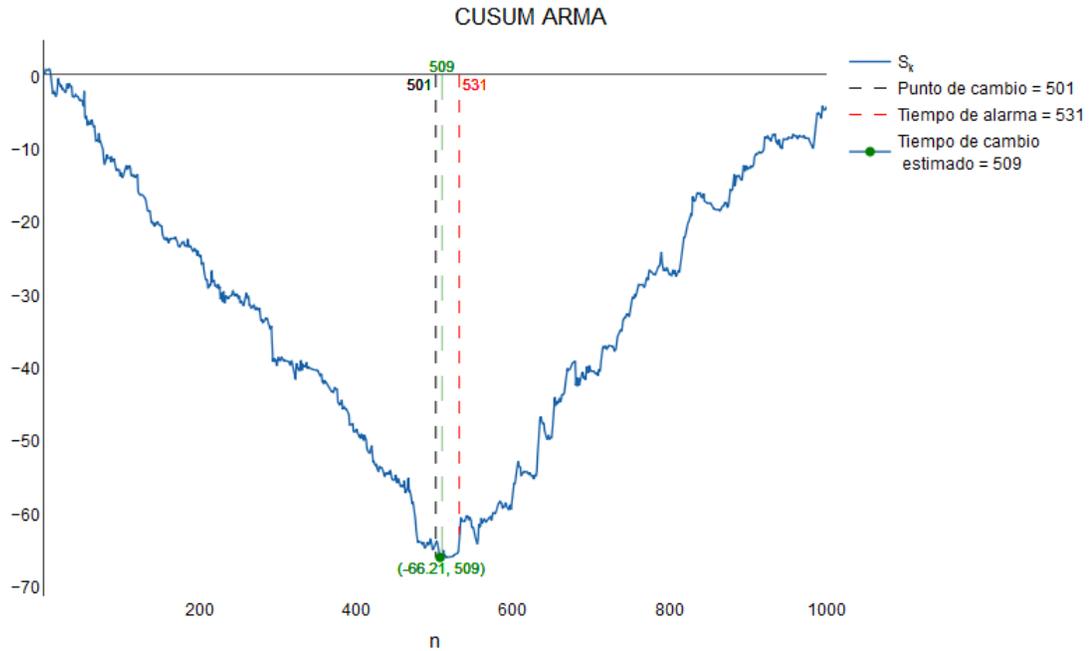


Figura 2.2: Comportamiento de la función S_t del CUSUM con tiempo de cambio 501.

Cabe resaltar que este algoritmo supone θ_0 y θ_1 conocidos, que se empieza con un proceso con parámetro θ_0 y se busca saber si este cambió y, en su caso, saber en qué momento lo hizo. Existen extensiones y variaciones de este algoritmo, para cuando, a priori, no se conocen dichos parámetros, estas extensiones se pueden encontrar en [2] Sección 7.3.1. Ya que el CUSUM es el mejor desde el punto de vista del peor retraso medio 1.3.4 para el caso de la detección en directo y también cuando se trata de una muestra fija ya que no es otra cosa que una prueba de razón de verosimilitudes, este algoritmo sirve como punto de comparación para medir el funcionamiento de otros algoritmos para la detección de cambios abruptos, en este caso con el CSSM.

2.2. Teoría ergódica

En esta sección se darán definiciones y teoremas que sirven para probar que el CUSUM es el mejor algoritmo para la detección de cambios abruptos respecto al peor retraso medio, esta prueba se encuentra en *Procedures for Reacting to a Change in Distribution*, [11] y aquí se presenta con una notación más sencilla.

Definición 2.2.1. Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de v.a.. Se define la σ -álgebra cola como:

$$\mathcal{C} = \bigcap \mathcal{F}_n^\infty = \lim_{n \rightarrow \infty} \mathcal{F}_n^\infty,$$

donde \mathcal{F}_n^∞ es la σ -álgebra generada por las variables X_n, X_{n+1}, \dots . Los eventos en esta σ -álgebra solo dependen de la “cola” de la sucesión (X_n) .

Si $C \in \mathcal{C}$ decimos que C es un evento cola, y si una variable es medible respecto a \mathcal{C} decimos que es una variable cola.

Definición 2.2.2. Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, la transformación $T : \Omega \mapsto \Omega$ es una transformación que preserva la medida si $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ para todo $A \in \mathcal{F}$.

Definición 2.2.3. Un conjunto Y es llamado T -invariante si $T^{-1}(Y) = Y$.

Observación 2.2.4. El conjunto $\mathcal{F}_{inv(T)}$ de todos los conjuntos T -invariantes es una σ -álgebra.

Definición 2.2.5. Una transformación $T : \Omega \mapsto \Omega$ que preserva la medida, es llamada *ergódica* si para cada $Y \in \mathcal{F}_{inv(T)}$, es tal que $\mathbb{P}(Y) = 0$ o $\mathbb{P}(Y^c) = 0$.

Definición 2.2.6. Una función T del espacio $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ en sí mismo es llamada *lineal* si:

$$T(cf + g) = cT(f) + T(g),$$

para cualquier $c \in \mathbb{R}$ y cualesquiera f y $g \in \mathcal{L}^1$. Y es llamada *positiva* si para cualquier $f \in \mathcal{L}^1$ y $f > 0$ se cumple: $T(f) > 0$.

Teorema 2.2.7. Sea f cualquier función medible de Ω en $[-\infty, \infty]$. Entonces

$$\int f d(\mu \circ T^{-1}) = \int f \circ T d\mu,$$

si cualquiera de las dos integrales está definida.

Demostración (La demostración de este teorema se encuentra en [7] Teo. 4.1.11, pág. 121).

Observación 2.2.8. Cualquier transformación que preserva la medida es lineal positiva, esto por el Teorema 2.2.7.

2.2.1. Teorema Ergódico de Birkhoff-Khinchin

El siguiente teorema es una generalización de la Ley de los Grandes Números y juega un papel muy importante para probar que el CUSUM es óptimo. La versión que aquí se presenta se puede encontrar en [7], Teorema 8.4.1.

Teorema 2.2.9. Birkhoff-Khinchin. ([7] Teo. 8.4.1, pág. 268)

Sea T una transformación que preserva la medida en $(\Omega, \mathcal{F}, \mu)$. Entonces para cualquier función $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ existe una función $\varphi \in \mathcal{L}^1(\Omega, \mathcal{F}_{inv(T)}, \mu)$ tal que $\lim_{n \rightarrow \infty} \frac{S_n(f(\omega))}{n} = \varphi(\omega)$ con $\int |\varphi| d\mu \leq \int |f| d\mu$. Si T es ergódica, entonces φ es una constante. Si $\mu(\Omega) < \infty$, $\frac{S_n(f)}{n}$ converge a φ en \mathcal{L}^1 así que $\int \varphi d\mu = \int f d\mu$, donde $S_n(f) = \sum_{j=0}^{n-1} f_j$ con $f_0 := f$, $f_j := f \circ T^j$, $j = 1, 2, \dots$

Por practicidad, en adelante se escribirá solo $S_n(f)$ en lugar de $S_n(f(\omega))$.

Para la demostración del teorema se encuentra en [7], pág.268-270. Aquí se presenta la misma demostración desarrollando utilizará el siguiente lema:

Lema 2.2.10. ([7] Lema 8.4.1, pág. 268) *Desigualdad Maximal.*

Sea $f \in \mathcal{L}^1$, se definen; $S_0(f) := 0$ y $S_n(f) := f + f \circ T + \dots + f \circ T^{n-1}$ para $n \geq 1$. Sean $S_n^+(f) := \max_{0 \leq j \leq n} S_j(f)$ y $A_n := \{\omega \in \Omega : S_n^+(f) > 0\}$. Entonces:

$$\int_{A_n} f d\mu \geq 0.$$

Demostración del Lema:

Debido a que T preserva la medida, por la Observación 2.2.8 T es lineal positiva por lo que se tiene:

$$f + S_r(f) \circ T = f + (f + f \circ T + \dots + f \circ T^{r-1}) \circ T = S_{r+1}(f),$$

con $r = 0, 1, \dots$. Además, para $j = 1, \dots, n$, se cumple:

$$f + S_{j-1}(f) \circ T \leq f + S_n^+ \circ T, \quad (2.4)$$

para todo $\omega \in A_n$. Uniendo las últimas dos ecuaciones se obtiene:

$$S_j(f) \leq f + S_n^+(f) \circ T,$$

para $j = 1, \dots, n$. Por lo tanto:

$$S_n^+(f) \leq f + S_n^+(f) \circ T,$$

lo que implica:

$$f \geq S_n^+(f) - S_n^+(f) \circ T.$$

Observación 2.2.11. Dado que T es lineal positiva, para $j = 1, \dots, n$ se cumple:

$$S_j(f) \circ T \leq S_n^*(f) \circ T.$$

Partiendo de la Ecuación (2.4), debido a que $S_n^+(f) \geq 0$ sobre Ω , $S_n^+(f) = 0$ fuera de A_n y por la observación anterior se tiene:

$$\begin{aligned} \int_{A_n} f \, d\mu &\geq \int_{A_n} S_n^+(f) - S_n^+(f) \circ T \, d\mu \\ &= \int_{\Omega} S_n^+(f) \, d\mu - \int_{A_n} S_n^+(f) \circ T \, d\mu \\ &\geq \int_{\Omega} S_n^+(f) \, d\mu - \int_{\Omega} S_n^+(f) \circ T \, d\mu \\ &= 0. \end{aligned}$$

La última igualdad se da porque T preserva la medida.

Corolario 2.2.12. ([17] Cor. 10.5.2, pág. 94) Sea $g \in \mathcal{L}^1$, $B_\alpha = \left\{ \omega \in \Omega : \sup_{n \geq 0} \frac{S_n(g)}{n} > \alpha \right\}$. Entonces para todo $E \in \mathcal{F}_{inv}$ se tiene que:

$$\int_{B_\alpha \cap E} g \, d\mu \geq \alpha \mu(B_\alpha \cap E). \quad (2.5)$$

Demostración del Corolario 2.2.12 (Ver [17], pág. 94):

Suponga que $E = \Omega$ y sea $f = g - \alpha$. Entonces:

$$B_\alpha = \bigcup_{n=1}^{\infty} \{\omega \in \Omega : S_n(g) > n\alpha\} = \bigcup_{n=1}^{\infty} \{\omega \in \Omega : S_n(f) > 0\} = \bigcup_{n=1}^{\infty} \{\omega : S_n^+(f) > 0\}.$$

Sea $A_n = \{\omega \in \Omega : S_n^+(f) > 0\}$, note que $A_n \subset A_{n+1}$, por lo que:

$$|f \mathbb{1}_{\omega \in A_n}| \leq |f \mathbb{1}_{\omega \in A_{n+1}}| \cdots \leq |f \mathbb{1}_{\omega \in B_\alpha}|,$$

además:

$$\lim_{n \rightarrow \infty} f \mathbb{1}_{\omega \in A_n} = f \mathbb{1}_{\omega \in B_\alpha},$$

entonces por el Teorema de Convergencia Dominada A.0.3 se tiene que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{A_n} f \, d\mu &= \lim_{n \rightarrow \infty} \int_{\Omega} f \mathbb{1}_{\omega \in A_n} \, d\mu \\ &= \int_{\Omega} f \mathbb{1}_{x \in B_\alpha} \, d\mu \\ &= \int_{B_\alpha \cap \Omega} f \, d\mu \\ &= \int_{B_\alpha \cap \Omega} g - \alpha \, d\mu \geq 0. \end{aligned}$$

Por lo tanto:

$$\int_{B_\alpha \cap E} g \, d\mu \geq \alpha \mu(B_\alpha \cap E).$$

Esto por la Desigualdad Maximal (2.2.10).

Ahora, para el caso general, si se restringe T a E , $T|_E : E \rightarrow E$ y se aplica la desigualdad maximal sobre este subconjunto se obtiene:

$$\int_{B_\alpha \cap E} g \, d\mu \geq \alpha \mu(B_\alpha \cap E),$$

porque E es T -invariante.

Demostración del Teorema Ergódico de B-K, 2.2.9 ([7], pág.268-270)

Sea $Y := Y(a, b) := \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_n}{n} < a < b < \limsup_{n \rightarrow \infty} \frac{S_n}{n} \right\}$, donde $a, b \in \mathbb{R}$, $a < b$.

Cambiando S_n por $S_n \circ T$ el conjunto Y se convierte en $T^{-1}(Y)$. Esto es:

$$T^{-1}(Y) = \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_n \circ T}{n} < a < b < \limsup_{n \rightarrow \infty} \frac{S_n \circ T}{n} \right\}.$$

Ahora note que $S_n \circ T = S_{n+1} - f$ y $\lim_{n \rightarrow \infty} \frac{f}{n} = 0$, así que se puede reemplazar $S_n \circ T$ por S_{n+1} .

$$T^{-1}(Y) = \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_{n+1}}{n} < a < b < \limsup_{n \rightarrow \infty} \frac{S_{n+1}}{n} \right\},$$

pero $\lim_{n \rightarrow \infty} \frac{S_{n+1}}{n} = \lim_{n \rightarrow \infty} \frac{S_{n+1}}{n+1}$. Entonces:

$$\begin{aligned} Y &= \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_{n+1}}{n+1} < a < b < \limsup_{n \rightarrow \infty} \frac{S_{n+1}}{n+1} \right\} \\ &= \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_n}{n} < a < b < \limsup_{n \rightarrow \infty} \frac{S_n}{n} \right\} \\ &= T^{-1}(Y). \end{aligned}$$

Por lo tanto Y es un conjunto T -invariante.

Sea $B := \left\{ \omega \in \Omega : \sup_{n \geq 1} \frac{S_n^+(f)(\omega)}{n} > b \right\}$. Por el Corolario 2.2.12, tomando a Y como el conjunto invariante, se tiene:

$$\int_{B \cap Y} f \, d\mu \geq b\mu(B \cap Y),$$

observe que $Y \subset B$. Entonces:

$$\int_Y f \, d\mu \geq b\mu(Y),$$

Por otro lado, si se considera $-f$ y $-a$, y se define $A := \left\{ \omega \in \Omega : \sup_{n \geq 1} \frac{S_n^+(-f)(\omega)}{n} > -a \right\}$ se obtiene:

$$-f \, d\mu \geq -a\mu(Y),$$

lo que implica que:

$$\int_Y f \, d\mu \leq a\mu(Y).$$

Juntando las desigualdades se obtiene,

$$b\mu(Y) \leq \int_Y f \, d\mu \leq a\mu(Y).$$

Por lo tanto:

$$\mu(Y) = 0.$$

Tomando todas la parejas $a, b \in \mathbb{Q}$ tal que $a < b$ se puede concluir que $\frac{S_n(f)}{n}$ converge casi en todas partes a alguna función φ con valores en $[-\infty, \infty]$. Como T preserva la medida se da la siguiente igualdad, para cada n .

$$\int |f \circ T^n| \, d\mu = \int |f| \circ T^n \, d\mu = \int |f| \, d\mu,$$

$$\int \left| \frac{S_n(f)}{n} \right| d\mu \leq \int n \left| \frac{f}{n} \right| d\mu = \int |f| d\mu,$$

como $\left| \frac{S_n(f)}{n} \right|$ es medible y no negativa, por el Lema de Fatou, A.0.2, se tiene:

$$\int \liminf_{n \rightarrow \infty} \left| \frac{S_n(f)}{n} \right| d\mu = \int |\varphi| d\mu \leq \liminf_{n \rightarrow \infty} \int \left| \frac{S_n(f)}{n} \right| d\mu \leq \int |f| d\mu,$$

entonces $\varphi \in \mathcal{L}^1$.

Además φ es T -Invariante, es decir, $\varphi = \varphi \circ T$, esto por la misma razón que Y es T -invariante. Así que para cualquier conjunto boreliano B , $\varphi^{-1}(B) = (T \circ \varphi)^{-1}(B) = T^{-1}(\varphi^{-1}(B))$, y $\varphi^{-1}(B) \in \mathcal{F}_{Inv(T)}$, por lo tanto φ es medible respecto de $\mathcal{F}_{Inv(T)}$.

Ahora se demostrará que si $\mu(X) < \infty$ entonces: $\int \varphi d\mu = \int f d\mu$.

Suponga que $|f| \leq k$ c.t.p entonces: $\frac{|S_n|}{n} \leq k$, para todo $n \in \mathbb{N}$. Por el Teorema de Convergencia Dominada se tiene que:

$$\lim_{n \rightarrow \infty} \int \left| \varphi - \frac{S_n}{n} \right| d\mu = \int \lim_{n \rightarrow \infty} \left| \varphi - \frac{S_n}{n} \right| d\mu = 0$$

Para el caso general defina $f_k := \max\{-k, \min\{f, k\}\}$, entonces cada f_k es acotada y $\lim_{k \rightarrow \infty} \int |f_k - f| d\mu = 0$.

Sean $\varepsilon > 0$ y $g = f_k$ para k suficientemente grande tal que $\int |f - g| d\mu < \frac{\varepsilon}{3}$. Entonces:

1. $\int \left| \frac{S_n(f) - S_n(g)}{n} \right| d\mu < \frac{\varepsilon}{3}$, para toda n .
2. $\frac{S_n(g)}{n}$ converge c.t.p. y en \mathcal{L}^1 a alguna función j (como en el caso no general). Entonces:

$$\int \left| \frac{S_n(g)}{n} - j \right| d\mu < \frac{\varepsilon}{3}.$$

para alguna n .

Por el Lema de Fatou:

$$\begin{aligned} \int \lim_{n \rightarrow \infty} \left| \frac{S_n(f) - S_n(g)}{n} \right| d\mu &= \int |\varphi - j| d\mu \\ &\leq \liminf_{n \rightarrow \infty} \int \left| \frac{S_n(f) - S_n(g)}{n} \right| d\mu \\ &\leq \sup_n \int \left| \frac{S_n(f) - S_n(g)}{n} \right| d\mu \\ &< \frac{\varepsilon}{3}, \end{aligned}$$

luego

$$\begin{aligned} \int \left| \frac{S_n(f)}{n} - \varphi \right| d\mu &\leq \int \left| \frac{S_n(f)}{n} - \frac{S_n(g)}{n} \right| + \left| \frac{S_n(g)}{n} - j \right| + |j - \varphi| d\mu \\ &= \int \left| \frac{S_n(f)}{n} - \frac{S_n(g)}{n} \right| d\mu + \int \left| \frac{S_n(g)}{n} - j \right| d\mu + \int |j - \varphi| d\mu \\ &= \varepsilon. \end{aligned}$$

Por lo tanto:

$$\int \varphi d\mu = \int f d\mu.$$

Por último solo falta ver que si T es ergódica entonces φ es una constante.

Sea $D := \{y \in \mathbb{R} : \mu(\{x : \varphi(x) > y\}) > 0\}$. Como T es ergódica, es decir cualquier conjunto T -invariante tiene medida 0 o el complemento de este conjunto tiene medida 0, y además φ es T -invariante, si $y \in D$, $\mu(\{x : \varphi(x) > y\}) > 0$ entonces la medida de su complemento tiene medida 0, esto es: $\mu(\{x : \varphi(x) \leq y\}) = 0$. Sea $C = \sup D$, entonces para $n = 1, 2, \dots$

$\mu(\{x : \varphi(x) > C + \frac{1}{n}\}) = 0$, es decir, $\varphi \leq C$ c.t.p.

Sea $y_n \in D$ tal que $y_n < C$ y $y_n \uparrow C$. Para cada $n \in \mathbb{N}$ se tiene:

$$\mu(\{x : \varphi(x) \leq y_n\}) = 0,$$

entonces $\varphi \geq y_n$ c.t.p para todo n , por lo tanto $\varphi \geq C$ c.t.p..

$$\therefore \varphi = C \text{ c.t.p.}$$

Que es lo que se deseaba demostrar.

Definición 2.2.13. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de v.a. respecto a $(\Omega, \mathcal{F}, \mathbb{P})$. A la transformación $T : \Omega \rightarrow \Omega$ tal que:

$$X_{n-1}(T(\omega)) = X_n(\omega).$$

se le conoce como Shift.

Teorema 2.2.14. *Sea T la transformación Shift, si $\{X_n\}$ es una sucesión de variables aleatorias independientes idénticamente distribuidas, T siempre es ergódica.*

La demostración de este teorema se encuentra en *Real Analysis and Probability*, [7] Teorema 8.4.5, pág. 271.

2.3. Optimalidad del algoritmo CUSUM

A continuación se presenta la demostración de que el CUSUM es óptimo respecto al peor retraso medio cuando el tiempo medio entre falsas alarmas tiende a infinito, esto que es muy conveniente, porque lo que se desea es una tasa baja de falsas alarmas y una detección lo más rápida posible, la demostración original se encuentra en [11].

Desigualdad y aproximación de Wald

Los siguientes resultados se pueden encontrar en [18] y serán utilizados en los teoremas subsecuentes, 2.3.2 y 2.3.3.

Teorema 2.3.1. *([18] Teo 2.1, pág. 16) Sean $-a < 0$ y $0 < h$ las barreras inferior y superior respectivamente de un SPRT con Probabilidades de Error $0 < \alpha$ y $0 \leq \beta$. Entonces:*

$$\ln \frac{\beta}{1-\alpha} \leq \text{mín}(0, -a),$$

$$\ln \frac{1-\beta}{\alpha} \geq \text{máx}(0, h).$$

Si se supone que la detección se da inmediatamente después de que la suma S_n tocó alguna de las barreras, entonces:

$$\ln \frac{\beta}{1-\alpha} \approx -a, \tag{2.6}$$

$$\ln \frac{1-\beta}{\alpha} \approx h, \tag{2.7}$$

y además:

$$\mathbf{E}_\theta(T) \approx \tilde{\mathbf{E}}_\theta(T),$$

donde:

$$\tilde{\mathbf{E}}_{\theta_0}(T) = \frac{(1 - \alpha) \ln \frac{1-\alpha}{\beta} - \alpha \ln \frac{1-\beta}{\alpha}}{-\mathbf{E}_{\theta_0}(s)}, \quad (2.8)$$

$$\tilde{\mathbf{E}}_{\theta_1}(T) = \frac{(1 - \beta) \ln \frac{1-\beta}{\alpha} - \beta \ln \frac{1-\alpha}{\beta}}{\mathbf{E}_{\theta_1}(s)}. \quad (2.9)$$

Teorema 2.3.2. Teorema de Lorden ([11] Teo 1, pág. 1899)

Sea $\{T(\alpha) : 0 < \alpha < 1\}$ el conjunto de los tiempos de paro referentes a los SPRT con barrera inferior $a = -\infty$ y barrera superior $h > 0$ tales que la probabilidad de detectar cambio cuando no ocurre en un tiempo finito es menor o igual a α , es decir:

$$\mathbb{P}_{\theta_0}(T(\alpha) < \infty) \leq \alpha,$$

Defina T_k de la siguiente manera:

$$T_k = \left\{ \begin{array}{l} \text{mín} \left\{ n \geq k : \sum_{i=k}^n \ln \frac{f_{\theta_1}(y_i)}{f_{\theta_0}(y_i)} \geq h \right\} \\ \infty \end{array} \right\} \quad \text{si no existe tal } n. \quad (2.10)$$

Sea T^* el tiempo de paro, definido por:

$$T^* = \text{mín}\{T_k : k = 1, 2, \dots\}$$

Entonces:

$$\alpha^{-1} \leq \mathbf{E}_{\theta_0}(T^*), \quad (2.11)$$

y para todo θ_1 , T^* minimiza el peor retraso medio, $\bar{\mathbf{E}}_{\theta_1}(\bar{T})$ de entre todos los tiempos de paro \bar{T} que satisfacen 2.11. Más aún:

$$\lim_{\alpha \rightarrow 0} \bar{\mathbf{E}}_{\theta_1}(T^*) \approx \frac{\ln \alpha^{-1}}{\mathbf{E}_{\theta_1}(s)}, \quad (2.12)$$

donde $s = \ln \frac{f_{\theta_1}(y)}{f_{\theta_0}(y)}$ y $\bar{\mathbf{E}}_{\theta_1}(T^*)$ es el Peor Retraso Medio de la Definición 1.3.4:

$$\bar{\mathbf{E}}_{\theta_1}(T^*) = \sup_{k \geq 1} \sup_{s \in \mathcal{S}} \mathbf{E}_k((T^* - k + 1)^+ | y_1, \dots, y_{k-1}).$$

El teorema anterior muestra lo óptimo del algoritmo CUSUM desde un punto de vista asintótico. En otras palabras el CUSUM es óptimo, respecto al peor retraso medio, cuando el tiempo promedio entre falsas alarmas tiende a infinito. La demostración de este teorema se encuentra más adelante ya que sigue directamente de los siguientes dos teoremas.

Teorema 2.3.3. ([11] Teo. 2, pág. 1900)

Sea Y_1, Y_2, \dots una sucesión de v. a. i. con función de densidad f_{θ_0} para Y_1, \dots, Y_m y f_{θ_1} de Y_m en adelante. Sea T tiempo de paro respecto a Y_1, \dots , i.e. T es \mathcal{F}_1^∞ -medible y el evento $\{T = n\} \in \mathcal{F}_1^n$, tal que:

$$\mathbb{P}_{\theta_0}(T < \infty) \leq \alpha. \quad (2.13)$$

Sea L la transformación Shift de la Definición 2.2.13, esto es:

$$Y_{n-1}(L(\omega)) = Y_n(\omega),$$

para todo $\omega \in \Omega$.

Defina \tilde{T}_k como el tiempo de paro T respecto a $Y_k, Y_{k+1} \dots$ es decir:

$$\tilde{T}_k(\omega) = T(L^{k-1}(\omega))$$

Sea $T_k = \tilde{T}_k + k - 1$, $k = 1, 2, \dots$

Defina ahora T^* como:

$$T^* = \min\{T_k : k = 1, 2, \dots\}.$$

Entonces T^* es tiempo de paro respecto a Y_1, \dots , y cumple:

$$\alpha^{-1} \leq \mathbf{E}_{\theta_0}(T^*), \quad (2.14)$$

$$\bar{\mathbf{E}}_{\theta_1}(T^*) \leq \mathbf{E}_{\theta_1}(T), \quad (2.15)$$

para alguna densidad f_{θ_1} , donde:

$$\bar{\mathbf{E}}_{\theta_1}(T^*) = \sup_{k \geq 1} \text{ess sup } \mathbf{E}_k((T^* - k + 1)^+ | y_1, \dots, y_{k-1}).$$

\mathbf{E}_k es la esperanza bajo la densidad generada cuando el tiempo de cambio es k .

Observación 2.3.4. \tilde{T}_k es \mathcal{F}_k^∞ -medible y el evento $\{\tilde{T}_k = n\} \in \mathcal{F}_k^{k+n-1}$.

Observación 2.3.5. Para toda k , $T_k \geq k$ y $T_k = k$ si y solo si $\tilde{T}_k = 1$.

Demostración

Por la Observación 2.3.4 y la Observación 2.3.5, el evento $\{T_k = k\} \in \mathcal{F}_k^{k+1-1} \subseteq \mathcal{F}_1^k$. Además ya que $\{T^* \leq n\} = \bigcup_{i=1}^n \{T_i \leq n\}$, se tiene que $\{T^* \leq n\} \in \mathcal{F}_1^n$. Por lo tanto T^* es tiempo de paro, respecto a Y_1, Y_2, \dots

Ahora se demostrará que $\mathbf{E}_{\theta_0}(T^*) \geq \alpha^{-1}$.

Ya que \mathbf{E}_{θ_0} es la esperanza bajo θ_0 , se tiene que Y_1, Y_2, \dots son *v.a.i.i.d* con densidad f_{θ_0} .

Defina

$$\xi_k = \begin{cases} 1 & \text{si } \tilde{T}_k < \infty \\ 0 & \text{si } \tilde{T}_k = \infty \end{cases}$$

Note que $\xi_k(\omega) = \xi_{k-1}(L(\omega))$ y que ξ_k es \mathcal{F}_k^∞ -medible, esto por cómo se definió \tilde{T}_k .

Dado que (Y_k) es una sucesión de *v.a.i.i.d.*, por el Teorema 2.2.14, la transformación shift L es ergódica, lo que implica que (ξ_k) cumple con el Teorema Ergódico 2.2.9, presentado en la sección anterior, por lo tanto:

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \xi_k}{n} = \mathbf{E}_{\theta_0}(\xi_1) \leq \alpha, \quad (2.16)$$

la última desigualdad se da porque

$$\mathbb{P}(\tilde{T}_1 < \infty) = \mathbb{P}(T_1 < \infty) = \mathbb{P}(T < \infty),$$

y porque por hipótesis $\mathbb{P}(T < \infty) \leq \alpha$.

Por otro lado defina de forma recursiva $T_0^* < T_2^* < T_3^* \dots$ de la siguiente manera:
 $T_0^* = 0$, para $m = 1, 2, \dots$, si $T_{m-1}^* = t$:

$$T_m^* = \min\{T_r : r = t + 1, t + 2, \dots\}.$$

Note que $(T_1^* - T_0^*), (T_2^* - T_1^*), \dots$ son *v.a.i.i.d.*

y que

$$1 \leq \xi_{T_m^*+1} + \xi_{T_m^*+2}, \dots, \xi_{T_{m+1}^*},$$

ya que tiene que existir algún $T_k < \infty$ para que la variable T_{m+1}^* exista. Por lo tanto:

$$m \leq \xi_1 + \xi_2, \dots, \xi_{T_m^*},$$

por lo tanto:

$$\lim_{m \rightarrow \infty} \frac{m}{T_m^*} \leq \lim_{m \rightarrow \infty} \frac{\xi_1 + \xi_2, \dots, \xi_{T_m^*}}{T_m^*}, \quad (2.17)$$

lo que implica que:

$$(T^*)^{-1} \leq \mathbf{E}_{\theta_0}(\xi_1)$$

luego por la Ecuación (2.16):

$$\alpha^{-1} \leq \mathbf{E}_{\theta_0}(T^*). \quad (2.18)$$

Falta demostrar que $\bar{\mathbf{E}}_{\theta_1}(T^*) \leq \mathbf{E}_{\theta_1}(T)$.

Para todo $m = 1, 2, \dots$ se tiene:

$$\begin{aligned} \mathbf{E}_m((T^* - m + 1)^+ | y_1, \dots, y_{(m-1)}) &\leq \mathbf{E}_m(\tilde{T}_m | y_1, \dots, y_{m-1}) \\ &= \mathbf{E}_m(\tilde{T}_m) \\ &= \mathbf{E}_{\theta_1}(\tilde{T}_1) \\ &= \mathbf{E}_{\theta_1}(T), \end{aligned}$$

en particular

$$\bar{\mathbf{E}}_{\theta_1}(T^*) = \sup_{k \geq 1} \text{ess sup } \mathbf{E}_k((T^* - k + 1)^+ | y_1, \dots, y_{k-1}) \leq \mathbf{E}_{\theta_1}(T).$$

Con esto concluye la demostración.

Teorema 2.3.6. *Sea $t(\gamma)$ el ínfimo de los peores retrasos medios $\bar{\mathbf{E}}_{\theta_1}(T)$ con T en el conjunto de los tiempos de paro extendido que satisfacen $\gamma \leq \mathbf{E}_{\theta_0}(T)$. Entonces, cuando γ tiende a infinito:*

$$t(\gamma) \approx \frac{\ln(\gamma)}{\mathbf{E}_{\theta_1}(s)}.$$

Demostración:

Ver [11] páginas 1901-1903.

Demostración del Teorema 2.3.2.

Por el Teorema 2.3.3

$$\begin{aligned}\alpha^{-1} &\leq \mathbf{E}_{\theta_0}(T^*), \\ \bar{\mathbf{E}}_{\theta_1}(T^*) &\leq \mathbf{E}_{\theta_1}(T)\end{aligned}$$

Por el teorema anterior, cuando α^{-1} tiende a infinito:

$$t \approx \frac{\ln \alpha^{-1}}{\mathbf{E}_{\theta_1}(s)}, \quad (2.19)$$

donde t es el ínfimo de los peores retrasos medios $\bar{\mathbf{E}}_{\theta_1}(T^*)$.

Por otro lado por la Aproximación de Wald (2.8), cuando T_k corresponde al de la Ecuación (2.10), se tiene que:

$$\mathbf{E}_{\theta_1}(T) \approx \frac{\ln \alpha^{-1}}{\mathbf{E}_{\theta_1}(s)},$$

ya que $\beta = 0$.

Por lo tanto:

$$\bar{\mathbf{E}}_{\theta_1}(T^*) \leq \frac{\ln \alpha^{-1}}{\mathbf{E}_{\theta_1}(s)},$$

por la Ecuación (2.19) y por la Aproximación de Wald (2.6), $\ln \alpha^{-1} \approx h$, cuando h tiende a infinito.

$$\bar{\mathbf{E}}_{\theta_1}(T^*) \approx \frac{\ln \alpha^{-1}}{\mathbf{E}_{\theta_1}(s)},$$

Observación 2.3.7. El teorema anterior prueba que el CUSUM es el mejor algoritmo para detectar cambios en directo ya que, con una barrera lo suficientemente grande, la tasa de falsas alarmas tiende a 0 y el peor retraso medio para la detección de un cambio es lo más pequeño posible.

Capítulo 3

Algoritmo CSSM

Para este capítulo se tomaron en cuenta principalmente dos artículos, *The CUSUM Test for Detecting Structural Changes in Strong Mixing Processes* [1] y *A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables* [9]. En el primero se presenta un algoritmo no paramétrico para detectar cambios abruptos estructurales en las series de tiempo, tanto en modelos lineales como no lineales por ejemplo *ARMA* o *GARCH*. Este algoritmo usa un estadístico de prueba que converge al máximo de la suma de cuadrados de Puentes Brownianos independientes. Para poder demostrar que tal estadístico converge, se usa un Teorema Funcional del Límite Central para cierto tipo de procesos, conocidos como Procesos Fuertemente Mezclantes o α –mezclantes (ver 3.1.3) el cual se encuentra en el artículo de Herrndorf *A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables* [9].

Definición 3.0.1. El proceso $W(t)$, $0 \leq t \leq T$, es un *Movimiento Browniano Estándar* si cumple las siguientes condiciones.

- $W(0) = 0$.
- $W(t)$ es continuo casi seguramente.
- $W(t)$ tiene incrementos independientes.
- $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ (para $0 \leq s \leq t$).

Definición 3.0.2. Sea $W(t)$, $0 \leq t \leq T$, un Movimiento Browniano Estándar, entonces el proceso estocástico $B(t)$ es un un *Puente Browniano* si:

$$B(t) = W(t) - \frac{t}{T}W(T).$$

3.1. Procesos fuertemente mezclantes (α – mixing)

Definición 3.1.1. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias. Se dice que $\{X_n\}_{n \in \mathbb{N}}$ es un *Proceso Estacionario* si

$$F(x_1, \dots, x_k) = F(x_{1+h}, \dots, x_{k+h}),$$

para todo $k, h \in \mathbb{N}$.

Definición 3.1.2. Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso estacionario. El *coeficiente α – mezclante de Rosenblatt* $\alpha_X(n, j)$ esta definido como:

$$\alpha_X(n, j) = \sup_{A, B} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

con $A \in \mathcal{F}_1^n$ y $B \in \mathcal{F}_{n+j}^\infty$.

Definición 3.1.3. Sea $\alpha_X(j) = \sup_n \alpha_X(n, j)$. Si $\{X_n\}_{n \in \mathbb{N}}$ es un proceso estacionario y la siguiente condición se cumple,

$$\lim_{j \rightarrow \infty} \alpha_X(j) = 0,$$

entonces $\{X_n\}_{n \in \mathbb{N}}$ es un *Proceso Fuertemente Mezclante (SM)*.

Los Procesos Fuertemente Mezclantes son procesos tales que entre más grande sea $|i - j|$, X_i y X_j se comportan “más” como variables independientes.

Lema 3.1.4. Desigualdad de Davydov ([20] Lema 1.2.4, pag. 10). Sea $\{X_n\}_{SM}$, sea ξ v.a. en L^p medible respecto a \mathcal{F}_1^n y $\eta \in L^q$ medible respecto a \mathcal{F}_{n+j}^∞ , con p y q tales que:

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1,$$

con $1 < r$. Entonces se cumple:

$$|\mathbf{Cov}(\xi, \eta)| \leq 10\alpha_X(j)^{\frac{1}{r}} \|\xi\|_p \|\eta\|_q.$$

Este lema es invocado en diferentes ocasiones en la demostración del siguiente teorema, la demostración de este se encuentra en [20] Lema 1.2.4, pag. 10.

La siguiente desigualdad se usa en la parte final de la demostración del siguiente teorema.

Lema 3.1.5. Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias, sea

$$\alpha = \sup_{1 \leq k \leq n-1} \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+1}^n \right\}.$$

Entonces para todo $\varepsilon > 0$:

$$\mathbb{P} \left(\max_{1 \leq r \leq n} \left| \sum_{i=1}^r X_i \right| > \varepsilon \right) \leq \frac{\mathbb{P} \left(\left| \sum_{i=1}^n X_i \right| > \frac{\varepsilon}{2} \right) + n\alpha}{\min_{1 \leq r \leq n-1} \mathbb{P} \left(\left| \sum_{i=r+1}^n X_i \right| \leq \frac{\varepsilon}{2} \right)}.$$

3.1.1. Teorema Funcional del Límite Central

En esta sección se presenta un teorema límite para procesos que son Fuertemente Mezclantes. El teorema será invocado más adelante para probar que el estadístico de prueba del CSSM converge al máximo de una suma de Puentes Brownianos independientes.

Definición 3.1.6. Sea $\{X_i\}_{i \in \mathbb{N}}$ una sucesión de variables aleatorias y sea $S_n = \sum_{i=1}^n X_i$. Considere el espacio $D = D[0, 1]$ de las funciones *cád-lág*, en el intervalo $[0, 1]$, con la topología de Skorokhod y defina las funciones aleatorias $W_n : \Omega \rightarrow D$ por:

$$W_n(\omega, t) = \frac{S(\omega)_{[nt]}}{\sigma\sqrt{n}}, \quad t \in [0, 1].$$

Por comodidad, en adelante se omitirá ω ,

$$W_n(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}}, \quad t \in [0, 1].$$

Teorema 3.1.7. Herrndorf 1968[9]. Sea $\beta \in (2, \infty)$ definimos $\gamma = \frac{2}{\beta}$. Sea $(a_n)_{n \in \mathbb{N}}$ una sucesión

en $[1, \infty]$. Y sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso tal que:

$$\begin{aligned}
 & i) \mathbf{E}(X_n) = 0 \text{ y } \mathbf{E}(X_n^2) < \infty, \text{ para todo } n \in \mathbb{N}. \\
 & ii) \text{ Sea } S_n = \sum_{i=1}^n X_i, \quad \lim_{n \rightarrow \infty} \mathbf{E} \left(\frac{S_n^2}{n} \right) = \sigma^2, \text{ para algún } \sigma \in \mathbb{R}. \\
 & iii) \sup_{m, n} \left\{ \frac{\mathbf{E}[(S_{m+n} - S_m)^2]}{n} \right\} < \infty. \\
 & iv) \lim_{n \rightarrow \infty} \sup_{i \leq n} \|X_i\|_{\beta}^2 \left(\sum_{i \geq a_n} \alpha_X(n, i)^{1-\gamma} + \frac{a_n^{2-\gamma}}{n^{1-\gamma}} \right) = 0.
 \end{aligned} \tag{3.1}$$

Entonces W_n converge en distribución a W , un Movimiento Browniano Estándar.

Demostración:

Defina la sucesión $(b_n)_{n \in \mathbb{N}}$ por:

$$\frac{b_n^{2-\gamma}}{n^{1-\gamma}} = \sum_{i \geq a_n} \alpha_X(n, i)^{1-\gamma} + \frac{a_n^{2-\gamma}}{n^{1-\gamma}}. \tag{3.2}$$

Despejando a_n de la ecuación anterior se puede notar que $a_n \leq b_n$ para toda $n \in \mathbb{N}$. Por 3.1 iv) y por 3.2, las siguientes condiciones se cumplen:

$$\begin{aligned}
 & a) \lim_{n \rightarrow \infty} \frac{b_n^{2-\gamma}}{n^{1-\gamma}} = 0. \\
 & b) \sum_{i \geq b_n} \alpha_X(n, i)^{1-\gamma} \leq \frac{b_n^{2-\gamma}}{n^{1-\gamma}}. \\
 & c) \sup_{i \leq n} \|X_i\|_{\beta}^2 = o \left(\frac{n^{1-\gamma}}{b_n^{2-\gamma}} \right).
 \end{aligned} \tag{3.3}$$

La Ecuación c) significa que existe $n_0 \in \mathbb{N}$ tal que:

$$\sup_{i \leq n} \|X_i\|_{\beta}^2 < C \frac{n^{1-\gamma}}{b_n^{2-\gamma}},$$

para toda $C > 0$ y para toda $n > n_0$. Por lo tanto existe $(r_n)_{n \in \mathbb{N}}$ una sucesión de enteros positivos tal que:

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} r_n = \infty. \\
 & r_n b_n = o(n). \\
 & \sup_{i \leq n} \|X_i\|_{\beta}^2 = o \left(\frac{n^{1-\gamma}}{(a_n b_n)^{2-\gamma}} \right).
 \end{aligned} \tag{3.4}$$

Sean:

$$q_n = 2 \inf\{i \in \mathbb{N} : i \geq b_n\}, \quad (3.5)$$

$$p_n = r_n b_n. \quad (3.6)$$

Las siguientes condiciones se cumplen:

$$\begin{aligned} \text{a)} \quad & \lim_{n \rightarrow \infty} \frac{q_n}{p_n} = 0. \\ \text{b)} \quad & \lim_{n \rightarrow \infty} \frac{p_n}{n} = 0. \\ \text{c)} \quad & \lim_{n \rightarrow \infty} \sum_{i=q_n} \alpha_X(n, i)^{1-\gamma} \sup_{i \leq n} \|X_i\|_\beta^2 = 0. \\ \text{d)} \quad & \lim_{n \rightarrow \infty} \frac{p_n^{2-\gamma}}{n^{1-\gamma}} \sup_{i \leq n} \|X_i\|_\beta^2 = 0. \\ \text{e)} \quad & \lim_{n \rightarrow \infty} \frac{n}{p_n} \alpha_X(n, q_n) = 0. \end{aligned} \quad (3.7)$$

Las ecuaciones 3.7 a) y 3.7 b) se siguen de la definición de p_n y q_n . Para 3.7 c) basta notar que $p_n > b_n$ y utilizar 3.3. La condición 3.7 d) se sigue de la definición de p_n y de la Ecuación (3.4).

Como $(\alpha_X(n, k))_{k \in \mathbb{N}}$ es no creciente:

$$\begin{aligned} \left(\frac{n}{p_n} \alpha_X(n, q_n) \right)^{1-\gamma} &= \frac{n^{1-\gamma}}{p_n^{1-\gamma}} \frac{1}{[b_n]} \sum_{i=1}^{[b_n]} \alpha_X(n, 2[b_n])^{1-\gamma} \\ &\leq \frac{n^{1-\gamma}}{p_n^{1-\gamma} b_n} \sum_{i=1}^{[b_n]} \alpha_X(n, 2[b_n] - i)^{1-\gamma} \\ &\leq \frac{n^{1-\gamma}}{p_n^{1-\gamma} b_n} \sum_{i \geq b_n} \alpha_X(n, i)^{1-\gamma} \\ &\leq \frac{b_n^{1-\gamma}}{p_n^{1-\gamma}}, \end{aligned}$$

la última desigualdad se da por 3.3 b), luego

$$\lim_{n \rightarrow \infty} \frac{b_n^{1-\gamma}}{p_n^{1-\gamma}} = \lim_{n \rightarrow \infty} \frac{b_n^{1-\gamma}}{(r_n b_n)^{1-\gamma}} = 0,$$

con esto se concluye 3.7 e).

El siguiente paso de la demostración será descomponer la suma S_n . Sea $k_n = \lfloor \frac{n}{p_n + q_n} \rfloor$, por comodidad, en adelante se escribirá $p = p_n$, $q = q_n$ y $k = k_n$ sin olvidar que estas siguen dependiendo de n .

Sean:

$$\begin{aligned}\xi_j &= \sum_{i=j(p+q)+1}^{j(p+q)+p} X_i, \\ \eta_j &= \sum_{i=j(p+q)+p+1}^{(j+1)(p+q)} X_i, \\ \eta_k &= \sum_{i=k(p+q)+p+1}^n X_i,\end{aligned}\quad \text{para } 0 \leq j \leq k-1.$$

defina también:

$$\begin{aligned}S'_n &= \sum_{j=0}^{k-1} \xi_j, \\ S''_n &= \sum_{j=0}^k \eta_j.\end{aligned}$$

Note que $S_n = S'_n + S''_n$ para toda $n \in \mathbb{N}$.

El Lema 3.1 del artículo de Withers, [19], establece que para demostrar que S_n cumple con el Teorema Central del Límite es suficiente probar:

$$\begin{aligned}\text{a) } & \lim_{n \rightarrow \infty} \frac{\mathbf{E}(S_n''^2)}{n} = 0. \\ \text{b) } & \lim_{n \rightarrow \infty} \frac{\sum_{0 \leq i < j \leq k} |\mathbf{E}(\xi_i \xi_j)|}{n} = 0. \\ \text{c) } & \lim_{n \rightarrow \infty} k = \infty. \\ \text{d) } & \lim_{n \rightarrow \infty} |\mathbf{E}(\exp^{iuS'_n}) - \prod_{j=0}^{k-1} \mathbf{E}(\exp^{iu\xi_j})| = 0. \\ \text{e) } & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{k-1} \mathbf{E}(\xi_j^2 1_{\{|\xi_j| > \varepsilon \sigma \sqrt{n}\}}) = 0 \quad \text{para todo } \varepsilon > 0.\end{aligned}\tag{3.8}$$

A continuación se mostrará que S_n cumple con estas condiciones.

Por 3.1 iii), $\mathbf{E}(\eta_j^2) < Cq$.

Utilizando la Desigualdad de Davydov 3.1.4 sobre X_i y X_j , se tiene:

$$\begin{aligned}
\mathbf{E}(S''_n)^2 &= \mathbf{E}\left(\sum_{j=0}^k \eta_j^2 + 2 \sum_{0 \leq i < j \leq k} \eta_i \eta_j\right) \\
&\leq C(kq + p + q) + 2 \sum_{0 \leq i < j \leq k} |\mathbf{E}(\eta_i \eta_j)| \\
&\leq C(kq + p + q) + 2 \sum_{i=1}^n \sum_{j=i+p+1}^n |\mathbf{E}(X_i X_j)| \\
&\leq C\left[\left(\frac{nq}{p+q}\right) + p + q\right] + 2(10) \sum_{i=1}^n \sum_{j=i+p+1}^n \alpha_X(n, j-i)^{1-\gamma} \sup_{t \leq n} \|X_t\|_\beta^2 \\
&\leq C\left[\left(\frac{nq}{p+q}\right) + p + q\right] + 20 \sum_{i=1}^n \sum_{j=1+p+1}^n \alpha_X(n, j)^{1-\gamma} \sup_{t \leq n} \|X_t\|_\beta^2 \\
&\leq C\left[\left(\frac{nq}{p+q}\right) + p + q\right] + 20n \sum_{j>p} \alpha_X(n, j)^{1-\gamma} \sup_{t \leq n} \|X_t\|_\beta^2,
\end{aligned}$$

dividiendo esto entre n y utilizando 3.7 se obtiene 3.8 a). La condición 3.8 b) se obtiene usando de manera análoga la desigualdad de Davydov y por 3.7 c):

$$\sum_{0 \leq i < j \leq k-1} |\mathbf{E}(\xi_i \xi_j)| \leq 20n \sum_{j>q} \alpha_X(n, j)^{1-\gamma} \sup_{t \leq n} \|X_t\|_\beta^2.$$

La condición 3.8 c) se sigue de la definición de k y de las condiciones 3.7 a), b). La condición d) se sigue de la siguiente desigualdad:

$$\begin{aligned}
|\mathbf{E}[e^{itS'_n}] - \prod_{j=0}^{k-1} \mathbf{E}[e^{it\xi_j}]| &\leq |\mathbf{E}[e^{it \sum_{j=0}^{k-1} \xi_j}] - \mathbf{E}[e^{it \sum_{j=0}^{k-2} \xi_j}] \mathbf{E}[e^{it\xi_{k-1}}]| \\
&\quad + |\mathbf{E}[e^{it \sum_{j=0}^{k-2} \xi_j}] \mathbf{E}[e^{it\xi_{k-1}}] - \mathbf{E}[e^{it \sum_{j=0}^{k-3} \xi_j}] \mathbf{E}[e^{it\xi_{k-1}}] \mathbf{E}[e^{it\xi_{k-1}}]| + \\
&\quad \dots + |\mathbf{E}[e^{it \sum_{j=0}^1 \xi_j}] \prod_{j=2}^{k-1} \mathbf{E}[e^{it\xi_j}] - \mathbf{E}[e^{it \sum_{j=0}^0 \xi_j}] \prod_{j=1}^{k-1} \mathbf{E}[e^{it\xi_j}]| \\
&\leq \sum_{r=1}^{k-1} |\mathbf{E}[e^{it \sum_{j=0}^r \xi_j}] - \mathbf{E}[e^{it \sum_{j=0}^{r-1} \xi_j}] \mathbf{E}[e^{it\xi_r}]|,
\end{aligned}$$

la última desigualdad se da porque $|e^{ir}| \leq 1$ para todo $r \in \mathbb{R}$. Para todo $s \in \mathbb{N}$; la variable $e^{it \sum_{j=0}^{s-1} \xi_j}$ es medible respecto a $\mathcal{F}_1^{sp+(s-1)q}$, mientras que la variable $e^{it\xi_s}$ es medible respecto a $\mathcal{F}_{sp+(s)q+1}^\infty$, por lo que utilizando la desigualdad de Davydov para *v. a.* complejas y 3.7 e) se tiene:

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \sum_{r=1}^{k-1} |\mathbf{E}[e^{it \sum_{j=0}^s \xi_j}] - \mathbf{E}[e^{it \sum_{j=0}^{s-1} \xi_j}] \mathbf{E}[e^{it \xi_s}]| \\
&= \lim_{n \rightarrow \infty} \sum_{r=1}^{k-1} |\mathbf{Cov}(e^{it \sum_{j=0}^{s-1} \xi_j}, e^{it \xi_s})| \\
&\leq \lim_{n \rightarrow \infty} 20k \alpha_X(n, q_n) = 0.
\end{aligned}$$

Para probar la condición $e)$ se utilizará el siguiente resultado. Sea Z una *v.a.* y sea $\lambda = \beta - 2$. Entonces:

$$\begin{aligned}
\int_{|z|>a} z^2 dF(z) &\leq \int_{|z|>a} |z|^2 \frac{|z|^\lambda}{a^\lambda} dF(z) \\
&= \int_{|z|>a} \frac{|z|^{2+\lambda}}{a^\lambda} dF(z) \\
&\leq \frac{\|Z\|_\beta^\beta}{a^{\beta-2}}.
\end{aligned}$$

Por lo tanto:

$$\begin{aligned}
\frac{1}{n} \sum_{j=0}^{k-1} \mathbf{E}(\xi_j^2 1_{\{|\xi_j| > \varepsilon \sigma \sqrt{n}\}}) &\leq \frac{(k) \sup_{0 \leq j \leq k-1} \|\xi_j\|_\beta^\beta}{n (\varepsilon \sigma \sqrt{n})^{\beta-2}} \\
&\leq \frac{(k)(p^\beta)}{n^{\frac{\beta}{2}} (\varepsilon \sigma)^{\beta-2}} \sup_{1 \leq j \leq n} \|X_j\|_\beta^\beta \\
&\leq (\varepsilon \sigma)^{2-\beta} (p^\beta) \left(\frac{n}{p}\right) (n^{-\frac{\beta}{2}}) \sup_{1 \leq j \leq n} \|X_j\|_\beta^\beta \\
&= (\varepsilon \sigma)^{2-\beta} (p^{2-\gamma} n^{\gamma-1}) \sup_{1 \leq j \leq n} \|X_j\|_\beta^{\frac{\beta}{2}},
\end{aligned}$$

que tiende a 0 cuando n tiende a infinito.

La demostración de Withers [19] utiliza las primeras 3 condiciones para probar que S_n'' es asintóticamente despreciable *i.e.* $\lim_{n \rightarrow \infty} |e^{it S_n} - e^{it S_n'}| = 0$. La condición $d)$ muestra que S' es una suma de variables aleatorias asintóticamente independientes y conjuntando esto con la condición $e)$ se cumplen las hipótesis del Teorema del Límite Central de Lindeberg, la prueba de este teorema se puede encontrar en *Central Limit Theorems and Proofs*. [15]. Por lo anterior se tiene que $W_n(t) = \frac{S_n}{\sigma \sqrt{n}}$ converge en distribución a $W(t)$ un Movimiento Browniano para todo $t \in [0, 1]$, es decir:

$$\lim_{n \rightarrow \infty} W_n(t) \stackrel{\mathcal{D}}{=} W(t).$$

Ahora se probará que dados $0 \leq t_1 < t_2 \cdots < t_k \leq 1$ la distribución a la que converge $(W_n(t_1), W_n(t_2), \dots, W_n(t_k))$ tiene incrementos independientes. Sea $r_n = \frac{2 + a_n}{n}$. Entonces:

$$\lim_{n \rightarrow \infty} \mathbf{E}[(W_n(t_i + r_n) - W(t_i))^2] = \lim_{n \rightarrow \infty} \mathbf{E} \left[\frac{\left(\sum_{i=1}^{[n(t_i + r_n)]} X_i - \sum_{i=1}^{[n(t_i)]} X_i \right)^2}{n\sigma^2} \right] \quad (3.9)$$

$$= \lim_{n \rightarrow \infty} \frac{[nr_n]}{n\sigma^2} \frac{\mathbf{E} \left[\left(\sum_{i=1}^{[n(t_i) + n(r_n)]} X_i - \sum_{i=1}^{[n(t_i)]} X_i \right)^2 \right]}{[nr_n]} \quad (3.10)$$

$$= 0, \quad (3.11)$$

porque $\lim_{n \rightarrow \infty} r_n = 0$ y por la hipótesis 3.1 iii), $\lim_{n \rightarrow \infty} \frac{\mathbf{E} \left[\left(\sum_{i=1}^{[n(t_i) + n(r_n)]} X_i - \sum_{i=1}^{[n(t_i)]} X_i \right)^2 \right]}{[nr_n]} < \infty$. Además debido la ecuación anterior y a que:

$$\lim_{n \rightarrow \infty} \alpha_X([nr_n]) = 0,$$

$$W_n(t_i) - W_n(t_{i-1} + r_n) = \frac{\sum_{j=[n(t_{i-1} + r_n)] + 1}^{[nt_i]} X_j}{\sigma\sqrt{n}} \text{ es } \mathcal{F}_1^{n[t_i]} - \text{medible},$$

$$W_n(t_{i+1}) - W_n(t_i + r_n) = \frac{\sum_{j=[n(t_i + r_n)] + 1}^{[nt_{i+1}]} X_j}{\sigma\sqrt{n}} \text{ es } \mathcal{F}_{[n(t_i + r_n)] + 1}^\infty - \text{medible},$$

se tiene que:

$$\lim_{n \rightarrow \infty} (W_n(t_2) - W_n(t_1 + r_n), \dots, W_n(t_k) - W_n(t_{k-1} + r_n)) \stackrel{\mathcal{D}}{=} (W(t_2) - W(t_1), \dots, W(t_k) - W(t_{k-1})),$$

y además para $1 \leq i \leq k$ las variables aleatorias $(W(t_i + 1) - W(t_i))$ son independientes.

Por lo tanto

$$\lim_{n \rightarrow \infty} (W_n(t_1), W_n(t_2), \dots, W_n(t_k)) \stackrel{\mathcal{D}}{=} (W(t_1), W(t_2), \dots, W(t_k)).$$

Para completar la demostración, falta probar una propiedad conocida como tensión. Sin esta propiedad podría pasar que el límite de la distribución sea una función que acumule toda la probabilidad en un punto o que nunca llegue a ser 1, en *Convergence of Probability Measures*, [4] hay diversos ejemplos de lo que podría pasar si no se cumple esta propiedad.

Por la caracterización clásica de Prohorov, que se puede encontrar en *Convergence of Probability Measures* [4], la tensión en $W_n(t)$ implica tensión en $(W_n(t_1), W_n(t_2), \dots, W_n(t_k))$, y por el Teorema 13.2 de *Convergence of Probability Measures* [4] para probar esta propiedad basta demostrar que:

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}(w(W_n, \delta) \geq \varepsilon) = 0, \quad (3.12)$$

donde $w(W_n, \delta) = \sup_{\{0 \leq t \leq 1 - \delta\}} \sup_{\{s, r \in [t, t + \delta]\}} |W_n(s) - W_n(r)|$. Sea $\varepsilon > 0$. Para $\delta > 0$ y $n \in \mathbb{N}$ fijas.

$$\begin{aligned} \mathbb{P}(w(W_n, \delta) > \varepsilon) &= \mathbb{P} \left(\sup_{\{0 \leq t \leq 1 - \delta\}} \sup_{\{s, r \in [t, t + \delta]\}} \frac{|S_{[sn]} - S_{[rn]}|}{\sigma \sqrt{n}} > \varepsilon \right) \\ &\leq \sum_{a=0}^{\lfloor 1/\delta \rfloor} \mathbb{P} \left(\max_{[nd\delta] < r \leq [n(a+1)\delta]} |S_r - S_{[na\delta]}| > \frac{\varepsilon}{3} \sigma \sqrt{n} \right). \end{aligned}$$

Sea $a \in \{0, 1, \dots, \lfloor 1/\delta \rfloor\}$ fija. Defina la sucesión m de la siguiente forma:

$$m = m(n) = \max \{i \in \mathbb{N} : (p+q)i \leq [n(a+1)\delta] - [na\delta]\}, \quad (3.13)$$

donde p y q son las sucesiones definidas anteriormente. Como $[n(a+1)\delta] - [na\delta] \approx [n\delta]$. Entonces:

$$m+1 \approx \frac{n\delta}{p+q}. \quad (3.14)$$

Para $j \in \{0, \dots, m-1\}$ se definen las siguientes variables:

$$\varphi_j = \sum_{i=[na\delta]+j(p+q)+1}^{[na\delta]+j(p+q)+p} X_i, \quad \psi_j = \sum_{i=[na\delta]+j(p+q)+p+1}^{[na\delta]+(j+1)(p+q)+p} X_i.$$

Note que cada variable contiene p y q variables respectivamente y que:

$$\sum_{j=0}^{m-1} \varphi_j + \psi_j \approx \sum_{i=[na\delta]+1}^{[n(a+1)\delta]} X_i = S_{[n(a+1)\delta]} - S_{[na\delta]}. \quad (3.15)$$

Entonces:

$$\begin{aligned} &\mathbb{P} \left(\max_{[nd\delta] < r \leq [n(a+1)\delta]} |S_r - S_{[na\delta]}| > \frac{\varepsilon}{3} \sigma \sqrt{n} \right) \\ &\leq \mathbb{P} \left(\sup_{0 \leq j \leq m-1} \max_{1 < r \leq p+q} |S_{[na\delta]+(j+1)(p+q)+r} - S_{[na\delta]+j(p+q)}| < \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left(\max_{0 \leq r \leq m-1} \left| \sum_{j=0}^r \varphi_j \right| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) + \mathbb{P} \left(\max_{0 \leq r \leq m-1} \left| \sum_{j=0}^r \psi_j \right| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \\
& \leq (m+1) \max_{0 \leq b \leq n-(p+q)} \mathbb{P} \left(\max_{1 < r \leq p+q} |S_{b+r} - S_b| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) + \mathbb{P} \left(\max_{0 \leq r \leq m-1} \left| \sum_{j=0}^r \varphi_j \right| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \\
& + \mathbb{P} \left(\max_{0 \leq r \leq m-1} \left| \sum_{j=0}^r \psi_j \right| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) = \mathbf{I} + \mathbf{II} + \mathbf{III}.
\end{aligned}$$

Sea $\beta \in (2, \infty)$, $\gamma = \frac{2}{\beta}$. Sea $b \in \{0, \dots, n - (p+q)\}$. Entonces:

$$\begin{aligned}
\mathbb{P} \left(\max_{1 < r \leq p+q} |S_{b+r} - S_b| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) & \leq \mathbb{P} \left(\sum_{i=1}^{p+q} |X_{b+i}| \geq \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \\
& \leq \mathbb{P} \left((p+q) \sup_{i \leq n} |X_i| \geq \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \\
& \leq \frac{\mathbf{E}[\sup_{i \leq n} |X_i|^\beta] (p+q)^\beta}{\left(\frac{\varepsilon \sigma \sqrt{n}}{9}\right)^\beta} \\
& = \left(\frac{\varepsilon \sigma^\beta}{9} n^{-\beta/2} (p+q)^\beta \sup_{1 \leq i} \|X_i\|_\beta \right)^\beta,
\end{aligned}$$

la última desigualdad se obtiene al aplicar la desigualdad de Markov A.0.4 sobre $\sup_{i \leq n} |X_i|$. Utilizando la Ecuación (3.14) y las ecuaciones 3.7 a) y b) note que:

$$\begin{aligned}
(m+1) n^{-\beta/2} (p+q)^\beta \sup_{i \leq n} \|X_i\|_\beta^\beta & \approx \delta (n^{\frac{-\beta}{2}+1}) (p+q)^{\beta-1} \sup_{i \leq n} \|X_i\|_\beta^\beta \\
& = \delta \left(n^{\frac{2}{\beta}-1} (p+q)^{2-\frac{2}{\beta}} \sup_{i \leq n} \|X_i\|_\beta^2 \right)^{\beta/2} \\
& \leq 2\delta \left(n^{\gamma-1} (p)^{2-\gamma} \sup_{i \leq n} \|X_i\|_\beta^2 \right)^{\beta/2},
\end{aligned}$$

que tiende a 0 cuando n tiende a infinito por 3.7 d). Por lo tanto:

$$(m+1) \max_{0 \leq b \leq n-(p+q)} \mathbb{P} \left(\max_{1 < r \leq p+q} |S_{b+r} - S_b| > \frac{\varepsilon}{9} \sigma \sqrt{n} \right) \leq \left(\frac{\varepsilon \sigma}{9}\right)^\beta 2\delta \left(n^{\gamma-1} (p)^{2-\gamma} \sup_{i \leq n} \|X_i\|_\beta^2 \right)^{\beta/2}.$$

Por lo tanto I tiende a cero conforme n tiende a infinito.

Para II y III note lo siguiente:

$$\begin{aligned} \max_{0 \leq r \leq m-1} \frac{\mathbf{E} \left[\left(\sum_{j=0}^r \varphi_j \right)^2 \right]}{n\sigma^2} &\leq \frac{1}{n\sigma^2} \sum_{j=0}^{m-1} \mathbf{E}[\varphi_j^2] + \frac{2}{n\sigma^2} \sum_{i=1}^n \sum_{j=i+q+1}^n |\mathbf{E}[X_i X_j]| \\ &\leq \frac{Cmp}{n\sigma^2} + \frac{24}{\sigma^2} \sum_{i>q} \alpha_X(n, i)^{i-\gamma} \sup_{i \leq n} \|X_i\|_\beta^2. \end{aligned}$$

Observación 3.1.8.

$$\frac{Cmp}{n\sigma^2} \approx \frac{Cp\delta}{\sigma^2(p+q)}.$$

$$\lim_{n \rightarrow \infty} \frac{Cp\delta}{\sigma^2(p+q)} = \frac{C\delta}{\sigma^2}.$$

$$\lim_{n \rightarrow \infty} \frac{24}{\sigma^2} \sum_{i>q} \alpha_X(n, i)^{i-\gamma} \sup_{i \leq n} \|X_i\|_\beta^2 = 0.$$

Entonces para n suficientemente grande:

$$\begin{aligned} \min_{0 \leq r \leq m-1} \mathbb{P} \left(\left| \sum_{i=r+1}^{m-1} \varphi_i \right| \leq \frac{\varepsilon}{18} \sigma n^{1/2} \right) &\geq 1 - \max_{0 \leq r \leq m-1} \mathbb{P} \left(\left| \sum_{i=r+1}^{m-1} \varphi_i \right| \geq \frac{\varepsilon}{18} \sigma n^{1/2} \right) \\ &\geq 1 - \max_{0 \leq r \leq m-1} \frac{\mathbf{E} \left[\left(\sum_{j=0}^r \varphi_j \right)^2 \right]}{\left(\frac{\varepsilon}{18} \right)^2 n\sigma^2} \\ &\geq 1 - \left(\frac{18}{\varepsilon} \right)^2 \frac{C\delta}{\sigma^2}. \end{aligned}$$

Sea $\delta_\varepsilon > 0$ tal que: $\left(\frac{18}{\varepsilon} \right)^2 C\sigma^{-2}\delta_\varepsilon < \frac{1}{2}$. Tome $\delta < \delta_\varepsilon$, entonces:

$$\min_{0 \leq r \leq m-1} \mathbb{P} \left(\left| \sum_{i=r+1}^{m-1} \varphi_i \right| \leq \frac{\varepsilon}{18} \sigma n^{1/2} \right) \geq \frac{1}{2}.$$

Por lo tanto, usando el Lema 3.1.5 se tiene:

$$\mathbb{P} \left(\max_{\leq r \leq m-1} \left| \sum_{j=0}^{m-1} \varphi_j \right| > \frac{\varepsilon}{9} \sigma n^{1/2} \right) \leq \frac{\mathbb{P} \left(\left| \sum_{j=0}^{m-1} \varphi_j \right| > \frac{\varepsilon}{18} \sigma n^{1/2} \right) + m\alpha_X(n, q+1)}{\min_{0 \leq l \leq n-2} \mathbb{P} \left(\left| \sum_{i=r+1}^{m-1} \varphi_i \right| \leq \frac{\varepsilon \sigma n^{1/2}}{18} \right)}$$

$$\leq 2 \left(\mathbb{P} \left(\left| \sum_{j=0}^{m-1} \varphi_j \right| > \frac{\varepsilon}{18} \sigma n^{1/2} \right) + m\alpha_X(n, q+1) \right).$$

Observación 3.1.9.

$$\lim_{n \rightarrow \infty} m\alpha_X(n, q+1) = 0 \text{ por 3.7.}$$

$$\lim_{n \rightarrow \infty} \left\| \sum_{j=0}^{m-1} \psi_j / \sigma n^{1/2} \right\|_2 = 0 \text{ por 3.8.}$$

$$\lim_{n \rightarrow \infty} \frac{\|S_{n[(a+1)\delta]} - S_{[na\delta]} - \sum_{j=0}^{m-1} (\varphi_j + \psi_j)\|_2}{\sigma \sqrt{n}} = 0 \text{ por 3.15.}$$

Por lo anterior:

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\max_{\leq r \leq m-1} \left| \sum_{j=0}^{m-1} \varphi_j \right| > \frac{\varepsilon}{9} \sigma n^{1/2} \right) \leq 2 \limsup_{n \rightarrow \infty} \mathbb{P} \left(|S_{n[(a+1)\delta]} - S_{[na\delta]}| > \frac{\varepsilon}{18} \sigma n^{1/2} \right).$$

Usando análogamente la Desigualdad de Tschevichev y el Lema 3.1.5 se obtiene:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\max_{\leq r \leq m-1} \left| \sum_{j=0}^{m-1} \psi_j \right| > \frac{\varepsilon}{9} \sigma n^{1/2} \right) &\leq 2 \limsup_{n \rightarrow \infty} \left(\mathbb{P} \left(\left| \sum_{j=0}^{m-1} \psi_j \right| > \frac{\varepsilon}{18} \sigma n^{1/2} \right) + m\alpha(p+1) \right) \\ &= 0. \end{aligned}$$

Por lo tanto:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(\omega(W_n, \delta) > \varepsilon) &\leq 2 \sum_{a=0}^{[1/\delta]} \limsup_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_{n[(a+1)\delta]} - S_{[na\delta]}}{\sigma \sqrt{n}} > \frac{\varepsilon}{18} \right) \\ &\leq 2 \left(\frac{1}{\delta} + 1 \right) \frac{2}{\sqrt{2\pi\delta}} \int_{\varepsilon/18}^{\infty} e^{-x^2/2\delta} dx. \end{aligned}$$

Note que la última parte de la desigualdad es la probabilidad de que una variable aleatoria con distribución normal $(0, \delta)$ sea mayor que $\varepsilon/9$ multiplicado por $4 \left(\frac{1}{\delta} + 1 \right)$ y si δ tiende a 0, la normal tiende a la constante 0 por lo tanto se cumple 3.12.

Con esto queda demostrado el Teorema Funcional del Límite Central de Herndorf el cual se utilizará más adelante, en la demostración del Teorema 3.2.5, para probar la convergencia del estadístico del CSSM a la suma de Puentes Brownianos.

3.2. Detección de cambios en procesos α -mezclantes

Definición 3.2.1. Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso estacionario, la *función de autocovarianzas* $\gamma(h)$ esta definida por:

$$\gamma(h) = \gamma(-h) = \mathbf{Cov}(X_t, X_{t+h}).$$

Definición 3.2.2. Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso estacionario con $\mathbf{E}(X_n) = 0$, para toda n , las *autocovarianzas muestrales* $\hat{\gamma}_n(h)$, $h = 0, \dots, n-1$, se definen como:

$$\hat{\gamma}_n(h) = \hat{\gamma}_n(-h) = \frac{1}{n} \sum_{i=1}^{n-h} X_i X_{i+h}.$$

Definición 3.2.3. Defina $C = [c_{hk}]_{h,k=1}^{L+1}$, $1 \leq L < n$, la *Matriz de Covarianzas de las autocovarianzas muestrales* $\hat{\gamma}_n(h)$, con:

$$c_{hk} = \lim_{n \rightarrow \infty} \theta_n(h, k), \quad (3.16)$$

$$\theta_n(h, k) = n \mathbf{Cov}(\hat{\gamma}_n(h), \hat{\gamma}_n(k)). \quad (3.17)$$

Lema 3.2.4. Sea $\{X_n\}_{n \in \mathbb{N}}$ un proceso fuertemente mezclante con $\mathbf{E}(X_n) = 0$, para toda n , tal que:

$$\sum_{k=1}^{\infty} \alpha_X(k)^{\delta/(2+\delta)}, \quad \text{para alguna } \delta \in (0, \infty),$$

$$\sup_i \mathbf{E}(|X_i|^{2+\delta}) = M < \infty,$$

sea $S_n = \sum_{i=1}^n X_i$. Entonces $\frac{\mathbf{E}(S_n)^2}{n}$ es convergente.

Demostración:

Por la Desigualdad de Davydov (3.1.4), se tiene:

$$\begin{aligned} \mathbf{E}(S_n)^2 &= \sum_{t=1}^n \sum_{s=1}^n \mathbf{E}(X_t X_s) \\ &\leq 10 \sum_{t=1}^n \sum_{s=1}^n \mathbf{E}(|X_t|^p)^{1/p} \mathbf{E}(|X_s|^q)^{1/q} \alpha_X(s-t)^{1-1/p-1/q}, \end{aligned}$$

sea $p = q = 2 + \delta$. Entonces:

$$\frac{1}{n} \mathbf{E}(S_n)^2 \leq \frac{1}{n} M^{2/(2+\delta)} \sum_{i=-n}^n n \alpha_X(i)^{\delta/(2+\delta)}$$

$$\begin{aligned} &\leq 10M \sum_{i=-\infty}^{\infty} \alpha_X(i)^{\delta/(2+\delta)} \\ &< \infty. \end{aligned}$$

Teorema 3.2.5. Sea $\{X\}_{n \in \mathbb{N}}$ un proceso estacionario tal que:

$$\begin{aligned} (i) & \sup_i \mathbf{E}(|X_i|^{4+2\delta}) < \infty. \\ (ii) & \sum_{k=1}^{\infty} \alpha_X(k)^{\frac{\delta}{2+\delta}} < \infty, \text{ para alguna } \delta \in (0, \infty). \end{aligned} \tag{3.18}$$

y sea

$$g_n^*(t) := \frac{[nt]}{\sqrt{n}} C^{-\frac{1}{2}} (\hat{\gamma}_{[nt]}(0) - \hat{\gamma}_n(0), \hat{\gamma}_{[nt]}(1) - \hat{\gamma}_n(1), \dots, \hat{\gamma}_{[nt]}(L) - \hat{\gamma}_n(L))^T, \quad 0 \leq t \leq 1.$$

Entonces:

$$\lim_{n \rightarrow \infty} g_n^*(t)^T \cdot g_n^*(t) \stackrel{\mathcal{D}}{=} \sum_{j=0}^L (W_j^0(\cdot))^2, \tag{3.19}$$

donde $W_j^0(\cdot)$ son Puentes Brownianos independientes.

Demostración:

Sea

$$g_n(t) := \frac{[nt]}{\sqrt{n}} C^{-1/2} (\hat{\gamma}_{[nt]}(0) - \gamma(0), \dots, \hat{\gamma}_{[nt]}(L) - \gamma(L))^T.$$

Note que:

$$g_n^*(t) = g_n(t) - \frac{[nt]}{n} g_n(1).$$

Observación 3.2.6. Si en la definición de Puente Browniano 3.0.2 se toma $T = 1$ entonces el Puente Browniano tiene la siguiente forma; $B(t) = W(t) - tW(1)$, donde $W(t)$ es un Movimiento Browniano Estándar, $0 \leq t \leq 1$. Por lo tanto si $\lim_{n \rightarrow \infty} g_n(t) \stackrel{\mathcal{D}}{=} (W_0(t), \dots, W_L(t))^T$, entonces:

$$\lim_{n \rightarrow \infty} g_n^*(t) \stackrel{\mathcal{D}}{=} (W_0^0(t), \dots, W_L^0(t))^T.$$

Se probará que $\lim_{n \rightarrow \infty} g_n(t) \stackrel{\mathcal{D}}{=} (W_0(t), \dots, W_L(t))^T$. Se puede escribir a g_n de la siguiente forma:

$$g_n(t) = \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^{[nt]} X_i X_i - [nt] \gamma(0) \\ \vdots \\ \sum_{i=1}^{[nt]-L} X_i X_{i+L} - [nt] \gamma(L) \end{pmatrix}.$$

Sea

$$Y_{m,t} = X_t X_{t+m} - \gamma(m).$$

Entonces:

$$\begin{aligned} g_n(t) &= \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^{[nt]} \{X_i X_i - \gamma(0)\} - 0\gamma(0) \\ \vdots \\ \sum_{i=1}^{[nt]-L} \{X_i X_{i+L} - \gamma(L)\} - L\gamma(L) \end{pmatrix} \\ &= \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^{[nt]} Y_{0,i} \\ \vdots \\ \sum_{i=1}^{[nt]-L} Y_{L,i} \end{pmatrix} - \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} 0\gamma(0) \\ \vdots \\ L\gamma(L) \end{pmatrix}. \end{aligned}$$

Observación 3.2.7. Note que:

$$\sigma(Y_{m,t} : t < n) = \sigma(X_t X_{t+m} : t < n) \subseteq \mathcal{F}_1^{n+m},$$

$$\sigma(Y_{m,t} : t > n+j) = \sigma(X_t X_{t+m} : t > n+j) \subseteq \mathcal{F}_{n+j}^\infty,$$

por lo tanto:

$$\alpha_{Y_m}(j) \leq \alpha_X(j-m).$$

Además:

$$\begin{aligned} \sup_i \mathbf{E}(|X_i X_{i+m}|^{2+\delta}) &\leq \sup_i \mathbf{E}(|X_{i+m} X_{i+m}|^{2+\delta}) \\ &= \sup_i \mathbf{E}(|X_i|^{4+2\delta}) < \infty. \end{aligned}$$

Por lo tanto:

$$\sup_i \mathbf{E}(|Y_{m,i}|^{2+\delta}) < \infty.$$

Observe también que:

$$\begin{aligned} \mathbf{E}(Y_{m,t}) &= \mathbf{E}(X_t X_{t+m} - \gamma(m)) \\ &= \mathbf{E}(X_t X_{t+m}) - \mathbf{Cov}(X_t, X_{t+m}) \\ &= \mathbf{E}(X_t X_{t+m}) - (\mathbf{E}(X_t X_{t+m}) - \mathbf{E}(X_t) \mathbf{E}(X_{t+m})) \\ &= 0. \end{aligned}$$

Con lo anterior se ha probado que $Y_{m,t}$ cumple con las hipótesis del Lema 3.2.4, por lo que:

$$\lim_{n \rightarrow \infty} \frac{\text{var}(S_{m,n})}{n} = \sigma_m^2, \quad \text{para alguna } \sigma_m^2 < \infty,$$

$$\text{donde } S_{m,n} = \sum_{i=1}^n Y_{m,i}.$$

Por el Teorema Funcional del Límite Central de Herrndorf 3.1.7:

$$\lim_{n \rightarrow \infty} \frac{S_{m,[nt]}}{\sigma_m \sqrt{n}} \stackrel{\mathcal{D}}{=} W(t), \quad \text{para } 0 \leq m \leq L,$$

donde $W(t)$ es un Movimiento Browniano Estándar.

Ahora note lo siguiente. Para $0 \leq h$ y $k \leq L$,

$$\begin{aligned} \mathbf{Cov} \left(\sum_{i=1}^{[nt]-h} Y_{h,i}, \sum_{i=1}^{[nt]-k} Y_{k,i} \right) &= \mathbf{Cov} \left(\sum_{i=1}^{[nt]-h} Y_{h,i} - h\gamma(h), \sum_{i=1}^{[nt]-k} Y_{k,i} - k\gamma(k) \right) \\ &= \mathbf{Cov} \left(\sum_{i=1}^{[nt]-h} \{X_i X_{i+h} - \gamma(h)\} - h\gamma(h), \sum_{i=1}^{[nt]-k} \{X_i X_{i+k} - \gamma(k)\} - k\gamma(k) \right) \\ &= \mathbf{Cov} \left(\sum_{i=1}^{[nt]-h} X_i X_{i+h} - [nt]\gamma(h), \sum_{i=1}^{[nt]-k} X_i X_{i+k} - [nt]\gamma(k) \right) \\ &= \mathbf{Cov} \left([nt] \left(\frac{\sum_{i=1}^{[nt]-h} X_i X_{i+h}}{[nt]} - \gamma(h) \right), [nt] \left(\frac{\sum_{i=1}^{[nt]-k} X_i X_{i+k}}{[nt]} - \gamma(k) \right) \right) \\ &= [nt]^2 \mathbf{Cov} (\hat{\gamma}_{[nt]}(h) - \gamma_{[nt]}(h), \hat{\gamma}_{[nt]}(k) - \gamma_{[nt]}(k)) \\ &= [nt]^2 \mathbf{Cov} (\hat{\gamma}_{[nt]}(h), \hat{\gamma}_{[nt]}(k)). \end{aligned}$$

Recuerde que:

$$c_{h,k} = \lim_{n \rightarrow \infty} [nt] \mathbf{Cov}(\hat{\gamma}_{[nt]}(h), \hat{\gamma}_{[nt]}(k)).$$

Por lo tanto

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^{[nt]} Y_{0,i} \\ \vdots \\ \sum_{i=1}^{[nt]-L} Y_{L,i} \end{pmatrix} - \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} 0\gamma(0) \\ \vdots \\ L\gamma(L) \end{pmatrix} &= \lim_{n \rightarrow \infty} \frac{C^{-1/2}}{\sqrt{n}} \begin{pmatrix} \sum_{i=1}^{[nt]} Y_{0,i} \\ \vdots \\ \sum_{i=1}^{[nt]-L} Y_{L,i} \end{pmatrix} \\ &\stackrel{\mathcal{D}}{=} \begin{pmatrix} W_0(t) \\ \vdots \\ W_L(t) \end{pmatrix}, \end{aligned}$$

donde $W_j(t)$, $j = 0, \dots, L$, son Movimientos Brownianos Estándar independientes. Lo que implica que:

$$\lim_{n \rightarrow \infty} g_n(t) \stackrel{\mathcal{D}}{=} \begin{pmatrix} W_0(t) \\ \vdots \\ W_L(t) \end{pmatrix}.$$

Entonces:

$$\begin{aligned} \lim_{n \rightarrow \infty} g_n^*(t) &= \lim_{n \rightarrow \infty} g_n(t) - \frac{[nt]}{n} g_n(1) \\ &\stackrel{\mathcal{D}}{=} \begin{pmatrix} W_0(t) - tW_0(1) \\ \vdots \\ W_L(t) - tW_L(1) \end{pmatrix}, \\ &= \begin{pmatrix} W_0^0(t) \\ \vdots \\ W_L^0(t) \end{pmatrix}, \end{aligned}$$

donde $W_j^0(t)$, $j = 0, \dots, L$, son Puentes Brownianos Estándar independientes. Por lo tanto

$$\begin{aligned} \lim_{n \rightarrow \infty} g_n^*(t)^T \cdot g_n^*(t) &\stackrel{\mathcal{D}}{=} (W_0^0(t), \dots, W_L^0(t)) \cdot \begin{pmatrix} W_0^0(t) \\ \vdots \\ W_L^0(t) \end{pmatrix} \\ &= \sum_{j=0}^L (W_j^0(t))^2, \end{aligned}$$

que es lo que se quería demostrar.

3.2.1. CSSM

Definición 3.2.8. Se define el proceso CSSM de la forma siguiente:

$$G(n, k) := g_n^*\left(\frac{k}{n}\right)^T \cdot g_n^*\left(\frac{k}{n}\right). \quad (3.20)$$

Y al estadístico de prueba T_n como:

$$T_n := \max_{L \leq k < n} G(n, k), \quad (3.21)$$

$$1 \leq L < n$$

Por el Teorema de Mapeo Continuo:

$$\lim_{n \rightarrow \infty} T_n \stackrel{\mathcal{D}}{=} \sup_{0 \leq t < 1} \sum_{j=0}^L (W_j^0(t))^2.$$

Para detectar cambios estructurales en una serie con una muestra de tamaño fijo n , basado en el proceso CSSM, en [1] se propone la siguiente prueba: Sea

H₀ : Para $1 \leq k \leq N$; $f_\theta(y_k|y_1, \dots, y_N) = f_{\theta_0}(y_k|y_1, \dots, y_N)$.

vs

H₁ : Existe $1 \leq t_0 \leq N$, desconocido tal que:

Para $1 \leq k \leq t_0 - 1$; $f_\theta(y_k|y_1, \dots, y_{k-1}) = f_{\theta_0}(y_k|y_1, \dots, y_{k-1})$.

Para $t_0 \leq k \leq N$; $f_\theta(y_k|y_1, \dots, y_{k-1}) = f_{\theta_1}(y_k|y_1, \dots, y_{k-1})$.

Se rechaza H_0 cuando T_n rebase cierta barrera $C > 0$, que depende del nivel de confianza que se desee, este valor C , fue obtenido para diferentes valores de L por Lee S. mediante simulaciones Montecarlo y los resultados se encuentran reportados en [14] página 784, tabla 1. y como estimador del punto de cambio t_0 se toma $\max_{L \leq k < n} G(n, k)$.

Tabla 3.1: ([14] página 784, tabla 1) Cuantiles $(1 - \alpha)$ empíricos de T_n , para $L = 1, \dots, 5$.

α	L				
	1	2	3	4	5
0.01	3.269	3.904	4.478	4.946	5.471
0.05	2.408	3.004	3.452	3.899	4.375
0.10	2.054	2.576	3.018	3.432	3.845

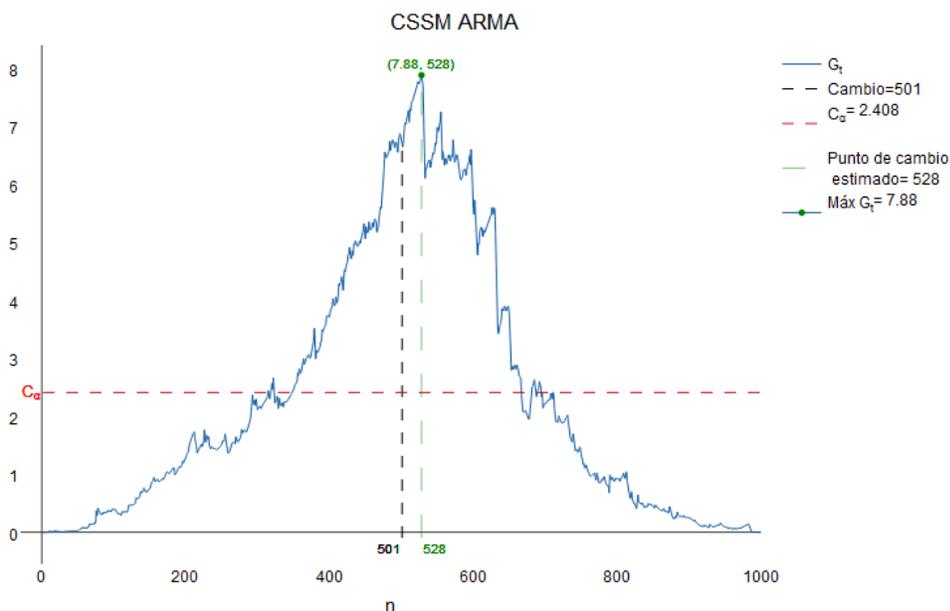


Figura 3.1: Comportamiento típico de la función $G(n, k)$ cuando el cambio se encuentra a la mitad del proceso.

Estimador consistente de la matriz de covarianzas

A continuación se presenta un estimador para $\theta_n(h, k)$.

$$\bar{\theta}_n(h, k) = \frac{1}{n} \sum_{l=0}^{h_n} \bar{\sigma}_{h,k}(l),$$

donde:

$$\bar{\sigma}_{h,k}(0) = \sum_{i=1}^{n-k} \left\{ \left(\frac{1}{n-k} \sum_{i=1}^{n-k} Y_{1i}^0 \right) - \hat{\gamma}_n(h) \hat{\gamma}_n(k) \right\},$$

$$\bar{\sigma}_{h,k}(l) = \sum_{i=1}^{n-k-l} \left\{ \left(\frac{1}{n-k-l} \sum_{i=1}^{n-k-l} \{Y_{1i}^l + Y_{2i}^l\} \right) - 2\hat{\gamma}_n(h) \hat{\gamma}_n(k) \right\}, \quad 1 \leq l < n,$$

con

$$Y_{1,t}^l = X_t X_{t+h} X_{t+l} X_{t+l+k},$$

$$Y_{2,t}^l = X_{t+l} X_{t+l+h} X_t X_{t+k},$$

y $h_n \in \mathbb{Z}^+$ tal que:

$$h_n = O(n^\beta) \quad \text{para algùn } \beta \in (0, 1/2). \quad (3.22)$$

Este estimador sirve para procesos que satisfagan las condiciones del Teorema 3.2.5.

Para comprobar que $\bar{\theta}$ es un estimador consistente, note primero lo siguiente:

Sin pérdida de generalidad se puede suponer $h < k$, ya que $\theta_n(h, k) = \theta_n(k, h)$. Defina:

$$\theta_{1,n}(h, k) = \frac{1}{n} \sum_{t=1}^{n-k} \sum_{s=1}^{n-k} \mathbf{E}[X_t X_{t+h} X_s X_{s+k}] - \gamma(h)\gamma(k),$$

$$\theta_{2,n}(h, k) = \frac{1}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \mathbf{E}[X_t X_{t+h} X_s X_{s+k}] - \gamma(h)\gamma(k),$$

entonces:

$$\theta_n(h, K) = \theta_{1,n}(h, k) + \theta_{2,n}(h, k).$$

Observación 3.2.9. $\theta_{1,n}$ puede ser reescrito de la siguiente forma:

$$\theta_{1,n}(h, k) = \frac{1}{n} \sum_{l=0}^{n-k-l} \sigma_{h,k}(l),$$

donde:

$$\begin{aligned} \sigma_{h,k}(0) &= \sum_{t=1}^{n-k} \left\{ \mathbf{E}[X_t X_{t+h} X_t X_{t+k}] - \gamma(h)\gamma(k) \right\}, \\ \sigma_{h,k}(l) &= \sum_{t=1}^{n-k-l} \left\{ \mathbf{E}[Y_{1t}^l] + \mathbf{E}[Y_{2t}^l] - 2\gamma(h)\gamma(k) \right\}, \quad 1 \leq l < n. \end{aligned}$$

Teorema 3.2.10. Sea $\{X\}_{n \in \mathbb{N}}$ un proceso estacionario tal que:

$$(i) \sup_i \mathbf{E}(|X_i|^{8+2\delta}) < \infty. \quad (3.23)$$

$$(ii) \sum_{k=1}^{\infty} \alpha_X(k)^{\frac{\delta}{2+\delta}} < \infty, \quad \text{para alguna } \delta \in (0, \infty). \quad (3.24)$$

Entonces:

$$\lim_{n \rightarrow \infty} \|\theta_{1,n}(h, k) - \bar{\theta}_n(h, k)\|_2 = 0.$$

Demostración

La demostración se hará en tres pasos.

I) Por la Observación 3.2.9 se tiene:

$$\begin{aligned}
|\sigma_{h,k}(l)| &= \left| \sum_{t=1}^{n-k-l} \left\{ \mathbf{E} \left(Y_{1,t}^l \right) + \mathbf{E} \left(Y_{2,t}^l \right) - 2\gamma(h)\gamma(k) \right\} \right| \\
&\leq \left| \sum_{t=1}^{n-k-l} \left\{ \mathbf{E} \left(X_t X_{t+h} X_{t+l} X_{t+l+k} \right) - \gamma(h)\gamma(k) \right\} \right| + \left| \sum_{t=1}^{n-k-l} \left\{ \mathbf{E} \left(X_{t+l} X_{t+l+h} X_t X_{t+k} \right) - \gamma(h)\gamma(k) \right\} \right| \\
&\leq \left| \sum_{t=1}^{n-k-l} \mathbf{Cov} \left(X_t X_{t+h}, X_{t+l} X_{t+l+k} \right) \right| + \left| \sum_{t=1}^{n-k-l} \mathbf{Cov} \left(X_{t+l} X_{t+l+h}, X_t X_{t+k} \right) \right| \\
&\leq \sum_{t=1}^{n-k-l} c_1 \left(\mathbf{E} \left(|X_t X_{t+h}| \right)^p \right)^{1/p} \left(\mathbf{E} \left(|X_{t+l} X_{t+l+k}| \right)^q \right)^{1/q} \left(\alpha_X(l-h) \right)^{1-\frac{1}{p}-\frac{1}{q}} \\
&\quad + \sum_{t=1}^{n-k-l} c_2 \left(\mathbf{E} \left(|X_{t+l} X_{t+l+h}| \right)^p \right)^{1/p} \left(\mathbf{E} \left(|X_t X_{t+k}| \right)^q \right)^{1/q} \left(\alpha_X(l+h) \right)^{1-\frac{1}{p}-\frac{1}{q}},
\end{aligned}$$

la última desigualdad se da por la desigualdad de Davydov 3.1.4. Haciendo $p = q = 2\delta$ y utilizando la hipótesis 3.23 i), $\sup_i \mathbf{E}(|X_i|^{8+2\delta}) < \infty$, se obtiene:

$$\begin{aligned}
&\sum_{t=1}^{n-k-l} c_1 \left(\mathbf{E} \left(|X_t X_{t+h}| \right)^p \right)^{1/p} \left(\mathbf{E} \left(|X_{t+l} X_{t+l+k}| \right)^q \right)^{1/q} \left(\alpha_X(l-h) \right)^{1-\frac{1}{p}-\frac{1}{q}} \\
&\quad + \sum_{t=1}^{n-k-l} c_2 \left(\mathbf{E} \left(|X_{t+l} X_{t+l+h}| \right)^p \right)^{1/p} \left(\mathbf{E} \left(|X_t X_{t+k}| \right)^q \right)^{1/q} \left(\alpha_X(l+h) \right)^{1-\frac{1}{p}-\frac{1}{q}} \\
&\leq M(n-k-l) \left(\alpha_X(l-h)^{\frac{\delta}{2+\delta}} + \alpha_X(l+h)^{\frac{\delta}{2+\delta}} \right),
\end{aligned}$$

por lo tanto, utilizando la hipótesis 3.18 ii), $\sum_{k=1}^{\infty} \alpha_X(k)^{\frac{\delta}{2+\delta}} < \infty$, se tiene:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^{n-k-l} |\sigma_{h,k}(l)| \leq \lim_{n \rightarrow \infty} \frac{M(n-k-l)}{n} \sum_{t=1}^{n-k-l} \left(\alpha_X(l-h)^{\frac{\delta}{2+\delta}} + \alpha_X(l+h)^{\frac{\delta}{2+\delta}} \right) < \infty.$$

II)

$$\|\sigma_{h,k}(l) - \bar{\sigma}_{h,k}(l)\|_2 = \left\| \sum_{t=1}^{n-k-l} \left\{ \mathbf{E} \left(Y_{1,t}^l \right) + \mathbf{E} \left(Y_{2,t}^l \right) - 2\gamma(h)\gamma(k) \right\} \right\|$$

$$\begin{aligned}
& - \sum_{t=1}^{n-k-l} \left\{ \left(\frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{1,i}^l \right) + \left(\frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{2,i}^l \right) - 2\hat{\gamma}_n(h)\hat{\gamma}_n(k) \right\} \Big\|_2 \\
& = \left\| \sum_{t=1}^{n-k-l} \left\{ \mathbf{E}(Y_{1,t}^l) + \mathbf{E}(Y_{2,t}^l) - 2\gamma(h)\gamma(k) \right. \right. \\
& \quad \left. \left. - \left(\frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{1,i}^l \right) - \left(\frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{2,i}^l \right) + 2\hat{\gamma}_n(h)\hat{\gamma}_n(k) \right\} \right\|_2 \\
& \leq \sum_{t=1}^{n-k-l} \left\| \left\{ \mathbf{E}(Y_{1,t}^l) - \frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{1,i}^l \right\} + \left\{ \mathbf{E}(Y_{2,t}^l) - \frac{1}{n-k-l} \sum_{i=1}^{n-k-l} Y_{2,i}^l \right\} \right. \\
& \quad \left. + \{2(\hat{\gamma}_n(h)\hat{\gamma}_n(k) - \gamma(h)\gamma(k))\} \right\|_2 \\
& \leq \sum_{i=1}^{n-k-l} a_1 + a_2 + 2a_3,
\end{aligned}$$

donde:

$$\begin{aligned}
a_1 &= \left\| \frac{1}{n-k-l} \sum_{i=1}^{n-k-l} \left\{ Y_{1,t}^l - \mathbf{E}(Y_{1,t}^l) \right\} \right\|_2, \\
a_2 &= \left\| \frac{1}{n-k-l} \sum_{i=1}^{n-k-l} \left\{ Y_{2,t}^l - \mathbf{E}(Y_{2,t}^l) \right\} \right\|_2, \\
a_3 &= \|\hat{\gamma}_n(h)\hat{\gamma}_n(k) - \gamma(h)\gamma(k)\|_2.
\end{aligned}$$

Por el Lema 3.2.4 y dado que $\sup_t \mathbf{E}(|Y_t|^{2+\delta/4}) < \infty$, se tiene:

$$a_1^2 = \frac{1}{(n-k-l)^2} \mathbf{E} \left[\left(\sum_{i=1}^{n-k-l} \left\{ Y_{1,t}^l - \mathbf{E}(Y_{1,t}^l) \right\} \right)^2 \right] = O(n^{-1}).$$

Por lo tanto $a_1 = O(n^{-1/2})$. Análogamente $a_2 = O(n^{-1/2})$. Ahora note que $a_3 = O(n^{-1/2})$, ya que:

$$\begin{aligned}
a_3 &= \|\hat{\gamma}_n(h)\hat{\gamma}_n(k) - \gamma(h)\gamma(k)\|_2 = \|\hat{\gamma}_n(h)\hat{\gamma}_n(k) - \hat{\gamma}_n(h)\gamma(k) + \hat{\gamma}_n(h)\gamma(k) - \gamma(h)\gamma(k)\|_2 \\
&\leq \|\hat{\gamma}_n(h)\|_2 \|\hat{\gamma}_n(k) - \gamma(k)\|_2 + \|\hat{\gamma}_n(h)\gamma(h)\|_2 |\gamma_n(k)|.
\end{aligned}$$

Por el Lema 3.2.4,

$$\|\hat{\gamma}_n(h)\|_2^2 = \left\| \frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+h} \right\|_2^2 = \frac{1}{n^2} \mathbf{E} \left[\left(\sum_{t=1}^{n-h} X_t X_{t+h} \right)^2 \right] < \infty,$$

y

$$\|\hat{\gamma}_n(k) - \gamma(k)\|_2^2 = \left\| \frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+k} - \gamma(k) \right\|_2^2 \leq \frac{1}{n^2} \mathbf{E} \left[\left(\sum_{t=1}^{n-h} X_t X_{t+k} - \gamma(k) \right)^2 \right] = O(n^{-1}).$$

Por las últimas desigualdades se tiene que $a_3 = O(n^{-1/2})$. Usando la misma lógica, es fácil verificar que $\|\bar{\sigma}_{h,k}(0) - \sigma_{h,k}(0)\|_2 = O(n^{-1/2})$. Por lo tanto,

$$\|\bar{\sigma}_{h,k}(l) - \sigma_{h,k}(l)\|_2 = O(n^{-1/2}). \quad (3.25)$$

Para el último paso se usarán los dos pasos anteriores además de la desigualdad de Minkowski:

$$\begin{aligned} \|\theta_{1,n}(h,k) - \bar{\theta}_n(h,k)\|_2 &= \left\| \frac{1}{n} \sum_{l=1}^{n-k-1} \sigma_{h,k}(l) - \frac{1}{n} \sum_{l=1}^{h_n} \bar{\sigma}_{h,k}(l) \right\|_2 \\ &\leq \frac{1}{n} \sum_{l=1}^{h_n} \|\sigma_{h,k}(l) - \bar{\sigma}_{h,k}(l)\|_2 + \frac{1}{n} \sum_{l=h_n+1}^{n-k-1} |\sigma_{h,k}(l)|. \end{aligned}$$

Luego por las ecuaciones (3.22) y (3.25)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=1}^{h_n} \|\sigma_{h,k}(l) - \bar{\sigma}_{h,k}(l)\|_2 \leq \frac{1}{n} h_n O(n^{1/2}) = 0.$$

Como $\lim_{n \rightarrow \infty} h_n = \infty$:

$$\lim_{n \rightarrow \infty} \sum_{h_n+1 < l < n-k-1} \frac{1}{n} |\sigma_{h,k}(l)| = 0.$$

Por lo tanto

$$\lim_{n \rightarrow \infty} \|\theta_{1,n}(h,k) - \bar{\theta}_n(h,k)\|_2 = 0.$$

La demostración de este teorema se deja en el apéndice.

Teorema 3.2.11. *Bajo las mismas hipótesis del Teorema 3.2.10 se tiene que:*

$$\lim_{n \rightarrow \infty} \|\theta_n(h,k) - \bar{\theta}_n(h,k)\|_2 = 0.$$

Demostración:

Como

$$\begin{aligned} \|\theta_n(h, k) - \bar{\theta}_n(h, k)\|_2 &= \|\theta_{1,n}(h, k) + \theta_{2,n}(h, k) - \bar{\theta}_n(h, k)\|_2 \\ &\leq \|\theta_{1,n}(h, k) - \bar{\theta}_n(h, k)\|_2 + \|\theta_{2,n}(h, k)\|_2, \end{aligned}$$

por el teorema anterior, la primera parte tiende a 0 cuando n tiende a infinito. Para la segunda parte usando la desigualdad de Davydov 3.1.4, se tiene que:

$$\begin{aligned} \|\theta_{2,n}(h, k)\| &= \left\| \frac{1}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \mathbf{E}[X_t X_{t+h} X_s X_{s+k}] - \gamma(h)\gamma(k) \right\| \\ &= \left\| \frac{1}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \mathbf{E}[X_t X_{t+h} X_s X_{s+k}] - \mathbf{Cov}(X_t, X_{t+h}) \mathbf{Cov}(X_s, X_{s+k}) \right\| \\ &= \left\| \frac{1}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \mathbf{E}[X_t X_{t+h} X_s X_{s+k}] - \mathbf{E}[X_t X_{t+h}] \mathbf{E}[X_s X_{s+k}] \right\| \\ &= \left\| \frac{1}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \mathbf{Cov}(X_t X_{t+h}, X_s X_{s+k}) \right\| \\ &\leq \frac{10}{n} \sum_{t=n-k+1}^{n-h} \sum_{s=1}^{n-k} \|X_t X_{t+h}\|_p \|X_s X_{s+k}\|_q \alpha_X(t+h-s)^{1-\frac{1}{p}-\frac{1}{q}} \\ &\leq \frac{k-h}{n} c \|X_t X_{t+h}\|_p \|X_s X_{s+k}\|_q \sum_{r=1}^{\infty} \alpha_X(r)^{1-\frac{1}{p}-\frac{1}{q}}, \end{aligned}$$

entonces, por los supuestos del Teorema 3.2.5 y haciendo $p = q = 2 + \delta$ se tiene que:

$$\lim_{n \rightarrow \infty} \|\theta_{2,n}(h, k)\|_2 = 0.$$

Corolario 3.2.12. Sea $\hat{C} = [\hat{c}_{hk}]_{(L+1) \times (L+1)}$, donde $\hat{c}_{hk} = \bar{\theta}_n(h, k)$. Entonces bajo los supuestos del teorema anterior, \hat{C} es un estimador consistente de la matriz de covarianzas C .

Corolario 3.2.13. El estadístico CSSM.

Bajo las hipótesis del Teorema 3.2.11, sea

$$\hat{T}_n = \max_{1 \leq k \leq n} \hat{g}^*_{*n}(k/n)^T \cdot \hat{g}^*_{*n}(k/n), \quad (3.26)$$

donde

$$\hat{g}^*_{*n}(s) := \frac{[ns]}{\sqrt{n}} \hat{C}^{-1/2} (\hat{\gamma}_{[ns]}(0) - \hat{\gamma}_n(0), \dots, \hat{\gamma}_{[ns]}(L) - \hat{\gamma}_n(L))^T, \quad (3.27)$$

con $\hat{C} = [\hat{c}_{hk}]_{(L+1) \times (L+1)}$, donde $\hat{c} = \bar{\theta}_n(h, h)$.

Capítulo 4

Simulaciones y Resultados

En este capítulo se muestra el funcionamiento de los algoritmos CUSUM y CSSM para la detección de cambios abruptos en las series de tiempo descritos en los capítulos previos. Primero se compara el rendimiento de cada uno de ellos ante un proceso $ARMA(1, 1)$ y un $ARMA(2, 3)$, posteriormente se muestran los resultados del CSSM cuando se tiene un proceso $GARCH$. El CUSUM solo se aplica en el primer proceso ya que en este caso es relativamente fácil calcular la función de log verosimilitudes, sin embargo, a pesar de que son simulaciones y se cuenta con los valores reales del vector de parámetros tanto antes como después del cambio, no es sencillo obtener las funciones de verosimilitud de los procesos $GARCH$.

Se realizan diferentes niveles de cambio en los parámetros, esto con el fin de captar qué tan grande tiene que ser el cambio para poder ser detectado, también se movió el punto de cambio para saber si este afectaba o no en la respuesta del algoritmo. Se realizaron 1000 simulaciones por cada cambio. Todas las simulaciones, así como la programación de los algoritmos, se realizaron en R y se pueden encontrar en los anexos B.1 y B.2 o en el siguiente repositorio:

<https://gist.github.com/Eduardo-NR90>

Antes de presentar los resultados se da una definición formal de los procesos a los que se les aplica el algoritmo y se describe como fueron implementados.

4.1. Comparación de los algoritmos *CSSM* y *CUSUM* en un proceso *ARMA*(1, 1)

Definición 4.1.1. Una Serie de Tiempo $(\varepsilon_t)_{t \in \mathbb{Z}}$ es un *Ruido Blanco* si para todo $t \in \mathbb{Z}$:

1. $\mathbf{E}(\varepsilon_t) = 0$.
2. $\mathbf{Var}(\varepsilon_t) = \sigma^2$.
3. $\mathbf{Corr}(\varepsilon_i, \varepsilon_j) = 0$ para todo $i \neq j$.

El proceso *ARMA*(p, q), con el cual se probarán los algoritmos, está definido de la siguiente forma:

Definición 4.1.2. Una Serie de Tiempo $(Y_t)_{t \in \mathbb{Z}}$ es un proceso *Autorregresivo de Promedios Móviles ARMA*(p, q), por sus siglas en inglés, si para todo $t \in \mathbb{Z}$:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

con $(\varepsilon_t)_{t \in \mathbb{Z}}$ un Ruido Blanco.

Primero se muestran los resultados para el *CUSUM*, tanto con una muestra fija como la detección en directo. Posteriormente se presentan los resultados del *CSSM*.

Utilización del *CUSUM*

Para la utilización de este algoritmo es necesario fijar el tamaño de error tipo I, la probabilidad de optar por cambio cuando no lo hubo, en este caso: $\alpha = .05$. Posteriormente, utilizando la Ecuación (2.6), $h = \ln \frac{1-\beta}{\alpha}$, se obtiene una barrera para la función de decisión, en este caso: $h = -\ln(.05) = 2.995732$.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1, 1) 63

El proceso *ARMA* se puede ver en términos del Ruido Blanco de la siguiente manera: Suponiendo que el Ruido Blanco es una sucesión de *v.a.i.i.d.* normales $(0, \sigma^2)$. Sea $(Y_k)_{k \in \mathbb{N}}$ un proceso *ARMA*(p, q), entonces:

$$\begin{aligned} f_{\theta}(y_k | y_{k-1}, \dots, y_1) &= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (y_k - \sum_{i=1}^p \phi_i y_{k-i} - \sum_{j=1}^q \theta_j \varepsilon_{k-j})^2} \\ &= \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} \varepsilon_k^2}, \end{aligned}$$

así que los incrementos de la razón de log verosimilitudes s_k tienen la siguiente forma:

$$s_k = \frac{1}{2} \ln \left(\frac{\sigma_0^2}{\sigma_1^2} \right) + \frac{(\varepsilon_k^0)^2}{2\sigma_0^2} - \frac{(\varepsilon_k^1)^2}{2\sigma_1^2},$$

por lo que la función S_k del CUSUM tiene la expresión explícita:

$$S_k = \sum_{i=1}^k \frac{1}{2} \left(\ln \left(\frac{\sigma_0^2}{\sigma_1^2} \right) + (\varepsilon_k^0)^2 + (\varepsilon_k^1)^2 \right).$$

En el caso del algoritmo CUSUM, para saber si hubo algún cambio en la serie con todas las observaciones, se toma k igual al total de observaciones que se tengan, mientras que para la detección en directo, k va creciendo. En ambos casos la función de decisión g_k cumple que:

$$g_k = \max_{1 \leq j \leq k} S_j^k,$$

y se decide que hubo cambio cuando $g_k > h = 2.996$. Para el caso en directo se decide que hubo cambio cuando $g_K > h = 2.996$.

4.1.1. Resultados CUSUM ante un ARMA(1,1)

Se realizaron simulaciones de un proceso *ARMA*(1, 1) con 1000 observaciones y cambio en los parámetros (ϕ, θ) al tiempo $n = 501$. Los parámetros iniciales del proceso $(\phi_0 = 0.1, \theta_0 = 0.2)$, ambos parámetros se movieron a diferentes valores (ϕ_1, θ_1) , para cada cambio se realizaron 1000 simulaciones.

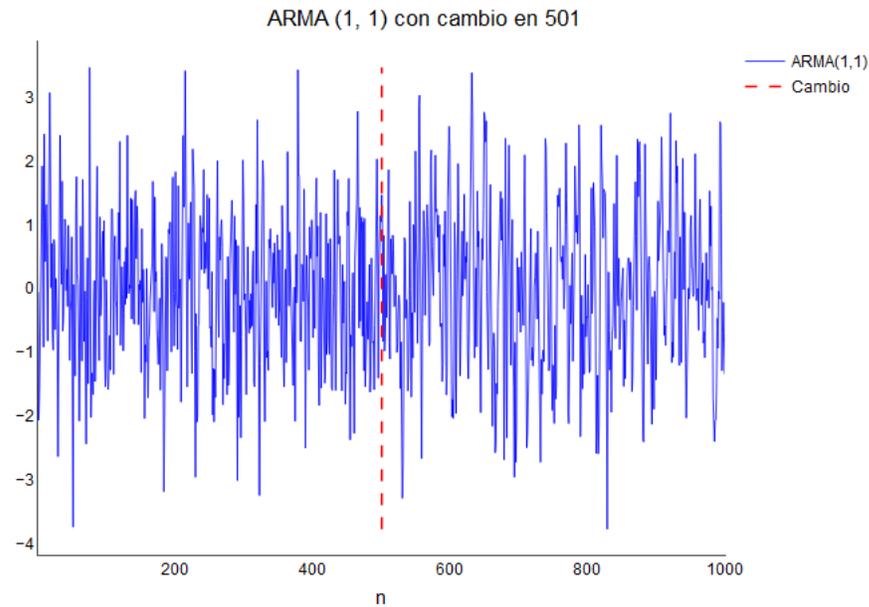


Figura 4.1: Serie simulada con 1000 observaciones de un $ARMA(1,1)$ con parámetros iniciales $\phi_0 = .1$ y $\theta_0 = .2$ tiempo de cambio 501 y parámetros después de cambio $\phi_1 = .2$ y $\theta_1 = .35$.

La siguiente tabla muestra la efectividad que tuvo el algoritmo *CUSUM* en las simulaciones descritas, entendiendo por efectividad el porcentaje de veces que se detectó el cambio.

Tabla 4.1: Porcentaje de detección CUSUM $ARMA(1,1)$ con parámetros iniciales ($\phi_0 = .1, \theta_0 = .2$).

$\theta_1 \backslash \phi_1$	0.1	0.15	0.2	0.25	0.3	0.35
0.2		5.60	56.10	87.50	97.70	99.80
0.25	9.10	60.40	89.20	96.50	99.90	99.90
0.3	65.20	90.30	98.20	99.40	100.00	100.00
0.35	91.00	97.00	100.00	99.80	100.00	100.00
0.4	98.40	99.90	100.00	100.00	100.00	100.00
0.45	99.80	99.80	100.00	100.00	100.00	100.00

Es claro que entre más grande sea el cambio el algoritmo tiene una tasa de detección mayor. Lo anterior implica que la barrera $h = \ln(.05) = 2.995732$ funciona bien cuando los cambios no son muy pequeños.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1)⁶⁵

Para conocer el error que tiene el algoritmo cuando no hay cambios, se simularon 1000 series con 1000 observaciones. Recuerde que el *CUSUM* necesita los parámetros antes y después del cambio. La siguiente tabla muestra el porcentaje de veces que el *CUSUM* detectó cambio sin que este existiera. En este caso ϕ_2 y θ_2 corresponden solo a los parámetros que recibe el algoritmo esperando un cambio.

Tabla 4.2: Porcentaje de falsos positivos del CUSUM con una muestra fija de 1000 observaciones de un ARMA(1,1) con parámetros ($\phi_0 = .1, \theta_0 = .2$).

$\theta_2 \backslash \phi_2$	0.1	0.15	0.2	0.25	0.3	0.35
0.2		1.20	3.10	3.60	2.90	2.60
0.25	1.60	3.60	3.40	2.70	2.70	2.30
0.3	3.40	2.80	2.80	2.60	2.20	2.10
0.35	3.20	3.00	2.70	2.50	2.10	2.20
0.4	3.00	2.70	2.50	2.30	2.20	1.90
0.45	2.90	2.30	2.30	2.20	1.80	1.60

El porcentaje de falsos positivos esta por debajo del 4% en todos los casos. Analizando los resultados de los cuadros 4.1 y 4.2, se puede intuir que cuando el cambio es pequeño, la barrera h podría ser más pequeña que la elegida para que el algoritmo tuviera una mayor eficiencia, es decir si la barrera aumentara cada que aumentara el tamaño del cambio, se tendrían menos falsas alarmas y se seguiría teniendo una tasa grande de detección.

A continuación se muestra el comportamiento de la función de decisión g_k del *CUSUM* en directo para la simulación de la Figura 4.1.

En este caso no hubo falsas alarmas antes del cambio ya que la función de decisión g_t se mantuvo por debajo de la barrera h hasta el punto 531, es decir 30 pasos después de haber ocurrido el cambio.

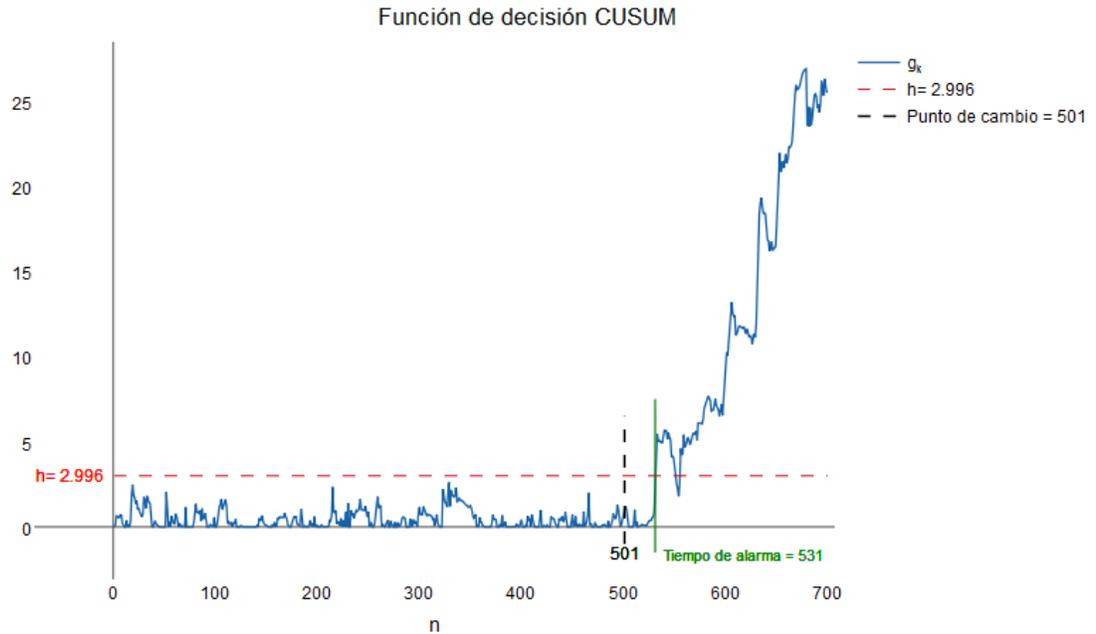


Figura 4.2: Función de decisión g_t del CUSUM para 1000 observaciones de un $ARMA(1, 1)$ con parámetros iniciales $\phi_0 = .1$ y $\theta_0 = .2$ tiempo de cambio 501 y parámetros después de cambio $\phi_1 = .2$ y $\theta_1 = .35$.

El siguiente cuadro muestra el promedio del tiempo (número de pasos) que tardó el algoritmo en detectar el cambio en las diferentes simulaciones.

Tabla 4.3: Número de pasos promedio que necesitó el CUSUM para la detección de cambio con un $ARMA(1,1)$ con parámetros iniciales ($\phi_0 = .1, \theta_0 = .2$).

$\theta_1 \backslash \phi_1$	0.1	0.15	0.2	0.25	0.3	0.35
1		347.00	232.57	154.47	96.33	65.00
2	327.03	220.67	151.94	94.44	67.28	47.71
3	223.06	143.63	92.73	67.62	47.37	37.52
4	137.42	92.13	62.68	47.67	36.64	29.28
5	86.95	61.53	44.84	36.11	28.62	23.92
6	54.80	40.64	34.09	26.45	23.96	20.85

Cuando el cambio es pequeño, el algoritmo tarda más tiempo en detectarlo, esto es normal

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1) 67

ya que, según la Ecuación (2.12), el número de falsas alarmas depende de la distancia entre los parámetros antes y después del cambio.

En este caso la barrera funciona bien para los cambios más grandes que se presentan, sin embargo, dados los resultados del Cuadro 4.3, si para cada cambio se da una barrera h diferente, haciendo la barrera h más pequeña conforme va disminuyendo el cambio, se puede mejorar el número de pasos que se necesitan para la detección.

La siguiente gráfica muestra el comportamiento que tiene la función S_k para la realización del proceso mostrado en la figura 4.1.

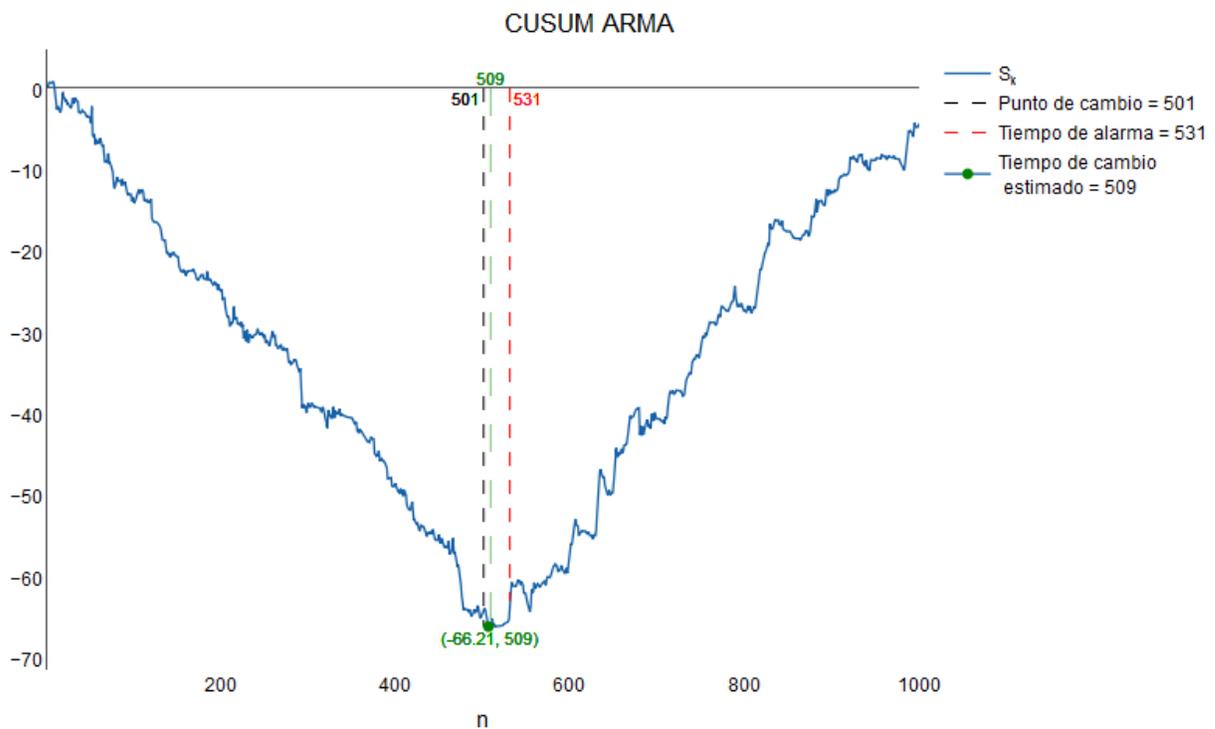


Figura 4.3: Comportamiento de la función S_k del CUSUM con 1000 observaciones de un proceso ARMA(1,1) con parámetros iniciales $\phi_0 = .1$ y $\theta_0 = .2$ tiempo de cambio 501 y parámetros después de cambio $\phi_1 = .2$ y $\theta_1 = .35$.

El tiempo estimado de cambio t_0 es el punto en que S_k alcanza el mínimo en este caso $t_0 = 509$, que difiere tan solo en 8 pasos del tiempo real de cambio. El siguiente gráfico es una comparación

de los tiempos de cambio estimados para tres de los cambios que se realizaron, este deja ver cómo entre más grande es el cambio, la estimación es más precisa.

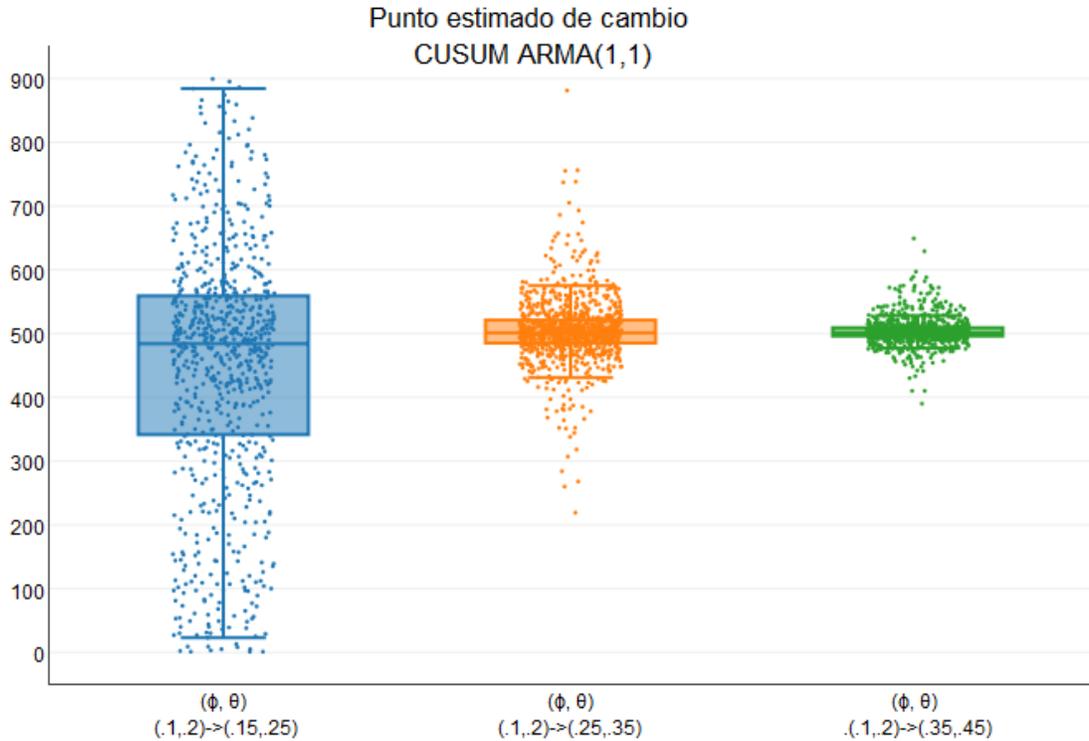


Figura 4.4: Comparación de la estimación del punto de cambio para diferentes valores de (ϕ, θ) . El cambio real ocurre en el tiempo 501.

Como puede apreciarse, entre más grande sea el cambio, el punto de cambio estimado es más preciso. En los tres casos presentados, el tiempo de cambio estimado de cambio está alrededor del 500 teniendo mucha menor varianza cuando el cambio es más grande.

El algoritmo resultó ser muy preciso tanto para la detección del cambio como para la estimación del punto donde este ocurrió. Esto era de esperarse ya que en el capítulo 2 se demostró que el CUSUM es el mejor algoritmo para estos propósitos. Sin embargo, como ya se ha mencionado, es difícil implementarlo aún cuando se cuenta con los parámetros de la serie por lo que es muy difícil de llevar a la práctica.

Utilización del CSSM

Para la implementación de este algoritmo se cuenta con los estadísticos g_n^* 3.19, $G(n, k)$, 3.19 y T_n 3.21 definidos en el capítulo anterior.

$$g_n^*(t) = \frac{[nt]}{\sqrt{n}} C^{-\frac{1}{2}} (\hat{\gamma}_{[nt]}(0) - \hat{\gamma}_n(0), \hat{\gamma}_{[nt]}(1) - \hat{\gamma}_n(1), \dots, \hat{\gamma}_{[nt]}(L) - \hat{\gamma}_n(L))^T, \quad 0 \leq t \leq 1.$$

$$G(n, k) = g_n^* \left(\frac{k}{n} \right)^T \cdot g_n^* \left(\frac{k}{n} \right).$$

$$T_n = \max_{L \leq k < n} G(n, k).$$

En la programación del algoritmo se consideró $L = 1$ y $h_n = n^{0.3}$. El valor crítico de este test estadístico a nivel $\alpha = 0.05$ es $c_\alpha = 2.408$, este valor fue obtenido para diferentes valores de L por Lee S. mediante simulaciones Montecarlo y los resultados se encuentran reportados en *The Cusum Test for Parameter Change in Time Series Models* [14] página 784, tabla 1.

4.1.2. Resultado CSSM ante un ARMA(1,1)

A continuación, se presentan los resultados del CSSM implementado para las mismas simulaciones del proceso ARMA(1, 1) que se utilizaron para el CUSUM. Cabe recalcar que el algoritmo solo necesita los datos para poder trabajar.

Cuando se tiene una muestra fija, el algoritmo CSSM, al igual que el CUSUM, se comporta mejor cuando el cambio es más grande, sin embargo, el CSSM tiene una tasa más grande de falsos positivos.

En el siguiente cuadro se observa que para cambios pequeños el algoritmo no tiene una buena tasa de detección, sin embargo, conforme el cambio va siendo mayor, el algoritmo tiene una mejor tasa de detección.

Por ejemplo, cuando el cambio en ambos parámetros es mayor a .1, la tasa de efectividad crece hasta el 90%. Cuando no hubo cambios, el CSSM falló el 21% de veces. A continuación, se presenta el porcentaje de efectividad de CSSM para un proceso ARMA(1, 1) con diferentes niveles de cambio en los parámetros ϕ_0 y θ_0 .

Tabla 4.4: Porcentaje de detección del CSSM con un $ARMA(1, 1)$ con parámetros iniciales ($\phi_0 = .1, \theta_0 = .2$).

$\theta_1 \backslash \phi_1$	0.1	0.15	0.2	0.25	0.3	0.35
0.2	21	22.3	29.7	46.5	62.7	81
0.25	23.7	31.3	49.5	67.6	81.7	91.1
0.3	37.2	49.7	72	83.5	94.1	97.3
0.35	55	70.3	87.9	94.5	98	99.1
0.4	78.9	90.7	96.2	98.1	99.9	100
0.45	90.9	97	99.4	99.9	100	100

Como se puede observar, este algoritmo llega a tener tasas muy altas de detección cuando el cambio en los parámetros de las series es grande, aunque para cambios medianos sigue manteniendo una tasa buena de efectividad.

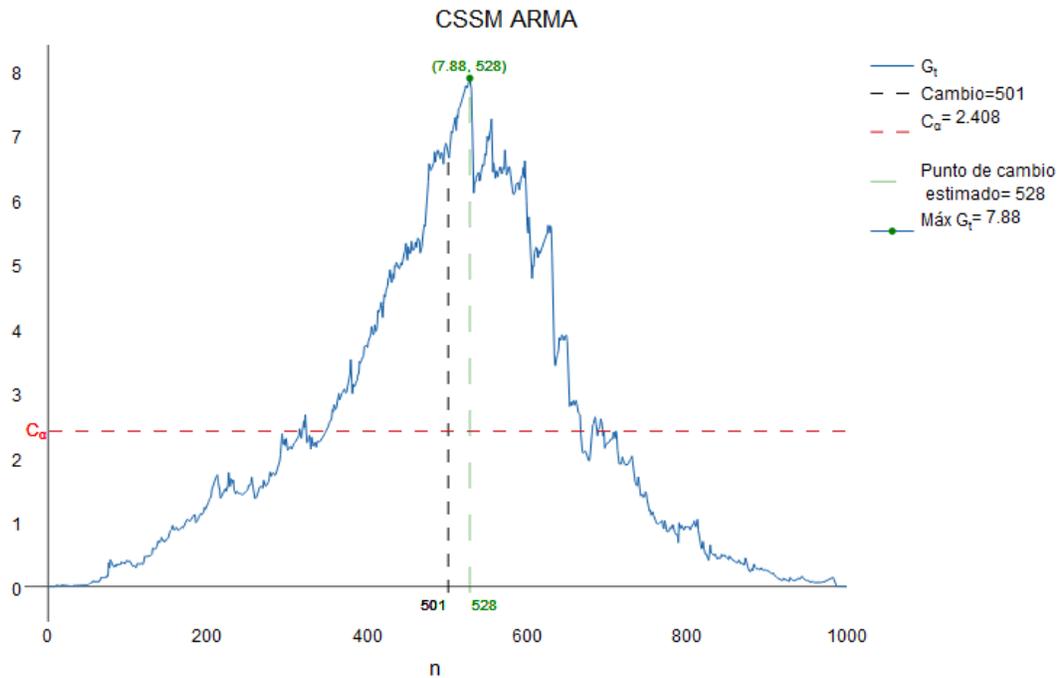


Figura 4.5: Función del estadístico del CSSM con un proceso $ARMA(1, 1)$ con parámetros iniciales $\phi_0 = .1$ y $\theta_0 = .2$ tiempo de cambio 501 y parámetros después de cambio $\phi_1 = .25$ y $\theta_1 = .35$.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1) 71

En la figura anterior se puede observar el comportamiento del algoritmo CSSM cuando recibe los datos del ARMA(1,1) presentado anteriormente, 4.1. El valor del estadístico rebasa por mucho la barrera C_α y el punto estimado de cambio es 529 que no se encuentra lejos del punto de cambio real, pero es menos preciso en comparación con el del CUSUM que, para esta serie, fue 509.

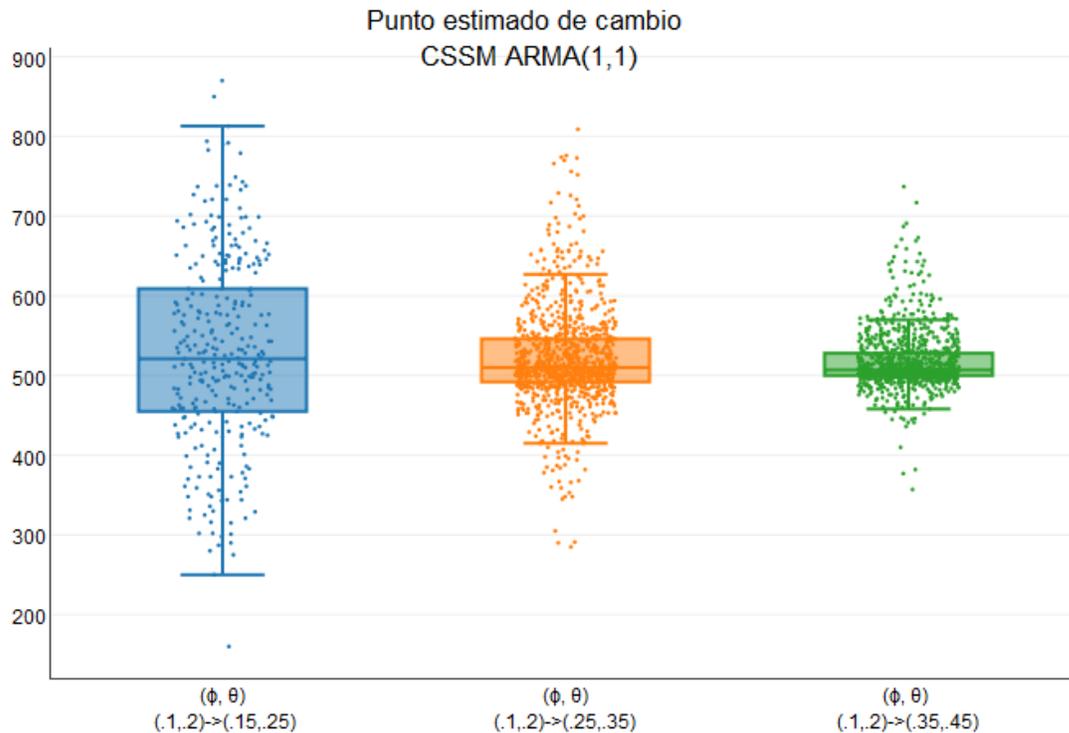


Figura 4.6: Comparación de la estimación del punto de cambio para diferentes valores de (ϕ_1, θ_1) . El cambio real ocurre en el tiempo 501.

Al igual que el CUSUM, el CSSM es más preciso cuando el cambio es mayor, tanto para la detección como para la estimación del punto de cambio. A continuación, se muestra la comparación entre ambos algoritmos respecto al punto de cambio estimado cuando los parámetros $\phi_0 = .1, \theta_0 = .2$ cambiaron a $\phi_0 = .1, \theta_0 = .2$.

Como se puede apreciar en la figura 4.1.2, los estimadores de ambos algoritmos se concentran alrededor del punto del cambio siendo el CUSUM el más preciso y el que presenta una varianza menor. Estos resultados son los que se esperaban, no hay que olvidar que el CUSUM es el mejor algoritmo desde el punto de vista teórico. El estimador de CSSM estuvo a menos de 50 pasos de

punto de cambio real el 50% de las veces, mientras que el del CUSUM a menos de 35 el mismo porcentaje de veces.

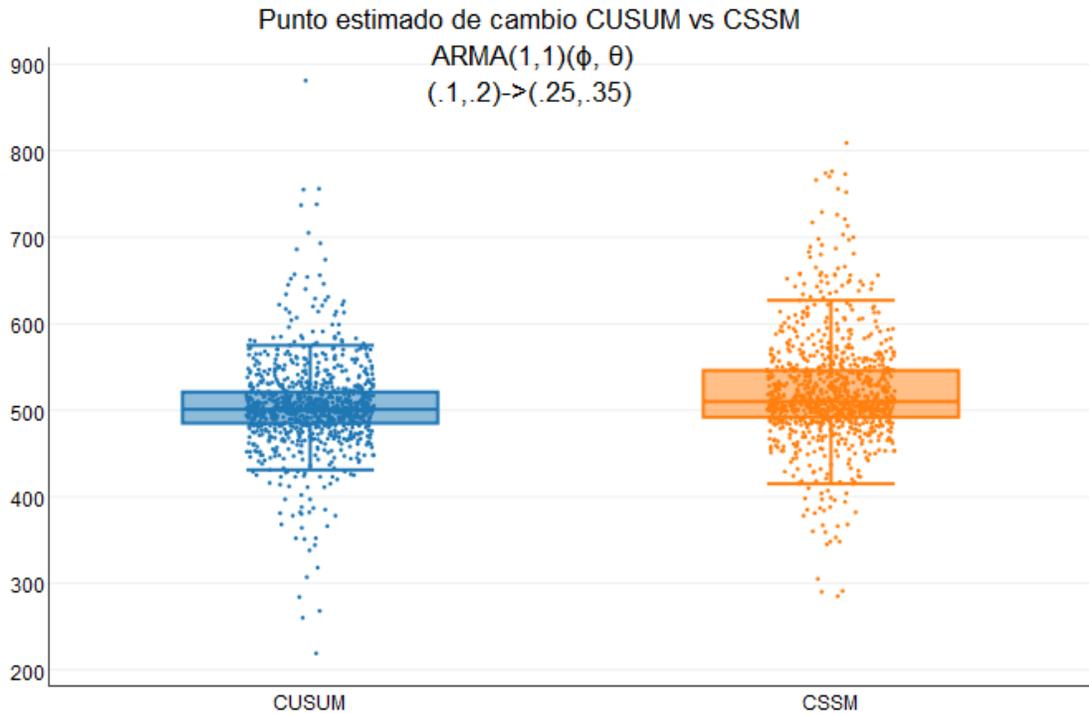


Figura 4.7: Comparación de la estimación del punto de cambio CUSUM vs CSSM

Cabe destacar que en el artículo consultado para este algoritmo [1] siempre realizan el cambio en el punto medio de la serie. Ambos algoritmos funcionan mejor entre más observaciones se tengan ya que se incrementa el porcentaje en la detección y baja el tamaño de error tipo II. Sin embargo, para el caso del CSSM también necesita una proporción parecida de observaciones antes y después del cambio para detectarlo, esto se puede apreciar en la tabla 4.8 que se encuentra más adelante.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1)⁷³

Tabla 4.5: Porcentaje de detección del CSSM ARMA(1,1) con observaciones de distinto tamaño, el cambio siempre a mitad de la serie

Algoritmo	Observaciones			
	200	400	600	800
CUSUM	64.5	94.5	98.9	100
CSSM	43.6	61.1	78	87.7

4.1.3. El CUSUM y el CSSM ante un ARMA(2,3)

Se realizaron 1000 simulaciones de un proceso ARMA(2,3) con 1000 observaciones y cambio en los parámetros (ϕ, θ) al tiempo $n = 501$. Los parámetros iniciales del proceso $(\phi_0 = (1, -0.24), \theta_0 = (0.4, 0.2, .1))$ y parámetros después de cambio $(\phi_1 = (.88, -.14), \theta_1 = (.26, -.31, .31))$.

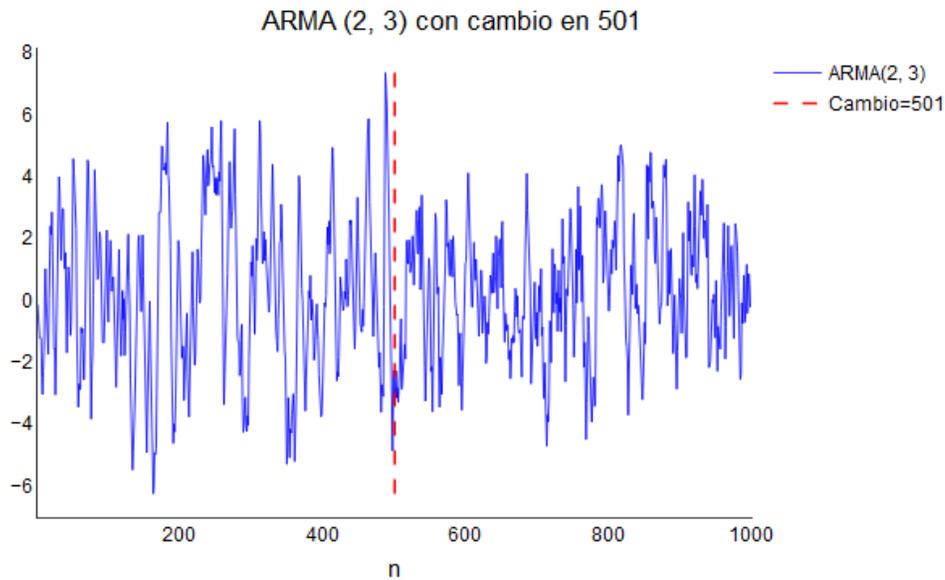


Figura 4.8: Serie simulada con 1000 observaciones de un ARMA(2,3) con parámetros iniciales $\phi_0 = (1, -0.24)$ y $\theta_0 = (0.4, 0.2, .1)$, tiempo de cambio 501 y parámetros después de cambio $\phi_1 = (.88, -.14)$ y $\theta_1 = (.26, -.31, .31)$.

En la siguiente tabla se muestran el porcentaje de efectividad de ambos algoritmos.

Tabla 4.6: Porcentaje de detección del CUSUM y del CSSM con un $ARMA(2,3)$

Algoritmo	Observaciones	
	sin cambio	con cambio
CUSUM	2.0	100
CSSM	4.1	95.4

Ambos algoritmos tuvieron un porcentaje de efectividad en la detección mayor de 95% y también ambos solo fallaron menos del 5% de las veces detectando cambio cuando en realidad no lo hubo.

Las figuras 4.9 y 4.10 muestran las gráficas de ambos algoritmos para una de la simulaciones del $ARMA(2,3)$ descrito al inicio de la subsección.

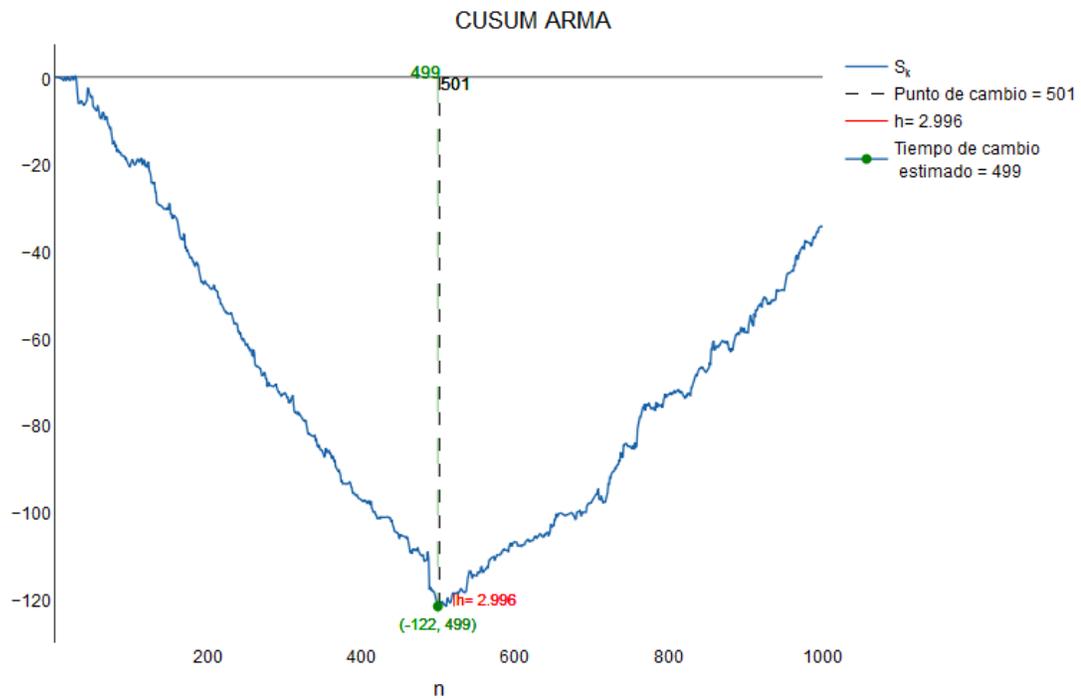


Figura 4.9: Serie simulada con 1000 observaciones de un $ARMA(2,3)$ con parámetros iniciales $\phi_0 = (1, -0.24)$ y $\theta_0 = (0.4, 0.2, .1)$, tiempo de cambio 501 y parámetros después de cambio $\phi_1 = (.88, -.14)$ y $\theta_1 = (.26, -.31, .31)$.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1)75

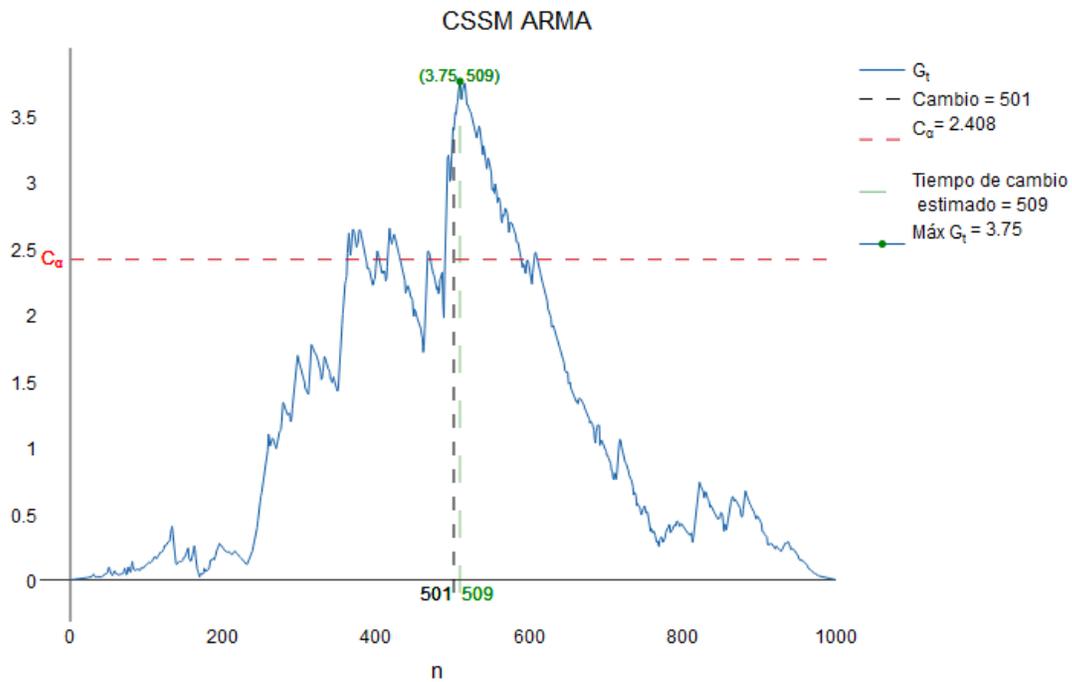


Figura 4.10: Función del estadístico del CSSM con un proceso $ARMA(2,3)$ con parámetros iniciales $\phi_0 = (1, -0.24)$ y $\theta_0 = (0.4, 0.2, .1)$, tiempo de cambio 501 y parámetros después de cambio $\phi_1 = (.88, -.14)$ y $\theta_1 = (.26, -.31, .31)$.

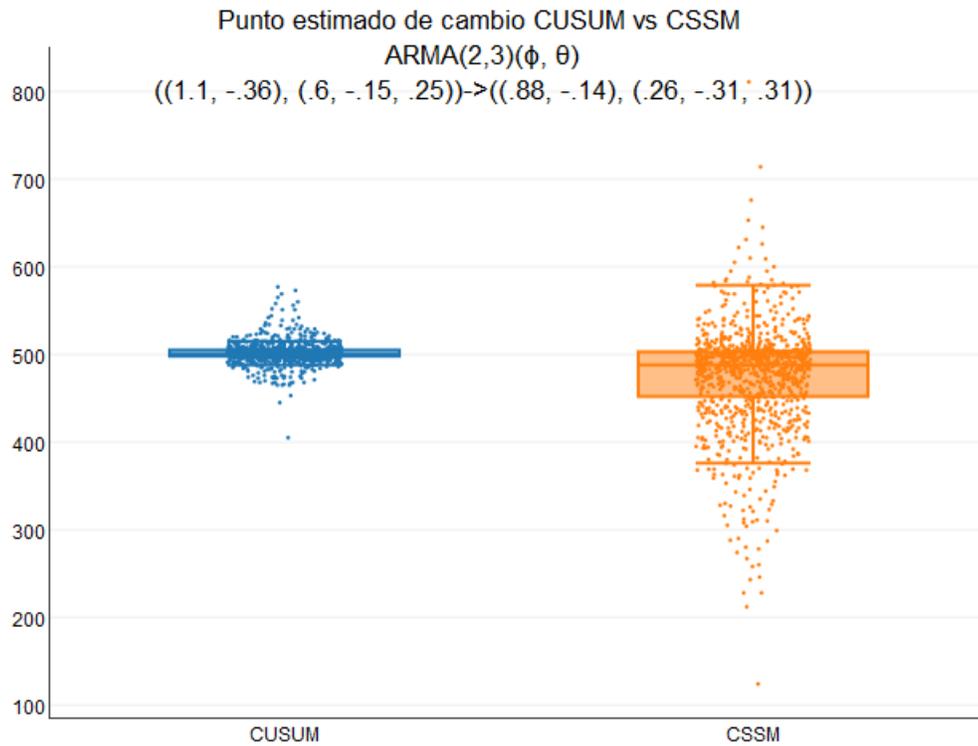


Figura 4.11: Comparación de la estimación del punto de cambio CUSUM vs CSSM

La gráfica 4.1.3 permite ver que el CUSUM tiene mucho más precisión al momento de estimar el punto de cambio.

Si se toma en cuenta el porcentaje de falsos positivos el CSSM tuvo un mejor desempeño con un $ARMA(2,3)$ que con un $ARMA(1,1)$, sin embargo el punto estimado de cambio fue mas cercano cuando se trató el $ARMA(1,1)$. Comparado con el CUSUM, el CSSM fue casi tan bueno en la detección pero a la hora de estimar el punto de cambio se pudo observar que el CUSUM es mucho mejor, cabe remarcar que el CSSM no necesita ningún parámetro de la serie para trabajar y es justo ahí donde radica su potencial.

4.1.4. El CSSM y un proceso GARCH(1,1)

A continuación se muestran los resultados del comportamiento del algoritmo CSSM con un proceso GARCH(1,1) el cual está definido de la siguiente manera.

Definición 4.1.3. Un proceso se dice *Generalizado Autoregresivo de Heterosedasticidad Condicional* de orden p, q , denotado como $GARCH(p, q)$, si es estrictamente estacionario, y si para toda $T \in \mathbb{N}$ y para algún proceso de valores positivos $(\sigma_t)_{t \in \mathbb{N}}$ satisface las siguientes condiciones:

$$X_t = h_t Z_t, \quad (4.1)$$

$$h_t^2 = \omega + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}^2, \quad (4.2)$$

para $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p$ y $\beta_i \geq 0$, $i = 1, \dots, q$.

Los modelos GARCH(1,1) son de los modelos más recurridos para la modelación de series financieras.

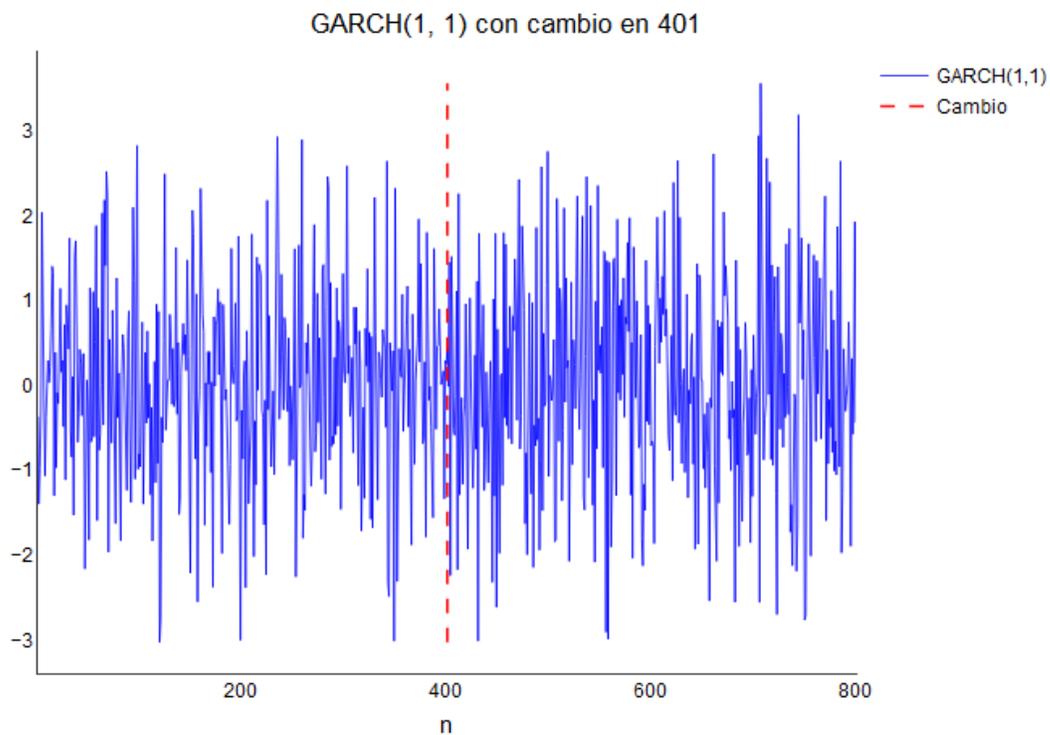


Figura 4.12: Serie simulada con 800 observaciones de un GARCH(1,1) con $(\alpha_0 = .1, \beta_0 = .2, \omega_0 = .8)$ tiempo de cambio 401 y $(\alpha_1 = .1, \beta_1 = .2, \omega_1 = 1.1)$.

Se realizaron múltiples simulaciones de un proceso $GARCH(1,1)$ con parámetros iniciales $\alpha_0 = .1$, $\beta_0 = .2$ y $\omega_0 = .8$, en los cuales se hicieron distintos cambios en el tiempo 401. Para cada cambio se realizaron 1000 simulaciones, cada una con 800 observaciones del proceso.

Tanto en el caso del $ARMA$ como en el del $GARCH$ no es posible apreciar un cambio a simple vista. Cuando existen cambios demasiado grandes en las series de tiempo estos sí son notables en la gráfica de la misma.

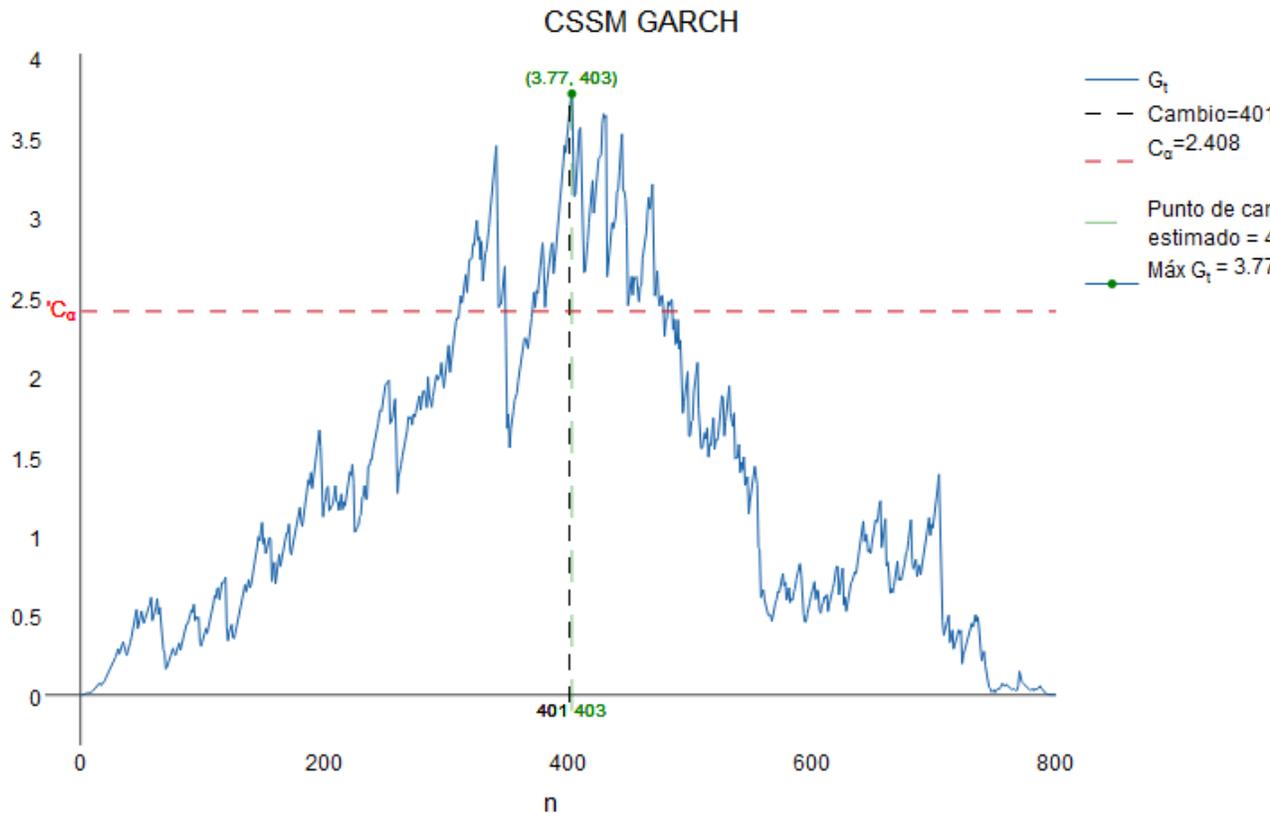


Figura 4.13: Función del estadístico del CSSM con un proceso $GARCH(1,1)$ con parámetros iniciales $\alpha_0 = .1$, $\beta_0 = .2$ y $\omega_0 = .8$ tiempo de cambio 401 y parámetros después de cambio $\alpha_1 = .1$, $\beta_1 = .2$ y $\omega_1 = 1.1$.

La función del CSSM tiene un comportamiento similar al que se tiene caso del $ARMA$. A continuación se presentan los resultados del comportamiento del CSSM, con cambio en los tres parámetros de la serie.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1) 79

Tabla 4.7: Porcentaje de detección del CSSM $GARCH(1,1)$

Sin cambio	α	β	ω	(α, β)	(α, β, ω)
	.1 → .3	.2 → .4	.8 → 1.1	(.1, .2) → (.3, .4)	(.1, .2, .8) → (.22, .33, .91)
38.10	87.30	93.30	89.20	99.80	99.90

Como en el caso del $ARMA$, entre más grande es el cambio la efectividad del algoritmo es mayor, sin embargo, para este caso el error tipo II es demasiado grande.

La siguiente gráfica muestra el comportamiento del CSSM cuando no existe cambio.

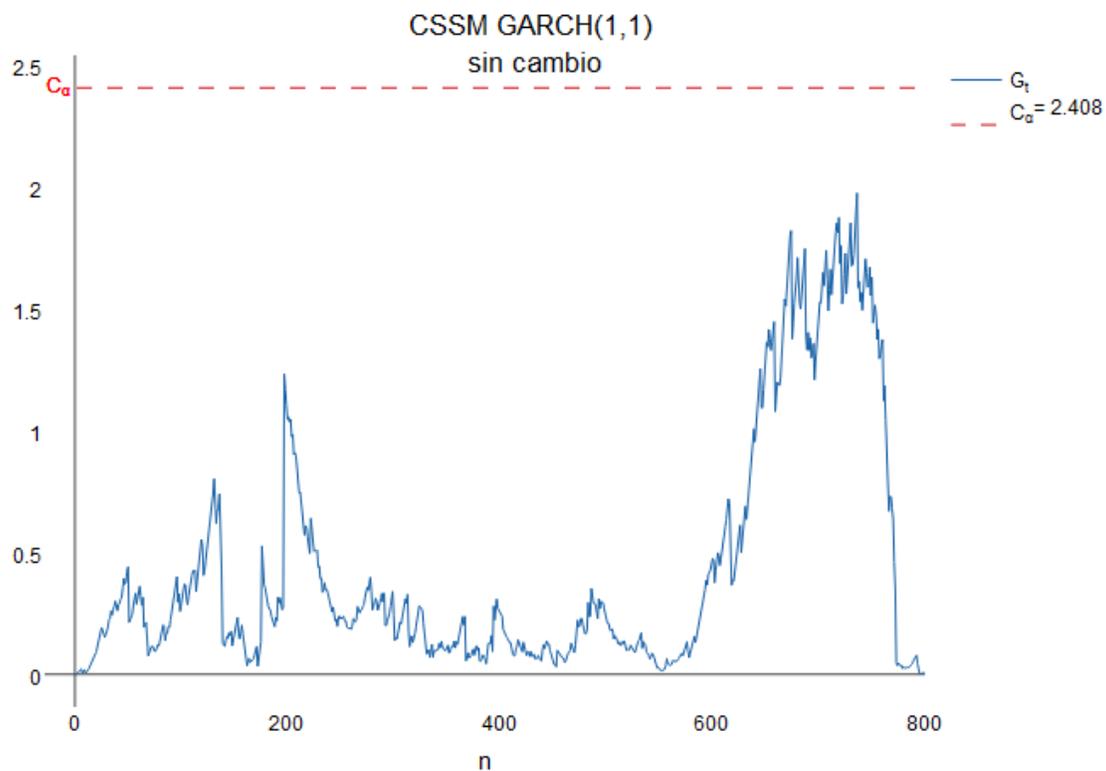


Figura 4.14: Función del estadístico del CSSM con un proceso $GARCH(1,1)$ con parámetros iniciales $\alpha_0 = .1$, $\beta_0 = .2$ y $\omega_0 = .8$ sin cambios.

En este caso, toda la gráfica queda por debajo de la barrera $C_{.05} = 2.408$, inclusive el máximo está muy por debajo de este nivel, por lo que se optó por H_0 , es decir no hubo cambio.

En la siguiente figura se puede observar que cuando el algoritmo detecta cambio y este no existe, los puntos de cambio estimados se encuentran bastante dispersos, mientras que cuando sí lo hubo, los puntos de cambio estimados se acumulan alrededor de 400 y conforme el cambio es más grande, la varianza disminuye.

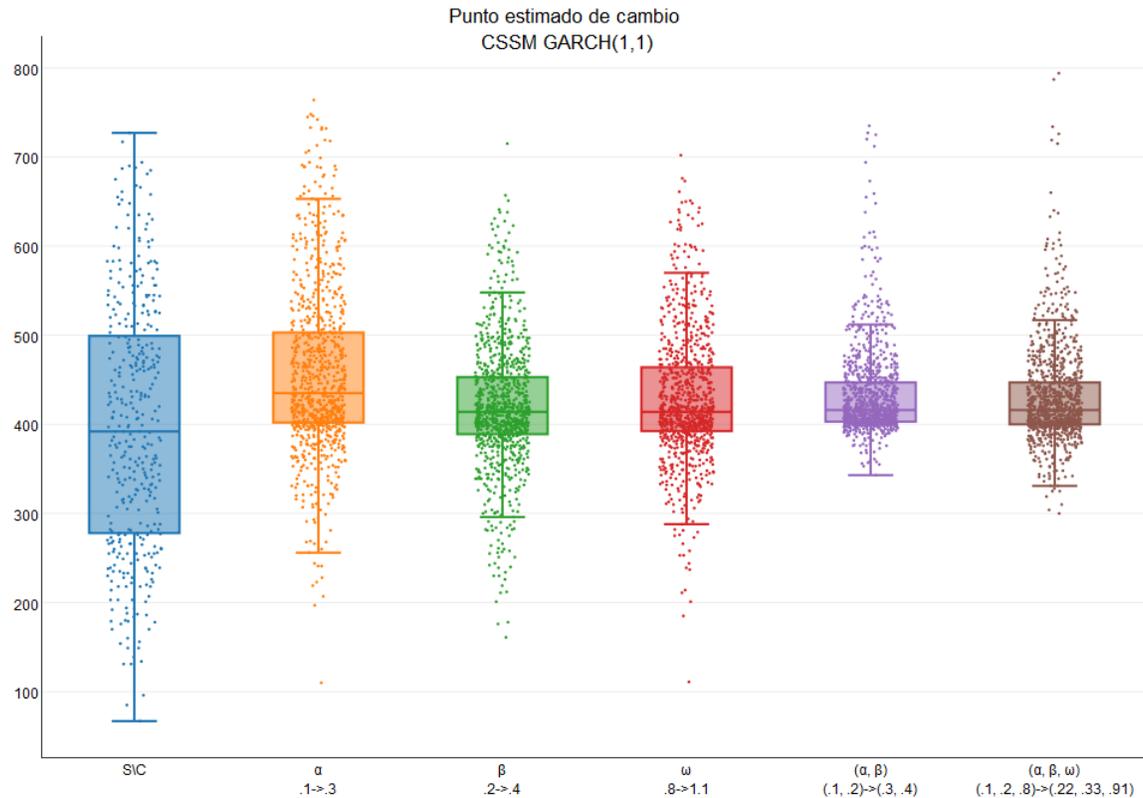


Figura 4.15: Punto de cambio estimado por el CSSM con un $GARCH(1, 1)$ con distintos cambios al tiempo 401.

Como ya se mostró el CSSM varía su efectividad dependiendo la magnitud del cambio, sin embargo, la ubicación del cambio y el tamaño de la muestra también afectan la efectividad de este algoritmo. Cuando el punto de cambio se encuentra a mitad de la serie pero se tienen diferentes tamaños de muestra, el Cuadro 4.5 deja ver que entre más grande sea la muestra, la tasa de detección es mayor, aunado a esto, el Cuadro 4.9 deja ver que el error tipo II aumenta cuando se disminuye la muestra. Además se realizaron simulaciones con el punto de cambio en distintos lugares; al tiempo 101, 201, 301, 401 y 501 esto para observar como varía la efectividad del algoritmo.

4.1. COMPARACIÓN DE LOS ALGORITMOS CSSM Y CUSUM EN UN PROCESO ARMA(1,1) 81

Tabla 4.8: Porcentaje de detección CSSM con 600 observaciones y cambio de parámetros en diferentes puntos.

Punto de cambio				
101	201	301	401	501
74.90	95.30	97.50	95.90	84.10

Tabla 4.9: Porcentaje de detección CSSM con diferentes tamaños de muestra y el cambio siempre a la mitad de la serie.

Tamaño de muestra		
200	300	400
77.60	92.70	97.80

En ambas tablas los parámetros cambiaron de $(\omega_0 = .8, \alpha_0 = .1, \beta_0 = .2)$ a $(\omega_1 = .85, \alpha_1 = .2, \beta_1 = .35)$.

El Cuadro 4.8 nos muestra que el algoritmo CSSM tiene mayor efectividad cuando el cambio ocurre a la mitad de la serie que cuando ocurre en alguno de los extremos, tomando en cuenta solo los extremos, el algoritmo tiene mayor efectividad cuando el cambio ocurre en el extremo final que en el extremo inicial. En el Cuadro 4.9, se puede observar que con una muestra de 300 observaciones el algoritmo ya tiene una tasa alta de detección, esto cuando el cambio está a la mitad de la serie.

Conclusiones

En el trabajo se presentó el Teorema Ergódico de Birkhoff-Khinchin que es una generalización de la Ley de los Grandes Números para variables aleatorias no necesariamente independientes. Se utilizaron resultados como el Teorema de Convergencia Monótona, el de Convergencia Dominada y el Lema de Fatou, los cuales son invocados frecuentemente cuando se quiere probar que tienen que ver con convergencia; todos estos resultados se utilizaron para fundamentar el algoritmo CUSUM. Con las simulaciones realizadas se mostró la eficacia de este algoritmo no solo en cuanto al peor retraso medio, sino también en cuanto a la efectividad en la detección y precisión en la estimación del punto de cambio. Sin embargo, al ser simulaciones, se contó con los valores reales de los parámetros tanto antes como después del cambio, algo que no se tiene en la práctica.

También se exhibió el Teorema Funcional del Límite Central de Herrndorf, el cual no solo lo cumplen las *v.a.i.i* sino una clase de variables conocidas como α – *mezclantes*. La demostración que se presentó sigue la misma idea que siguen las demostraciones de otros Teoremas Funcionales Límite. Este resultado se utilizó para probar que el CSSM converge a la suma de Puentes Brownianos independientes.

Las simulaciones mostraron que el CSSM es un algoritmo bastante efectivo para la detección de cambios cuando se tiene una muestra fija siempre y cuando el cambio no ocurra a menos de 100 pasos del comienzo o del final de la serie. Cuando se tienen suficientes observaciones del proceso antes y después del cambio, el algoritmo tiene una tasa de más del 90 % de efectividad en la detección y la mayoría de las veces el punto de cambio estimado se encuentra a menos de 50 pasos del punto real de cambio. La anterior fue resultado de probarlo con simulaciones de proceso $ARMA(1, 1)$, $ARMA(2, 3)$ y $GARCH(1, 1)$.

También se probó el CSSM con un $GARCH(3, 3)$ sin embargo porcentaje de detección cuando no hubo cambio fue mayor al 97 %, esto lo vuelve inservible.

Para la detección en directo del cambio teniendo una ventana de tiempo de 80 observaciones, el algoritmo CSSM tuvo una tasa del 65 % de detección, sin embargo esta no está cercana a la del CUSUM.

En este trabajo no se demostró que los procesos utilizados en el CSSM cumplieran las hipótesis necesarias para su funcionamiento, es decir que fueran α – *mezclantes*, esto se dio por hecho ya que en el artículo en el que se investigó usan este tipo de procesos para probar el algoritmo. Sin embargo, es importante verificar que se cumplan las hipótesis a la hora de implementar dicho algoritmo. Es aquí donde hay otra línea de estudio; cómo saber que una serie de tiempo cumple con las hipótesis para poder implementar el CSSM con los puros datos.

La gran ventaja del CSSM sobre otros algoritmos es que solo necesita los datos de la serie para poder trabajar y no depende de nada más. Que el algoritmo CSSM no dependa de ningún parámetro lo vuelve una herramienta muy poderosa, queda para trabajos futuros estudiar qué pasa cuando la serie contiene más de un cambio.

El CSSM es de gran utilidad cuando se quiere modelar una serie que contiene una gran cantidad de observaciones, pues en estos casos es difícil implementar un modelo que ajuste de forma adecuada a toda la serie, por lo que se toma solo una parte de la serie para ajustar un modelo, y ya que el algoritmo nos da una buena estimación del punto de cambio de la serie si es que lo hay, se puede usar este punto como referencia para hacer una modelación más precisa.

Apéndice A

Más teoremas

Teorema A.0.1. Teorema de convergencia monótona. Sea $\{f_n\}$ una sucesión de funciones medibles de (X, \mathcal{A}, μ) , un espacio con medida, en $[-\infty, \infty]$ tal que $f_n \uparrow f$ y $\int f_1 d\mu > -\infty$. Entonces $\int f_n d\mu \uparrow \int f d\mu$.

Lema A.0.2. Lema de Fatou.

Sea $\{f_n\}$ una sucesión de funciones medibles no negativas sobre (X, \mathcal{A}, μ) , un espacio con medida. Entonces $\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu$.

Teorema A.0.3. Teorema de convergencia dominada.

Sean $\{f_n\}$ y g en $\mathcal{L}_1(X, \mathcal{A}, \mu)$, tal que $|f_n(x)| \leq g(x)$ y $\lim_{n \rightarrow \infty} f_n(x) = f(x)$, para todo $x \in X$. Entonces $f \in \mathcal{L}_1(X, \mathcal{A}, \mu)$ y $\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$.

Teorema A.0.4. Desigualdad de Markov.

Sea φ una función monótonamente creciente no negativa, sea X una variable aleatoria, $0 \leq a$ y $0 < \varphi(a)$. Entonces:

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbf{E}(\varphi(|X|))}{\varphi(a)}.$$

El siguiente corolario se sigue inmediatamente,

Corolario A.0.5. Sea X una variable aleatoria no negativa, $0 \leq a$ y $0 < \varphi(a)$. Entonces:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbf{E}(X^n)}{a^n}.$$

Apéndice B

Códigos R

B.1. Código CUSUM

```
#CUSUM para un ARMA usando el codigo para generar cambios en un ARMA  
#e_k es para ver la serie de tiempo en terminos del ruido blanco  
e_k<-function(x, p, q, phi, th, e){  
  N<-length(x)  
  p<-length(phi)  
  q<-length(th)  
  s<-numeric(N)  
  for(i in (max(p,q)+1):N){  
    s[i]<-x[i]-(sum(x[(i-1):(i-p)]*phi)+sum(e[(i-1):(i-q)]*th))  
  }  
  return(s)  
}
```

```
#Es el CUSUM, recibe:  
#La serie de tiempo, los parametros antes del cambio,  
#los parametros despues del cambio y  
#el vector de Ruidos Blancos (para facilitar los calculos),  
 #(tambien recibe el punto de cambio, aunque este no se utiliza)  
#regresa:  
#1. La suma de logverosimilitudes en cada paso, es la funcion S_k.  
#2. El  $\min\{S_k\}$  y el punto donde ocurre este minimo.
```

#3. la funcion g_k para la prueba secuencial.

#4. la logverosimilitus en cada paso (lo que va sumando S_k).

```
S_k<-function(x, p1, phi1, q1, th1, s1,
              p2, phi2, q2, th2, s2, e, t){
  N<-length(x)
  s<-numeric(N)
  R<-max(p1, q1, p2, q2)
  e1<-e_k(x, p1, q1, phi1, th1, e)
  e2<-e_k(x, p2, q2, phi2, th2, e)
  for(i in R:N){
    s[i]<-((log((s1^2)/(s2^2)))+((e1[i]^2)/(s1^2))-
          ((e2[i]^2)/(s2^2)))/2
  }
  g<-c()
  g[1]<-0
  S<-numeric(length(s))
  for(i in 1:N){
    S[i]<-sum(s[1:i])
    #if(i<t && g[i-1]<(-3.01)){
    if(g[i]<0){
      g[i+1]<-max(0, s[i+1])
    }else{
      g[i+1]<-max(0, g[i]+s[i+1])
    }
  }
  length(g)=N
  return(list(S_k=S, res=c(min(S, na.rm=T),
                          which.min(S)), g_k=g, s_k=s))
}
```

B.2. Código CSSM

#Funcion para obtener el ACF muestral $\gamma_n(h)$.

#Recibe una serie "X" y las variable n y h .

```
sACF_1<-function(x, n, h){
  h<-abs(h)
  if((n-h)<=0){
    a<-"Error"
  }else{
    a<-sum(x[1:(n-h)]*x[(1+h):n])/n
  }
}
```

```

    }
    return(a)
}

#Funcion O grande de n, en este caso n^(0.3).
#Recibe una serie solo para tomar su tamaño.
fO_n<-function(x){
  a<-(length(x))^0.3
  return(a)
}

#Funcion para obtener el vector de diferencias entre
# las autocovarianzas y las autocovarianza muestrales.
dif_sacf_l<-function(x, s, L, n){
  a<-numeric(L+1)
  nt<-max(1,(as.integer(s*n)))
  for(i in 0:L){
    if(i<nt){
      a[i+1]<-sACF_l(x, nt, i)-sACF_l(x, n, i)
    }
    else {
      a[i+1]<-0
    }
  }
  return(a)
}

#:::*****:::*****:::
#Funcion Y_{l}^{l}
Yl_l<-function(x, l, h, k, n){
  y<-c()
  if(n>(k+1)){
    for(i in 1:(n-k-1)){
      y[i]<-x[i]*x[i+h]*x[i+1]*x[i+1+k]
    }
  } else {
    y=0
  }
  return(y)
}

```

```

#:::*****:::*****:::#
#Funcion  $Y_{-2}^{\{l\}}$ 
Y2_l<-function(x, l, h, k, n){
  y<-c()
  if(n>k){
    for(i in 1:(n-k-1)){
      y[i]<-x[i+1]*x[i+1+h]*x[i]*x[i+k]
    }
  } else {
    y=0
  }
  return(y)
}

#Funcion  $\bar{\sigma}_{hk}(l)$ .
sigma_bar_hk<-function(x, l, h, k, n){
  s<-0
  if(l==0){
    s<-(n-k)*(mean(Y1_l(x, l, h, k, n), na.rm = T)-
               (sACF_l(x, n, h)*sACF_l(x, n, k)))
  } else {
    s<-(n-k)*(mean(Y1_l(x, l, h, k, n), na.rm = T)+
               mean(Y2_l(x, l, h, k, n), na.rm = T)
               -(2*sACF_l(x, n, h)*sACF_l(x, n, k)))
  }
  return(s)
}

#Funcion  $\bar{\theta}_{hk}$ .
theta_hk<-function(x, h, k, n){
  O<-c()
  for(i in 1:(as.integer(fO_n(x)))){
    O[i]<-sigma_bar_hk(x, i-1, h, k, n)
  }
  return(sum(O, na.rm = T)/n)
}

#Funcion para crear la matriz de varianzas y
#covarianzas estimadas  $\hat{C}$ .
C<-function(x, L, n){

```

```

C<-matrix(0, L+1, L+1)
for(i in 1:(L+1)){
  for(j in 1:(L+1)){
    if(i<=j){
      C[i,j]<-theta_hk(x, i, j, n)
    }else{
      C[i,j]<-C[j,i]
    }
  }
}
return(C)
}

#Algoritmo CSSM, la funcion Tn recibe
#x, la serie de tiempo y L un numero natural
#para la simulaciones se uso L=1.
#Regresa:
#1. El proceso CSSM  $G(n,k)$ , es un vector con  $n$   $G(n,k)$ 
#corresponde a la  $k$ -esima entrada
#2. El  $\max\{G(n,k)\}$  y la posicion de este.
#3. La matriz de varianzas y autocovarianzas
#estimada  $\hat{C}$ .
T_n<-function(x, L){
  m<-length(x)
  g<-c()
  C<-C(x, L, m)
  Cinv<-solve(C)
  g[1]=0
  g[m]=0
  for(i in 2:(m-1)){
    #  $g[i]<-g_n(x, (i/n), L)$ 
    g[i]<-((i^2)/m)*
      (t(as.vector(dif_sacf_1(x, (i/m), L, m)))%*%
        Cinv%*%t(as.vector(dif_sacf_1(x, (i/m), L, m))))
  }
  return(list(t=g, res=data.frame(max_g=max(g, na.rm=T),
                                Change_point=which.max(g)), C=C))
}

```


Bibliografía

- [1] AZIZZADEHA, F., AND REZAKHAHAB, S. *The CUSUM Test for Detecting Structural Changes in Strong Mixing Processes. Communications in Statistics -Theory and Methods* (2014), 3733–3750.
- [2] BASSEVILLE, M., AND NIKIFOROV, I. V. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall Information and System Sciences, 1993.
- [3] BERKES, I., GOMBAY, E., AND HORVÁTH, L. *Testing for changes in the covariance structure of linear processes. Journal of Statistical Planning and Inference* 139 (2009), 2044–2063.
- [4] BILLINGSLEY, P. *Convergence of Probability Measures*, second edition ed. Wiley Series in Probability and Statistics. Wiley-Interscience, 1999. ISBN 0-471-19745-9.
- [5] BROCKWELL, P. J., AND DAVIS, R. A. *Introduction to Time Series and Forecasting*, second edition ed. Springer Texts in Statistics. Springer New York, 2006. ISBN 0-387-95351-5.
- [6] DAVYDOV, Y. A. *Convergence of Distributions Generated by Stationary Stochastic Processes. Theor. Probab. Appl.* 13, 4 (1968), 691–696.
- [7] DUDLEY, R. M. *Real Analysis and Probability*. Cambridge University Press, Massachusetts Institute of Technology, Cambridge, 1963. ISBN 0-511-04208-6.
- [8] GNEDENKO, B. V., AND KOLMOGOROV, A. N. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1954.
- [9] HERRNDORF, N. *A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables. The Annals of Probability* 12, 1 (1984), 141–153.
- [10] IBRAGIMOV, I. A., AND LINNIK, Y. V. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen, 1975.

- [11] LORDEN, G. *Procedures for Reacting to a Change in Distribution. The Annals of Mathematical Statistics* 42, 6 (1971), 1837–1908.
- [12] LOÈVE, M. *Probability Theory*, fourth edition ed. Graduate Text in Mathematics. Springer, 1963.
- [13] MOOD, A. M. *Introduction to the Theory of Statistics*, third edition ed. series in probability and statistics. McGraw-Hill, 1974.
- [14] SANGYEOL, L., JEONGCHEOL, H., OKYOUNG, N., AND SEONGRYONG, N. *The Cusum Test for Parameter Change in Time Series Models. Board of the Foundation of the Scandinavian Journal of Statistics* 30 (2003), 781–796.
- [15] SCHOLZ, F. W. *Central Limit Theorems and Proofs. Math/Stat* 39 (2011).
- [16] WALD, A. *Sequential Analysis*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. J. Wiley & Sons, Incorporated, 1947.
- [17] WALKDEN, C. *Ergodic Theory*. 2018.
- [18] WETHERILL, AND GLAZEBROOK. *Sequential Methods in Statistics*, third edition ed. Chapman and Hall, Bristol, Great Britain, 1986.
- [19] WITHERS, C. S. *Central Limit Theorems for Dependent Variables. I. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57, 4 (1981), 509–534.
- [20] ZHENGYAN, L., AND CHUANRONG, L. *Limit Theory for Mixing Dependent Random Variables*. Mathematics and Its Applications. Springer Netherlands, 1996.