



Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias Físico Matemáticas
Postgrado en Ciencias Matemáticas

Análisis de puntos de cambio en series de tiempo

TESIS

Que para obtener el grado de:

Doctor en Ciencias Matemáticas

Presenta:

Silvia Herrera Cortés

Directores de tesis:

**Dr. Bulmaro Juárez Hernández, Dr. Víctor Hugo Vázquez
Guevara, Dr. Hugo Adán Cruz Suárez**

Puebla, Puebla. Diciembre 2019



DRA. LIDIA AURORA HERNÁNDEZ REBOLLAR
SECRETARIA DE INVESTIGACIÓN Y
ESTUDIOS DE POSGRADO, FCFM-BUAP
P R E S E N T E:

Por este medio le informo que el(la) C:

SILVIA HERRERA CORTÉS

estudiante del Doctorado en Ciencias (Matemáticas), ha cumplido con las indicaciones que el Jurado le señaló en el Coloquio que se realizó el día 29 de noviembre de 2019, con la tesis titulada:

“ANÁLISIS DE PUNTOS DE CAMBIO EN SERIES DE TIEMPO”

Por lo que se le autoriza a proceder con los trámites y realizar el examen de grado en la fecha que se le asigne.

A T E N T A M E N T E.

H. Puebla de Z, a 2 de diciembre de 2019

DRA. PATRICIA DOMÍNGUEZ SOTO
COORDINADORA DEL POSGRADO
EN MATEMÁTICAS.



DRA.PDS/mrv

Facultad
de Ciencias
Físico Matemáticas

Av. San Claudio y 18 sur, edif. FM1
Ciudad Universitaria, Col. San
Manuel, Puebla, Pue. C.P. 72570
01 (222) 229 55 00 Ext. 7550 y 7552

Agradecimientos

La vida me dio a elegir entre dos caminos, la decisión fue difícil, sé que elegí el camino correcto, estoy en paz, tranquila. Sólo me queda pedir sabiduría, fortaleza y humildad para recorrer el camino que me falta.

Mi formación profesional ha sido gracias al apoyo de muchas personas, la lista es inmensa; sin embargo, cuento con un número finito de páginas, por tanto trataré de ser breve y mencionar a las personas más importantes.

Destaco a los integrantes del comité tutorial, a los doctores Bulmaro Juárez Hernández, Víctor Hugo Vázquez Guevara y Hugo Adán Cruz Suárez, por el tiempo dedicado en la revisión del trabajo desarrollado en este trayecto, una mención especial al Dr. Bulmaro Juárez Hernández por sus conocimientos, paciencia, humildad y por todo el apoyo y confianza que me ha brindado.

A todos los doctores que me dieron clase, compañeros y amigos del posgrado, gracias por las experiencias compartidas. Hago mención especial a Miriam y Xóchitl. Miriam, gracias por tu amistad, por las noches de desvelo donde no sólo compartimos ideas y conocimiento, me permitiste conocer tus aspiraciones personales y profesionales, eres una gran mujer. Xóchitl, me impresiona tu entusiasmo, gracias por las palabras de ánimo.

Mis mejores deseos para la señora Blanca Aurora, quien sin conocerme me brindó alojamiento durante seis años, gracias por esos minutos de charla amena, por sus consejos y atenciones hacia mi persona, su hospitalidad me brindó tranquilidad en la lejanía de mi hogar.

A mis papás, sin ellos mi existencia no hubiera sido posible, sé que están orgullosos de mis logros, a pesar no comulgar siempre con mis ideas, mi gratitud,

respeto y cariño para ustedes.

Mamá, tú fuiste parte fundamental de este proyecto, has sido testigo principal de mis inquietudes y de este caminar de seis años, ambas sabemos que no fue fácil. Te vivo agradecida por tus palabras de aliento, las horas de desvelo así como las angustias que pasaste por acompañarme a la parada del autobús a las cinco de la mañana.

Papá, con respeto dedico este triunfo de vida, sé que en la diferencia de ideas siempre estará el respeto.

A todos mis hermanos, en especial a Ana y Federico, son mi ejemplo. Aún tengo en mente las palabras “Lo que empiezas, lo terminas” y “Todo lo que se hace, se hace bien”, gracias por ello, sé que lo hicieron para que no dudara y retomara el camino. La vida nos ha separado en distancia pero no en pensamiento, sé que cuento con su apoyo.

Compañeros y amigos de trabajo: José Antonio Ortiz Zamora, Minerva Flores Morales, Maricela Hernández Vázquez, Carlos Vázquez Palma y Andrea Vera Reyes, gracias por escucharme en mis días de desesperación, enojo y frustración, valoro sus palabras de ánimo y la amistad que ha surgido entre nosotros, espero no defraudarlos.

A todas aquellas personas que omití sin intención, estas líneas son para ustedes, muchas gracias por todo.

Finalmente, agradezco el financiamiento económico de CONACyT, de no haber contado con ese invaluable apoyo, la meta aún estaría en proyecto.

Índice general

	Página
Índice de tablas	VI
Índice de figuras	VIII
Introducción	XI
1. Preliminares sobre puntos de cambio	1
1.1. Antecedentes del problema de puntos de cambio	2
1.2. Conceptos	2
1.3. Clasificación de puntos de cambio	4
1.4. Metodologías relevantes para resolver el problema de puntos de cambio	5
1.4.1. Metodologías para detectar puntos de cambio en forma secuencial	5
1.4.2. Metodologías a posteriori para detectar puntos de cambio	6
1.5. Áreas de aplicación	8
2. Series temporales	11
2.1. Conceptos	11
2.2. Media, función de autocovarianza y autocorrelación muestrales .	15
2.3. Modelos ARIMA	16
3. Puntos de cambio en series temporales	17
3.1. Metodologías supervisadas y no supervisadas para detectar puntos de cambio en series de tiempo	18
3.1.1. Metodologías supervisadas y no supervisadas	18

3.2. Metodologías paramétricas para detectar puntos de cambio en series de tiempo	19
3.3. Metodologías paramétricas para múltiples puntos de cambio . . .	34
4. Comparación de metodologías	39
4.1. Caracterización de la zona de estudio y análisis de calidad de datos	39
4.2. Ajuste de modelos	40
4.3. Aplicación de metodologías para detectar un sólo punto de cambio (AMOC)	43
4.3.1. Detección de múltiples puntos de cambio	47
4.3.2. Modelos ajustados después de detectar puntos de cambio en la media de temperatura máxima	50
Conclusiones	55
A. Convergencia Estocástica	59
A.1. Convergencia en distribución	59
A.2. Convergencia en Probabilidad	60
A.3. Convergencia casi segura	62
A.4. Orden de Convergencia	66
A.5. Teoremas importantes	67
A.6. Puente Browniano	69
A.7. Matriz inversa generalizada	69
B. Programas Desarrollados en R para la aplicación de las metodologías	71
B.1. Programa elaborado en R para la aplicación de la metodología de Davis <i>et al.</i> [17]	71
B.2. Programa elaborado en R para la aplicación de la metodología de Gombay E [37]	75

Índice de tablas

1.1. Bibliografía sobre puntos de cambio con distintos enfoques (<i>Elaboración propia</i>).	3
3.1. Pseudocódigo del método PELT.	36
3.2. Pseudocódigo del método BS.	37
4.1. Modelos utilizados para ajustar los datos.	42
4.2. Coeficientes del modelo ARIMA(3,1,0).	42
4.3. Valores máximos para probar un cambio en los parámetros μ , σ^2 , ϕ_1 , ϕ_2 and ϕ_3	44
4.4. Ubicación de un punto de cambio en la media y la varianza de la temperatura máxima utilizando la librería <i>changeoint</i> de <i>R</i>	44
4.5. Número de puntos de cambio para la media de temperatura máxima.	47
4.6. Ubicación de los puntos de cambio en la media de la temperatura máxima, determinada con el método PELT.	47
4.7. Número de puntos de cambio en la varianza de la temperatura máxima.	50
4.8. Ubicación de los puntos de cambio determinados con el método PELT.	50
4.9. Modelos ajustados después de detectar un sólo punto de cambio.	51
4.10. Ajuste de modelos después de aplicar el método SN.	51

Índice de figuras

1.1. Tipos de puntos de cambio.	5
2.1. Temperatura promedio de la Ciudad de México durante 2015 (la línea anaranjada corresponde al promedio y el área sombreada a la desviación estándar). Fuente: Calidad del aire en la Ciudad de México. http://www.aire.cdmx.gob.mx/descargas/publicaciones/flippingbook/informe_anual_calidad_aire_2015v3/files/downloads/Informe2015v3.pdf	12
3.1. Clasificación de las metodologías supervisadas para puntos de cambio en series temporales [1].	19
3.2. Clasificación de las metodologías no supervisadas para puntos de cambio en series temporales [1].	19
4.1. Ubicación geográfica del municipio de Tlaxco, Tlaxcala. Mapa extraído de INEGI.	40
4.2. Gráfico box-plot y serie de tiempo, para los registros semanales de temperatura máxima.	41
4.3. Serie de tiempo para los registros semanales.	41
4.4. Análisis de los residuos del modelo $ARIMA(3, 1, 0)$	43
4.5. Valores del estadístico $\Lambda_n(k)$, para $k = 1, \dots, 313$	44
4.6. Valores de $\widehat{\mathbf{B}}(k)$, $0 < k \leq n - 1$, para cada parámetro del modelo dado por (2.1).	45
4.7. Ubicación de los puntos de cambio para la media y la varianza de la temperatura máxima.	46
4.8. Ubicación de los puntos de cambio para la media de temperatura máxima.	48

4.9. Ubicación de los puntos de cambio para la varianza de temperatura máxima.	49
4.10. Segmentación de la serie temporal después de detectar un punto de cambio.	52
4.11. Series de tiempo particionadas después de aplicar el método SN.	54

Introducción

El pronóstico de series de tiempo tiene una historia extensa, su objetivo es recolectar y estudiar rigurosamente las observaciones pasadas de una serie temporal para desarrollar un modelo apropiado que pueda describir la estructura inherente de la serie, a fin de generar valores futuros. Bajo este enfoque, el área de investigación en series de tiempo es fructífera con muchas metodologías existentes, la más conocida es sin duda la metodología Box-Jenkins con los modelos autorregresivos integrados de promedios móviles (modelos ARIMA). Con la finalidad de contar con modelos apropiados, es necesario considerar la existencia de puntos de cambio y datos anómalos para brindar mejores modelos de predicción.

Desde el punto de vista de puntos de cambio, el problema es decidir si el modelo estadístico para una serie de observaciones no cambia o si el modelo cambia una o más veces, en este último caso, es necesario identificar cuándo se han producido los cambios y establecer modelos distintos para antes y después de haber detectado los cambios [24].

Se considera que los trabajos de Girshick [36] y Page [82], publicados en 1952 y 1954 respectivamente, sentaron las bases para el problema de puntos de cambio, el cual fue formulado hasta finales de la década de 1950 por A. N. Kolmogorov [6].

Aunque las primeras aplicaciones del problema de puntos de cambio se consideraron en el área de control de calidad, con el tiempo y la implementación de nuevas metodologías para detectar puntos de cambio tanto para el caso de variables aleatorias independientes como en el caso de variables aleatorias dependientes, las aplicaciones han permeado en áreas del conocimiento como climatología, finanzas, medicina, seguridad, genética y todas aquellas áreas del

conocimiento cuyo cambio en la estructura estadística de la información, repercute en la toma de decisiones [6, 56].

El calentamiento global, es la manifestación más evidente del cambio climático, se refiere al incremento promedio de la temperatura global, terrestre y marina [28], y se le atribuye al incremento en las emisiones de gases de efecto invernadero como CO_2 , CH_4 y N_2O . Si bien, los gases de efecto invernadero están presentes y se consumen de manera natural en la atmósfera, el consumo de combustibles fósiles para la producción de energía e insumos producto del desarrollo humano, provoca emisiones de estos gases de manera más rápida que en su ciclo natural [52, 79]. El Grupo Intergubernamental de Expertos sobre Cambio Climático (IPCC), estima impactos radicales que implican incrementos en la temperatura, aumento en el nivel del mar (derivado del deshielo de los polos) así como destrucción de los bienes y servicios ecosistémicos [53], por lo tanto, es importante monitorear y realizar análisis del comportamiento de datos climatológicos para establecer políticas, en caso de ser necesario, que coadyuven a disminuir los impactos al ambiente y a la sociedad.

Con base en los comentarios anteriores, se consideró pertinente realizar un análisis del comportamiento de la temperatura máxima en una región del estado de Tlaxcala, que cuenta con una estación climatológica que registra, entre otras variables, la temperatura máxima. Dicha estación está ubicada en el municipio de Tlaxco-Tlaxcala, es responsabilidad de la Comisión de Protección Civil del Estado de Tlaxcala y del Centro de Investigación del Cambio Climático de la Universidad Autónoma de Tlaxcala. Con los registros antes mencionados, se aplicaron metodologías para detectar uno y múltiples puntos de cambio y se compararon los resultados, para ello, se realizó un análisis de calidad de datos, se completaron los datos perdidos por regresión lineal y se ajustaron los modelos ARIMA. Para contrastar la bondad de ajuste de los modelos considerados, se consideraron el Criterio de Información de Akaike (AIC), así como el Criterio de Información Bayesiano (BIC).

Este trabajo está dividido en cuatro capítulos y dos apéndices. En el capítulo 1, se presentan los antecedentes del problema de puntos de cambio, la clasificación de los mismos así como las metodologías que se han utilizado para su solución. En el capítulo 2, se presentan los resultados más importantes de series de tiempo, los cuales serán necesarios para abordar las metodologías de puntos de cambio en series de tiempo. En el capítulo 3, se presentan las metodologías para detectar uno y múltiples puntos de cambio en series de tiempo, finalmente, en el capítulo 4, se presentan los resultados obtenidos del análisis de los datos y la aplicación de las metodologías descritas en el capítulo 3 y la teoría necesaria se presenta en el apéndice A. El apéndice B, contiene los programas desarrollados en R que se implementaron en el presente estudio.

CAPÍTULO 1

Preliminares sobre puntos de cambio

A partir de la revolución industrial que dio origen a la producción en masa, fue necesario implementar mecanismos que garantizaran procesos de producción satisfactorios, desde entonces, algunas técnicas puestas en práctica consisten en la revisión de los productos elaborados en intervalos regulares de tiempo para registrar medidas sobre la calidad del producto (cuando el producto no satisface los estándares de calidad establecidos, es necesaria la toma de decisiones, como parar el proceso de producción y reiniciar los equipos). En la década de 1950, Girshick y Page [36, 82] establecieron las primeras metodologías para detectar cambios en la estructura estadística de registros de un proceso de producción, a este problema se le conoce actualmente como el "*problema de puntos de cambio*", el cual, en sentido estricto, busca responder lo siguiente:

1. Dado un conjunto de observaciones muestrales, ¿Este conjunto es homogéneo en el sentido estadístico?
2. Si la respuesta es no, entonces ¿qué segmentos de homogeneidad existen en la muestra dada?

En este capítulo, se presentan los antecedentes del problema de puntos de cambio y algunas metodologías relevantes para su detección, para ello se define el concepto de punto de cambio y su clasificación, finalmente se abordan algunas metodologías relevantes.

1.1. Antecedentes del problema de puntos de cambio

Se considera que los trabajos de Girshik y Page publicados en 1952 y 1954 respectivamente, relacionados a encontrar cambios en un proceso de producción, sentaron las bases (con enfoques diferentes) para abordar un problema de punto de cambio en teoría de control de calidad, para ello, Girshick [36] modeló un proceso de producción usando la teoría de cadenas de Markov y Page [82] estableció el método de sumas acumuladas (CUSUM). Aunque ambos autores trabajaron con un problema de punto de cambio, éste no fue definido matemáticamente hasta finales de la década de 1950 por Kolmogorov [6].

Establecido el problema en sentido matemático, las metodologías propuestas para su solución se encaminaron en determinar cambios en la media y varianza de variables aleatorias independientes, principalmente de la distribución normal univariada con enfoques clásicos y no paramétricos, así lo muestran los trabajos de Chernoff y Zacks [12], Hinkley [47] y Lorden [73].

A finales de la década de 1970, se publicaron metodologías para detectar cambios en la varianza (dadas las aplicaciones directas en los modelos de un cambio en el stock del precio de mercado) [100], tiempo después, se establecieron metodologías para el caso de variables dependientes, en particular para series de tiempo.

La necesidad de plantear procedimientos más eficientes, dio paso a analizar el problema de puntos de cambio con enfoques como el análisis Bayesiano, análisis no paramétrico además de hacer uso de las herramientas computacionales aplicando algoritmos genéticos, entropía de la información, Wavelets o estadística difusa [19, 59, 62, 64, 76, 77, 106].

Existe una amplia bibliografía que se puede consultar para detectar puntos de cambio con distintos enfoques (Tabla 1.1).

1.2. Conceptos

Definición 1.1 Sea $\mathbf{X}_1, \dots, \mathbf{X}_n$ una sucesión de vectores aleatorios independientes cada uno con función de densidad $p(\mathbf{x}|\boldsymbol{\theta}_i)$, $\boldsymbol{\theta}_i$ desconocidos para $i = 1, 2, \dots, n$. Supóngase que $\boldsymbol{\theta}_1 \leq \boldsymbol{\theta}_2 \leq \dots \leq \boldsymbol{\theta}_n$, se dice que existe un punto de cambio en \mathbf{X}_i para $1 < i < n$, si existe un entero desconocido q con $1 < q \leq n - 1$, tal que $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_q < \boldsymbol{\theta}_{q+1} = \dots = \boldsymbol{\theta}_n$.

Definición 1.2 Sea $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ una sucesión de vectores (variables) alea-

Capítulo 1. Preliminares sobre puntos de cambio

Tabla 1.1: Bibliografía sobre puntos de cambio con distintos enfoques (*Elaboración propia*).

Tipo de problema	Biografía
Puntos de cambio múltiples	Battacharya [4], Csörgó and SzyszKowicz [16], Maboudou and Hawkins [75], Tartakovsky and Moustakides [97]
Pruebas de rango en problemas de puntos de cambio	Lombard [70], Hawkins [43], Elnaser <i>et al.</i> [30]
Principios de invarianza de puntos de cambio	Csörgó and Horváth [13], Csörgó and Horváth, [14], Tartakovsky [96]
Intervalos de confianza para variables aleatorias independientes	Worsley [103], Seigmund [89]
Problemas de puntos de cambio en modelos de regresión y cambio en la varianza	Chow [11], Sprent [94], Müller [78], Poluchenko <i>et al.</i> [85], Sen and Srivastava [90], Yong Zhou <i>et al.</i> [106], Davis [20]
Métodos no paramétricos	Darkhovski [21], Brodsky and Darkhovsky [6], Giriatis <i>et al.</i> [35], Nikiforov [81], Sharma <i>et al.</i> [93]
Problemas secuenciales	Bhattacharya [5], Brodsky and Darkhovsky [7], Lai and Xing [64], Gozdek <i>et al.</i> [40]
Problemas de puntos de cambio por medio de Análisis Bayesiano and Mohammad	Atherton <i>et al.</i> [2], Loschi <i>et al.</i> [72], Féron [76], Mostafa and Ghorbal [77], Giordani and Kohn [34], Lai and Xing [63]

torios independientes con funciones de distribución F_1, F_2, \dots, F_n respectivamente, el problema de puntos de cambio consiste en probar la hipótesis estadística

$$\begin{aligned}
 H_0 : F_1 = F_2 = \dots = F_n \\
 \text{vs} \\
 H_1 : F_1 = F_2 \dots = F_{k_1} \neq F_{k_1+1} = \dots = F_{k_2} \neq F_{k_2+1} \\
 \dots = F_{k_q} \neq F_{k_q+1} = \dots = F_n
 \end{aligned}$$

donde $1 < k_1 < k_2 < \dots < k_q < n$, q es el número desconocido de puntos de cambio en las posiciones k_1, k_2, \dots, k_q .

Si las funciones de distribución pertenecen a una familia paramétrica común $F(\boldsymbol{\theta})$, donde $\boldsymbol{\theta} \in \mathbb{R}^p$, entonces el problema de puntos de cambio consiste en probar la hipótesis

$$\begin{aligned}
 H_0 = \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_n \\
 \text{vs} \\
 H_1 = \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_{k_1} \neq \boldsymbol{\theta}_{k_1+1} \\
 \neq \boldsymbol{\theta}_{k_q-1} = \dots = \boldsymbol{\theta}_{k_q} \neq \boldsymbol{\theta}_{k_q+1} = \dots = \boldsymbol{\theta}_n
 \end{aligned} \tag{2.1}$$

donde q y k_1, k_2, \dots, k_q deben ser estimados.

Del juego de hipótesis (2.1) resaltan dos aspectos importantes de la inferencia en puntos de cambio:

- a) Determinar la existencia de puntos de cambio,
- b) Estimar las posiciones de los puntos de cambio.

Davis *et al.*, Gombay, Horváth *et al.* entre otros autores, han demostrado que en pruebas paramétricas, el estadístico de prueba utilizado converge a un puente Browniano o a un movimiento Browniano; por otra parte, al utilizar constantes de normalización, los mismos autores, han demostrado la convergencia exacta del estadístico de prueba a una función de distribución Gumbel, aunque con parámetros distintos [17, 37, 50].

Para Killick [60], dada una sucesión ordenada de datos $\mathbf{x} = (x_1, \dots, x_n)$ y suponiendo que existen m puntos de cambio en las posiciones $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$, con $\tau_0 = 0$ y $\tau_{m+1} = n$ (se asume que los puntos de cambio están ordenados como $\tau_i < \tau_j$ si y sólo si $i < j$), los m puntos de cambio, particionan los datos en $m + 1$ segmentos, donde el i -ésimo segmento contiene las observaciones $(x_{\tau_{i-1}+1}, \dots, x_{\tau_i})$, proponen minimizar

$$\sum_{1 \leq i \leq m+1} C(x_{\tau_{i-1}+1:\tau_i}) + \beta f(m),$$

donde $C(x_{\tau_{i-1}+1:\tau_i})$ es una función de costo para un segmento y $\beta f(m)$ es una penalización, para identificar los puntos de cambio.

1.3. Clasificación de puntos de cambio

Autores como Aminikhanghahi *et al.* [1] y Brodsky [6], clasifican los puntos de cambio en *a posteriori* (offline) y *secuenciales* (online) .

En un problema de punto de cambio *a posteriori*, se considera que la adquisición de la información se completa hasta el momento en que se verifica la homogeneidad; mientras que para el problema de punto de cambio *secuencial*, la hipótesis de homogeneidad de los datos se verifica de manera simultánea con el proceso de adquisición de la información

Se considera punto de cambio *abrupto*, cuando las propiedades estocásticas permanecen constantes por un periodo de tiempo y después cambian repentinamente a otro valor; un punto de cambio se considera *gradual*, cuando las propiedades estocásticas son (aproximadamente) constantes por un tiempo y luego alcanzan un punto en el tiempo donde comienza lentamente el cambio (Figura 1.1).

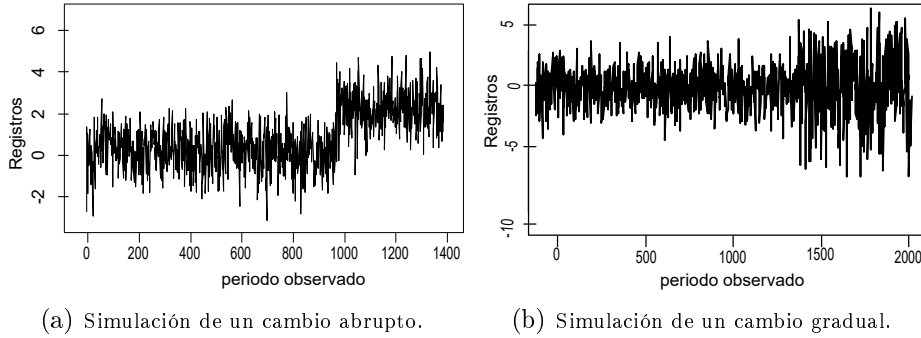


Figura 1.1: Tipos de puntos de cambio.

1.4. Metodologías relevantes para resolver el problema de puntos de cambio

1.4.1. Metodologías para detectar puntos de cambio en forma secuencial

Históricamente, los problemas de punto de cambio han venido desde el punto de vista secuencial. En esta sección se presentan algunas de las ideas más importantes del problema de puntos de cambio secuenciales. Si bien, la revisión no es exhaustiva, deja una idea clara de lo que buscan este tipo de metodologías. Se omiten los procedimientos bayesianos, por estar fuera del alcance de este manuscrito, sin embargo se pueden revisar los trabajos de Fearnhead y Liu [31] así como las referencias allí contenidas para una introducción a este punto de vista.

Cartas de control

Sea X_1, \dots, X_n observaciones secuenciales de un proceso aleatorio con valores reales que sufre un cambio en la media para cualquier $i \in \{1, 2, \dots, n\}$ y se define

$$\mu^* = \mathbb{E}(X_i).$$

La idea más inmediata para detectar cambios en la media es, sin duda, establecer un umbral y decidir si hay un cambio siempre que cierta estadística cruce el umbral. Para la elección de ese umbral, Shewhart en 1931, hizo el primer intento, suponiendo por ahora que la verdadera media de las observaciones es μ^* así como la varianza σ^2 , se elige un tamaño de muestra N y se elige

$$\bar{X}(K) = \frac{1}{N} \sum_{i=NK}^{N(K-1)+1} X_i, \quad K = 2, 3, \dots$$

El propósito es detectar un cambio tan pronto como

$$|\bar{X}(K) - \mu^*| > \kappa \frac{\sigma}{\sqrt{N}},$$

donde κ es una constante.

Es necesario tener en cuenta que en el contexto de control de calidad en el que se introdujeron gráficos de control, μ^* es un valor conocido prescrito por los artículos industriales de primera necesidad. Por otra parte, σ^2 puede estimarse fácilmente (si es desconocido), como

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_i)^2.$$

Sumas Acumuladas (CUSUM)

El método CUSUM propuesto por Page [82], consiste en tomar muestras de tamaño N para después asignar un puntaje (score) a la k -ésima muestra, sea $S_m = \sum_{t=1}^m Y(K)$, Page consideró la existencia de un punto de cambio si

$$S_m - \min_{0 \leq i < m} S_i \geq h,$$

con h una constante positiva. Aunque en Page [82] no está explícito cómo elegir los puntajes $Y(K)$ es común utilizar la razón del logaritmo de verosimilitud. Esto es, suponiendo que los X_i siguen una distribución F_θ con $\theta = \theta_0 \in \Theta$ antes del cambio y $\theta = \theta_1$ después del cambio, se establece

$$Y(K) = \frac{1}{N} \sum_{i=N(k-1)+1}^{NK} \log \left(\frac{F_{\theta_1}(X_i)}{F_{\theta_0}(X_i)} \right).$$

Aquí $\Theta = \{\theta : \theta \text{ es un valor tal que } F_\theta \text{ está definida}\}$ y $1 < m \leq N$.

1.4.2. Metodologías a posteriori para detectar puntos de cambio

Bajo el enfoque de determinar puntos de cambio a posteriori, la primera idea es interpretar el problema de punto de cambio como una prueba de hipótesis estadística, decidiendo la existencia de un punto de cambio si este resulta estadísticamente significativo. Otro enfoque de un problema a posteriori, es visualizar el problema de puntos de cambio como un problema de estimación, donde los puntos de cambio son parámetros del modelo que se desean estimar. Entre los métodos de estimación, las metodologías estudiadas hacen énfasis en

las estimaciones obtenidas a través de minimización por el criterio de mínimos cuadrados o por máxima verosimilitud.

Para cualquier entero n se llaman enteros de segmentación $1 \leq D \leq n$ y $\tau_0 = 0 < \tau_1 < \dots < \tau_{D-1} < \tau_D = n$ y se establecen los conjuntos de segmentación $\{1, \dots, \tau_1\}, \{\tau_1 + 1, \dots, \tau_2\}, \dots, \{\tau_{D-1} + 1, \dots, \tau_D = n\}$. Dadas las observaciones X_1, \dots, X_n , se dice que hay una segmentación τ^* con D^* segmentos tales que la distribución de X_i es constante para los segmentos, pero distinta para segmentos consecutivos (en particular la media $\mu_i^* = \mathbb{E}(X_i)$ es constante sobre los segmentos), otra suposición paramétrica es que las observaciones X_i tienen función de densidad f_θ , con $\theta = \theta_l$ sobre el segmento $l + 1$ el cual está dado por el segmento $\{\tau_l + 1, \dots, \tau_{l+1}\}$.

Pruebas de hipótesis

Una primera idea para detectar un único punto de cambio es considerar este problema como una prueba de hipótesis. Se puede probar la hipótesis

$$H_0 : F_1 = F_2 = \dots = F_n$$

vs

$$H_1 : \exists t \in \{1, 2, \dots, n - 1\}, \text{ tal que } F_1 = F_2 \dots = F_t \neq F_{t+1}, \dots, F_n.$$

CUSUM y extensiones

Sea $\mathcal{L}(X_i) \sim f_\theta$ para alguno $\theta \in \Theta$ (θ puede depender de i ya que los X_i pueden no estar distribuidos de manera idéntica). La situación más simple, es cuando se conocen los parámetros iniciales y finales θ_0 y θ_1 respectivamente, bajo este esquema, la idea es construir una estadística de prueba a partir de la función de probabilidad asociada con las observaciones X_1, \dots, X_n . En Page [83], se propone elegir el H_t que maximiza la función de verosimilitud de la hipótesis, es decir, τ_1 es estimada por

$$\hat{\tau}_1 \in \arg \max_{1 \leq t \leq n} \left\{ \sum_{j=1}^t \log f_{\theta_0}(x_j) + \sum_{j=t+1}^n \log f_{\theta_1}(x_j) \right\}.$$

En el caso de un cambio unilateral en la media de las observaciones gaussianas con una varianza conocida σ^2 , Page propuso la suma acumulativa del algoritmo CUSUM

$$S_t := \sum_{j=1}^t (X_j - \theta_0 + \delta\sigma),$$

y rechazar H_0 si $S_n - \max_{t < n} S_t < -h$, donde h es un número positivo y se elige $\hat{\tau}_1$ como el primer t tal que $S_n - \max_{t < n} S_t \geq 0$ y δ es una constante. Las ideas

de Page fueron estudiadas por Hinkley [47] quien logró obtener la distribución asintótica del estimador bajo la suposición Gaussiana.

Si el parámetro inicial θ_0 se desconoce (pero la varianza σ^2 es conocida), Sen y Srivastava [90] propusieron la estadística

$$U^* = \frac{1}{n^2} \sum_{i=1}^{n-1} \left(\sum_{j=1}^{n-1} (X_{j+1} - \bar{X}) \right)^2,$$

obteniendo la función de distribución acumulada exacta de U^* .

Prueba de la razón de verosimilitud

Otra posibilidad como estadística de prueba, es utilizar la estadística de razón de verosimilitud. La idea es utilizar como estadística el logaritmo de la razón entre la verosimilitud bajo la hipótesis alternativa (H_1) y la verosimilitud bajo la hipótesis nula (H_0). Suponiendo σ conocida, sea

$$Q_k = \sum_{j=1}^k (X_j - \bar{X}_{1:k})^2 + \sum_{j=k+1}^n (X_j - \hat{X}_{(k+1):n})^2,$$

donde $\bar{X}_{a:b}$ denota la media de las $b - a$ observaciones X_{a+1}, \dots, X_b , luego el estadístico de prueba de razón de verosimilitud para la media desconocida de observaciones Gaussianas antes y después del cambio está dada por $Q_n - Q_{k^*}$, donde k^* es la posición del punto de cambio. En Garreau D. [27], se puede consultar la descripción de los procedimientos de estimación los cuales incluyen la estimación por máxima verosimilitud y la estimación por mínimos cuadrados.

1.5. Áreas de aplicación

Las áreas de aplicación de los puntos de cambio son muy amplias; a lo largo del tiempo, los investigadores han buscado mejorar las metodologías propuestas para detectar el punto de cambio; el objetivo es reducir el costo computacional; en este sentido, se pueden encontrar artículos de la teoría difusa que se han utilizado para analizar puntos de cambio en series de tiempo no lineales. En 2001, Kumar y Wu [62], propusieron una metodología estadística difusa para detectar puntos de cambio en series de tiempo a través de un concepto conocido como entropía difusa, además, los autores compararon la metodología con otras estadísticas paramétricas. Finalmente, el método de Segmentación binaria se ha mejorado recientemente por un nuevo método propuesto por Fryzlewicz [32].

En climatología, Jarušková [55], describió una metodología basada en el logaritmos de verosimilitud en busca de un punto de cambio en un modelo de

regresión aplicado a promedios mensuales de la presión del aire. El autor asumió la normalidad y la independencia para determinar la distribución de la prueba estadística, este autor hace énfasis sobre la dificultad de aplicar las nuevas metodologías de detección de puntos de cambio, ya que solo se conocen y comparten dentro de la comunidad matemática y rara vez se transmiten a otras áreas de conocimiento. Horváth *et al.* [50], propusieron una estadística basada en la metodología de relación de probabilidad y determinan puntos de cambio en los promedios mensuales de las temperaturas de Praga. Otros documentos sobre detección de puntos de cambio en climatología se pueden consultar en [10, 26, 74]. Los documentos de Csörgó [14] e Inclán y Tiao [51], se utilizaron en el trabajo de Wei *et al.* [102], para detectar cambios en el problema de recuperación de imágenes. Xu *et al.* [104], asumieron un modelo lineal y usaron la prueba CUSUM y el método PELT propuesto en [60], para encontrar cambios en vigilancia post mercado. En medicina, se han utilizado métodos bayesianos para detectar el punto de cambio en la emergencia diaria de enfermedades similares a la influenza [58]. Para detectar cambios en los datos del electrocardiograma, Qi JP *et al.* [107], usaron las ideas de Brodsky [6], Darkhovsky [21] y Wang *et al.* [101]. Técnicas para detectar el punto de cambio también se han aplicado en problemas de sueño, epilepsia, interpretación de imágenes de resonancia magnética, así como en la comprensión de las actividades cerebrales.

Los métodos de detección de puntos de cambio se han aplicado en el reconocimiento de voz para la segmentación de audio y en el reconocimiento de los límites entre las palabras de silencio y el ruido. En el área de análisis de imágenes, la detección de eventos abruptos como la seguridad se puede formular como un problema de punto de cambio [86].

En el área económica, existen varios artículos que utilizan metodologías de detección de puntos de cambio, Joo-Oh y Han [57], aplicaron metodologías paramétricas y no paramétricas de detección de puntos de cambio, así como redes neuronales para agrupar el sistema de pronóstico en economía financiera, Takayasu [95], utiliza la teoría del punto de cambio para modelar series de tiempo del mercado financiero, otros trabajos en esta área, se pueden consultar en [33, 84].

CAPÍTULO 2

Series temporales

Para fines de análisis, es importante recopilar información y concentrarla en bases de datos. Contar con bases de datos actualizadas es muy importante, pues de su análisis es posible lograr planeaciones y la toma de decisiones acertadas. En este capítulo se abordan los conceptos básicos de series de tiempo para después abordar el problema de puntos de cambio en series de tiempo.

2.1. Conceptos

Definición 2.1 *Un Proceso Estocástico es una familia de variables aleatorias asociadas a un conjunto índice de números reales de forma tal que cada elemento del conjunto le corresponda una y sólo una variable aleatoria y se denota como $\{X_t; t \in T\}$, donde T es el conjunto índice y X_t es la variable aleatoria correspondiente al elemento t de T .*

Si T es un intervalo de números reales abierto o cerrado, se dice que el proceso es continuo, si T es un conjunto finito o infinito numerable, se dice que T es discreto [41].

Definición 2.2 *Se define una serie de tiempo como la sucesión de observaciones generadas por un proceso estocástico $\{X_t\}$ cuyo conjunto índice se toma en relación al tiempo.*

Definición 2.3 *Una serie de tiempo discreta es aquella en la que el conjunto índice T sobre el que se hacen las observaciones es un conjunto discreto (por ejemplo cuando las observaciones se hacen en intervalos regulares de tiempo).*

En la Figura 2.1, se presenta un ejemplo de una serie de tiempo.

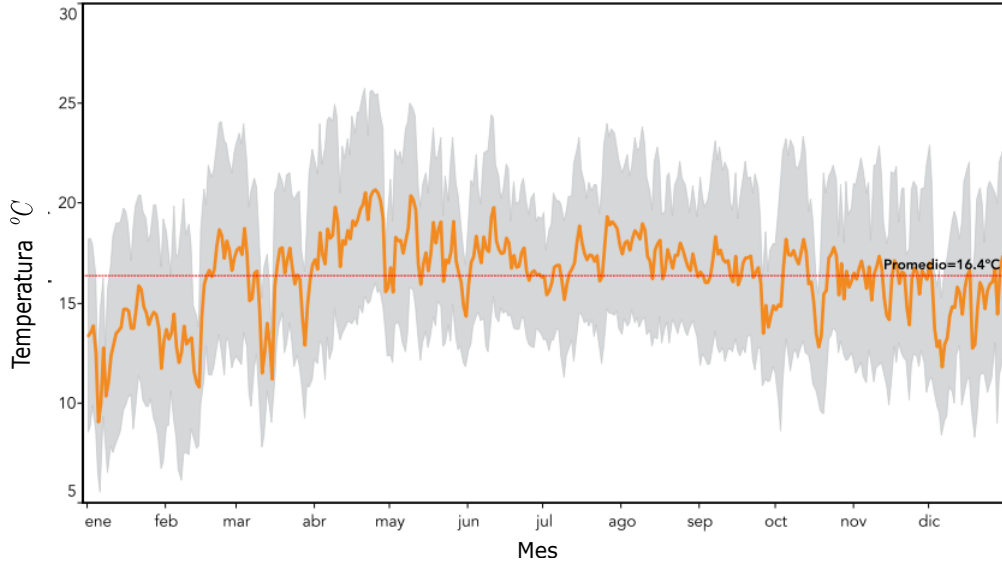


Figura 2.1: Temperatura promedio de la Ciudad de México durante 2015 (la línea anaranjada corresponde al promedio y el área sombreada a la desviación estándar). Fuente: Calidad del aire en la Ciudad de México. http://www.aire.cdmx.gob.mx/descargas/publicaciones/flippingbook/informe_anual_calidad_aire_2015v3/files/downloads/Informe2015v3.pdf.

Definición 2.4 Si $\{X_t\}$ es un proceso tal que $E(X_t^2) < \infty$ para cada $t \in T$, entonces la función de autocovarianza $\gamma_X(\cdot, \cdot)$ de $\{X_t\}$ se define por

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - E[X_r])(X_s - E[X_s])], \quad r, s \in T.$$

Definición 2.5 La serie de tiempo $\{X_t\}$ es débilmente estacionaria si

1. $E|X_t^2| < \infty$.
2. $E[X_t] = m$, para todo $t \in \mathbb{Z}$.
3. $\gamma_X(r, s) = \gamma_X(r + t, s + t)$, para todo $r, s, t \in \mathbb{Z}$.

Observación 2.1 Si $\{X_t\}$ es estacionaria, entonces $\gamma_X(r, s) = \gamma_X(r - s, 0)$ para todo $r, s \in \mathbb{Z}$. En efecto, por ser $\{X_t\}$ estacionaria, se cumple la Condición 3 de la Definición 2.5, luego tomando $t = -s$, se tiene $\gamma_X(r, s) = \gamma_X(r - s, 0)$. Por lo tanto, es conveniente redefinir la función de autocovarianza de un proceso como función de una sola variable.

Definición 2.6 Si $\{X_t\}$ es una serie de tiempo débilmente estacionaria, se define

$$\gamma_X(h) = \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t), \quad \forall h \in \mathbb{Z},$$

como la función de autocovarianza de $\{X_t\}$ y a

$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} ,$$

como la función de autocorrelación de $\{X_t\}$.

Observación 2.2 Si $\{X_t\}$ es una serie de tiempo débilmente estacionaria, entonces la función de autocovarianza es independiente de t , y está dada por $Cov(X_{t+h}, X_t)$.

Proposición 2.1 La función $\gamma_x(\cdot)$ tiene las siguientes propiedades:

1. $\gamma_x(0) \geq 0$.
2. $|\gamma_x(h)| \leq \gamma_x(0)$.
3. $\gamma_x(h) = \gamma_x(-h)$.

Demostración:

1. $\gamma_x(0) = E[(X_t - E[X_t])^2] = Var(X_t) \geq 0$.
2. $|\frac{\gamma_x(h)}{\gamma_x(0)}| \leq 1 \iff |\gamma_x(h)| \leq \gamma_x(0)$, si $\gamma_x(0) > 0$.
3. $\gamma_x(h) = Cov(X_{t+h}, X_t) = Cov(X_t, X_{t+h}) = \gamma_x(-h)$.

Definición 2.7 Una función κ con valores reales definida sobre los enteros, es definida no negativa si

$$\sum_{i,j=1}^n a_i \kappa(i-j) a_j \geq 0 .$$

Para todo entero positivo n y vectores $\mathbf{a} = (a_1, a_2, \dots, a_n)'$.

Teorema 2.1 Una función con valores reales definida sobre los enteros es la función de autocovarianza de una serie de tiempo estacionaria si y sólo si es par y definida no negativa.

Demostración: \Rightarrow) Sea $\gamma_x(\cdot)$ una función tal que $\gamma_x(\cdot)$ es la función de autocovarianza de una serie de tiempo $\{X_t\}$, por propiedades de la función, $\gamma_x(\cdot)$ es par. Sea $\mathbf{a} = (a_1, \dots, a_n)'$ cualquier vector con componentes reales, $\mathbf{X}_n = (X_1, \dots, X_n)'$ y $\Gamma_n = (\sigma_{ij})$, $i, j = 1, 2, \dots, n$, la matriz de covarianza de

\mathbf{X}_n .

$$\text{Var}(\mathbf{a}'\mathbf{X}_n) = \mathbf{a}'\Gamma_n\mathbf{a} = \sum_{i,j=1}^n a_i\gamma(i-j)a_j \geq 0 ,$$

de lo anterior, $\gamma_x(\cdot)$ es definida no negativa.

\Leftrightarrow) Sea $\kappa : \mathbb{Z} \rightarrow \mathbb{R}$ una función par y definida no negativa. Se debe demostrar que existe un proceso estacionario con $\kappa(\cdot)$ como su función de autocovarianza, para esto se usará el Teorema de Kolmogorov ([9] p.11). Para cada entero n y cada $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{Z}^n$ tal que $t_1 < t_2 < \dots < t_n$. Sea $F_{\mathbf{t}}$ la función de distribución sobre \mathbb{R}^n con función característica $\Phi_{\mathbf{t}}(\mathbf{u}) = e^{\mathbf{u}'\mathbf{K}\mathbf{u}/2}$ donde $\mathbf{u}' = (u_1, \dots, u_n) \in \mathbb{R}^n$ y $\mathbf{K} = [\kappa(t_i - t_j)]_{i,j=1}^n$, como κ es definida no negativa, la matriz \mathbf{K} también es definida no negativa y consecuentemente $\phi_{\mathbf{t}}$ es la función característica de una distribución multinormal con n variables con media cero y matriz de covarianza \mathbf{K} , luego utilizando la notación del Teorema de Kolmogorov,

$$\lim_{\mathbf{u}_{\mathbf{t}} \rightarrow \mathbf{0}} \phi_{\mathbf{t}}(\mathbf{u}) = \phi_{t(i)}(u(i))$$

es decir, las funciones $F_{\mathbf{t}}$ son consistentes, luego por el Teorema de Kolmogorov ([9] p.11), existe una serie de tiempo $\{X_t\}$ con funciones de distribución $F_{\mathbf{t}}$ y funciones características $\phi_{\mathbf{t}}, t \in \mathcal{T}$. En particular la distribución conjunta de X_i y X_j es la distribución normal bivariada con media $\mathbf{0}$ y matriz de covarianza

$$\begin{bmatrix} \kappa(0) & \kappa(i-j) \\ \kappa(i-j) & \kappa(0) \end{bmatrix} ,$$

lo cual demuestra que $\text{Cov}(X_i, X_j) = \kappa(i-j)$.

Definición 2.8 *La serie de tiempo $\{X_t\}$ se dice estacionaria estricta si las distribuciones conjuntas de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ son las mismas para todos los enteros positivos k y para todo $t_1, \dots, t_k, h \in \mathbb{Z}$.*

Proposición 2.2 *Una serie de tiempo estrictamente estacionaria, tiene las siguientes propiedades*

1. *Las variables aleatorias $\{X_t\}$ son idénticamente distribuidas.*
2. *$(X_t, X_{t+h})' = (X_1, X_{1+h})'$ en distribución, $\forall t \in \mathbb{Z}, h \in \mathbb{Z}$.*
3. *Si $E(X_t^2) < \infty$, $\{X_t\}$ es débilmente estacionaria.*
4. *Una sucesión independiente e idénticamente distribuida (iid) es estacionaria estricta.*

Demostración: Sea $\{X_t\}$ una serie estacionaria estricta.

1. Por la definición de estacionaridad estricta, las variables aleatorias $\{X_t\}$ son idénticamente distribuidas.
2. Por la Propiedad 1, $(X_t, X_{t+h})' = (X_{t+h'}, X_{t+h+h'})'$ en distribución, en particular si $h' = 1 - t$, entonces, $(X_t, X_{t+h})' = (X_{t+1-t}, X_{t+h+1-t})' = (X_1, X_{1+h})'$.
3. Supongamos que $E(X_t^2) < \infty$, $\forall t \in \mathbb{Z}$, como $\{X_t\}$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas, $E(X_t)$ es independiente de t y por la Propiedad 2, $Cov(X_t, X_{t+h}) = Cov(X_1, X_{1+h})$ también es independiente de t , de donde se tiene que $\{X_t\}$ es débilmente estacionaria.
4. Sea $\{X_t\}$ una sucesión *iid* de variables aleatorias con distribución común F , entonces la distribución conjunta de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ claramente son las mismas, independientemente de h , esto es, la sucesión es estacionaria estricta.

2.2. Media, función de autocovarianza y autocorrelación muestrales

El primer momento muestral de un proceso estacionario es la media muestral

$$\bar{X}_n = n^{-1}(X_1 + \dots + X_n) ,$$

el cual es un estimador insesgado de la media μ del proceso. El error cuadrado medio de \bar{X}_n es

$$\begin{aligned} E(\bar{X}_n - \mu)^2 &= Var(\bar{X}_n) \\ &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) \\ &= n^{-2} \sum_{i-j=1}^n (n - |i - j|) \gamma(i - j) \\ &= n^{-1} \sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \gamma(h) , \end{aligned}$$

Ahora, si $\gamma(h) \rightarrow 0$ cuando $h \rightarrow \infty$, $E(\bar{X}_n - \mu)^2$ converge a 0 y si

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty ,$$

entonces $\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \sum_{|h| < \infty} \gamma(h)$.

Estimación de $\gamma(\cdot)$ y $\rho(\cdot)$

Se define la función de autocovarianza muestral y autocorrelación muestral como

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n) (X_t - \bar{X}_n)$$

y

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

respectivamente, en donde n es el tamaño de la muestra. Finalmente, se define la matriz de autocovarianza muestral como la matriz de tamaño k definida por

$$\hat{\Gamma}_k = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(k-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(k-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(k-1) & \hat{\gamma}(k-2) & \cdots & \hat{\gamma}(0) \end{bmatrix},$$

2.3. Modelos ARIMA

Los procesos ARIMA son una generalización de los modelos ARMA, son útiles cuando los procesos ARMA son no estacionarios. El objetivo de los procesos ARIMA es transformar un proceso ARMA no estacionario en estacionario mediante diferencias [48, 71].

Definición 2.9 *Un proceso autorregresivo e integrado de promedios móviles ARIMA(p, d, q) se define como*

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d y_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t$$

donde $p, d, q \in \mathbb{Z}, p, d, q \geq 0$ hacen referencia a la parte autorregresiva, integrada y de promedios móviles respectivamente y L es el operador diferencia.

Un proceso ARIMA($p, 0, 0$) es un modelo AR(p), el modelo ARIMA($0, 0, q$) es un proceso MA(q), finalmente, si $d = 0$, un proceso ARIMA(p, d, q) se transforma en un modelo ARMA(p, q).

En este trabajo se omite la descripción de los modelos ARMA así como el procedimiento para su ajuste. En [46], se puede consultar la metodología Box-Jenkins para el ajuste de modelos y pronóstico de series temporales estacionarias.

CAPÍTULO 3

Puntos de cambio en series temporales

Las metodologías utilizadas para detectar puntos de cambio en el caso de variables independientes no se pueden aplicar de manera directa en series de tiempo, debido a la dependencia entre las variables. Para el caso de series temporales, para determinar puntos de cambio, se han desarrollado metodologías paramétricas, no paramétricas y Bayesianas cada una con sus desarrollos particulares e importantes; por ejemplo, las metodologías paramétricas han utilizado la estadística de razón de verosimilitud; los métodos Bayesianos, han utilizado la distribución uniforme como distribución apriori, otras metodologías incluyen la aplicación de algoritmos genéticos, entropía de la información, estadística difusa, el uso de waveletes y recientemente se ha propuesto el método de sumas móviles (MOSUM) [2, 3, 18, 19, 22, 23, 29, 34, 38, 42, 63, 65, 66, 67, 68, 69, 72, 76, 77, 92, 99, 106, 108].

En este capítulo, se da una breve descripción de las metodologías paramétricas para determinar puntos de cambio en series temporales, poniendo énfasis en los trabajos desarrollados por Davis *et al.* [17], Gombay [37] y los métodos Pruned Exact Linear Time (PELT), Segment Neighbourhood (SN) y Binary segmentation (BS), que fueron aplicadas para el desarrollo del capítulo cuatro.

3.1. Metodologías supervisadas y no supervisadas para detectar puntos de cambio en series de tiempo

Las metodologías aplicadas para determinar puntos de cambio en series de tiempo, autores como Aminikhanghahi y Cook [1] las clasifican en metodologías supervisadas y no supervisadas. A continuación se da una breve descripción de estas metodologías.

3.1.1. Metodologías supervisadas y no supervisadas

Los algoritmos de aprendizaje supervisados son algoritmos de aprendizaje automático que aprenden una asignación de información de entrada a un atributo objetivo de los datos, que generalmente es una etiqueta de clase. Si bien este enfoque tiene una fase de capacitación más simple, se debe proporcionar una cantidad y diversidad suficiente de datos de aprendizaje para representar a todas las clases. Por otro lado, detectar cada clase por separado proporciona suficiente información para encontrar tanto la naturaleza como la cantidad de cambios. Una alternativa es tratar la detección del punto de cambio como un problema de clase binaria, donde todas las posibles secuencias de transición de estado (punto de cambio) representan una clase y todas las secuencias dentro del estado representan una segunda clase. Mientras solo dos clases necesitan ser aprendidas. En este caso, este es un problema de aprendizaje mucho más complejo si el número de posibles tipos de transiciones es grande. Al igual que con el tipo anterior de metodologías supervisadas, en este aprendizaje cada enfoque en el vector de entrada indica una fuente de posible cambio. Por lo tanto, cualquier algoritmo de aprendizaje supervisado que genere un modelo interpretable (como un árbol de decisión o un aprendiz de reglas) no solo identificará un cambio sino que también describirá la naturaleza del cambio [1].

Los algoritmos de aprendizaje no supervisados se usan típicamente para descubrir patrones en datos no etiquetados. En el contexto de la detección de puntos de cambio, dichos algoritmos se pueden usar para segmentar datos de series de tiempo encontrando así puntos de cambio basados en características estadísticas de los datos. La segmentación no supervisada es atractiva porque puede manejar una variedad de situaciones diferentes sin requerir capacitación previa para cada situación. Los primeros métodos informados utilizan la razón de probabilidad basada en la observación de que la densidad de probabilidad de dos intervalos consecutivos es la misma si pertenecen al mismo estado. Otra solución tradicional es el modelado del subespacio, que presenta una serie temporal que utiliza espacios de estado y, por lo tanto, detecta puntos de cambio al predecir los parámetros del espacio de estado. Los métodos probabilísticos

estiman las distribuciones de probabilidad del nuevo intervalo en función de los datos que se han observado desde el candidato a punto de cambio anterior. Finalmente, los métodos agrupan los datos de series de tiempo en sus respectivos estados y encuentran cambios al identificar diferencias entre las características de los estados. Las Figuras 3.1 y 3.2 presentan un resumen de las metodologías que se pueden utilizar para detectar puntos de cambio en series de tiempo.

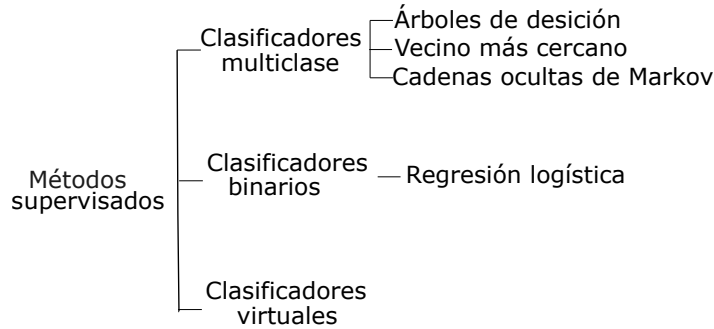


Figura 3.1: Clasificación de las metodologías supervisadas para puntos de cambio en series temporales [1].

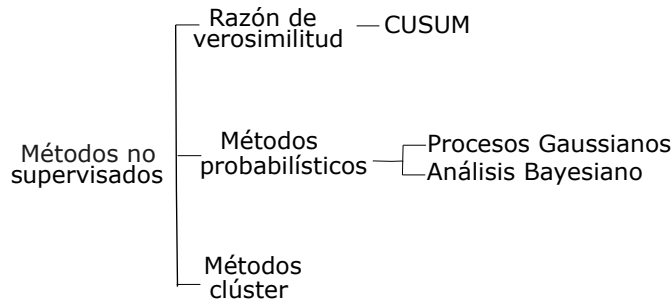


Figura 3.2: Clasificación de las metodologías no supervisadas para puntos de cambio en series temporales [1].

Dentro de las metodologías no supervisadas, se encuentran los métodos paramétricos. En la siguiente sección se aborda el problema de puntos de cambio desde este punto de vista.

3.2. Metodologías paramétricas para detectar puntos de cambio en series de tiempo

En esta sección se pone énfasis en las metodologías de Davis *et al.* [17] y Gombay [37], las cuales se aplican en el Capítulo 4.

Sea X_1, \dots, X_n , n observaciones consecutivas del proceso

$$X_t = \begin{cases} \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t; & -\infty < t \leq \tau, \\ \alpha_0 + \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t, & t \geq \tau + 1; \end{cases} \quad (2.1)$$

donde $\tau \in (p, n]$, $\{\varepsilon_t\}$ es una sucesión de ruido blanco, con cuarto momento finito, es decir

$$E(\varepsilon_i) = 0, \quad (2.2)$$

$$E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & \text{si } i = j; \\ 0, & \text{si } i < j. \end{cases} \quad (2.3)$$

$$E(\varepsilon_i \varepsilon_j \varepsilon_k) = \begin{cases} \mu_3, & \text{si } i = j = k; \\ 0, & \text{en otro caso.} \end{cases} \quad (2.4)$$

$$E(\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l) = \begin{cases} \mu_4, & \text{si } i = j = k = l; \\ \sigma^4, & \text{si } i = j < k = l; \\ 0, & \text{en otro caso.} \end{cases} \quad (2.5)$$

Además, $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0$ para todo z , tal que $|z| \leq 1$, se quiere probar

$$H_0 : \tau = n \text{ vs } H_1 : \tau < n. \quad (2.6)$$

Bajo la hipótesis nula, la razón de verosimilitud Gaussiana condicionada sobre las primeras p observaciones después de considerar menos dos veces el logaritmo de la función de verosimilitud y suponiendo que $\sigma^2 = 1$, queda expresada como:

$$\begin{aligned} \Lambda_n(k) &= \min_{\phi} \sum_{t=p+1}^n (X_t - \phi_0 - \sum_{j=1}^p \phi_j X_{t-j})^2 \\ &\quad - \min_{\phi} \sum_{t=p+1}^k (X_t - \phi_0 - \sum_{j=1}^p \phi_j X_{t-j})^2 \\ &\quad - \min_{\alpha} \sum_{t=k+1}^n (X_t - \alpha_0 - \sum_{j=1}^p \alpha_j X_{t-j})^2. \end{aligned} \quad (2.7)$$

La expresión (2.7), puede ser reescrita de manera matricial de la siguiente manera:

$$\Lambda_n(k) := Q_1 - Q_2 - Q_3,$$

donde

$$\begin{aligned}
 Q_1 &= \mathbf{X}_n^T \mathbf{X}_n - \mathbf{X}_n^T \mathbf{M}_n (\mathbf{M}_n^T \mathbf{M}_n)^{-1} \mathbf{M}_n^T \mathbf{X}_n, \\
 Q_2 &= \mathbf{X}_k^T \mathbf{X}_k - \mathbf{X}_k^T \mathbf{M}_k (\mathbf{M}_k^T \mathbf{M}_k)^{-1} \mathbf{M}_k^T \mathbf{X}_k, \\
 Q_3 &= \tilde{\mathbf{X}}_k^T \tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_k^T \tilde{\mathbf{M}}_k (\tilde{\mathbf{M}}_k^T \tilde{\mathbf{M}}_k)^{-1} \tilde{\mathbf{M}}_k^T \tilde{\mathbf{X}}_k, \\
 \mathbf{X}_k &= (X_{p+1}, \dots, X_k)^T, \\
 \tilde{\mathbf{X}}_k &= (X_{k+1}, \dots, X_n), \\
 \mathbf{M}_k &= \begin{bmatrix} 1 & X_p & X_{p-1} & \cdots & X_1 \\ 1 & X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{k-1} & X_{k-2} & \cdots & X_{k-p} \end{bmatrix}, \quad \tilde{\mathbf{M}}_k = \begin{bmatrix} 1 & X_k & X_{k-1} & \cdots & X_{k-p+1} \\ 1 & X_{k+1} & X_k & \cdots & X_{k-p+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n-1} & X_{n-2} & \cdots & X_{n-p} \end{bmatrix}.
 \end{aligned}$$

Al escribir

$$\boldsymbol{\varepsilon}_n = (\varepsilon_{p+1}, \dots, \varepsilon_n)^T \text{ y } \boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)^T,$$

se tiene que bajo H_0 ,

$$\mathbf{X}_n = \mathbf{M}_n \boldsymbol{\phi} + \boldsymbol{\varepsilon}_n,$$

por tanto

$$\begin{aligned}
 Q_1 &= \boldsymbol{\varepsilon}_n^T \boldsymbol{\varepsilon}_n - \boldsymbol{\varepsilon}_n^T \mathbf{M}_n (\mathbf{M}_n^T \mathbf{M}_n)^{-1} \mathbf{M}_n^T \boldsymbol{\varepsilon}_n, \\
 Q_2 &= \boldsymbol{\varepsilon}_k^T \boldsymbol{\varepsilon}_k - \boldsymbol{\varepsilon}_k^T \mathbf{M}_k (\mathbf{M}_k^T \mathbf{M}_k)^{-1} \mathbf{M}_k^T \boldsymbol{\varepsilon}_k, \\
 Q_3 &= \tilde{\boldsymbol{\varepsilon}}_k^T \tilde{\boldsymbol{\varepsilon}}_k - \tilde{\boldsymbol{\varepsilon}}_k^T \tilde{\mathbf{M}}_k (\tilde{\mathbf{M}}_k^T \tilde{\mathbf{M}}_k)^{-1} \tilde{\mathbf{M}}_k^T \tilde{\boldsymbol{\varepsilon}}_k.
 \end{aligned}$$

Considérese $\mathbf{S}_k = \mathbf{M}_k^T \boldsymbol{\varepsilon}_k$, $\mathbf{P}_k = (\mathbf{M}_k^T \mathbf{M}_k)^{-1}$ y $\tilde{\mathbf{P}}_k = (\tilde{\mathbf{M}}_k^T \tilde{\mathbf{M}}_k)^{-1}$, así

$$\Lambda_n(k) = -\mathbf{S}_n^T \mathbf{P}_n \mathbf{S}_n + \mathbf{S}_k^T \mathbf{P}_k \mathbf{S}_k + (\mathbf{S}_n - \mathbf{S}_k)^T \tilde{\mathbf{P}}_k (\mathbf{S}_n - \mathbf{S}_k). \quad (2.8)$$

Teorema 3.1 *Sea $\{X_t\}$ un proceso autorregresivo de orden p (AR(p)) que satisfice*

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad -\infty < t < \infty,$$

donde $\{\varepsilon_t\}$ satisfice las condiciones (2.2), (2.3), (2.4) y (2.5) con $\sigma^2 = 1$ y $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ para algún $0 < \delta \leq 1$. Además supóngase que $\{X_t\}$ es ρ -dependiente, con ρ la correlación de las observaciones y $\rho(n) \ll n^{-(1+\varepsilon)(1+4/\delta)}$ para algún $\varepsilon > 0$. Entonces para la estadística de razón de verosimilitud $Q_n(t) = \Lambda_n([nt])$, donde $\Lambda_n(k)$ está definida en (2.8), se tiene que para $0 < t_1 < t_2 < 1$,

$$Q_n(t) \xrightarrow{d} \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)}, \quad (2.9)$$

en $D[t_1, t_2]$, donde $\mathbf{W}(t)$ es un movimiento Browniano $(p+1)$ -dimensional y $D[0, 1]$ es el espacio de funciones reales sobre $[0, 1]$ continuas por la derecha y con límite por la izquierda.

Demostración: Considérese $Q_n(t) := \Lambda_n([nt])$ ($[nt]$ representa la parte entera de nt) para $0 \leq t \leq 1$. Supóngase también que $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ para algún $0 < \delta \leq 1$ y que el proceso está fuertemente relacionado con $\rho(n) \ll n^{-(1+\epsilon)(1+4/\delta)}$ para algún $\epsilon > 0$.

\mathbf{S}_n es la suma parcial de $n-p$ términos no correlacionados de una sucesión $p+1$ dimensional débilmente estacionaria fuertemente relacionada con la matriz de covarianzas $\Gamma_{p+1} = [\gamma_{ij}]_{i,j=1,2,\dots,p}$, donde

$$\gamma_{ij} = \begin{cases} 1, & \text{si } i = j = 1, \\ E(X_i), & \text{si } i = 1, j > 1 \text{ ó } i > 1, j = 1, \\ E(X_i X_j), & \text{si } i > 1, j > 1. \end{cases}$$

Sea $\xi_n = S_n - S_{n-1}$, aplicando el Teorema A.12 del apéndice A, existe una sucesión $\{\mathbf{Z}_n\}$ de vectores aleatorios Guassianos, definidos posiblemente sobre un nuevo espacio de probabilidad con media cero y matriz de covarianza Γ_{p+1} tal que

$$\mathbf{S}_k - \mathbf{U}_k = O(k^{1/2-\lambda}) \text{ conforme } k \rightarrow \infty \text{ a.s.}, \quad (2.10)$$

para algún $\lambda > 0$, donde $\mathbf{U}_k = \sum_{i=p+1}^k \mathbf{Z}_i$, por tanto

$$\frac{1}{\sqrt{n}} \Gamma_{p+1}^{1/2} \mathbf{S}_{[n]} \xrightarrow{d} \mathbf{W}(\cdot) \quad (2.11)$$

en $D^{p+1}[0, 1]$, donde $\mathbf{W}(\cdot)$ es un movimiento Browniano $p+1$ -dimensional con matriz de covarianza I_{p+1} .

Se puede ver que la componente (i, j) ($i, j > 1$) de P_n^{-1} es $\sum_{s=1}^{n-p} X_{p+1-i+s} X_{p+1-j+s}$, aplicando el Teorema A.12 del apéndice A, se tiene

$$\sum_{s=1}^{n-p} (X_{p+1-i+s} X_{p+1-j+s} - E(X_i X_j)) - V(n-p) = O(n^{1/2-\lambda}) \text{ conforme } n \rightarrow \infty \text{ a.s.}, \quad (2.12)$$

Para algún $\lambda > 0$, donde $V(t)$ es un movimiento Browniano con varianza $\text{Var}(X_i, X_j) + 2 \sum_{s=1}^{\infty} \text{Cov}(X_{-i} X_{-j}, X_{s-i} X_{s-j})$.

aplicando este argumento a cada componente de P_n^{-1} , se tiene

$$nP_{[nt]} \rightarrow t^{-1} \Gamma_{p+1}^{-1} \quad (2.13)$$

uniformemente sobre $[t_1, t_2]$, $0 < t_1 < t_2 \leq 1$. Luego de las expresiones (2.11) y (2.13), del teorema del mapeo continuo y de $\tilde{P}_{[nt]}^{-1} = P_n^{-1} - P_{[nt]}^{-1}$, se cumple

$$Q_n(t) \xrightarrow{d} \frac{\|\mathbf{W}(t)\|^2}{t} + \frac{\|\mathbf{W}(1) - \mathbf{W}(t)\|^2}{1-t} - \|\mathbf{W}(1)\| = \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)},$$

en $D[t_1, t_2]$, ($\|\cdot\|$ representa la longitud Euclidea del vector).

Para determinar un punto de cambio en $\tau \in [nt_1, nt_2]$, se considera la razón de verosimilitud restringida

$$\max_{k \in [nt_1, nt_2]} \Lambda_n(k) = \max_{t \in [t_1, t_2]} Q_n(t) \xrightarrow{d} \max_{t \in [t_1, t_2]} \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)}. \quad (2.14)$$

En [17] se comenta que el proceso límite en (2.9) toma el valor de ∞ en las dos vecindades de 0 y 1. Es posible determinar la convergencia exacta de Λ_n con unas constantes de normalización, esto se establece en el Teorema 3.2.

Teorema 3.2 *Sea $\{X_t\}$ el proceso definido en (2.1), tal que bajo H_0 , $\{\varepsilon_t\}$ satisface las condiciones (2.2) a (2.5) con $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ para algún $0 < \delta \leq 1$ y $\{X_t\}$ fuertemente relacionada con la función $\rho(n) \ll n^{-(1+\varepsilon)(1+4/\delta)}$ para algún $\varepsilon > 0$. Entonces bajo H_0 ,*

$$P \left[\frac{\sigma^{-2}\Lambda_n - b_n(p+1)}{a_n(p+1)} \leq x \right] \xrightarrow{L} \exp(-2e^{-x/2}),$$

donde $\Lambda_n = \max_{p < k \leq n} \Lambda_n(k)$ es la estadística de la razón de verosimilitud,

$b_n(d) = \frac{(2 \ln \ln n + (d/2) \ln \ln \ln n - \ln \Gamma(d/2))^2}{2 \ln \ln n}$, $a_n(d) = \sqrt{\frac{b_n(d)}{2 \ln \ln n}}$ son constantes de normalización y $\Gamma(\cdot)$ es la función Γ .

Demostración: Como la distribución de $\sigma^{-2}\Lambda_n$ no depende de σ^2 , se asumirá en toda la demostración que $\sigma^2 = 1$. Por otro lado la demostración se centra en la idea de probar que para todo $\delta' > 0$, se cumplen las expresiones (2.15-2.18).

$$\lim_{n \rightarrow \infty} \sup P \left[\left| \frac{\max_{p < k \leq \epsilon' n} \Lambda_n(k) - \max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k}{a_n} \right| > \delta' \right] \rightarrow 0 \text{ a.s. } \epsilon' \rightarrow 0 \quad (2.15)$$

$$\lim_{n \rightarrow \infty} \sup P \left[\left| \frac{\max_{(1-\epsilon')n < k \leq n} \Lambda_n(k) - \max_{(1-\epsilon')n < k \leq n} (\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k)}{a_n} \right| > \delta' \right] \rightarrow 0 \text{ a.s., } \epsilon' \rightarrow 0, \quad (2.16)$$

para todo $\epsilon' > 0$.

$$P \left[\frac{\max_{p < k \leq n} \mathbf{S}'_k P_k \mathbf{S}_k - b_n}{a_n} \leq x \right] \rightarrow \exp(-e^{-x/2}), \quad (2.17)$$

$$P \left[\frac{\max_{(1-\epsilon')n < k \leq n} (\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k) - b_n}{a_n} \leq x \right] \rightarrow \exp(-e^{-x/2}), \quad (2.18)$$

donde $a_n = a_n(p+1)$ y $b_n = b_n(p+1)$.

Para demostrar (2.15), se cumple

$$\begin{aligned} & \left| \frac{\max_{p < k \leq \epsilon' n} \Lambda_n(k) - \max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k}{a_n} \right| \leq \max_{p < k \leq \epsilon' n} \left| \frac{\Lambda_n(k) - \mathbf{S}'_k P_k \mathbf{S}_k}{a_n} \right| \\ & = \max_{p < k \leq \epsilon' n} \left| \frac{(\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k) - \mathbf{S}'_n P_n \mathbf{S}_n}{a_n} \right| \\ & \xrightarrow{d} \max_{t \leq \epsilon'} \left| \frac{\|\mathbf{W}(1) - \mathbf{W}(t)\|^2}{1-t} - \|\mathbf{W}(1)\|^2 \right| \quad (\text{a.s. } n \rightarrow \infty) \\ & \rightarrow 0 \text{ a.s. conforme } \epsilon' \rightarrow 0 \text{ conforme } a \rightarrow 1. \end{aligned}$$

La demostración de (2.16) es similar a la demostración de (2.15) y por tanto se omite.

La demostración de (2.17), estará dividida en tres pasos.

Paso 1. Para la sucesión $\{U_n\}$ dada en (2.10),

$$\mathbf{S}'_k \Gamma_{p+1}^{-1} \mathbf{S}_k - \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k = O(k^{1-\lambda'}),$$

conforme $k \rightarrow \infty$ a.s para algún $\lambda' > 0$ en efecto,

$$\mathbf{S}'_k \Gamma_{p+1}^{-1} \mathbf{S}_k - \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k = \mathbf{S}'_k \Gamma_{p+1}^{-1} (\mathbf{S}_k - \mathbf{U}_k) + (\mathbf{S}'_k - \mathbf{U}'_k) \Gamma_{p+1}^{-1} \mathbf{U}_k \quad (2.19)$$

además, por la ley del logaritmo iterado (apéndice A.13),

$$\mathbf{U}'_k \Gamma_{p+1}^{-1} = O((k \ln_2 k)^{1/2}),$$

así de (2.10), $\mathbf{S}'_k \Gamma_{p+1}^{-1} = O((k \ln_2 k)^{1/2})$. Utilizando las propiedades de la suma y el producto de $O(g)$ y de (2.10) y (2.19), se tiene que $\mathbf{S}'_k \Gamma_{p+1}^{-1} \mathbf{S}_k - \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k = O(k^{1-\lambda'})$.

Paso 2. $\mathbf{S}'_k P_k \mathbf{S}_k - \frac{1}{k} \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k \rightarrow 0$, conforme $k \rightarrow \infty$ a.s..

Del paso 1, es suficiente demostrar que

$$\mathbf{S}'_k P_k \mathbf{S}_k - \frac{\mathbf{S}'_k \Gamma_{p+1}^{-1} \mathbf{S}_k}{k} = \frac{\mathbf{S}_k}{k^{1/2}} k P_k \left(\Gamma_{p+1} - \frac{P_k^{-1}}{k} \right) \Gamma_{p+1}^{-1} \frac{\mathbf{S}_k}{k^{1/2}} \rightarrow 0 \text{ a.s.} \quad (2.20)$$

Además por (2.12) y la ley del logaritmo iterado (apéndice A.13), $\Gamma_{p+1} - \frac{1}{k}P_k^{-1} = O((\ln_2 k/k)^{1/2})$ conforme $k \rightarrow \infty$, por lo tanto (2.20) converge a cero conforme $k \rightarrow \infty$ a.s.

Paso 3. Para todo $\epsilon' > 0$, conforme $n \rightarrow \infty$,

$$P \left[\max_{p < k \leq n} \mathbf{S}'_k P_k \mathbf{S}_k = \max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k \right] \rightarrow 1. \quad (2.21)$$

En efecto, de la demostración del Teorema 3.1,

$$\max_{\epsilon' n \leq k \leq n} \mathbf{S}'_k P_k \mathbf{S}_k \xrightarrow{d} \max_{t \in [\epsilon', 1]} \frac{\|\mathbf{W}(t)\|^2}{t},$$

(2.21) sigue de $\max_{\epsilon' n \leq k \leq n} \mathbf{S}'_k P_k \mathbf{S}_k = O_p(1)$ y $\max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k \xrightarrow{P} \infty$. Por otro lado el paso 2, implica

$$\begin{aligned} & \left| \max_{M \leq k \leq n} \mathbf{S}'_k P_k \mathbf{S}_k - \max_{M \leq k \leq n} \frac{1}{k} \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k \right| \\ & \leq \sup_{k \geq M} \left| \mathbf{S}'_k P_k \mathbf{S}_k - \frac{1}{k} \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k \right| \rightarrow 0 \end{aligned} \quad (2.22)$$

conforme $M \rightarrow \infty$ a.s. Como los vectores aleatorios $\Gamma_{p+1}^{-1/2} \mathbf{Z}_k$ satisfacen las suposiciones del Corolario A.2 del apéndice A, se sigue que

$$P \left[\left(\max_{p < k \leq n} \frac{1}{k} \mathbf{U}'_k \Gamma_{p+1}^{-1} \mathbf{U}_k \right) / a_n \leq x \right] \rightarrow \exp(-e^{-x/2})$$

esta relación junto con (2.21) y (2.22), prueban la relación (2.17).

La demostración de (2.18) se omite, al ser similar a la demostración de (2.17).

Una vez que se ha verificado (2.15-2.18), resta ver que la suposición de relación, implica que $\max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k$ y $\max_{(1-\epsilon')n < k \leq n} (\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k)$ son asintóticamente independientes (para $\epsilon' < 1/2$), por lo tanto de (2.15)-(2.18),

$$\begin{aligned} & P \left[\frac{\max_{p < k \leq \epsilon' n} \mathbf{S}'_k P_k \mathbf{S}_k - b_n}{a_n} \leq x, \frac{\max_{(1-\epsilon')n < k \leq n} (\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k)}{a_n} \leq x \right] \\ & \rightarrow \exp(-2e^{-x/2}), \end{aligned}$$

también por (2.14), $\left(\max_{\epsilon' n \leq k \leq (1-\epsilon')n} \Lambda_n(k) - b_n \right) / a_n \xrightarrow{P} -\infty$. Aplicando el Teorema A.8 del apéndice A, se tiene el resultado.

Partiendo del proceso autorregresivo

$$X_i - \mu = \phi_1(X_{i-1} - \mu) + \dots + \phi_p(X_{i-p} - \mu) + \varepsilon_i, \quad i \geq p + 1, \quad (2.23)$$

con $\{\varepsilon_i\}$ un ruido blanco, $E(\varepsilon_1^2) = \sigma^2$, y $\boldsymbol{\xi} = (\mu, \sigma^2, \phi_1, \dots, \phi_p)^T$ el vector de parámetros, Gombay [37], propuso utilizar el vector score como prueba estadística, para esto, se considera el logaritmo de la función de verosimilitud.

Bajo la suposición Gaussiana, basada sobre las primeras k observaciones, las componentes del vector score eficiente $\nabla_{\boldsymbol{\xi}^k} l_k(\boldsymbol{\xi}) = \nabla_{\boldsymbol{\xi}^k} l_k(\boldsymbol{\xi})$ son:

$$\begin{aligned} \frac{\partial}{\partial \mu} l_k(\boldsymbol{\xi}) &= \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \sum_{i=1}^k \left[Y_i - \mu - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right], \\ \frac{\partial}{\partial \sigma^2} l_k(\boldsymbol{\xi}) &= -\frac{k}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^k \left[Y_i - \mu - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right]^2, \\ \frac{\partial}{\partial \phi_s} l_k(\boldsymbol{\xi}) &= \frac{1}{\sigma^2} \sum_{i=1}^k \left[Y_i - \mu - \sum_{j=1}^p \phi_j (Y_{i-j} - \mu) \right] (Y_{i-s} - \mu), \quad s = 1, \dots, p. \end{aligned}$$

Donde l_k denota la función logaritmo de verosimilitud basada en las observaciones Y_1, Y_2, \dots, Y_k , se asumen como cero los valores de $Y_0, Y_{-1}, \dots, Y_{-p}$.

Considérese

$$\mathbf{I}(\boldsymbol{\xi}) = \mathbf{I}(\mu, \sigma^2, \phi_1, \dots, \phi_p) = \begin{bmatrix} \frac{(1 - \sum_{j=1}^p \phi_j)^2}{\sigma^2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma^4} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2} \Gamma \end{bmatrix},$$

la matriz de información de tamaño $(p+2) \times (p+2)$ donde Γ es la matriz de covarianza del vector (Y_1, \dots, Y_p) y $\hat{\mu}_n, \hat{\sigma}_n^2$ y $\hat{\boldsymbol{\phi}}_n$ los estimadores de máxima verosimilitud de los parámetros del proceso, sea

$$\hat{\mathbf{B}}(u) = n^{-1/2} \mathbf{I}(\hat{\boldsymbol{\xi}}) \begin{pmatrix} \frac{\partial}{\partial \mu} l_{[nu]}(\hat{\boldsymbol{\xi}}_n) \\ \frac{\partial}{\partial \sigma^2} l_{[nu]}(\hat{\boldsymbol{\xi}}_n) \\ \frac{\partial}{\partial \boldsymbol{\phi}} l_{[nu]}(\hat{\boldsymbol{\xi}}_n) \end{pmatrix}. \quad (2.24)$$

Teorema 3.3 *Supóngase que la sucesión de observaciones $\{Y_i\}$ satisface el modelo (2.23) con $\{\varepsilon_i\}$ una sucesión Gaussiana de ruido blanco independiente e idénticamente distribuida, $\text{Var}(\varepsilon_i) = \sigma^2$ y polinomio característico $\phi(z) =$*

$1 - \phi_1 z - \dots - \phi_p z^p$ con raíces fuera del círculo unitario. Entonces existe un proceso Gaussiano $(p+2)$ -dimensional $\mathbf{B}(u)$ cuyas componentes son puentes Brownianos independientes $B^{(j)}(u)$, $j = 1, 2, \dots, p+2$ tal que

$$\max_{1 \leq j \leq p+2} \sup_{0 \leq u \leq 1} |\widehat{\mathbf{B}}^{(j)}(u) - \mathbf{B}^{(j)}(u)| = o_p(1).$$

Demostración: La demostración se realizará para cada uno de los diferentes parámetros.

1. Considérese el caso de probar un cambio en μ . La componente del vector score eficiente es

$$\begin{aligned} \frac{\partial l_k}{\partial \mu} &= \frac{1 - \sum_{j=1}^p \widehat{\phi}_{jn}}{\widehat{\sigma}_n^2} \sum_{j=1}^k \left[(X_i - \widehat{\mu}_n) - \sum_{j=1}^p \widehat{\phi}_{jn} (X_{i-j} - \widehat{\mu}_n) \right] \\ &\pm \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \sum_{j=1}^k \left[(X_i - \bar{X}_n) - \sum_{j=1}^p \phi_j (X_{i-j} - \bar{X}_n) \right]. \end{aligned} \quad (2.25)$$

Es claro que

$$\begin{aligned} &\frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \sum_{i=1}^k \left[(X_i - \bar{X}_n) - \sum_{j=1}^p \phi_j (X_{i-j} - \bar{X}_n) \right] \\ &= \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \left\{ \sum_{i=1}^k (X_i - \bar{X}_n) - \sum_{j=1}^p \phi_j \sum_{i=1}^k (X_{i-j} - \bar{X}_n) \right\}. \end{aligned} \quad (2.26)$$

En [39], se demuestra

$$\left| \sum_{i=1}^{[t]} (X_i - \mu) - \sigma^* W(t) \right| = o(t^{1/v}) \text{ a.s. para algún } v > 2, W(t) \text{ un movimiento Browniano y } \sigma^* > 0. \quad (2.27)$$

De lo anterior se obtiene

$$\sup_{0 \leq u \leq 1} \left| n^{-1/2} \left(\sum_{i=1}^{[nu]} X_i - \frac{[nu]}{n} \sum_{i=1}^n X_i \right) - \sigma^* (W(u) - uW(1)) \right| = o_p(1). \quad (2.28)$$

Para la diferencia

$$n^{-1/2} \left(\sum_{i=1}^k X_{i-j} - \sum_{i=1}^k X_i \right) = n^{-1/2} \sum_{i=1}^j X_{i-j} - n^{-1/2} \sum_{i=k-j+1}^k X_i. \quad (2.29)$$

Considerando que

$$\lim_{k \rightarrow \infty} \sup_{0 \leq s \leq p} |W(k-s) - W(k)| = O((\log k)^{1/2}) \text{ a.s.}, \quad (2.30)$$

(ver [15]) entonces, del término del error de (2.29), utilizando (2.27) y (2.30), se tiene que conforme $n \rightarrow \infty$

$$\max_{1 < k \leq n} \max_{1 \leq j \leq p} n^{-1/2} \left| \sum_{i=k-j+1}^k X_i \right| = o_p(1). \quad (2.31)$$

De (2.25) a (2.31) y después de estandarizar, se tiene

$$\begin{aligned} \sup_{0 \leq u \leq 1} \left| \left(1 - \sum_{j=1}^p \phi_j \right)^{-1} \sum_{i=1}^{[nu]} \left[(X_i - \bar{X}_n) - \sum_{j=1}^p \phi_j (X_{i-j} - \bar{X}_n) \right] - \sigma^* \mathbf{B}(u) \right| \\ = o_p(1) \end{aligned}$$

Gombay [37], demuestra que

$$|\hat{\mu}_n - \mu| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ a.s.}, \quad (2.32)$$

$$|\hat{\sigma}_n^2 - \sigma^2| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ a.s.}, \quad (2.33)$$

y para $j = 1, 2, \dots, p$

$$|\hat{\phi}_{jn} - \phi_j| = O \left(\sqrt{\frac{\log \log n}{n}} \right) \text{ a.s.} \quad (2.34)$$

Por (2.33) y (2.34) el error cometido de reemplazar $(1 - \sum \hat{\phi}_j)/\hat{\sigma}^2$ por $(1 - \sum \phi_j)/\sigma^2$ es $O(n^{1/2}(\log \log n)^{1/2})$ a.s. Por tanto se puede escribir

$$\begin{aligned}
& \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \left\{ \sum_{i=1}^k (X_i - \bar{X}_n) - \sum_{j=1}^p \phi_j (X_{i-j} - \bar{X}_n) \right\} \\
& - \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \left\{ \sum_{i=1}^k (X_i - \bar{X}_n) - \sum_{j=1}^p \hat{\phi}_j (X_{i-j} - \bar{X}_n) \right\} \\
& = \frac{1 - \sum_{j=1}^p \phi_j}{\sigma^2} \left\{ \sum_{i=1}^k \sum_{j=1}^p (\hat{\phi}_{jn} - \phi_j) (X_{i-j} - \bar{X}_n) \right\} \\
& \leq \max_{1 \leq j \leq p} |\hat{\phi}_{jn} - \phi_j| \left| \sum_{i=1}^k (X_{i-j} - \bar{X}_n) \right|.
\end{aligned}$$

Luego por (2.28), (2.31) y (2.34)

$$\sup_{0 \leq u \leq 1} \max_{1 \leq j \leq p} |\hat{\phi}_{jn} - \phi_j| n^{-1/2} \left| \sum_{i=1}^{\lfloor nu \rfloor} (X_{i-j} - \bar{X}_n) \right| = O_p \left(\sqrt{\frac{\log \log n}{n}} \right).$$

Finalmente como

$$\hat{\mu}_n = \frac{1}{n \left(1 - \sum_{j=1}^p \phi_j \right)} \sum_{i=1}^n \left[X_i - \sum_{j=1}^p \phi_j X_{i-j} \right], \text{ utilizando (2.27) y (2.30), se}$$

tiene

$$|\hat{\mu}_n - \bar{X}_n| = o(n^{1/v-1}) \text{ a.s.}, \quad (2.35)$$

lo cual lleva a un error que se puede omitir al reemplazar $\hat{\mu}_n$ por \bar{X}_n en el estadístico de prueba.

2. Para detectar cambios en ϕ_s , $s = 1, \dots, p$, se considera

$$\frac{\partial}{\partial \phi_s} l_k(\hat{\mu}_n, \hat{\sigma}_n^2, \hat{\phi}_n) = \frac{1}{\hat{\sigma}_n^2} \sum_{i=1}^k (X_{i-s} - \hat{\mu}_n) \left[X_i - \hat{\mu}_n - \sum_{j=1}^p \hat{\phi}_{nj} (X_{i-j} - \hat{\mu}_n) \right].$$

Al cambiar σ^2 por $\hat{\sigma}^2$ en (2.33), no se cambia la distribución asintótica.

Sea $\hat{\varepsilon} = (X_i - \hat{\mu}_n) - \sum_{j=1}^p \hat{\phi}_{nj} (X_{i-j} - \hat{\mu}_n)$, lo que estima $\varepsilon_i = (X_i - \mu) -$

$\sum_{j=1}^p \phi_j (X_{i-j} - \mu)$. Además $\sum_{i=1}^n (X_{i-s} - \hat{\mu}_n) \hat{\varepsilon}_j \equiv 0$. Luego por el principio de invarianza para la sucesión $\sum_{i=1}^k (X_{i-s} - \mu) \varepsilon_i$ probado en [39],

$$\sup_{0 \leq u \leq 1} \left| n^{-1/2} \left[\sum_{i=1}^{[nu]} (X_{i-s} - \mu) \varepsilon_i - \frac{[nu]}{n} \sum_{i=1}^n (X_{i-s} - \mu) \varepsilon_i - \sigma_1 \mathbf{B}(u) \right] \right| = o_p(1) \quad (2.36)$$

con algún $\sigma_1^2 > 0$. A continuación, se demostrará que el error cometido al reemplazar los parámetros por sus respectivos estimadores de máxima verosimilitud se pueden omitir. El error se puede escribir como una constante por tiempos

$$\begin{aligned} & \left[\sum_{i=1}^k \left(X_{i-s} - \hat{\mu}_n \hat{\varepsilon}_j - \sum_{i=1}^k (X_{i-s} - \mu) \varepsilon_i \right) \right] \\ & + \left[-\frac{k}{n} \sum_{i=1}^n (X_{i-s} - \hat{\mu}_n) \hat{\varepsilon}_i + \frac{k}{n} \sum_{i=1}^n (X_{i-s} - \mu) \varepsilon_i \right] = B_{nk} + D_{nk}. \end{aligned}$$

Al analizar B_{nk} como

$$\begin{aligned} B_{nk} &= \sum_{i=1}^k (X_{i-s} - \hat{\mu}_n \pm \mu) (\hat{\varepsilon}_i \pm \varepsilon_i) - \sum_{i=1}^k (X_{i-s} - \mu) \varepsilon_i \\ &= (\mu - \hat{\mu}_n) \sum_{i=1}^k \varepsilon_i + \sum_{i=1}^k (X_{i-s} - \mu) (\hat{\varepsilon}_i - \varepsilon_i) + (\mu - \hat{\mu}_n) \sum_{i=1}^k (\hat{\varepsilon}_i - \varepsilon_i) \\ &= B_{nk}^{(1)} + B_{nk}^{(2)} + B_{nk}^{(3)}. \end{aligned} \quad (2.37)$$

Por (2.32) y la Ley del ogaritmo iterado (apéndice A.13), para la sucesión $\{\varepsilon_i\}$, $B_{nk}^{(1)} = O(\log \log n)$ a.s. como

$$\begin{aligned} \hat{\varepsilon}_i - \varepsilon_i &= (\hat{\mu}_n - \mu) \left(1 - \sum_{j=1}^p \phi_j \right) + \sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j) (X_{i-j} - \mu) \\ &+ (\mu - \hat{\mu}_n) \sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j), \end{aligned}$$

por otro lado, utilizando el principio de invarianza y la ley del logaritmo iterado para la sucesión $\sum (X_{i-s} - \mu)$, (2.32) y (2.34) se obtiene

$$B_{nk}^{(3)} = O(\log \log n) \text{ a.s.}, \quad (2.38)$$

Para $B_{nk}^{(2)}$ se realiza un análisis similar, aplicando el principio de invarianza para $\sum_{i=1}^k [(X_{i-s} - \mu)(X_{i-j} - \mu) - E((X_{i-s} - \mu)(X_{i-j} - \mu))]$, además

$$\begin{aligned} B_{nk}^{(2)} &= \sum_{i=1}^k (X_{i-s} - \mu) \left[(\hat{\mu}_n - \mu) \left(1 - \sum_{j=1}^p \phi_j \right) + (\mu - \hat{\mu}_n) \sum_{j=1}^p (\hat{\phi}_{nj} - \phi_j) \right] \\ &\quad + \sum_{i=1}^k (X_{i-s} - \mu) \sum_{j=1}^p (X_{i-j} - \mu) (\hat{\phi}_{nj} - \phi_j) \\ &\quad \pm k \sum_{j=1}^p E(X_{i-s} - \mu)(X_{i-j} - \mu) (\hat{\phi}_{nj} - \phi_j) \end{aligned} \quad (2.39)$$

$$\begin{aligned} &= k \sum_{j=1}^p E((X_{i-s} - \mu)(Y_{i-j} - \mu)) (\hat{\phi}_{nj} - \phi_j) + O(\log \log n) \text{ a.s.} \\ &\hspace{15em} (2.40) \end{aligned}$$

El término de error D_{nk} , tiene la misma estructura que B_{nk} , excepto por el valor k/n y que la suma es $\sum_{i=1}^n$. Por esta razón, todos los términos en (2.38) y (2.39) no se cancela $O(\log \log n)$, por tanto se tiene que $\sup_{1 \leq k \leq n} |B_{nk} + D_{nk}| = O(\log \log n)$ a.s., con lo que se concluye la demostración para el caso 2.

3. Para detectar el cambio en σ^2 , se considera

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l_k(\hat{\mu}_n, \hat{\sigma}_n^2, \hat{\phi}_n) &= \frac{1}{2\hat{\sigma}_n^4} \sum_{i=1}^k \hat{\varepsilon}_i^2 - \frac{1}{2\hat{\sigma}_n^4} \frac{k}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \\ &= \frac{1}{2\sigma^4} \left(\sum_{i=1}^k \varepsilon_i^2 - \frac{k}{n} \sum_{i=1}^n \varepsilon_i^2 \right) + \left[\frac{1}{2\hat{\sigma}_n^4} \sum_{i=1}^k \hat{\varepsilon}_i^2 - \frac{1}{2\sigma^4} \sum_{i=1}^k \varepsilon_i^2 \right] \\ &\quad - \frac{k}{n} \left[\frac{1}{\hat{\sigma}} \sum_{i=1}^n \hat{\varepsilon}_i^2 - \frac{1}{\sigma^4} \sum_{i=1}^n \varepsilon_i^2 \right] \\ &= A_{nk} + B_{nk} + D_{nk}. \end{aligned}$$

Para A_{nk} se obtiene por argumentos de principios de invarianza que

$$\sup_{0 \leq u \leq 1} \left| n^{-1/2} A_{[nu]} - \sigma_2 B(u) \right| = o_p(1),$$

para algún $\sigma_2^2 > 0$ y un puente Browniano $B(u)$. Dado que por el argumento en (2.33), σ^2 y $\hat{\sigma}^2$ no cambian la distribución asintótica, se puede considerar

$$\begin{aligned} \sum_{i=1}^k (\widehat{\varepsilon}_i^2 - \varepsilon_i) &= \sum_{i=1}^k \left[X_i - \widehat{\mu}_n - \sum_{j=1}^p \widehat{\phi}_j (X_{i-j} - \widehat{\mu}) \right]^2 \\ &\quad - \sum_{i=1}^k \left[X_i - \mu - \sum_{j=1}^p \phi_j (X_{i-j} - \mu) \right]^2 \\ &\quad \pm \sum_{i=1}^k \left[X_i - \mu - \sum_{j=1}^p \widehat{\phi}_j (X_{i-j} - \mu) \right]^2. \end{aligned}$$

Se tiene

$$\begin{aligned} \sum_{i=1}^k \varepsilon_i^2 - \sum_{i=1}^k \left[X_i - \mu - \sum_{j=1}^p \widehat{\phi}_j (X_{i-j} - \mu) \right]^2 \\ = k \left(-2(X_i - \mu) \sum_{j=1}^p (\phi_j - \widehat{\phi}_{nj})(X_{i-j} - \mu) \right. \\ \left. + \left[\sum_{j=1}^p \phi_j (X_{i-j} - \mu) \right]^2 - \left[\sum_{j=1}^p \widehat{\phi}_j (X_{i-j} - \mu) \right]^2 \right). \end{aligned}$$

De nuevo por los principios de invarianza para $\sum_{i=1}^k (X_i - \mu)(X_{i-j} - \mu)$, la ley del logaritmo iterado y (2.34), se obtienen términos del tipo

$$\begin{aligned} (\phi_j - \widehat{\phi}_{nj}) \sum_{i=1}^k (X_i - \mu)(X_{i-j} - \mu) &= (\phi_j - \widehat{\phi}_{nj}) k E((X_i - \mu)(X_{i-j} - \mu)) \\ &\quad + O(\log \log n) \text{ a.s., } 0 \leq j \leq p. \end{aligned}$$

Ahora, considérese

$$\begin{aligned}
& \sum_{i=1}^k \left\{ \left[X_i - \hat{\mu}_n - \sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \hat{\mu}_n) \right]^2 - \left[X_i - \mu - \sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \mu) \right]^2 \right\} \\
&= \sum_{i=1}^k \left(2(X_i - \mu)(\mu - \hat{\mu}_n) + (\mu - \hat{\mu}_n)^2 - 2(X_i - \mu) \sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \hat{\mu}_n) \right. \\
&\quad - 2(\mu - \hat{\mu}_n) \sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \hat{\mu}_n) + \left(\sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \hat{\mu}_n) \right)^2 \\
&\quad \left. + 2(X_i - \mu) \sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \mu) - \left(\sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \mu) \right)^2 \right).
\end{aligned}$$

Obsérvese que en la relación anterior se tiene

$$\begin{aligned}
(\mu - \hat{\mu}_n) \sum_{i=1}^k (X_i - \mu) &= O(\log \log n) \text{ a.s.}, \\
k(\mu - \hat{\mu}_n) &= O(\log \log n) \text{ a.s. y}
\end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^k \left[\left(\sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \hat{\mu}_n) \right)^2 - \left(\sum_{j=1}^p \hat{\phi}_{nj}(X_{i-j} - \mu) \right)^2 \right] \\
&= \sum_{i=1}^k \left[\sum_{j=1}^p \sum_{l=1}^p \hat{\phi}_{nj} \hat{\phi}_{nl}(X_{i-j} - \hat{\mu}_n)(X_{i-l} - \hat{\mu}_n) \right. \\
&\quad \left. - \sum_{j=1}^p \sum_{l=1}^p \hat{\phi}_{nj} \hat{\phi}_{nl}(X_{i-j} - \mu)(X_{i-l} - \mu) \right] \\
&= \sum_{i=1}^k \left\{ \sum_{j=1}^p \sum_{l=1}^p \hat{\phi}_{nj} \hat{\phi}_{nl} [(\hat{\mu}_n - \mu)((X_{i-l} - \mu) + (X_{i-j} - \mu)) + (\hat{\mu}_n - \mu_n)^2] \right\} \\
&= O(\log \log n) \text{ a.s.}
\end{aligned}$$

Dado que los términos $(\phi_j - \hat{\phi}_{nj})kE((X_i - \mu)(X_{i-j} - \mu))$ en B_{nk} y D_{nk} se cancelan cada uno para $0 \leq j \leq p$ y el resto de los términos son $O(\log \log n)$ a.s., el apartado 3 del teorema queda demostrado.

Para pruebas simultáneas del cambio en d parámetros para tener un nivel de significancia α exacto, se sugiere utilizar $\alpha^* = 1 - (1 - \alpha)^{1/d}$ para cada componente, así para pruebas de un solo lado, se tiene,

Prueba 1: Si $\sup_{0 \leq u \leq 1} \widehat{\mathbf{B}}^{(j)}(u) \geq C_1(\alpha^*)$, entonces se concluye que existe un cambio en el parámetro ξ_j , ($j = 1, \dots, p+2$) sobre la sucesión Y_1, \dots, Y_n . El valor crítico $C_1(\alpha^*)$ se obtiene de la relación $P \left(\sup_{0 \leq u \leq 1} \mathbf{B}^{(1)}(u) \geq x \right) = e^{-2x^2}$.

Para pruebas de dos lados, se tiene

Prueba 2: Si $\sup_{0 \leq u \leq 1} \widehat{\mathbf{B}}^{(j)}(u) \geq C_2(\alpha^*)$, entonces se concluye que existe un cambio en el parámetro ξ_j , ($j = 1, \dots, p+2$) sobre la sucesión Y_1, \dots, Y_n . El valor crítico $C_2(\alpha^*)$ se obtiene de la relación $P \left(\sup_{0 \leq u \leq 1} |\mathbf{B}^{(1)}(u)| > x \right) = \sum_{k \neq 0} (-1)^{k+1} \exp^{-2k^2 x^2}$.

Una vez que se asegura el cambio para ξ_j , $j = 1, \dots, p+2$, en $\tau = [\rho n]$, $0 \leq \rho \leq 1$, el estimador para τ en la prueba unilateral es

$$\widehat{\tau} = \min_k \left\{ \frac{\partial}{\partial \xi_j} l_k(\widehat{\boldsymbol{\xi}}_n) = \max_{1 < m \leq n} \frac{\partial}{\partial \xi_j} l_m(\widehat{\boldsymbol{\xi}}_n) \right\}, \quad (2.41)$$

o bien

$$\widehat{\tau} = \min_n \left\{ \frac{\partial}{\partial \xi_j} l_k(\widehat{\boldsymbol{\xi}}_n) = \min_{1 < m \leq n} \frac{\partial}{\partial \xi_j} l_m(\widehat{\boldsymbol{\xi}}_n) \right\}.$$

Para una prueba bilateral el estimador de τ es

$$\widehat{\tau}_n = \min_k \left\{ \left| \frac{\partial}{\partial \xi_j} l_k(\widehat{\boldsymbol{\xi}}_n) \right| = \max_{1 < m \leq n} \left| \frac{\partial}{\partial \xi_j} l_m(\widehat{\boldsymbol{\xi}}_n) \right| \right\}. \quad (2.42)$$

3.3. Metodologías paramétricas para múltiples puntos de cambio

Sea $\{x_t, t = 1, \dots, n\}$ una sucesión ordenada de datos, supóngase que existen m puntos de cambio ordenados en tiempos aleatorios $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$, con $\tau_0 = 1$, $\tau_{m+1} = n$ por tanto cada punto de cambio, induce una partición sobre los datos en $m+1$ segmentos, donde el i -ésimo segmento contendrá las observaciones $(x_{\tau_{i-1}+1}, \dots, x_{\tau_i})$. Algo comúnmente utilizado para identificar múltiples puntos de cambio es, minimizar

$$\sum_{1 \leq i \leq m+1} [C(x_{\tau_{i-1}+1}, \dots, \tau_i)] + \beta f(m),$$

con respecto a m y $\tau_1, \tau_2, \dots, \tau_m$. Aquí $C(x_{\tau_{i-1}+1}, \dots, \tau_i)$ es una función de costo para un segmento y $\beta f(m)$ es un término de penalización para evitar un sobre

ajuste [87]. Algunas funciones de costo comúnmente utilizadas, son el inverso aditivo de la función de probabilidad de registro, las sumas acumuladas [51], otros métodos incluyen la aplicación de información de entropía, para la penalización $\beta f(m)$ [19]. La opción más común de penalización es lineal en el número de puntos de cambio, es decir $\beta f(m) = \beta m$. Ejemplos de tales sanciones incluyen el Criterio de información de Akaike (AIC) y el Criterio de información de Schwarz (BIC) [25].

Para los métodos que se verán en esta sección, se quiere probar la hipótesis

$$H_0 : \text{No existe un punto de cambio en los datos.}$$

vs

$$H_1 : \text{Existe un cambio en los datos.}$$

Método Pruned Exact Linear Time (PELT)

El método *Pruned Exact Linear Time*, está basado en el método propuesto por Jackson [54] y Yao[105]. Sea $\{x_t, t = 1, \dots, n\}$ un conjunto de observaciones, la idea del método es la búsqueda de una partición óptima de los datos en un intervalo de tiempo I cuyos elementos son conjuntos de celdas de datos no provocarán una pérdida significativa de información.

Una partición \mathbf{P} de un intervalo I es un conjunto de bloques $\{\tau_1, \dots, \tau_m\}$, $\mathbf{P}(I) = \{B_k, k \in \mathcal{M}\}$, $\mathcal{M} \equiv \{1, 2, \dots, m\}$, donde los bloques son conjuntos de celdas de datos, definidos por conjuntos de índices \mathcal{N}_k : $B_k = \{x_t, t \in \mathcal{N}_k\}$ que satisfacen las condiciones $\bigcup_k B_k = I$ y $B_k \cap B_{k'} = \emptyset$ si $k \neq k'$, el número de bloques debe satisfacer $0 \leq m \leq n$. Sea P^* el conjunto (finito) de todas las particiones de I en bloques. Considérese como una función de ajuste, que asigna un valor a cualquier partición $\mathbf{P} \in P^*$ en la forma

$$V(\mathbf{P}) = \sum_{1 \leq i \leq m+1} [C(x_{\tau_{i-1}+1}, \dots, x_{\tau_i}) + \beta], \quad (3.43)$$

donde $C(x_{\tau_{i-1}+1}, \dots, x_{\tau_i})$ es el bloque de ajuste B_k y debe ser minimizado con la finalidad de encontrar un punto de cambio. De acuerdo con [87], el método PELT, propone un paso de poda, el cual incrementa la eficiencia computacional asegurando además un mínimo global de la función de costo, la cual es lineal en n (3.43). La segmentación óptima es $F(n)$, donde

$$F(n) = \min_{\tau} \left\{ \sum_{1 \leq i \leq m+1} [C(x_{\tau_{i-1}+1}, \dots, x_{\tau_i}) + \beta] \right\}.$$

El código para programar el método PELT se muestra en la Tabla 3.1.

Tabla 3.1: Pseudocódigo del método PELT.

<p>Datos de entrada: Conjunto de observaciones (x_1, \dots, x_n), $x_i \in \mathbb{R}$ Medida de ajuste $C(\cdot)$ que depende de los datos. Constante de penalización β que no depende del número de o ubicación de los puntos de cambio. Constante K que satisface la ecuación $C(x_{(t+1):s}) + C(x_{(s+1):n}) + K \leq C(x_{(t+1):n})$, para todo $t < s < n$ Proceso: Sea n el total de datos, $F(0) = -\beta$, $cp(0) = Null$, $R_1 = 0$. Iteraciones: Para $\tau^* = 1, 2, \dots, n$, 1. Calcular $F(\tau^*) = \min_{\tau \in R_{\tau^*}} [F(\tau) + C(x_{(\tau+1):\tau^*}) + \beta]$ 2. Sea $\tau^1 = arg \left\{ \min_{\tau \in R_{\tau^*}} [F(\tau) + C(x_{(\tau+1):\tau^*}) + \beta] \right\}$. 3. Considerar $cp(\tau^*) = [cp(\tau^1, \tau^1)]$ 4. Sea $R_{\tau^*+1} = \{\tau \in R_{\tau^*} \cup \{\tau^*\} : F(\tau) + C(x_{(\tau+1):\tau^*}) + K \leq F(\tau^*)\}$. Salida: Los puntos de cambio almacenados en $cp(n)$</p>

Método de Segmentación Binaria (BS)

El método de segmentación binaria (BS), considera el modelo $x_t = f_t + \xi_t$, $t = 1, \dots, n$, donde $f(t)$ es una señal determinista, unidimensional, constante por partes con puntos de cambio cuyo número N y sus ubicaciones τ_1, \dots, τ_n son desconocidos y la secuencia aleatoria $\{\xi_t\}_1^n$ es Gaussiana i.i.d. con media cero y varianza 1.

Este método utiliza la estadística CUSUM definido por el producto interno entre el vector (x_s, \dots, x_e) y un vector particular de contraste dado por:

$$\tilde{x}_{s,e}^b = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^b x_t - \sqrt{\frac{b-s+1}{n-(e-b)}} \sum_{t=b+1}^e x_t,$$

donde $s \leq b < e$, con $n = e - s + 1$. En este primer paso, el algoritmo BS calcula $\tilde{x}_{1,n}^b$ y después toma $b_{1,1} = \arg \max_{b:1 \leq b < n} |\tilde{x}_{1,n}^b|$ como el primer candidato para punto de cambio, cuya importancia debe juzgarse según un determinado criterio. Si se considera significativo, el dominio $[1, n]$ se divide en dos subintervalos a la izquierda y a la derecha de $b_{1,1}$ y la recursión continúa calculando $\tilde{x}_{1,b_{1,1}}^n$ y $\tilde{x}_{b_{1,1}+1,n}^b$ posiblemente resultando en más divisiones [32].

Este método, extiende cualquier método de punto de cambio único para múltiples puntos de cambio repitiendo iterativamente el método en diferentes subconjuntos de la secuencia para esto, comienza aplicando inicialmente el método de punto de cambio único a todos los datos y prueba si existe τ tal que

$$C(x_{1:\tau}) + C(x_{\tau+1:n}) + \beta < C(x_{1:n}). \quad (3.44)$$

Si (3.44) no se cumple, se garantiza que no hay punto de cambio a lo largo de $\mathbf{x} = (x_1, x_2, \dots, x_n)$ y el proceso finaliza; de lo contrario, los datos se dividen

Tabla 3.2: Pseudocódigo del método BS.

<p>Datos de entrada: Conjunto de observaciones (x_1, \dots, x_n), $x_i \in \mathbb{R}$ Estadística de prueba $\lambda(\cdot)$ que depende de los datos. Estimador de puntos de cambio $\hat{\tau}(\cdot)$ Región de rechazo β Proceso: Sea $C = \emptyset$ y $S = [1, n]$ Iteraciones: Mientras $S \neq \emptyset$, 1. Elegir un elemento de S y denotar este elemento como $[s, t]$. 2. Si $\lambda(x_{s:t}) < \beta$, eliminar $[s, t]$ 3. Si $\lambda(x_{s:t}) \geq \beta$: a) Quitar $[s, t]$ de S; b) Calcular $\hat{\tau}(x_{s:t}) + s - 1$ y agregar r a C; c) Si $r \neq s$ agregar $[s, r]$ a S; d) Si $r \neq t - 1$ agregar $[r + 1, t]$ a S; Salida: El conjunto de puntos de cambio almacenados en C.</p>

en dos conjuntos, que consisten en la secuencia de datos antes y después del punto de cambio, τ_α , el proceso se aplica nuevamente en cada partición hasta que no se encuentren más puntos de cambio.

El pseudocódigo del método BS, se presenta en la Tabla 3.2

Método de Segmentación del vecino más cercano (SN)

El método del vecino más cercano (SN), utiliza programación dinámica para detectar el punto de cambio, comienza estableciendo un límite superior en el tamaño del espacio de segmentación (es decir, el número máximo de puntos de cambio) que se requiere, este método continúa calculando la función de costo para todos los segmentos posibles.

Toma el caso restringido que segmenta los datos hasta t , para $t \geq m + 1$, en $m + 1$ segmentos (usando k puntos de cambio), y denota el valor mínimo del costo por $C_{m,t}$. La idea del método SN es buscar una relación entre $C_{t,m}$ y $C_{s,m-1}$ para $s < t$:

$$\begin{aligned}
 C_{m,t} &= \min_{\tau} \sum_{1 \leq j \leq m+1} C(\mathbf{x}_{\tau_{j-1}:\tau_j}) \\
 &= \min_{\tau \in \{m, \dots, t-1\}} [C_{m-1,\tau} + C(\mathbf{x}_{\tau:t})].
 \end{aligned}$$

Si esto se ejecuta para todos los valores de t hasta n y para $j = 2, \dots, m$, entonces se pueden adquirir las segmentaciones exactas con $1, \dots, m$ segmentos.

Para extraer la segmentación exacta, primero se deja que $\tau_l^*(t)$ denote la posición óptima del último punto de cambio si segmentamos los datos $y_{1:t}$ usando l puntos de cambio. Esto se puede calcular como

$$\tau_l^*(t) \arg \min_{\tau \in \{l, \dots, t-1\}} [C_{l-1, \tau} + C(x_{\tau+1:t})].$$

CAPÍTULO 4

Comparación de metodologías

Se cuenta con registros de temperatura máxima de una estación climatológica ubicada en el municipio de Tlaxco- Tlaxcala, la estación está a cargo de Protección Civil del Estado de Tlaxcala y del Centro de Investigación en Cambio Climático de la Universidad Autónoma de Tlaxcala. Con los registros, se aplicaron y compararon metodologías paramétricas para determinar la existencia de un sólo punto de cambio (AMOC) y de múltiples puntos de cambio explicadas en el capítulo anterior. Para ello, se realizó un análisis de calidad de datos, se completaron datos perdidos mediante regresión lineal y se ajustaron modelos SARIMA antes y después de identificar los puntos de cambio, finalmente para el contraste de la bondad del ajuste de los modelos, se consideró el Criterio de Información de Akaike (AIC).

4.1. Caracterización de la zona de estudio y análisis de calidad de datos

La Comisión de Protección Civil del Estado de Tlaxcala y el Centro de Investigación en Cambio Climático de la Universidad Autónoma de Tlaxcala, proporcionaron información de 6 años (2 de septiembre de 2011 a 2 de septiembre de 2017) de una estación climatológica, ubicada en el centro del municipio de Tlaxco-Tlaxcala (Figura 4.1).

Motivados por el interés de entender el comportamiento de la temperatura máxima de la región en estudio, se realizó un gráfico box-plot de los registros; ver Figura 4.2, para su construcción, se consideró la temperatura máxima por semana, para ello, se tomó como primer semana del año a semana 1 del mes de enero (se considera como primer semana de enero, aquella semana que

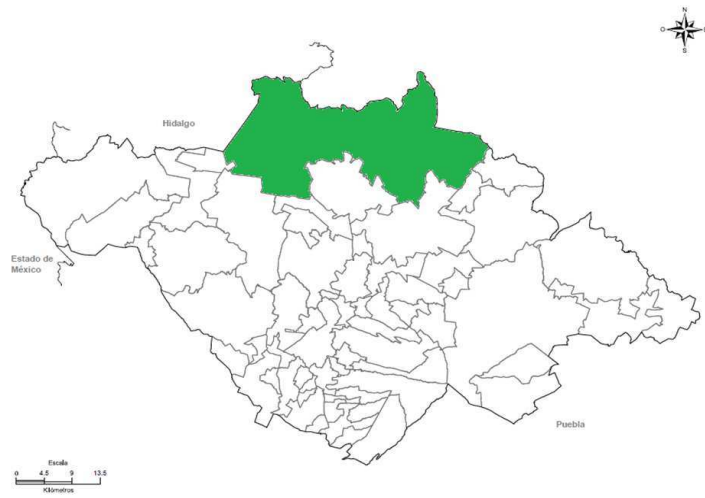


Figura 4.1: Ubicación geográfica del municipio de Tlaxco, Tlaxcala. Mapa extraído de INEGI.

inicia con el primer domingo del mes de enero) y la semana 52 como la última semana del mes de diciembre.

En la Figura 4.2, se observan los incrementos de la temperatura máxima hasta alcanzar sus valores máximos en las semanas 11 a 23, las cuales coinciden con el periodo de primavera, con la llegada de las lluvias (mes de junio, semana 24), se observa un descenso en la temperatura, en el box-plot también se destaca el efecto de la canícula o sequía intraestival, evento climático que consiste en una disminución de la cantidad de precipitación a la mitad de la temporada de lluvias (ocurre principalmente en los meses cálidos de julio y agosto), este evento se caracteriza además de la disminución de la lluvia por un aumento de la temperatura (semana 28 a 33).

4.2. Ajuste de modelos

El ajuste de los modelos que a continuación se presenta, fue realizado después de eliminar los últimos 5 datos de la serie original y aplicando la metodología Box-Jenkins [8]. Se analizó la función de autocorrelación (fac) y la función de autocorrelación parcial (fap) de la serie para verificar estacionaridad y estacionalidad. La Figura 4.3, presenta la serie de tiempo de los registros de temperatura en la cual se observa que la serie de tiempo no es estacionaria. Se procedió entonces a realizar una diferencia a la serie y se observó que la fac decrece de manera exponencial en forma rápida por lo que se pudo garantizar

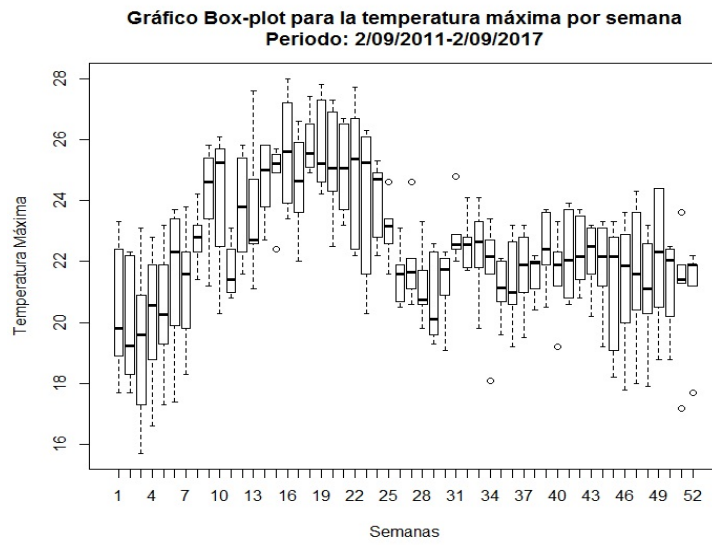


Figura 4.2: Gráfico box-plot y serie de tiempo, para los registros semanales de temperatura máxima.

la estacionaridad de la serie con una diferencia.

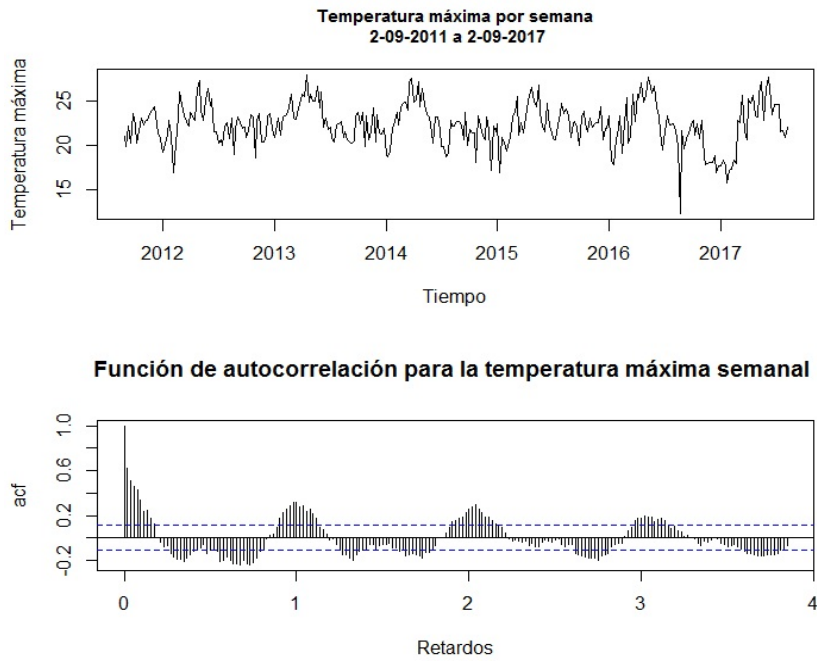


Figura 4.3: Serie de tiempo para los registros semanales.

Tabla 4.1: Modelos utilizados para ajustar los datos.

Modelo	AIC
ARIMA(3,1,2)	1281.79
ARIMA(3,1,1)	1281.79
ARIMA(3,1,0)	1279.85
ARIMA(2,1,2)	1282.04
ARIMA(2,1,1)	1281.01

Tabla 4.2: Coeficientes del modelo ARIMA(3,1,0).

	ar1	ar2	ar3	Constatn
	-0.4872	-0.2981	-0.1795	0.0036
s.e.	0.0559	0.0600	0.0559	0.0548
σ^2 estimado: 3.563, aic = 1279.85				

Con el análisis de la *fac* y la *fap* de la serie, se ajustaron varios modelos, considerando como mejor modelo a aquel con el menor número de Akaike. En la Tabla 4.1, se presentan todos los modelos que se ajustaron a los registros de temperatura con su correspondiente valor de Akaike. De acuerdo al criterio de Akaike, el mejor modelo para ajustar los datos, fue un modelo *ARIMA*(3, 1, 0) cuyos coeficientes se muestran en la Tabla 4.2. Una vez ajustado el modelo, se realizó un análisis de residuos. Con la prueba de Ljung-Box se verificó la no correlación de los residuos y con la prueba de Shapiro se verificó su normalidad (Figura 4.4).

Con los coeficientes del modelo obtenidos en la Tabla 4.2, el modelo *ARIMA*(3, 1, 0), que resultó ser elegido para modelar la temperatura máxima, es

$$\begin{aligned}
 Tmax_t = & 0.0036 + Tmax_{t-1} - 0.4872(Tmax_{t-1} - Tmax_{t-2}) \\
 & - 0.2981(Tmax_{t-2} - Tmax_{t-3}) - 0.1795(Tmax_{t-3} - Tmax_{t-4}) + \varepsilon_t,
 \end{aligned}
 \tag{2.1}$$

donde $Tmax_t$ denota a la Temperatura al instante t , ε_t es un proceso de Ruido Blanco.

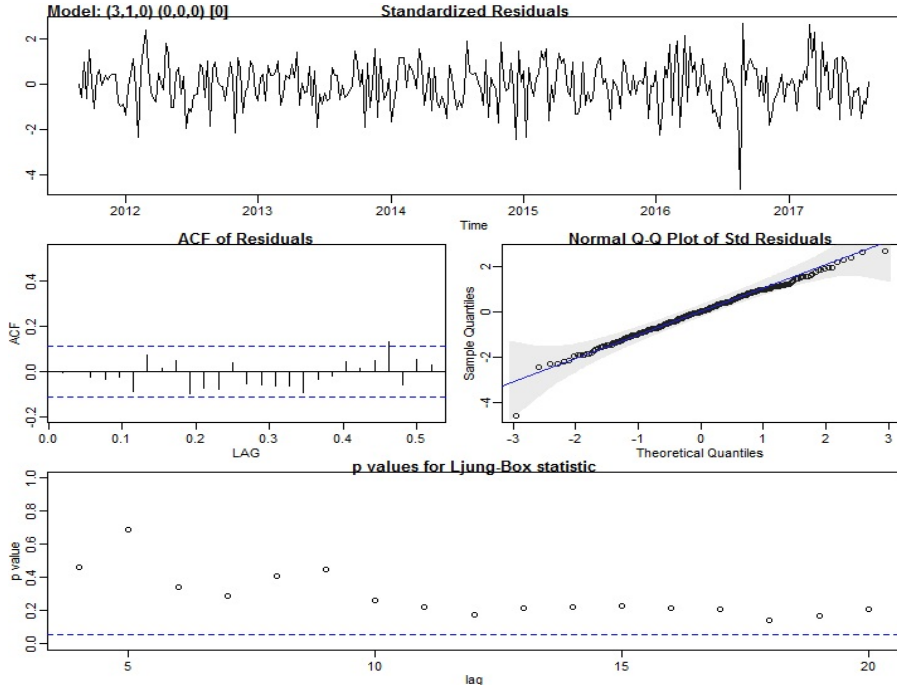


Figura 4.4: Análisis de los residuos del modelo $ARIMA(3,1,0)$.

4.3. Aplicación de metodologías para detectar un sólo punto de cambio (AMOC)

Después de verificar las suposiciones del Teorema 3.2, se calcularon los valores del estadístico de prueba $\Lambda_n(k)$ para $k = 1, \dots, 313$ (Figura 4.5), el valor máximo para $\Lambda_n(k)$ fue de $\Lambda_n(k)$ para $k = 36$.

Considerando $\alpha = 0.05$, el valor crítico para aceptar o rechazar un cambio en los parámetros del modelo expresado por (2.1) es 15.76829, por lo tanto, existe evidencia suficiente para garantizar un cambio en algunos parámetros.

Para aplicar la metodología de Gombay, se estimó y graficó cada componente del vector $\hat{\mathbf{B}}(u)$ (Figura 4.6). Además, en la Tabla 4.3, se presentan los valores máximos de $\hat{\mathbf{B}}(u)$ para $0 < u < 1$ así como el valor de α , de lo que se puede concluir que no se considera suficiente evidencia para garantizar un cambio en cualquier parámetro del modelo dado en (2.1).

Aplicando la librería *changepoint* de *R*, se detectó un único punto de cambio (AMOC) en la media y la varianza de la base de datos.

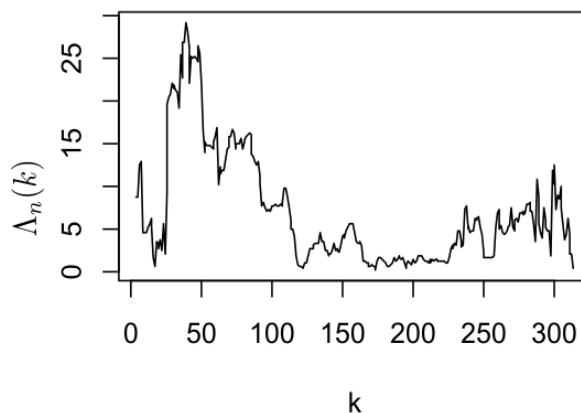


Figura 4.5: Valores del estadístico $\Lambda_n(k)$, para $k = 1, \dots, 313$.

Tabla 4.3: Valores máximos para probar un cambio en los parámetros μ , σ^2 , ϕ_1 , ϕ_2 and ϕ_3 .

Para probar un punto de cambio en:	k	$\sup_{0 \leq u \leq 1} \widehat{\mathbf{B}}(u)$	$C_1(0.01021614)$
μ	245	0.3325948	1.5139
σ^2	300	0.1475097	1.5139
ϕ_1	48	0.5952468	1.5139
ϕ_2	52	0.258343	1.5139
ϕ_3	263	0.4006972	1.5139

Tabla 4.4: Ubicación de un punto de cambio en la media y la varianza de la temperatura máxima utilizando la librería *changeoint* de *R*.

Puntos de cambio estimado en la media	Semana	Media estimada antes del punto de cambio	Media estimada después del punto de cambio
	250	22.64560	21.50152

Punto de cambio estimado en la varianza	Semana	Varianza estimada antes del punto de cambio	Varianza estimada después del punto de cambio
	27	4.078351	9.970822

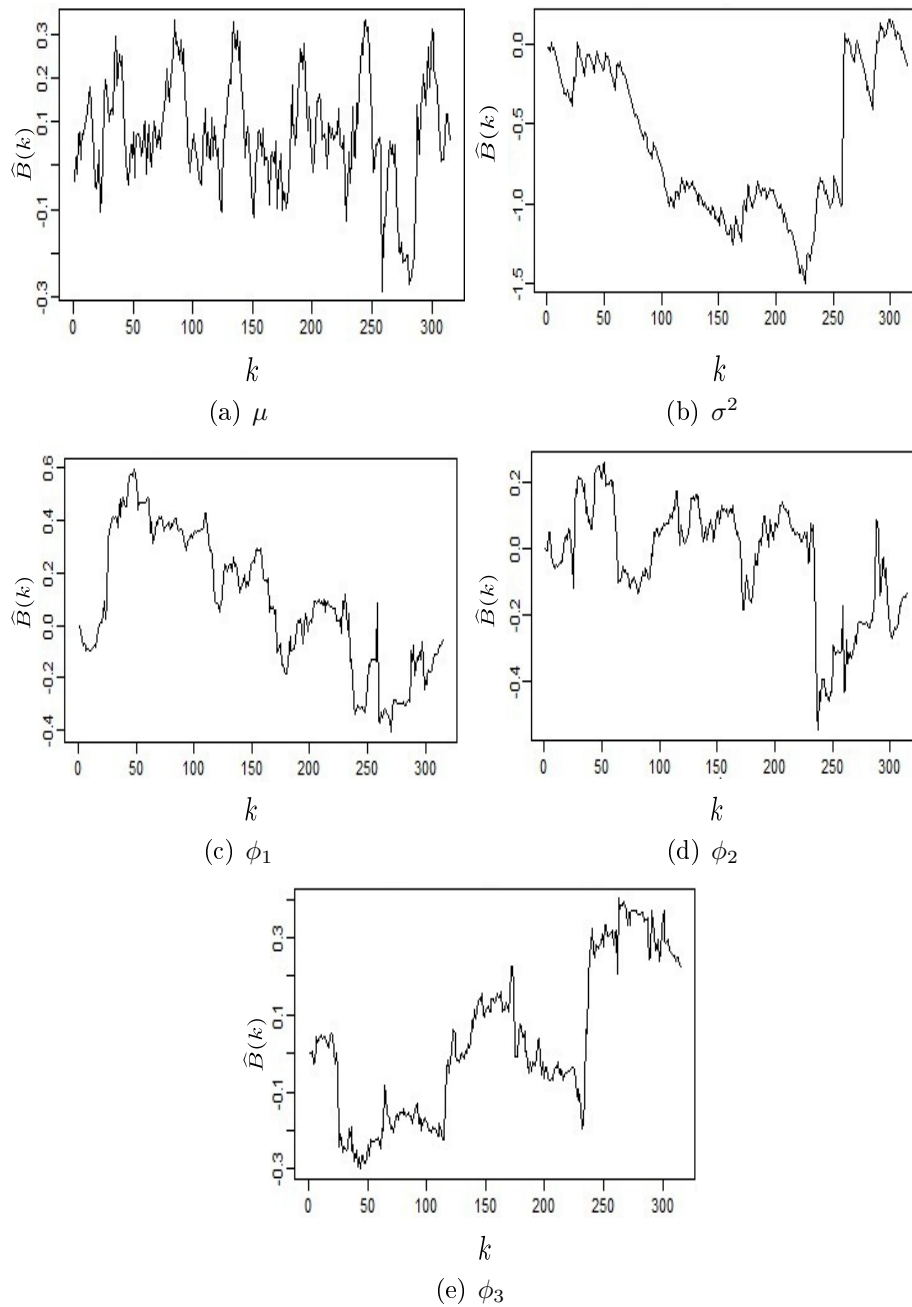
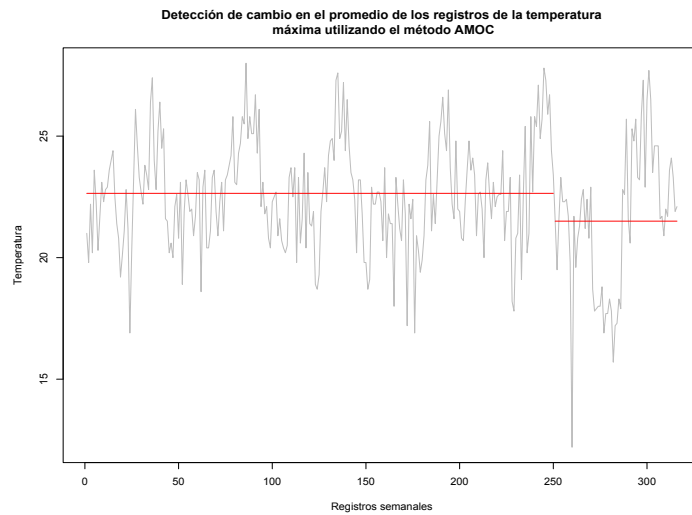
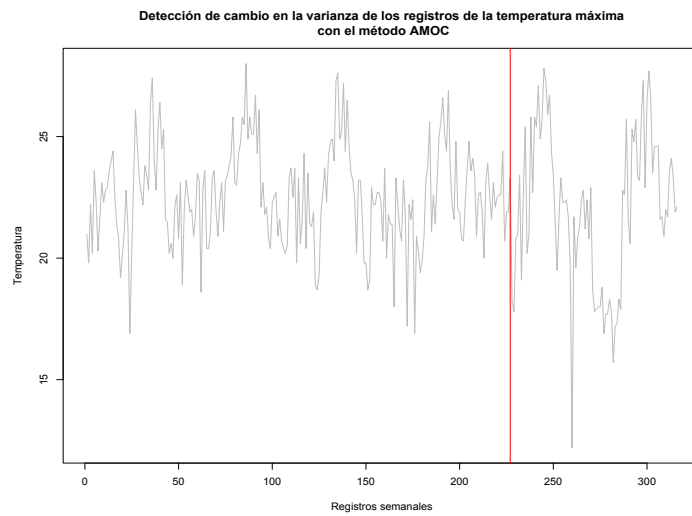


Figura 4.6: Valores de $\hat{\mathbf{B}}(k)$, $0 < k \leq n - 1$, para cada parámetro del modelo dado por (2.1).



(a) Punto de cambio para la media.



(b) Punto de cambio para la varianza.

Figura 4.7: Ubicación de los puntos de cambio para la media y la varianza de la temperatura máxima.

4.3.1. Detección de múltiples puntos de cambio

Con la librería *changepoint* de *R*, se realizó el análisis para determinar múltiples puntos de cambio usando los métodos PELT, BS y SN. Las tablas 4.5 y 4.7, presentan el número de puntos de cambio, así como su estimación de la media y la varianza de la temperatura máxima, respectivamente.

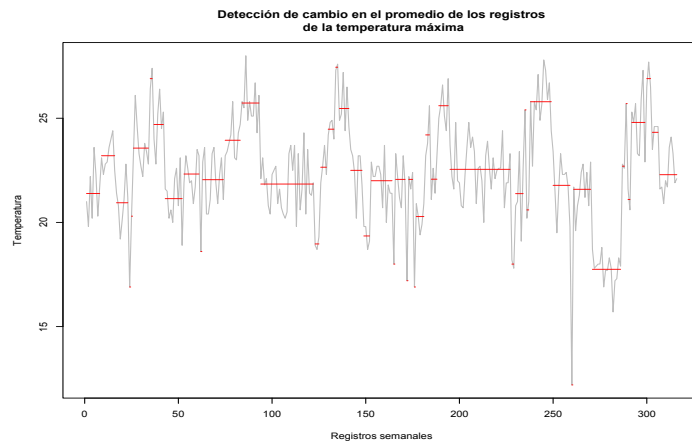
Las Tablas 4.6 y 4.8 presentan todos los puntos de cambio de media y varianza detectados con el método PELT. Para tener una mejor concepción de la ubicación de los puntos de cambio identificados mediante los métodos PELT, BS y SN y, para interpretar la ubicación de los puntos de cambio, se presentan las Figuras 4.8 y 4.9.

Tabla 4.5: Número de puntos de cambio para la media de temperatura máxima.

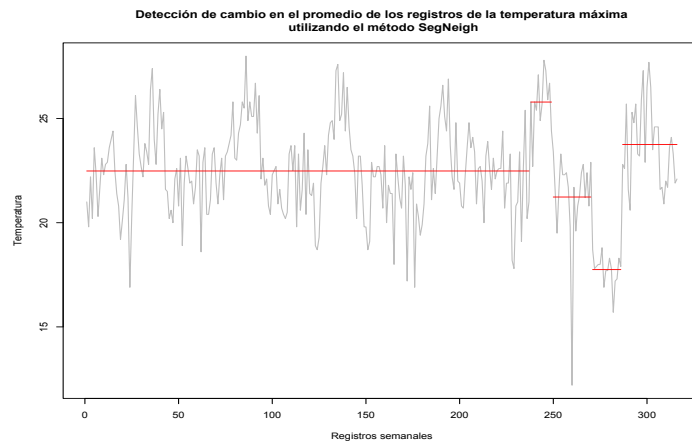
Método	Número de puntos de cambio	Punto de cambio estimado	Punto de cambio Media
PELT	48	Ver Tabla 4.6	Ver Tabla 4.6
SN	4	237, 249, 270, 286	22.48312, 25.79167, 21.23333, 17.75000, 23.75333
BS	5	237, 250, 270, 286, 306,	22.48312, 25.60769, 21.12500, 17.75000, 24.48500, 22.29000

Tabla 4.6: Ubicación de los puntos de cambio en la media de la temperatura máxima, determinada con el método PELT.

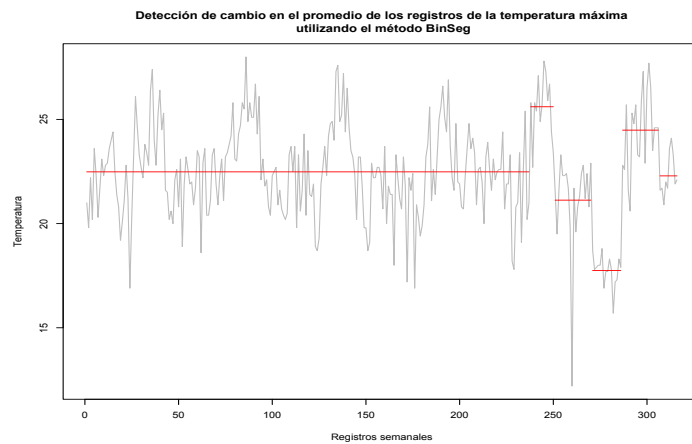
Semana	8	16	23	24	25	34	36	42	52
Promedio	21.38750	23.20000	20.94286	16.90000	20.30000	23.56667	26.90000	24.70000	21.14000
Semana	61	62	74	83	93	122	125	129	133
Promedio	22.32222	18.60000	22.05000	23.94444	25.73000	21.84138	18.96667	22.65000	24.47500
Semana	135	141	148	152	164	165	171	172	175
Promedio	27.45000	25.46667	22.50000	19.35000	22.00000	18.00000	22.06667	17.20000	22.06667
Semana	176	181	184	188	194	227	229	234	235
Promedio	16.90000	20.28000	24.20000	22.07500	25.60000	22.54848	18.00000	21.38000	25.40000
Semana	237	249	259	260	270	286	288	289	291
Promedio	20.60000	25.79167	21.78000	12.20000	21.59000	17.75000	22.70000	25.70000	21.10000
Semana	299	302	306						
Promedio	24.80000	26.90000	24.32500	22.29000					



(a) Método PELT.

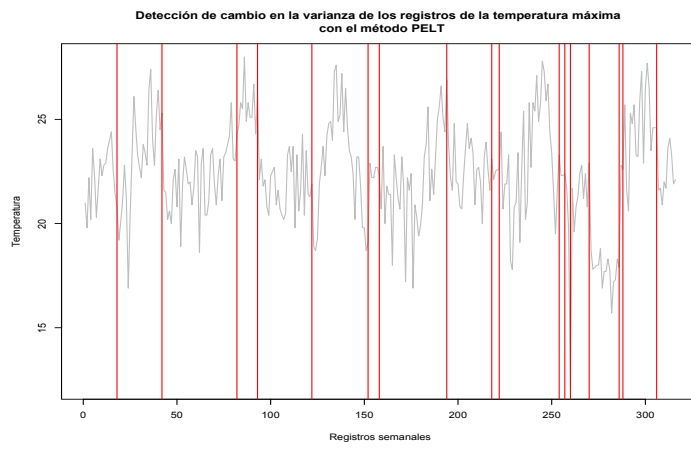


(b) Método SN.

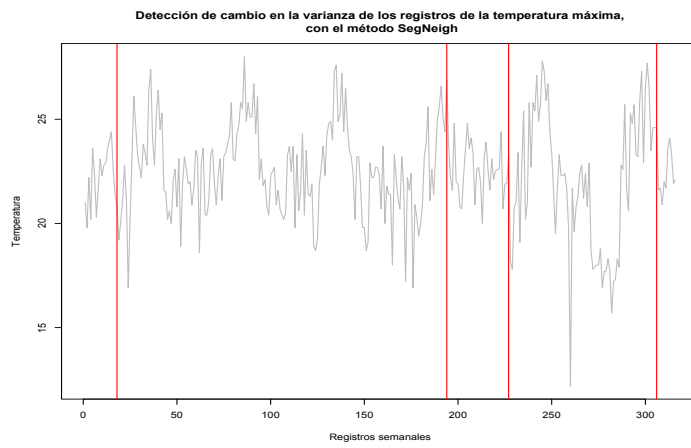


(c) Método BS.

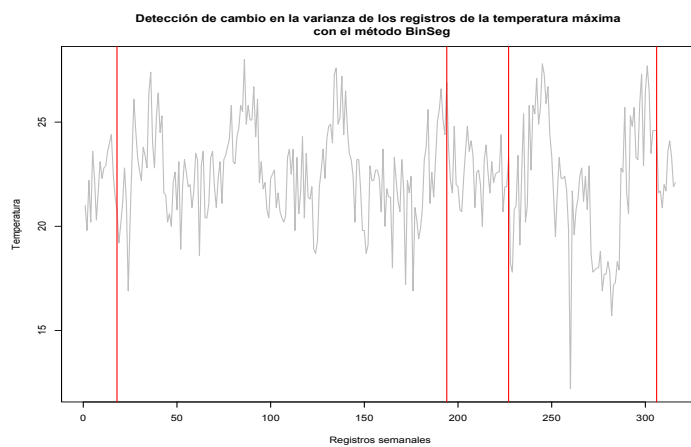
Figura 4.8: Ubicación de los puntos de cambio para la media de temperatura máxima.



(a) Método PELT.



(b) Método SN.



(c) Método BS.

Figura 4.9: Ubicación de los puntos de cambio para la varianza de temperatura máxima.

Tabla 4.7: Número de puntos de cambio en la varianza de la temperatura máxima.

Método	Número de puntos de cambio	Puntos de cambio estimados	Varianzas estimadas
PELT	17	Ver Tabla 4.8	Ver Tabla 4.8
SN	4	18 194 227 306	1.847222, 4.804260 1.454451, 11.119685, 1.043222
BS	4	18, 194, 227, 306	1.847222, 4.804260, 1.454451, 11.119685, 1.043222

Tabla 4.8: Ubicación de los puntos de cambio determinados con el método PELT.

Semana	18	42	82	93	122	152
Varianza	1.847222222	6.236503623	2.305583333	1.078545455	1.564655172	7.142298851
Semana	158	194	218	222	254	257
Varianza	0.092000000	5.700571429	1.654184783	0.056666667	7.507006048	0.003333333
Semana	260	270	286	288	306	
Varianza	25.270000000	1.105444444	0.548000000	0.020000000	3.610882353	1.043222222

4.3.2. Modelos ajustados después de detectar puntos de cambio en la media de temperatura máxima

Después de aplicar los métodos AMOC, BS, PELT y SN para la media, observamos que el método PELT realiza varias segmentaciones en los datos, por lo tanto, no es posible un ajuste ARIMA. El método BS realiza seis particiones en la base de datos; aunque, la última segmentación tiene solo siete datos (de la semana 307 a la 313) y luego no es posible un ajuste SARIMA en este segmento.

Para modelar las particiones de series temporales y comparar sus resultados, en este documento se consideraron particiones de los métodos AMOC y SN. En el método SN, consideramos, como una segmentación, las particiones de las semanas 276 a 286 (ver Figura 4.8), se ajustaron tres modelos ARIMA para la temperatura máxima en lapsos: semanas 1 a 237, 238 a 286 y 287 a 313 aplicando la metodología de Box-Jenkins.

La tabla 4.9 muestra los modelos ARIMA ajustados; La figura 4.10 presenta las series de tiempo particionadas y su función de autocorrelación para cada segmento.

La Tabla 4.10, presenta modelos ARIMA ajustados para tres particiones consideradas con el método SN; La figura 4.11 muestra cada partición para la serie de tiempo.

Tabla 4.9: Modelos ajustados después de detectar un sólo punto de cambio.

	Modelo ajustado Primer bloque de datos	AIC	Modelo ajustado Segundo bloque de datos	AIC
AMOC Method	ARIMA(2,1,1)	987.06	ARIMA(1,1,1)	273.61
	ARIMA(3,1,0)	987.74	ARIMA(2,1,0)	274.79
	ARIMA(2,1,2)	988.47	ARIMA(2,1,1)	275.61
	ARIMA(3,1,1)	988.85	ARIMA(2,1,2)	277.52
	ARIMA(3,1,2)	990.47		
	ARIMA(2,2,2)	991.34		
	ARIMA(2,1,0)	994.72		

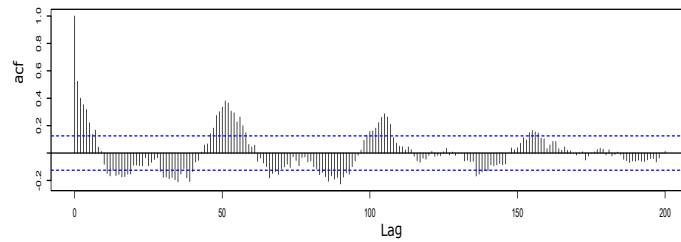
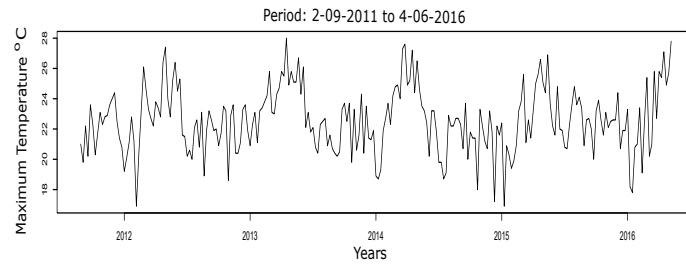
Tabla 4.10: Ajuste de modelos después de aplicar el método SN.

	Modelo ajustado Primer segmentación	AIC	Modelo ajustado Segunda segmentación	AIC	Modelo ajustado Tercera segmentación	AIC
Método SN	ARIMA(2,1,1)	914.29	ARIMA(1,1,1)	191.99	MA(1)	108.29
	ARIMA(3,1,2)	917.13	ARIMA(2,1,2)	194.44	ARMA(1,1)	110.21
	ARIMA(2,1,2)	924.6	ARIMA(1,1,0)	194.84	MA(2)	110.23
	ARIMA(3,1,1)	924.64	ARIMA(1,1,2)	197.58	ARMA(1,2)	111.71
				ARMA(2,2)	113.5	

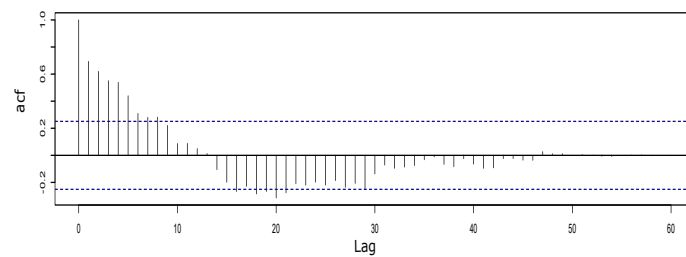
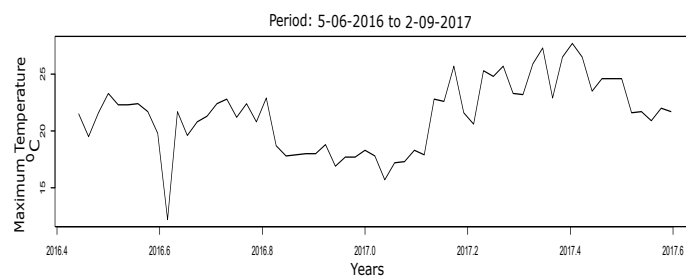
Para elegir el mejor modelo para la serie temporal dividida, se consideraron los Criterios de Información de Akaike, después de aplicar el método AMOC, el ajuste a la serie temporal se presenta en (3.2), la ecuación (3.3), representa el ajuste después de aplicar el método SN.

$$Tmax_t = \begin{cases} Tmax_{t-1} + 0.0213 - 0.7750(Tmax_{t-1} - Tmax_{t-2}) \\ + 0.0732(Tmax_{t-2} - Tmax_{t-3}) - 0.2464\varepsilon_{t-1} + 0.5789\varepsilon_{t-2} \\ + \varepsilon_t, & \text{if } 1 \leq t \leq 250 \\ 0.0102 + Tmax_{t-1} + 0.0419(Tmax_{t-1} - Tmax_{t-2}) + 0.5162\varepsilon_{t-1} \\ + \varepsilon_t, & \text{if } t \geq 250 \end{cases} \quad (3.2)$$

$$Tmax_t = \begin{cases} -0.0008 + Tmax_{t-1} + 0.4547(Tmax_{t-1} - Tmax_{t-2}) \\ + 0.1698(Tmax_{t-2} - Tmax_{t-3}) - \varepsilon_{t-1} + \varepsilon_t, & \text{si } 1 \leq t \leq 237 \\ -0.2044 + Tmax_{t-1} + 0.3740(Tmax_{t-1} - Tmax_{t-2}) + \varepsilon_{t-1} \\ + \varepsilon_t, & \text{si } 238 \leq t \leq 286 \\ 23.8714 + 0.3968\varepsilon_{t-1}, & \text{si } t \geq 287 \end{cases} \quad (3.3)$$

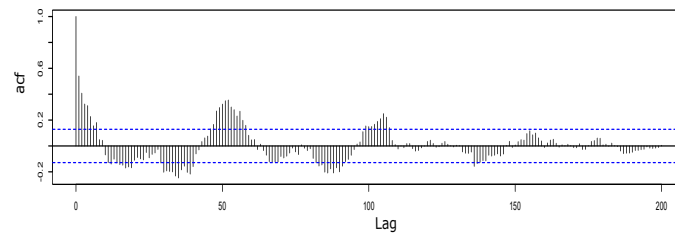
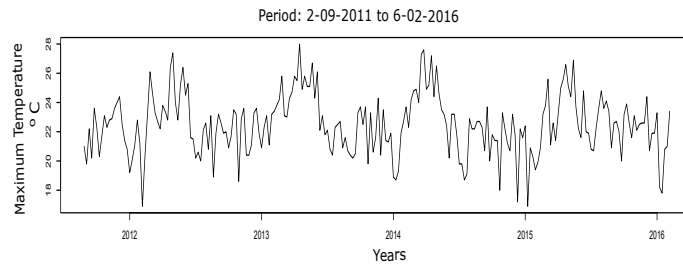


(a) Primer segmentación.

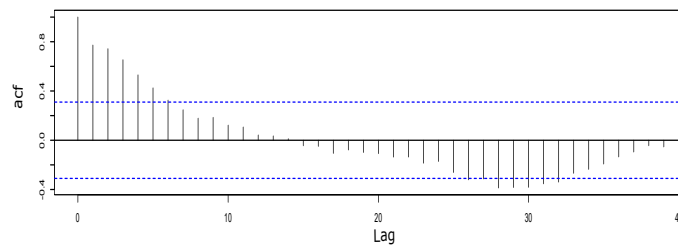
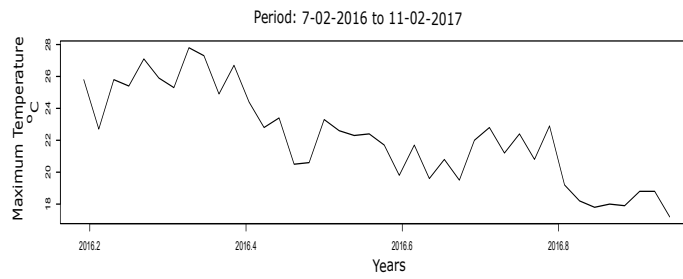


(b) Segunda segmentación.

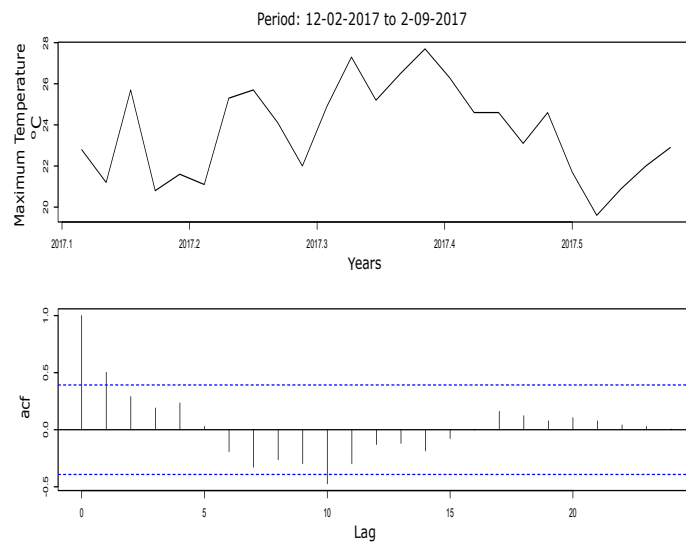
Figura 4.10: Segmentación de la serie temporal después de detectar un punto de cambio.



(a) Primer segmentación.



(b) Segunda segmentación.



(c) Tercer segmentación.

Figura 4.11: Series de tiempo particionadas después de aplicar el método SN.

Conclusiones

Con la finalidad de determinar cambios en la temperatura máxima de un municipio del norte del Estado de Tlaxcala, mediante metodologías estadísticas, se realizó un estudio comparativo sobre metodologías para detectar puntos de cambio en series de tiempo. Para la implementación de las metodologías, el Centro de Investigación en Cambio Climático de la Universidad Autónoma del Estado de Tlaxcala junto con Protección Civil del Estado, proporcionaron una base de datos (2 de septiembre de 2011 a 2 de septiembre de 2017) de una estación climatológica ubicada en Tlaxco, Tlaxcala. Esta estación, registra entre otras variables climáticas, la temperatura máxima en intervalos de cada media hora. Para la implementación de las metodologías descritas en este trabajo de investigación, fue necesario un análisis de calidad de datos que incluyó la estimación de datos perdidos (se tuvo una pérdida de información del 2%), para estimar los datos perdidos de un año en particular, se consideró hacer un proceso de regresión lineal con el año en que se obtuvo la correlación más alta. La base de datos original, puede solicitarse a Dr. Tomás Morales Acoltzi vía el siguiente correo electrónico acoltzi@atmosfera.unam.mx.

Para la implementación de las metodologías de Davis *et al.* y Gombay E., se desarrollaron programas en R (ver Apéndice B), dado que las metodologías consideran matrices inversas y dado que se realizó una diferencia a la base de datos para que se cumplieran todos los supuestos de las metodologías, computacionalmente, se provocó que R no pudo calcular la inversa de las matrices correspondientes, por lo que se implementó en las metodologías, el cálculo de la matriz inversa generalizada.

Las metodologías de Davis *et al.* [17] y Gombay [37] nos permiten probar si existe un único punto de cambio o no en algún parámetro del modelo autorregresivo de orden p ; sin embargo, la metodología de Davis *et al.* no nos permite

estimar el tiempo en que ocurre. Al aplicar estas metodologías en la base de datos de temperatura máxima, los resultados encontrados muestran que con la metodología de Davis *et al.*, la evidencia de un cambio en los parámetros del modelo autorregresivo propuesto, por otro lado, la metodología de Gombay, no proporciona evidencia suficiente para considerar un cambio, esto se atribuye al hecho de que para ajustar el modelo autorregresivo a la base de datos, hubo que hacer una diferencia.

Las otras metodologías aplicadas no consideran la estructura de los datos, además de estar diseñadas para estimar los puntos de cambio en la media y la varianza. El método AMOC estima un único punto de cambio, mientras que el método PELT estima 48 puntos de cambio en la base de datos considerada. La segmentación de esta última metodología hace que sea difícil proponer modelos que se ajusten a la base de datos. Para el cambio en la media de la temperatura máxima, los métodos AMOC, PELT, SN y BS detectaron el mismo punto de cambio en la semana 250.

La metodología de Gombay no se presentó evidencia suficiente para decidir un cambio en la varianza, el método AMOC estima un cambio en la varianza en la semana 27 y los métodos PELT, SN y BS coinciden con un cambio en las semanas 18, 194, 227 y 306. Con el método AMOC, el cambio se informa en la semana 250, que representa la tercera semana de junio de 2016, revisando los episodios de frío y calor por temporada informados por el Centro de Predicción Climática [80], en el mes de junio de 2016, se observa que el océano cambió de un calentamiento a un enfriamiento, esto nos permite validar el cambio reflejado con el método AMOC, de manera similar, para el método SN, los puntos de cambio estimados coinciden con la transición de un calentamiento al enfriamiento del océano (semana 237, que representa la última semana del mes de febrero de 2016) y un período sin calentamiento del océano hasta un enfriamiento (semana 287, que representa el mes de junio de 2017).

Aunque los métodos PELT y BS detectan un cambio en la semana 306 (Tablas 4.5, 4.6, 4.7 y 4.8), esta semana es de las últimas observaciones por lo tanto, éste cambio detectado puede no ser un punto de cambio confiable.

El estudio de los puntos de cambio en los datos climáticos a nivel regional es esencial para los tomadores de decisiones y para el desarrollo local de las comunidades a analizar, ya que el nivel de impacto en su crecimiento económico puede dimensionarse al mismo tiempo que podrá establecer medidas para adaptarse a estos cambios con la sociedad y los sectores político-económicos.

A lo largo del desarrollo de esta investigación, se participó en congresos nacionales, internacionales además de tener un artículo que ha sido aceptado, a

continuación, se lista los congresos y el artículo que sustentan este proyecto.

1. *Distribución asintótica de la prueba de razón de verosimilitud para verificar la existencia de puntos de cambio*. XXXI Foro Internacional de la Probabilidad y Estadística (2015).
2. *Puntos de cambio en series de tiempo*. 4 CIMA FCFM-BUAP (2016).
3. *Distribución asintótica de una prueba de razón de verosimilitud para verificar la existencia de puntos de cambio*. XVI Encuentro de la Participación de la Mujer en la Ciencia (2017).
4. *Change point analysis of maximum temperature databases in Tlaxco, Tlaxcala, México*. International Congress of Mathematicians. Río de Janeiro Brasil (2018).
5. *Detección de cambios en la temperatura máxima de Tlaxco, Tlaxcala utilizando metodologías de puntos de cambio*. XV Encuentro de la Participación de la Mujer en la Ciencia (2019) [45].
6. *Parametric methodologies for detecting changes in maximum temperature of Tlaxco-Tlaxcala, México* (artículo aceptado en el Journal of Probability and Statistics) [44].

Para los estudios posteriores, se tiene la intención de considerar modelos estacionales, así como modelos multivariados que pueden estar relacionados para describir el clima de alguna región.

La aplicación de las metodología paramétricas descritas en el capítulo 3, permitió identificar trabajos futuros con los que se puede continuar a partir de este estudio, a continuación se presenta una lista de los posibles trabajos.

- Dado que la mayoría de las series de tiempo que se analizan en el área de climatología son estacionales, es importante implementar metodologías paramétricas y no paramétricas para detectar puntos de cambio en modelos SARIMA.
- De acuerdo con la metodología del IPCC, es necesario comparar los resultados con series temporales históricas para verificar los cambios obtenidos y poder identificar las causas de esos cambios (las causas pueden ser por variabilidad natural o bien por actividades antropogénicas).
- Con base en los resultados de la distribución exacta obtenida en el Teorema 3.2 del capítulo 3 que utiliza resultados del teorema de Darling-Erdős, se considera que es posible lograr una mejora proponiendo una sucesión

de constantes a_n y b_n de este teorema, utilizando métodos de Montecarlo y la teoría de valores extremos.

- Algunas metodologías estudiadas que no se presentaron en este trabajo, determinan un estadístico de prueba cuya distribución asintótica es proporcional a una distribución χ^2 , por lo cual es importante analizar el comportamiento de las ditribuciones de este tipo.

APÉNDICE A

Convergencia Estocástica

A.1. Convergencia en distribución

Definición A.1 Sea $\{F_n\}$ una sucesión de funciones de distribución. Si existe una función de distribución F , tal que $F_n(x) \rightarrow F(x)$ conforme $n \rightarrow \infty$ en cada punto x en el que F es continua, se dice que F_n converge en ley (o débilmente a F) y se denota por $F_n \xrightarrow{d} F$.

La definición A.1, significa que para todo $\varepsilon > 0$ existe $N(\varepsilon, x) \in \mathbb{N}$, tal que para todo $n \geq N(\varepsilon, x)$ y todo punto de continuidad x de F ,

$$|F_n(x) - F(x)| < \varepsilon.$$

Definición A.2 Si $\{X_n\}$ es una sucesión de variables aleatorias y F_n es la correspondiente sucesión de funciones de distribución, se dice que X_n converge en distribución a X si existe una variable aleatoria X con función de distribución F tal que $F_n \xrightarrow{d} F$ y se denota por $X_n \xrightarrow{d} X$.

Teorema A.1 Sea $\{X\}_{n \geq 1}$ una sucesión de variables aleatorias con valores reales, considérese $f_n(k) = P\{X_n = k\}$, $k = 0, 1, \dots$ la función de probabilidad de X_n , $n = 1, 2, \dots$ y $f(k) = P\{X = k\}$ como la función de densidad de X . Entonces,

$$f_n(x) \rightarrow f(x) \text{ para todo } x \Leftrightarrow X_n \xrightarrow{d} X.$$

Demostración: \Rightarrow) Como para cada k , $f_n(k) \rightarrow f(k)$, se cumple que $\sum_{j=-\infty}^k f_n(j) \rightarrow \sum_{j=-\infty}^k f(j) \Rightarrow F_n(k) \rightarrow F(k)$.

\Leftrightarrow) Supóngase que $X_n \xrightarrow{d} X$, entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{j=-\infty}^k f_n(k) &= \sum_{j=-\infty}^k f(k) \Rightarrow \\ \lim_{n \rightarrow \infty} \int_{(-\infty, k]} X_n dP &= \int_{(-\infty, k]} X dP. \end{aligned}$$

Teorema A.2 Sea $\{X_n\}$ una sucesión de variables aleatorias tales que $X_n \xrightarrow{d} X$ y sea c una constante, entonces

- $X_n + c \xrightarrow{d} X + c$,
- $cX_n \xrightarrow{d} cX$.

Demostración:

- Sea $\varepsilon > 0$, como $X_n \xrightarrow{d} X$ existe $N(\varepsilon, x)$ tal que para todo $n \geq N(\varepsilon, x)$ y todo punto de continuidad x de F ,

$$\begin{aligned} \left| P\{X_n + c \leq x\} - P\{X + c \leq x\} \right| &= \left| P\{X_n \leq x - c\} - P\{X \leq x - c\} \right| \\ &= \left| F_n(x - c) - F(x - c) \right| < \varepsilon. \end{aligned}$$

- Sea $c < 0$, $\varepsilon > 0$, como $X_n \xrightarrow{d} X$ existe $N(\varepsilon, x)$ tal que para todo $n \geq N(\varepsilon, x)$ y todo punto de continuidad x de F ,

$$\begin{aligned} \left| P\{cX_n \leq x\} - P\{cX \leq x\} \right| &= \left| P\left\{X_n \geq \frac{x}{c}\right\} - P\left\{X \geq \frac{x}{c}\right\} \right| \\ &= \left| 1 - P\left\{X_n < \frac{x}{c}\right\} - 1 + P\left\{X < \frac{x}{c}\right\} \right| \\ &= \left| F_n\left(\frac{x}{c}\right) - F\left(\frac{x}{c}\right) \right| < \varepsilon. \end{aligned}$$

Para $c > 0$, se prueba de manera similar.

A.2. Convergencia en Probabilidad

Definición A.3 Sea $\{X_n\}$ una sucesión de variables aleatorias definidas en algún espacio de probabilidad (Ω, \mathcal{F}, P) . Se dice que la sucesión $\{X_n\}$ converge en probabilidad a la variable aleatoria X si para cualquier $\varepsilon > 0$

$$P\{|X_n - X| > \varepsilon\} \rightarrow 0$$

conforme $n \rightarrow \infty$,
o equivalentemente $P\{|X_n - X| \leq \varepsilon\} = 1$ conforme $n \rightarrow \infty$ y se escribe $X_n \xrightarrow{P} X$.

Teorema A.3 Sea $X_n \xrightarrow{P} X$ y g una función continua definida sobre \mathbb{R} , entonces $g(X_n) \xrightarrow{P} g(x)$ conforme $n \rightarrow \infty$.

Demostración:

Como X es una variable aleatoria, dado $\varepsilon > 0$ se puede encontrar una constante $k = k(\varepsilon)$ tal que $P\{|X| > k\} < \frac{\varepsilon}{2}$.

Como g es una función continua sobre \mathbb{R} , se cumple que g es uniformemente continua sobre $[-k, k]$, así, es posible encontrar un $\delta = \delta(\varepsilon, k)$ tal que $|g(x_n - g(x))| < \varepsilon$ siempre que $|x| < k$ y $|x_n - x| < \delta$.

Sea $A = \{|X| \geq k\}$, $B = \{|X_n - X| \geq \delta\}$, $C = \{|g(X_n) - g(X)| \geq \varepsilon\}$, de forma que $A \cap B \subseteq C$ y de donde se sigue que $C^c \subseteq A^c \cup B^c$ y por lo tanto $P(C^c) \leq P(A^c) + P(B^c)$.

Así, $P\{|g(x_n - g(x))| \geq \varepsilon\} \leq P\{|x| > k\} + P\{|x_n - x| \geq \delta\} < \varepsilon$.

Para $n \geq N(\varepsilon, \delta, k)$, donde $N(\varepsilon, \delta, k)$ es elegido de tal forma que

$$P\{|X_n - X| \geq \delta\} < \varepsilon/2, \text{ para } n \geq N(\varepsilon, \delta, k).$$

Teorema A.4 Si $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{d} X$.

Demostración:

Sean F_n y F las funciones de distribución de X_n y X respectivamente.

$$\begin{aligned} \{\omega : X(\omega) \leq x'\} &= \{\omega : X(\omega) \leq x \cap X(\omega) \leq x'\} \cup \{\omega : X(\omega) > x \cap X(\omega) \leq x'\} \\ &\subseteq \{X_n \leq x\} \cup \{X_n > x \cap X_n \leq x'\}, \end{aligned}$$

$$\text{Así, } F(x') \leq F_n(x) + P\{X_n > x, X_n \leq x'\}.$$

Como $X_n - X \xrightarrow{P} 0$, se cumple para $x' \leq x$,

$P\{X_n > x, X \leq x'\} \leq P\{|X_n - X| > x - x'\} \rightarrow 0$ conforme $n \rightarrow \infty$, por lo tanto, $F(x') \leq \lim_{n \rightarrow \infty} F_n(x)$, $x' < x$.

Similarmente, intercambiando X y X_n y x y x' se tiene

$$\overline{\lim}_{n \rightarrow \infty} F_n(x) \leq F(x''), \quad x < x''.$$

Así para $x' < x < x''$, se tiene

$$F(x') \leq \underline{\lim} F_n(x) \leq \overline{\lim} F(x) \leq F(x'').$$

Como F tiene un número contable de puntos de discontinuidad, se elige x un punto de continuidad de F tal que $x'' \downarrow x$ y $x' \uparrow x$, así se tiene

$$F(x) = \lim_{n \rightarrow \infty} F_n(x),$$

en todos los puntos de continuidad de F .

A.3. Convergencia casi segura

Definición A.4 Sea $\{X_n\}$ una sucesión de variables aleatorias. Se dice que X_n converge casi seguramente (a.s) a una variable aleatoria X sí y sólo si

$$P \left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} = 1,$$

Teorema A.5 $X_n \xrightarrow{a.s} X$ sí y sólo si $\lim_{n \rightarrow \infty} P \left\{ \sup_{m \geq n} |X_m - X| > \varepsilon \right\} = 0$.

Demostración:

Como $X_n \xrightarrow{a.s} X$, $X_n - X \xrightarrow{a.s} 0$, por lo que será suficiente demostrar la equivalencia de $X_n \xrightarrow{a.s} 0$ y $\lim_{n \rightarrow \infty} P \left\{ \sup_{m \geq n} |X_m - X| > \varepsilon \right\} = 0$. \Rightarrow) Supongamos que $X_n \xrightarrow{a.s} 0$. Sea $\varepsilon > 0$ y considérese

$$A_n(\varepsilon) = \left\{ \sup_{m \geq n} |X_m| > \varepsilon \right\}, \quad C = \left\{ \lim_{n \rightarrow \infty} X_n = 0 \right\} \text{ y } B_n(\varepsilon) = C \cap A_n(\varepsilon).$$

Se puede verificar que $B_{n+1}(\varepsilon) \subset B_n(\varepsilon)$ y por tanto $\bigcap_{n=1}^{\infty} B_n(\varepsilon) = \emptyset$, así que $\lim_{n \rightarrow \infty} P \{B_n(\varepsilon)\} = P \{\bigcap_{n=1}^{\infty} B_n(\varepsilon)\} = P\{\emptyset\} = 0$.

Como $P(C) = 1$ (pues $X_n \xrightarrow{a.s} X$), $P(C^c) = 0$, entonces,

$$\begin{aligned} P(B_n(\varepsilon)) &= P(A_n(\varepsilon) \cap C) = 1 - P(C^c \cup A_n^c(\varepsilon)) \\ &= 1 - P(C^c) - P(A_n^c(\varepsilon)) + P(C^c \cap A_n^c(\varepsilon)) \\ &= P(A_n(\varepsilon)) + P(C^c \cap A_n^c(\varepsilon)) \\ &= P(A_n(\varepsilon)). \end{aligned}$$

$\therefore \lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = \lim_{n \rightarrow \infty} P(A_n(\varepsilon)) \Rightarrow \lim_{n \rightarrow \infty} P \left\{ \sup_{m \geq n} |X_m| > \varepsilon \right\} = 0$. Así, si $X_n \xrightarrow{a.s}$

$$0 \Rightarrow \lim_{n \rightarrow \infty} P \left\{ \sup_{m \geq n} |X_m| > \varepsilon \right\}.$$

\Leftrightarrow Supóngase que $\lim_{n \rightarrow \infty} P \left\{ \sup_{m \geq n} |X_m| > \varepsilon \right\} = \lim_{n \rightarrow \infty} A_n(\varepsilon) = 0$ y sea $D(\varepsilon) = \left\{ \overline{\lim}_{n \rightarrow \infty} |X_n| > \varepsilon > 0 \right\}$. $D(\varepsilon) \subset A_n(\varepsilon)$, para $n = 1, 2, \dots$, por lo tanto $0 \geq P(D(\varepsilon)) \leq P(A_n(\varepsilon)) \xrightarrow{n \rightarrow \infty} 0 \Rightarrow P(D(\varepsilon)) = 0$.

También se cumple

$C^c = \left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq 0 \right\} \subset \bigcup_{n=1}^{\infty} \left\{ \overline{\lim}_{n \rightarrow \infty} |X_n(\omega)| > \frac{1}{k} \right\}$, así,

$P(C^c) = 1 - P(C) \leq \sum_{n=1}^{\infty} P \left\{ D \left(\frac{1}{k} \right) \right\} = 0$, por lo tanto

$P(C) = 1$, es decir;

$$P \left\{ \omega : \lim_{n \rightarrow \infty} X_n \right\} = 0 \Rightarrow X_n \xrightarrow{a.s.} 0.$$

Observación A.1 $X_n \xrightarrow{a.s.} 0$ significa que para todo $\varepsilon > 0$, $\eta > 0$ arbitrario, se puede encontrar η_0 tal que

$$P \left\{ \sup_{\eta \geq \eta_0} |X_n| > \varepsilon \right\} < \eta,$$

en efecto se puede escribir equivalentemente que

$$\lim_{\eta_0 \rightarrow \infty} P \left\{ \bigcup_{\eta \geq \eta_0} \{|X_n| > \varepsilon\} \right\} = 0.$$

Teorema A.6 Si $X_n \xrightarrow{a.s.} X$ entonces $X_n \xrightarrow{P} X$.

Demostración: Por la observación anterior, para $\varepsilon > 0$ y $\eta > 0$ se puede encontrar un $\eta_0 = \eta_0(\varepsilon, \eta)$ tal que $P \left\{ \bigcap_{\eta=\eta_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \right\} \geq 1 - \eta$.

$\bigcap_{\eta=\eta_0}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\} \subset \{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon\}$ para $\eta \geq \eta_0$, así

para $\eta \geq \eta_0$,

$$P \{|X_n - X| \leq \varepsilon\} \geq P \left\{ \bigcap_{\eta=\eta_0}^{\infty} \{|X_n - X| \leq \varepsilon\} \right\} \geq 1 - \eta \text{ entonces,}$$

$$P \{|X_n - X| \leq \varepsilon\} \geq 1 - \eta \Rightarrow P \{|X_n - X| > \varepsilon\} < 1 - (1 - \eta) = \eta \quad \forall \eta \geq \eta_0 \Rightarrow$$

$$\lim_{n \rightarrow \infty} P \{|X_n - X| > \varepsilon\} = 0 \Rightarrow X_n \xrightarrow{P} X.$$

Teorema A.7 Sean $\{X_n, Y_n\}_{n \geq 1}$ una sucesión de variables aleatorias tales que $|X_n - Y| \xrightarrow{P} 0$ y $Y_n \xrightarrow{d} Y$ entonces $X_n \xrightarrow{d} Y$.

Demostración: Sea x un punto de continuidad de la función de distribución de Y y $\varepsilon > 0$. Entonces,

$$\begin{aligned} P\{X_n \leq x\} &= P\{X_n \leq x + Y_n - Y_n\} \\ &= P\{Y_n \leq x + Y_n - X_n\} \\ &= P\{Y_n \leq x + Y_n - X_n, Y_n - X_n \leq \varepsilon\} \\ &\quad + P\{Y_n \leq x + Y_n - X_n, Y_n - X_n > \varepsilon\} \\ &\leq P\{Y_n \leq x + \varepsilon\} + P\{Y_n - X_n > \varepsilon\}. \end{aligned}$$

Como $\lim_{n \rightarrow \infty} P\{|Y_n - X_n| > \varepsilon\} = 0$ entonces

$$\overline{\lim}_{n \rightarrow \infty} P\{X_n \leq x\} \leq \underline{\lim}_{n \rightarrow \infty} P\{Y_n \leq x + \varepsilon\}, \text{ similarmente,}$$

$$\begin{aligned} P\{Y_n \leq x - \varepsilon\} &= P\{Y_n \leq x + X_n - X_n\} \\ &= P\{X_n \leq x - \varepsilon + X_n - Y_n\} \\ &= P\{X_n \leq x - \varepsilon + X_n - Y_n, X_n - Y_n \leq \varepsilon\} \\ &\quad + P\{X_n \leq x - \varepsilon + X_n - Y_n, X_n - Y_n > \varepsilon\} \\ &\leq P\{X_n - Y_n > \varepsilon\} + P\{X_n \leq x\}. \end{aligned}$$

Luego,

$$\overline{\lim}_{n \rightarrow \infty} P\{Y_n \leq x - \varepsilon\} \leq \underline{\lim}_{n \rightarrow \infty} P\{X_n - Y_n > \varepsilon\} + \underline{\lim}_{n \rightarrow \infty} P\{X_n \leq x\}, \text{ así,}$$

$$\overline{\lim}_{n \rightarrow \infty} P\{Y_n \leq x - \varepsilon\} \leq \underline{\lim}_{n \rightarrow \infty} P\{X_n \leq x\} \leq \overline{\lim}_{n \rightarrow \infty} P\{X_n \leq x\} \leq \underline{\lim}_{n \rightarrow \infty} P\{Y_n \leq x + \varepsilon\}$$

como $\varepsilon > 0$, el resultado se obtiene si $\varepsilon \rightarrow 0$.

Corolario A.1 Si $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{d} X$.

Demostración: Considerando $Y_n = X_n$, se tiene que $X_n - X \xrightarrow{P} 0$ y $Y_n = X_n \xrightarrow{d} X = Y$, luego por el teorema anterior, $X_n \xrightarrow{d} X$.

Teorema A.8 (Teorema de Slutsky) Sea $\{X_n, Y_n\}_{n \geq 1}$ una sucesión de pares de variables aleatorias y c una constante. Entonces

1. $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} c \Rightarrow X_n + Y_n \xrightarrow{d} X + c.$

2. $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} c \Rightarrow$

$$\begin{cases} X_n Y_n \xrightarrow{d} cX & \text{si } c \neq 0, \\ X_n Y_n \xrightarrow{P} 0 & \text{si } c = 0, \end{cases}$$

3. $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} c \Rightarrow \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}.$

Demostración:

1. Como $Y_n \xrightarrow{P} c$ entonces $(Y_n + X_n) - (X_n + c) \xrightarrow{P} 0$, si $X_n \xrightarrow{d} X$ implica que $X_n + c \xrightarrow{d} X + c$, luego por el teorema anterior, $Y_n + X_n \xrightarrow{d} X + c$.
2. a) Supóngase que $c \neq 0$, como X_n es una variable aleatoria y $Y_n - c \xrightarrow{P} 0$ entonces $X_n Y_n - c X_n = X_n (Y_n - c) \xrightarrow{P} 0$, además si $X_n \xrightarrow{d} X$ entonces $c X_n \xrightarrow{d} c X$, luego por el teorema anterior, $X_n Y_n \xrightarrow{d} c X$.
- b) Sea $c = 0$, para cualquier $k > 0$ fijo,

$$P\{|X_n Y_n| > \varepsilon\} = P\left\{|X_n Y_n| > \varepsilon, |Y_n| \leq \frac{\varepsilon}{k}\right\} \\ + P\left\{|X_n Y_n| > \varepsilon, |Y_n| > \frac{\varepsilon}{k}\right\}$$

$$= P\left\{|X_n| > \frac{\varepsilon}{|Y_n|}, \frac{\varepsilon}{|Y_n|} \geq k\right\} + P\left\{|X_n Y_n| > \varepsilon, |Y_n| > \frac{\varepsilon}{k}\right\}$$

$$\text{luego } P\{|X_n Y_n| > \varepsilon\} \leq P\{|X_n| > k\} + P\left\{|Y_n| > \frac{\varepsilon}{k}\right\},$$

como $Y_n \xrightarrow{P} 0$ y $X_n \xrightarrow{d} X$, se sigue que para k fijo y positivo,

$$\overline{\lim}_{n \rightarrow \infty} P\{|X_n Y_n| > \varepsilon\} \leq \overline{\lim}_{n \rightarrow \infty} P\{|X_n| > k\} + \overline{\lim}_{n \rightarrow \infty} P\left\{|Y_n| > \frac{\varepsilon}{k}\right\} \\ = P\{|X| > k\},$$

como k es arbitrario, es posible hacer $P\{|X| > k\}$ tan pequeño como sea posible, tomando k suficientemente grande, así se sigue que $X_n Y_n \xrightarrow{P} 0$.

3. como $Y_n \xrightarrow{P} c$, $c \neq 0$, $Y_n^{-1} \xrightarrow{P} c^{-1} \Rightarrow Y_n^{-1} - c^{-1} \xrightarrow{P} 0$, como X_n es una variable aleatoria, $X_n (Y_n^{-1} - c^{-1}) = X_n Y_n^{-1} - c^{-1} X_n \xrightarrow{P} 0$, y $c^{-1} X_n \xrightarrow{d} c^{-1} X$, luego por el teorema anterior, $X_n Y_n^{-1} \xrightarrow{d} c^{-1} X$.

Teorema A.9 (Teorema del mapeo continuo) Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ continua en cada punto de un conjunto C tal que $P(\mathbf{X} \in C) = 1$.

1. Si $\mathbf{X}_n \xrightarrow{a.s.} \mathbf{X}$ entonces $g(\mathbf{X}_n) \xrightarrow{a.s.} g(\mathbf{X})$.
2. Si $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ entonces $g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{X})$.
3. Si $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ entonces $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$.

Demostración:

1. Supóngase que $\mathbf{X}_n \xrightarrow{a.s.} \mathbf{X}$. Para $\omega \in \Omega$ tal que $\mathbf{X}_n(\omega) \rightarrow \mathbf{X}(\omega)$ y tal que g es continua en $\mathbf{X}(\omega)$, se tiene que $g(\mathbf{X}_n(\omega)) \rightarrow g(\mathbf{X}(\omega))$ conforme $n \rightarrow \infty$. Por suposición, el conjunto de ω tiene probabilidad 1, así $g(\mathbf{X}_n(\omega)) \xrightarrow{a.s.} g(\mathbf{X}(\omega))$
2. Sea $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ y supóngase que $g(\mathbf{X}_n) \not\xrightarrow{P} g(\mathbf{X})$, entonces, para algún $\varepsilon > 0$ y $\Delta > 0$ existe una subsucesión n_k para la cual

$$(*) P(|g(\mathbf{X}_{n_k}) - g(\mathbf{X})| > \varepsilon) > \Delta, \text{ para todo } k = 1, 2, \dots$$

pero $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ implica que $\mathbf{X}_{n_k} \xrightarrow{P} \mathbf{X}$, luego es posible garantizar la existencia de una subsucesión $\{n_{k_j}\}$ de $\{n_k\}$ para la que

$$\mathbf{X}_{n_j} \xrightarrow{a.s.} \mathbf{X}, j \rightarrow \infty$$

y como convergencia casi segura implica convergencia en probabilidad, se tiene $g(\mathbf{X}_{n_{k_j}}) \xrightarrow{P} g(\mathbf{X})$, contradiciendo (*), por lo tanto $g(\mathbf{X}_n) \xrightarrow{P} g(\mathbf{X})$.

3. Supóngase que $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, es posible construir sobre un espacio de probabilidad $(\Omega', \mathcal{A}', P')$ algunos vectores aleatorios $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, y \mathbf{Y} tales que $(Y_1) = (X_1), (Y_2) = (X_2), \dots, (Y) = (X)$ y $\mathbf{Y}_n \rightarrow \mathbf{Y}$ con probabilidad P' de 1. Sea D el conjunto de discontinuidad de la función g , entonces,

$$\begin{aligned} P'(\{\omega' : g \text{ es discontinua en } \mathbf{Y}(\omega')\}) &= P'(\mathbf{Y}^{-1}(D)) \\ &= P'_{\mathbf{Y}}(D) = P_{\mathbf{X}}(D) = P(\mathbf{X}^{-1}(D)) \\ &= 0 \end{aligned}$$

por el inciso 1, $g(\mathbf{Y}_n) \rightarrow g(\mathbf{Y})$ con probabilidad P' de 1, así $g(\mathbf{Y}_n) \xrightarrow{d} g(\mathbf{Y})$, que es lo mismo que $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$.

A.4. Orden de Convergencia

Definición A.5 Sean f y g dos funciones, tales que $f, g : \mathbb{N} \rightarrow \mathbb{R}^+$. Se dice que $f(n) \in O(g(n))$ si existe una constante $c \in \mathbb{R}$ y $n_0 \in \mathbb{N}$, tal que para cada $n \geq n_0$

$$f(n) \leq cg(n)$$

Teorema A.10 Si $f_1(n) \in O(g_1(n))$ y $f_2(n) \in O(g_2(n))$, entonces $f_1(n) + f_2(n) \in O(\max(g_1(n), g_2(n)))$

Demostración: Como $f_1(n) = O(g_1(n))$ y $f_2(n) = O(g_2(n))$, $\exists c_1, c_2 \in \mathbb{R}^+$ y $n_1, n_2 \in \mathbb{R}$ tal que $f_1(n) \leq c_1 g_1(n)$, $\forall n \geq n_1$ y $f_2(n) \leq c_2 g_2(n)$, $\forall n \geq n_2$. Sea $C = \max(c_1, c_2)$ y $N = \max(n_1, n_2)$, entonces $f_1(n) + f_2(n) \leq 2C \max(g_1, g_2)$.

Observación A.2 $f(n) = O(g(n))$ es equivalente a escribir $\frac{f(n)}{g(n)} = O(1)$

Definición A.6 Para una sucesión de variables aleatorias R_n ,

1. $X_n = o_P(R_n)$ significa que $X_n = Y_n R_n$ y $Y_n \xrightarrow{P} 0$.
2. $X_n = O_P(R_n)$ significa que $X_n = Y_n R_n$ y $Y_n O_P(1)$.

Propiedades

- | | |
|---------------------------------|-----------------------------------|
| 1. $o_p(1) + o_p(1) = o_p(1)$, | 4. $(1 + o_p(1))^{-1} = O_p(1)$, |
| 2. $O_p(1) + O_p(1) = O_p(1)$, | 5. $o_p(R_n) = R_n o_p(1)$, |
| 3. $o_p(O_p(1)) = o_p(1)$, | 6. $O_p(R_n) = R_n O_p(1)$. |

A.5. Teoremas importantes

Teorema A.11 (Teorema de Donsker) Si $\varepsilon_1, \varepsilon_2, \dots$ son independientes e idénticamente distribuidos con media cero y varianza σ^2 , y si

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{\sigma}{\sqrt{n}} \varepsilon_{[nt]+1}(\omega),$$

entonces, $X^n \xrightarrow{d} W$

Demostración: Como se ha demostrado la existencia de la medida de Wiener y la correspondiente función aleatoria W , se puede concluir que $(X_s^n, X_t^n - X_s^n) \xrightarrow{d} (W_s, W_t - W_s)$ esto implica que $(X_s^n, X_t^n) \xrightarrow{d} (W_s, W_t)$. Por el Teorema 1, $X^n \xrightarrow{d} W$.

Teorema A.12 Sea $\{\xi_n, n \geq 1\}$ una sucesión de vectores aleatorios de dimensión d y con valores reales tal que la sucesión es débil en sentido estacionario, es centrada en esperanza con momento $(2 + \delta)$ para $0 < \delta \leq 1$, uniformemente acotada por 1. Supóngase que $\{\xi_n, n \geq 1\}$ satisface la condición

$$|P(AB) - P(A)P(B)| \leq \rho(n) \downarrow 0, \quad \forall n, k \geq 1$$

y todo $A \in \mathcal{M}_1^k$ y $B \in \mathcal{M}_{k+1}^\infty$, con \mathcal{M}_a^b la σ -álgebra generada por los vectores $\xi_a, \xi_{a+1}, \dots, \xi_b$ y $\rho(n)$ una sucesión no creciente con $\rho(n) \ll n^{-(1+\epsilon)(1+2/\delta)}$, escribiendo $\xi_n = (\xi_{n1}, \dots, \xi_{nd})$, entonces las dos series en

$$\gamma_{ij} = E(\xi_{1i}\xi_{1j}) + \sum_{k \geq 2} E(\xi_{1i}\xi_{kj}) + \sum_{k \geq 2} E(\xi_{ki}\xi_{1j}),$$

converge absolutamente.

Escribiendo la matriz (γ_{ij}) , $1 \leq i, j \leq d$ por Γ , entonces se puede redefinir la sucesión $\{\xi_n, n \geq 1\}$ sobre un nuevo espacio de probabilidad junto con un movimiento Browniano $X(t)$ con matriz de covarianza Γ , tal que

$$\sum_{n \leq t} \xi_n - X(t) \ll t^{1/2-\lambda} \text{ a.s.},$$

para algún $\lambda > 0$ dependiendo sólo de ϵ, δ y d

Demostración: La demostración de este teorema se puede consultar en [61].

Proposición A.1 Sea $\mathbf{Z}, t \geq 1$, una sucesión de vectores aleatorios de dimensión d , independientes e idénticamente distribuidos con $E(\mathbf{Z}) = 0$ y $E(\mathbf{Z}\mathbf{Z}') = I$. Si todas las componentes de \mathbf{Z}_1 tienen r -ésimo momento finito para $r > 2$, entonces

$$P \left[\alpha(\ln n) \max_{1 \leq k \leq n} \frac{\|\mathbf{S}_k\|}{k^{1/2}} - \beta_d(\ln n) \leq x \right] \rightarrow \exp(e^{-x}),$$

conforme $n \rightarrow \infty$, donde $\mathbf{S}_k = \sum_{j=1}^k \mathbf{Z}_j$, $\alpha(x) = (2 \ln x)^{1/2}$, $\beta_d(x) = 2 \ln x + (d/2) \ln \ln x - \ln \Gamma(d/2)$, $\Gamma(\cdot)$ es la función Γ .

Demostración: Ver demostración en [49].

Corolario A.2 Bajo las suposiciones de la Proposición A.1,

$$P \left[\left(\max_{1 \leq k \leq n} \frac{\|\mathbf{S}_k\|^2}{k} - B_d(\ln n) \right) / A_d(\ln n) \leq x \right] \rightarrow \exp(-e^{x/2}),$$

donde $A_d(x) = \frac{B_d(x)}{\alpha^2(x)}$ y $B_d(x) = \left(\frac{\beta_d(x)}{\alpha(x)} \right)^2$.

Demostración: Ver demostración en [49].

Teorema A.13 (Ley del Logaritmo Iterado) Si $\{X_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza, entonces la sucesión de sumas parciales $S_n = \sum_{i=1}^n X_i$, satisfice

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1 \text{ a.s.}$$

Demostración: Ver demostración en [91].

A.6. Puente Browniano

Definición A.7 *El movimiento Browniano estándar condicionado en $W(1) = 0$ es llamado **Puente Browniano Estándar**, $B(t)$.*

Considerando dos variables aleatorias independientes W_t y $W_1 - W_t$, su distribución de densidad conjunta para $W_t = x$ y $W_1 - W_t = y$ es

$$\frac{1}{\sqrt{2\pi t}} \frac{1}{\sqrt{2\pi(1-t)}} e^{-\frac{x^2}{2t}} e^{-\frac{y^2}{2(1-t)}}$$

Considerando el teorema de la transformación, la densidad conjunta de (W_1, W_t) es

$$f_{W_1, W_t}(x, y) = \frac{1}{\sqrt{2\pi t(1-t)}} e^{-\left(\frac{1}{1-t}\right)\left(\frac{x^2}{2t} - xy + \frac{y^2}{2}\right)}$$

Luego la función de densidad condicional para W_t dato $W_1 = 0$, es

$$f_{B_t}(x, t) = f_{W_t|W_1=0}(x, t) = \frac{1}{\sqrt{2\pi t(1-t)}} e^{-\frac{x^2}{2t(1-t)}}$$

A.7. Matriz inversa generalizada

Definición A.8 *Sea A una matriz de tamaño $m \times n$. Si M es una matriz de tamaño $n \times m$ tal que:*

1. AM es una matriz simétrica.
2. MA es una matriz simétrica.
3. $AMA = A$.
4. $MAM = M$.

entonces se dice que M es una inversa generalizada (pseudoinversa) de A o simplemente que M es una g -inversa de A .

Observación A.3

1. Si A es invertible, entonces la matriz A^{-1} es una g -inversa de A .
2. Si $A = \mathbf{0}_{m \times n}$ entonces la matriz $M = \mathbf{0}_{n \times m}$ es una g -inversa de A .

Teorema A.14 (Existencia de la matriz inversa generalizada) *Toda matriz A de tamaño $m \times n$ tiene una inversa generalizada.*

Demostración:

Si A es una matriz de ceros, por la observación A.3 inciso 2, la demostración es trivial.

Suponiendo que $A \neq \mathbf{0}$ con rango $r > 0$, existen matrices de tamaño B $m \times r$ y C de tamaño $r \times n$ tal que $A = BC$. Como B y C tienen rango r , las matrices $B'B$ y CC' son invertibles. Finalmente, al considerar la matriz

$$M = C'(CC')^{-1}(B'B)^{-1}B' = B(B'B)^{-1}B'$$

Al verificar las condiciones de la definición de la matriz inverza generalizada, se tiene

$$MA = C'(CC')^{-1}(B'B)^{-1}B'BC = C'(CC')^{-1}C,$$

$$AM = BCC'(CC')^{-1}(B'B)^{-1}B',$$

además $AMA = B(BB')^{-1}B'BC = BC = A$ y

$$\begin{aligned} MAM &= C'(CC')^{-1}CC'(CC')^{-1}(B'B)^{-1}B' \\ &= C'(CC')^{-1}(B'B)^{-1}B' = M \end{aligned}$$

por lo tanto, M definida anteriormente es una g -inversa de A .

Teorema A.15 (Unicidad de la matriz inverza generalizada) *Toda matriz A tiene una única g -inversa.*

Demostración: Supóngase que M_1 y M_2 son dos g -inversas de una matriz A . Utilizando la definición de g -inversa de una matriz se obtiene:

$$\begin{aligned} AM_2 &= (AM_1A)M_2 = (AM_1)(AM_2) = (AM_1)'(AM_2)' \\ &= ((AM_2)(AM_1))' = ((AM_2A)M_1)' = (AM_1)' = AM_1. \end{aligned}$$

De aquí, $AM_2 = AM_1$. De forma análoga se obtiene que $M_2A = M_1A$, por lo tanto

$$\begin{aligned} M_1 &= M_1AM_1 = (M_1A)M_1 = (M_2A)M_1 = M_2(AM_1) \\ &= M_2(AM_2) = M_2AM_2 = M_2. \end{aligned}$$

APÉNDICE B

Programas Desarrollados en *R* para la aplicación de las metodologías

B.1. Programa elaborado en *R* para la aplicación de la metodología de Davis *et al.* [17]

```
# Metodología Davis 1995.
#Cada columna de la matriz vetorescore,
# representa el vector score para k = No. de columna de la matriz.
# Se debe introducir el vector de parámetros en el vector xi
# bajo el siguiente orden:
#media, varianza, estimación de los parámetros del proceso AR(p).
#Cargar las observaciones como un archivo con extensión .csv
serie=read.table("C://Users//Silvia//Documents
//DOCTORADO//APLICACION//serie_estacionaria.csv",sep=",",header=TRUE);
datos=serie[,1];
library(matlib)
#####
media=mean(datos)
dimensionxi=length(xiestimado)
p=3;
n=length(datos)

# Cálculo de la matriz MN
renglones=n-p;
matrizmn=matrix(data=0,nrow=renglones, ncol=p+1);
for(j in 1:renglones){
```

```

    if(j==1){
    for(i in 1:(p+1)){
      if(i==1){matrizmn[j,i]=1}
      else
        if(i!=1){
          pos=p+2-i;
          matrizmn[j,i]=datos[pos]
        }
      i=i+1;
    }
  }
  if(j>1){
    for(i in 1:(p+1)){
      if(i==1){matrizmn[j,i]=1}
      else
        if(i!=1){
          pos=p+j-i+1;
          matrizmn[j,i]=datos[pos]
        }
      i=i+1;
    }
  }

  j=j+1;
}
# Cálculo de la matriz Mk
tope=n-p-1
matrizdeltak=matrix(data=0, nrow=1, ncol=n-k-1);
for(k in 1:tope){
  renglones=k;
  matrizmk=matrix(data=0,nrow=renglones, ncol=p+1);
  for(j in 1:renglones){
    if(j==1){
      for(i in 1:(p+1)){
        if(i==1){matrizmk[j,i]=1}
        else
          if(i!=1){
            pos=p+2-i;
            matrizmk[j,i]=datos[pos]
          }
        i=i+1;
      }
    }
  }
}

```

```

    if(j>1){
      for(i in 1:(p+1)){
        if(i==1){matrizmk[j,i]=1}
        else
          if(i!=1){
            pos=p+j-i+1;
            matrizmk[j,i]=datos[pos]
          }
        i=i+1;
      }
    }

    j=j+1;
  }

# Cálculo de la matriz MKtilde
ren=n-k-p
matrizmktilde=matrix(data=0,nrow=ren, ncol=p+1);
for(j in 1:ren){
  if(j==1){
    for(i in 1:(p+1)){
      if(i==1){matrizmktilde[j,i]=1}
      else
        if(i!=1){
          pos=p+k+j-i+1;
          matrizmktilde[j,i]=datos[pos]
        }
      i=i+1;
    }
  }
  if(j>1){
    for(i in 1:(p+1)){
      if(i==1){matrizmktilde[j,i]=1}
      else
        if(i!=1){
          pos=p+k+j-i+1;
          matrizmktilde[j,i]=datos[pos]
        }
      i=i+1;
    }
  }
  j=j+1;
}

```

```

vectorxn=matrix(data=0, nrow=n-p,ncol=1)
vectorxk=matrix(data=0, nrow=k,ncol=1)
vectorxktilde=matrix(data=0, nrow=n-k-p,ncol=1);
for(z in 1:n-p){
vectorxn[z,1]=datos[z+p];
  z=z+1;
}
for(z in 1:k){
  vectorxk[z,1]=datos[p+z];
  z=z+1;
}

contador=1;
contador2=k+p+1;
for(z in contador2:n){
  vectorxktilde[contador,1]=datos[z];
  z=z+1;
  contador=contador+1;
}

producto1=t(matrizmk)%*%matrizmk;
producto2=t(matrizmktilde)%*%matrizmktilde;
if(det(producto1)<=0.05){
  mkinversa=Ginv(t(matrizmk)%*%matrizmk);
}
if(det(producto1)>0.05){
  mkinversa=solve(t(matrizmk)%*%matrizmk);
}

if(det(producto2)<=0.05){
  mktildeinversa=Ginv(t(matrizmktilde)%*%matrizmktilde);
}

if(det(producto2)>0.05){
  mktildeinversa=solve(t(matrizmktilde)%*%matrizmktilde);
}

q1=t(vectorxn)%*%vectorxn-t(vectorxn)%*%matrizmn)%*%
solve(t(matrizmn)%*%matrizmn)%*%t(matrizmn)%*%vectorxn;
q2=t(vectorxk)%*%vectorxk-t(vectorxk)%*%matrizmk)%*%mkinversa)%*%
t(matrizmk)%*%vectorxk;
q3=t(vectorxktilde)%*%vectorxktilde-t(vectorxktilde)%*%matrizmktilde)%*%
mktildeinversa)%*%t(matrizmktilde)%*%vectorxktilde;

```

```

matrizdeltak[k]=q1-q2-q3;
k=k+1;
}
maximodeltak=max(matrizdeltak)
for(j in 1:length(matrizdeltak)){
  if(matrizdeltak[j]==maximodeltak){
    posicion=j;
  }
  j=j+1;
}
x=seq(from=(p+1), to=(n-1), by=1);
y=matrizdeltak;
plot(x,y, type="l", xlab="k", ylab="f(k)")

print(maximodeltak)
print(posicion)

# Cálculo de los cuantiles de la Distribución Gumbel
d=p+1;
bnd=(2*log(log(n))+(d/2*log(log(log(n))))-log(gamma(d/2)))^2/(2*log(log(n))));
an=sqrt(bnd/(2*log(log(n))));
print(bnd)
print(an)

probabilidades=matrix(data=0,nrow=4000, ncol=2)
x=4
for(j in 1:4000){
  probabilidades[j,1]=x*an+bnd;
  probabilidades[j,2]=exp(-2*exp(-x/2));
  x=x+0.01;
  j=j+1;
}
print(probabilidades)

```

B.2. Programa elaborado en *R* para la aplicación de la metodología de Gombay E [37]

```

# Metodología Gombay 2008.
#Cada columna de la matriz vetorescore, representa el vector score
para k = No. de columna de la matriz..

```

```

# Se debe introducir el vector de parámetros en el vector xi
# bajo el siguiente orden:
#media, varianza, estimación de los parámetros del proceso AR(p)

xiestimado=c(0.0013360338573,3.563,-0.4872,-0.2981,-0.1995);

#Cargar las observaciones como un archivo con extensión .csv
#serie=read.csv("C://Users//SAMSUNG//Documents//DoCTORADO//APLICACION//
temperatura_maxima_estandarizada
_por_numero_de_semana_al_anio.csv",header=TRUE, sep=",")
serie=read.table("C://Users//Silvia//Documents//
DOCTORADO//APLICACION//serie_estacionaria.csv",sep=",",header=TRUE);
datos=serie[,1];
#####
media=mean(datos)
dimensionxi=length(xiestimado);
p=length(xiestimado)-2;
n=length(datos);

#Se genera la matriz cuyas columnas son elementos del vector Score
vectorscore=matrix(data=0,nrow=length(xiestimado), ncol=n);

for(w in 1:n){
  s=1;
  suma1=0;
  suma6=0;
  for(i in 1:p){
    suma1=suma1+xiestimado[i+2]
    i=i+1;
  }
  factor1=(1-suma1)/xiestimado[2];
  suma2=0;
  suma4=0;
  for(j in 1:w){
    suma3=0;
    suma5=0;
    for(c in 1:p){
      if((j-c) <=0){y=0;}
      else
      if((j-c)>0){y=datos[j-c];}
      suma3=suma3+xiestimado[c+2]*(y-xiestimado[1]);
      c=c+1;
    }
  }
}

```

```

        suma2=suma2+(datos[j]-xiestimado[1])-suma3;
        suma5=suma5+(datos[j]-xiestimado[1])-suma3;
        suma4=suma4+(suma5*suma5);

        j=j+1;

    }
vectorscore[1,w]=factor1*suma2;
vectorscore[2,w]=(-w/(2*xiestimado[2]))+((1/(2*xiestimado[2]*xiestimado[2]))*suma4)
    w=w+1;
}
for(s in 1:p){
    for(w in 1:n){
        suma7=0;
        for(j in 1:w){
            suma6=0;

            for(c in 1:p){
                if((j-c) <=0){y=0;}
                else
                    if((j-c)>0){y=datos[j-c];}
                suma6=suma6+xiestimado[c+2]*(y-xiestimado[1]);
                c=c+1;
            }
            if((j-s)<=0){h=0;}
            else
                if ((j-s)>0){h=datos[j-s];}
            suma7=suma7+((datos[j]-xiestimado[1])-suma6)*(h-xiestimado[1]);
            j=j+1;
        }
        vectorscore[s+2,w]=suma7/xiestimado[2];
        w=w+1;
    }
    s=s+1;
}
# Matriz de varianzas y covarianzas.
#Extracción de las primeras p observaciones:

pob=matrix(data=0,nrow=p, ncol=1);
for(i in 1:p){
    pob[i]=datos[i];
}
media1=mean(pob);

```

```

veccov=matrix(data=0,nrow=p, ncol=1);
for(k in 0:p-1){
  suma=0;
  for(j in 1:(p-k)){
    suma=suma+(pob[j]-media1)*(pob[j+k]-media1)
    j=j+1;
  }
  suma=suma/p;
  veccov[k+1]=suma
  k=k+1;
}
mcov=matrix(data=0,nrow=p+2, ncol=p+2);
r=1;
for(j in 1:p){
  for(f in 1:p){
    if(j==f){mcov[j+2,f+2]=veccov[1];}
    if(j!=f){
      maximo=max(j,f);
      mcov[j+2,f+2]=veccov[maximo]
    }
  }
  j=j+1;
}

mcov[1,1]=(1-suma1)^2/xiestimado[2]
mcov[2,2]=1/(2*xiestimado[2]^2);
valpropio=eigen(mcov);
matrizdelta=matrix(data=0,nrow=length(xiestimado), ncol=length(xiestimado));
matrizgama=valpropio$vectors;
for(contador in 1:(p+2)){
  for(c in 1:(p+2)){
    if(contador==c){
      matrizdelta[c,c]=1/(sqrt(valpropio$values[c]));
    }
    c=c+1;
  }
  contador=contador+1;}
matrizIsqrt=matrizgama%%matrizdelta%% t(matrizgama);
vectorB=matrix(data=0,nrow=length(xiestimado),ncol=n);
for(j in 1:n){
  B=(1/sqrt(n))*matrizIsqrt%%vectorscore[,j];
  vectorB[,j]=B;
  j=j+1;
}

```

```

}
print(vectorB)

#Detección de máximos y su ubicación
##El primer renglón indica el valor máximo.
#El segundo renglón indica la posición del máximo en el vector B.

maximos=matrix(data=0,nrow=2,ncol=length(xiestimado))
par(mfrow=c(3,2))
for(j in 1:length(xiestimado)){
matrizauxiliar=vectorB[j,];

x=seq(1, length(datos), by=1);
plot(x,matrizauxiliar, type="l", ylab="valor ", xlab="k" )

maximo=max(matrizauxiliar);
for(z in 1:length(datos)){
if(maximo==matrizauxiliar[z]){
posicion=z;
}
z=z+1;
}
maximos[1,j]=maximo;
maximos[2,j]=posicion;
j=j+1;
}
print(matrizIsqrt)
print(vectorscore)
print(vectorB)
print(maximos)

#Para la prueba de un solo lado
#Cuantiles C1
cuantil1=0;
alpha0=0.05
alpha=1-((1-alpha0)^(1/(p+2)));## Este valor se puede modificar
probabilidad=exp(-2*(cuantil1*cuantil1));
tolerancia=0.00001 # Este valor se puede modificar
while(probabilidad-alpha>tolerancia){
  cuantil1=cuantil1+0.00001;
  probabilidad=exp(-2*(cuantil1*cuantil1));
}

```

```
print(probabilidad);
print(cuantil1)

#p-valores
pvalores=matrix(data=0,nrow=length(xiestimado), ncol=1)
for(i in 1:length(xiestimado)){
  pvalores[i,1]=exp(-2*(maximos[1,i])^2)
}
print(pvalores)

#Prueba de dos colas
#Cuantiles C2
suma=0;
alpha0=0.005
alpha=1-(((1-alpha0)^(1/(p+2))));# Este valor se puede modificar

toerancia=0.001; # Este valor se puede modificar
cuantil2=0;
print(alpha)

for(k in 1:50001){
  suma=suma+((-1)^(k+1)*exp(-2*k*k*cuantil2*cuantil2));
  k=k+1;
}
probabilidad=suma;

while(probabilidad-alpha>tolerancia){
  cuantil2=cuantil2+0.0001;
  suma=0;
  for(k in 1:50001){
    suma=suma+((-1)^(k+1)*exp(-2*k*k*cuantil2*cuantil2));
    k=k+1;
  }
  probabilidad=suma;
}
print(probabilidad);
print(cuantil2)
```

Bibliografía

- [1] Aminikhanghahi, S., and Cook, J. D. (2017). *A survey of methods for time series change point detection*. Knowl. Inf. Syst. 51, 339–367. doi: 10.1007/s10115-016-0987-z.
- [2] Atherton, J., Charbonneau, B., Wolfson, D. B., Joseph, L. , Zhou, X., and Vandal, A. C. (2009). *Bayesian optimal design for change point problems*. Can. J. Statistics, 37, 495-513. doi:10.1002/cjs.10037.
- [3] Badagián, A., Kaiser, R., and Peña, D. (2015). *Time series segmentation procedures to detect, locate and estimate change-points*. Empirical Economic and Financial Research. Advanced Studies in Theoretical and Applied Econometrics, 48, 45-59. doi:10.1007/978-3-319-03122-4_3.
- [4] Bhattacharya, P. K. (1987). *Maximum likelihood estimation of a change point in the distribution of independent random variables: general multi-parameter case*. Journal of Multivariate Analysis, 23(2), 183-208.
- [5] Bhattacharya, P. K. (1994). *Some aspects of change point analysis*. Lecture Notes-Monograph Series, 23, 28-56.
- [6] Brodsky, B. S., and Darkhovsky, B. E.(1993). *Non-parametric methods in change point problems*. Kluwer Academic Publishers.
- [7] Brodsky, B. S., and Darkhovsky, B. E. (1993). *Asymptotically optimal sequential change point detection under composite hypotheses*. 4th IEE Conference on Decision and Control and the European Control Conference Seville Spain, 7346-7351.
- [8] Brockwell, J. P., and Davis, A. R. (2016). *Introduction to time series and forecasting*. Springer. Second edition.

- [9] Brockwell, J. P., and Davis, A. R. (1987) *Time series: Theory and Methods*. Springer- Verlag. ISBN: 0-387-96406-1.
- [10] Cahill, N., Stefan, R., and Parnell, C. A. (2015). *Change points of global temperature*. Environmental Research Letters, 10(8). doi:10.1088/1748-9326/10/8/084002.
- [11] Chow, G. C. (1960). *Tests of equality between sets of coeficientes in two linear regressions*, 28(3), 1-15.
- [12] Chernoff, H., and Zacks, S. (1964). *Estimating the current mean of a normal distribution which is subjected to changes in time*. Ann. Math. Statist., 35, 999-1018.
- [13] Csörgó, M., and Horváth, L. (1988). *Invariance principles for change point problems*. Journal of Multivariate Analysis, 27, 157-168.
- [14] Csörgó, M., and Horváth, L. (1997). *Limit theorems in change point analysis*. John Wiley.
- [15] Csörgó, M., and Révész, P. (1981). *Strong approximations in probability and statistics*, Academic Press, NewYork.
- [16] Csörgó, M., and SzyszKowicz, B. (1994). *Weighted multivariate empirical processes and contiguous change point analysis*. IMS Lecture Notes, 23, 1-6.
- [17] Davis, R. A, Dawei, H., and Yao, Y. (1995). *Testing for a change in the parameter values and order of an autoregressive model*. The Annals of Statistics, 23(1), 282-304.
- [18] Davis, R. A., Lee, C. M. T, and Rodríguez-Yam, A. G. (2008). *Break detection for a class of non-linear time series models*. Journal of Time Series Analysis, 29, 834-867. doi:10.1111/j.1467-9892.2008.00585.x.
- [19] Davis, R. A., Lee, C. M. T., and Rodríguez-Yam, A. G. (2006). *Structural break estimation for non-stationary time series models*. Journal of the American Statistical Association, 101(473), 223-239, doi: 10.1198/016214505000000745.
- [20] Davis, W. W. (1979). *Robust methods for detection of shifts of the innovation variance of a time series*, 21(3), 313-320. doi: 10.1080/00401706.1979.10489775.
- [21] Darkhovski, S. B. (1994). *Nonparametric methods in change point problems: A general approach and some concrete algorithms*. Lecture Notes-Monograph Series, 23, 99-107.

- [22] Dehling, H., Fried R., and Wendler, M. (2015). *A robust method for shift detection in time series*. doi:10.17877/DE290R-7443.
- [23] Dette, H., Wu, W., and Zhou, Z. (2015). *Change point analysis of second order characteristics in non stationary time series*. doi:10.17877/DE290R-2361.
- [24] Dibal, N. P., Mustapha, M., Adegoke, T. M., and Yahaya, A. M. (2017). *Statistical change point analysis in air temperature and rainfall time series for Cocoa Research Institute of Nigeria, Ibadan, Oyo State, Nigeria*. International Journal of Applied Mathematics and Theoretical Physics. 3(4), 92-96. doi: 10.11648/j.ijamtp.20170304.13.
- [25] Dorcas, W. G., Anthony, W. G., and Wanjoya, A. (2015). *The power of the Pruned Exact Linear Time (PELT) Test in multiple change point Detection*. American Journal of Theoretical and Applied Statistics, 4(6), 581-586. doi: 10.11648/j.ajtas.20150406.30.
- [26] Gallagher C., Lund, R., and Robbins, M. (2013). *Change point detection in climate time series with long-term trends*. J. Climate, 26, 4994–5006. doi:10.1175/JCLI-D-12-00704.1.
- [27] Garreau, D. (2017). *Change point detection and kernel methods*. Statistics [math.ST]. PSL Research University, English. ffnnt : 2017PSLEE061ff. fftel-01693360v2f. PhD thesis.
- [28] Guiaro, C. E. R., Gama, C. M. L., and Rangel, R. J. L. (2016). *Efecto del cambio climático sobre aves migratorias*. Kuxulkab. XXII(44), ISSN 2448-508X, 5-10.
- [29] Eichinger, B., and Kirch, C. (2018). *MOSUM procedure for the estimation of multiple random change points*. Bernoulli, 24(1), 526-564. doi: 10.3150/16-BEJ887.
- [30] Abd-Elnaser, S., Abd-Rabou, and Gad, M. A. (2007). *A new change point rank tests*. Journal of Data Science, 5, 379-392.
- [31] Fearnhead, P. and Liu, Z. (2007). *On-line inference for multiple change point problems*. Journal of the Royal Statistical Society. Series B. Statistical Methodology, 69(4), 589–605.
- [32] Fryzlewicz, P. (2014). *Wild binary segmentation for multiple change-point detection*. Ann. Statist. 42(6), 2243–2281. doi: 10.1214/14-AOS1245.
- [33] Georgescu, Vasile. (2012). *On-line change point detection in financial time series: challenges and experimental evidence with frequentist and Bayesian setups*. 131-145. doi:10.1142/9789814415774_0009.

- [34] Giordani, P., and Kohn, R. (2008). *Efficient Bayesian inference for multiple change point and mixture innovation models*. Journal of Business and Economic Statistics, 26(1), 66-77. doi: 10.1198/073500107000000241.
- [35] Giriatis, L., Leipus, R., and Surgailis, D., (1996). *The change point problem for dependent observations*. Journal of Statistical Planning and Inference, 53(3), 297-310. doi:10.1016/0378-3758(95)00148-4.
- [36] Girshick, M. A., and Rubin, H. (1952). *A Bayes approach to a quality control model*. Ann. Math. Statist., 23(1), 114-125.
- [37] Gombay, E. (2008). *Change detection in autoregressive time series*. Journal of Multivariate Analysis, 99, 451-464.
- [38] Gombay, E., and Horváth, L., (1994). *An applications of the maximum likelihood test on the change point problem*. Stochastic Processes and their Applications, 50(1), 161-171. doi:10.1016/0304-4149(94)90154-6.
- [39] Gombay, E., and Serban, D. (2005). *Monitoring parameter change in AR(p) time series models*. Statistics Centre Technical Reports 05.04, The University of Alberta, Edmonton, Canada.
- [40] Gozdek, O. A., Sarnowski, W., and Szajowski, K. J. (2014). *A precision of the sequential change point detection*. Applicationes Mathematicae 11(4), 1-8. 10.4064/am2278-5-2017.
- [41] Guerrero, V. (1987). *Modelos estadísticos para series de tiempo*. V Coloquio del Departamento de Matemáticas, Centro de Investigación y Estudios Avanzados del IPN.
- [42] Hušková, M., and Kirch, C. (2008), *Bootstrapping confidence intervals for the change point of time series*. Journal of Time Series Analysis, 29, 947-972. doi:10.1111/j.1467-9892.2008.00589.x.
- [43] Hawkins, M. D. (1977). *Testing a sequence of observations for a shift in location*. Journal of the American Statistical Association, 72(357), 180-186, doi: 10.1080/01621459.1977.10479935.
- [44] Herrera, C. S., Juárez, H. B., Vázquez, G. V. V and Cruz, S. H. A. *Parametric methodologies for detecting changes in maximum temperature of Tlaxco-Tlaxcala, México*. Journal of Probability and Statistics. (artículo aceptado el 5 de octubre de 2019).
- [45] Herrera, C. S., Juárez, H. B., Vázquez, G. V. V and Cruz, S. H. A. (2019). *Detección de cambios en la temperatura máxima de Tlaxco, Tlaxcala utilizando metodologías de puntos de cambio*. Desarrollo científico en México. Editorial CIO. ISBN:978-607-95228-9-6.

- [46] Herrera, C. S. (2015). *Análisis de variables climatológicas del estado de Tlaxcala utilizando series de tiempo y redes neuronales*. FCFM-BUAP. Tesis de maestría.
- [47] Hinkley, V. D. (1970). *Inference about the change-point in a sequence of random variables*. *Biometrika*, 57(1), 1-17. doi: 10.1093/biomet/57.1.1.
- [48] Hipel, K. W., and McLeod, A. I. (1994) *Time Series Modelling of Water Resources and Environmental Systems*. Amsterdam, Elsevier, 1 St. Edition, vol. 45. ISBN:9780080870366.
- [49] Horváth, L. (1993). *The maximum likelihood method for testing changes in the parameters of normal observations*. *Ann. Statist.* 21,671-680.
- [50] Horváth, L., Kokoska, P., and Steinebach, J. (1999). *Testing for changes in multivariate dependent observations with an application to temperature changes*. *Journal of Multivariate Analysis*, 68, 96-119.
- [51] Inclán, C., and Tiao, C. G. (1994). *Use of cumulative sums of squares of retrospective detection of changes of variance*. *Journal of the American Statistical Association*, 89(427), 913-923, doi: 10.1080/01621459.1994.10476824.
- [52] IPCC (Intergovernmental Panel on climate Change).(2013). *Summary for Policy makers Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. United Kingdom and USA.
- [53] IPCC (Intergovernmental Panel on climate Change). (2014). *Climate change 2014: Impacts, adaptation and vulnerability: Summary for policy makers*. pp. 32.
- [54] Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumoussis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005). *An algorithm for optimal partitioning of data on an interval*. *IEEE, Signal Processing Letters*, 12(2), 105-108. doi: 10.1109/LSP.2001.838216.
- [55] Jarušková, D.(1996). *Change point detection in meteorological measurement*. *American Meteorological Society*, 124, 1535-1543.
- [56] Jarušková, D. (1997). *Some problems with application of change point detection methods to environmental data*. *Environmetrics*, 8, 469-483.
- [57] Joo-Oh, K., and Han, I. (2002). *An intelligent clustering forecasting system based on change-point detection and artificial neural networks: application to financial economics*. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. doi: 10.1109/HICSS.2001.926304.

- [58] Kass-Hout T. A., Xu, Z., McMurray P., Park S., Buckeridge D. L., Brownstein J. S., and Groseclose S. L. (2012). *Application of change point analysis to daily influenza-like illness emergency department visits*. Journal of the American Medical Informatics Association: JAMIA, 19(6), 1075-1081. doi: 10.1136/amiajnl-2011-000793.
- [59] Kezim, B., and Abdelli, Z. (2004). *A Bayesian analysis of a structural change in the parameters of a time series*. Communications in Statistics - Theory and Methods, 33(8), 1863-1876. doi: 10.1081/STA-120037446.
- [60] Killick, R., Fearnhead, P. and Eckley, I., A. (2012). *Optimal detection of changepoints with a linear computational cost*. Journal of the American Statistical Association, 107(500), 1590-1598. doi: 10.1080/01621459.2012.737745.
- [61] Kuelbs, J., and Philipp W. (1980). *Almost sure invariance principles for partitioned sums of mixing B-valued random variables*. The Annals of probability, 8(6), 1003-1036.
- [62] Kumar, K., and Wu, B. (2001). *Detection of change points in time series analysis with fuzzy statistics*. International Journal of Systems Science, 32(9), 1185-1192. doi: 10.1080/00207720110034698
- [63] Lai, L. T., and Xing, H. (2011). *A simple Bayesian approach to multiple change-points*. Statistica Sinica, 21(2), 539-569. Recuperado de: <https://www.jstor.org/stable/24309531>.
- [64] Lai, L. T., and Xing, H.(2010). *Sequential change point detection when the pre-and post-change parameters are unknown*. Sequential Analysis, 29(2), 162-175. doi: 10.1080/07474941003741078.
- [65] Lavielle, M., Lebarbier, E.(2001). *An application of MCMC methods for the multiple change points problem*. Elsevier, 81(1), 39-53. [https://doi.org/10.1016/S0165-1684\(00\)00189-4](https://doi.org/10.1016/S0165-1684(00)00189-4).
- [66] Lee, S., Ha, J., Na, O., and Na, S.(2003). *The Cusum test for parameter change in time series models*. Scandinavian Journal of Statistics, 30, 781-796. doi:10.1111/1467-9469.00364.
- [67] Lee, S., and Na, O. (2005). *Test for parameter change in stochastic processes based on conditional least-squares estimator*. Journal of Multivariate Analysis, 93, 375-393.
- [68] Ling, S. (2007). *Testing for change points in time series models and limit theorems for NED sequences*. The Annals of Statistics, 35(3), 1213-1237.

- [69] Ling, S. (2014). *Estimation of change points in linear and non-linear time series models*. *Econometric Theory*, 32(2), 402-430. doi:10.1017/S0266466614000863.
- [70] Lombard, F. (1987). *Rank test for change point problems*. *Biometrika*, 74(3), 615-624. doi: 10.1093/biomet/74.3.615.
- [71] Lombardo, R., and Flaherty, J. (2000). *Modelling private new housing starts in Australia*. Pacific-Rim Real Estate Society Conference, University of Technology Sydney (UTS), January. 24-27.
- [72] Loschi, H.R., Moura, R. C., and Iglesias, L. P. (2005). *Bayesian analysis for Change Points in the volatility of Latin American Emerging Markets*. *Journal of Data Science*, 3(5), 101-122.
- [73] Lorden, G. (1971). *Procedures for reacting to a change in distribution*. *Ann. Math. Statist.* 42(6), 1897-1908.
- [74] Lund, R., Wang, X. L., Lu, Q. Q., Reeves, J., Gallagher, C., and Feng, Y. (2007) *Change point detection in periodic and autocorrelated time series*. *J. Climate*, 20, 5178–5190. doi: 10.1175/JCLI4291.1.
- [75] Maboudou-Tchao, M. E., and Hawkins, M. D. (2013). *Detection of multiple change points in multivariate data*. *Journal of Applied Statistics*, 40(9), 1979-1995. doi: 10.1080/02664763.2013.800471.
- [76] Mohammad, A., and Féron, O. (2006). *A Bayesian approach to change points detection in time series*. *Int. J. Imaging Syst. Technol.*, 16, 215-221. doi:10.1002/ima.20080.
- [77] Mostafa, A. A., and Ghorbal, B. (2011). *Bayesian and non-Bayesian analysis for random change point problem using standard computer packages*. *International Journal of mathematical Archive*, 2(10), 1963-1979. Available at SSRN: <https://ssrn.com/abstract=2554020>.
- [78] Müller, H. G. (1992). *Change points in nonparametric regression analysis*. *The Annals of Statistics*, 20(2), 737-761.
- [79] National Academy of Sciences. (2012). *Climate Change. Evidence, Impacts, and Choices. Answers to common questions about the science of climate change*. 1-40.
- [80] National Weather Service. *Cold and Warm Episodes by Season*. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

- [81] Nikiforov, V. I. (2000). *A simple recursive algorithm for diagnosis of abrupt changes in random signals*. IEEE Transactions on Information Theory, 46(7), 2740-2746. doi: 10.1109/18.887891
- [82] Page, E.S. (1954). *Continuous inspection schemes*. 41 (1/2), 100-115. doi: 10.2307/2333009.
- [83] Page, E. S. (1957). *On problems in which a change in a parameter occurs at an unknown point*. Biometrika, 44(1-2), 248-252.
- [84] Pepelyshev, A., and Polunchenko, S. A. (2015). *Real-time financial surveillance via quickest change-point detection methods*. Recuperado de <https://arxiv.org/pdf/1509.01570.pdf>.
- [85] Poluchenko, S. A., Sokolov, G., and Du, W. (2013), *Quickest change-point Detection: A Bird's Eye View*. Proceedings of the 2013 Joint Statistical Meetings (JSM-2013), Montreal, Quebec, Canada, 1-16. Recuperado de: <https://arxiv.org/pdf/1310.3285.pdf>.
- [86] Radke, R. J., Andra, S., Al-Kaofahi, O., and Roysam, B.(2005). *Image change detection algorithms: a systematic survey*. IEEE Trans. Image Process. 14(3), 294-307. doi:10.1109/TIP.2004.838698.
- [87] Killick, R., Fearnhead, P., and Eckley, I. A., (2012). *Optimal detection of change points with a linear computational cost*. Journal of the American Statistical Association, 107:500, 1590-1598. doi: 10.1080/01621459.2012.737745.
- [88] Reves, J., Chen, J.(2007). *A Review and Comparison of changepoint detection techniques for climate data*. J. Appl. Meteor. Climatol., 46, 900-915. doi: 10.1175/JAM2493.1.
- [89] Seigmund, D. (1988). *Confidence set in change-point problems*. International Statistical Review, 56(1), 31-48. doi: 10.2307/1403360.
- [90] Sen, A. and Srivastava, M. S. (1975a). *On tests for detecting change in mean*. The Annals of Statistics, 3(1), 98-108.
- [91] Serfling, J. R. (2002). *Approximation theorems of mathematical statistics*. John Wiley. ISBN: 0-471-21927-4.
- [92] Shao, X., and Zhang, X. (2010). *Testing for change points in time series*. Journal of the American Statistical Association, 105(491), 1228-1240. doi: 10.1198/jasa.2010.tm10103.
- [93] Sharma, S., Swayne, A. D., and Obimbo, C. (2016). *Trend analysis and change point techniques: a survey*. Ener. Ecol. Environ., 1(3),123-130. doi: 10.1007/s40974-016-0011-1.

- [94] Sprent, P. (1961). *Some hypotheses concerning two phase regression lines*, 17(4), 634-645. doi: 10.2307/2527862.
- [95] Takayasu, H. (2015). *Basic methods of change-point detection of financial fluctuations*. International Conference on Noise and Fluctuations (ICNF). doi: 10.1109/ICNF.2015.7288606.
- [96] Tartakovsky, A. (1995). *Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in a non homogeneous Gaussian process*. Math. Methods of stat., 4(4), 389-404.
- [97] Tartakovsky, G. A., and Moustakides, V. G. (2010). *State-of-the-art in Bayesian changepoint detection*. Sequential Analysis, 29(2), 125-145. doi: 10.1080/07474941003740997.
- [98] Tartakovsky, G. A. (2008). *Multi decision quickest change point detection: previous achievements and open problems*. Sequential Analysis, 27(2), 201-231. doi:10.1080/07474940801989202.
- [99] Tsay, R. S., (1990). *Testing and modelling threshold autoregressive processes*. Journal of the American Statistical Association, 84, 231- 240.
- [100] Tze-San, Lee. (2010). *Change point problems: Bibliography and Review*. Journal of Statistical Theory and Practice, 4(4), 643-662. doi: 10.1080/15598608.2010.10412010.
- [101] Wang, Y., Wu, C., Ji, Z., Wang, B., and Liang, Y. (2011). *Non-parametric change-point method for differential gene expression detection*. PloS one 6: e20060.
- [102] Wei D., Shi X., and Yuehua Wu Y. (2014). *Application of change point detection in image retrieval*. Data Analytics 2014 : The Third International Conference on Data Analytics. Recuperado de [www.data_analytics_2014_2_40_60067\%20\(1\).pdf](http://www.data_analytics_2014_2_40_60067\%20(1).pdf)
- [103] Worsley, K. J. (1986). *Confidence regions and test for a change point in a sequence of exponential family random variables*. Biometrika, 73(1), 91-104. doi: 10.1093/biomet/73.1.91.
- [104] Xu Z., Kass-Hout T., Anderson-Smits C., and Gray G. (2015). *Signal detection using change point analysis in postmarket surveillance*. Pharmacoepidemiology and drug safety, 24, 663-668.
- [105] Yao, Y. (1984). *Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches*. The Annals of Statistics, 12(4), 1434-447.

-
- [106] Yong Zhou, Wan T. K. A., Xie, S., and Wang, X. (2010). *Wavelet analysis of change points in a non-parametric regression with heteroscedastic variance*, 159(1), 183-201. doi: 10.1016/j.jeconom.2010.06.001.
- [107] Qi, J.P., Zhang, Q., and Qi, J. (2014). *A novel method for fast change-detection on simulated time series and electrocardiogram data*. PLoS ONE 9(4) :e93365. doi.10371/journal.pone.0093365.
- [108] Zoeter, O., and Heskes, T. (2005). *Change point problems in linear dynamical systems*. Journal of Machine Research, 6, 1999-2026.
- [109] Zou, C., Liu, Y., Qin, P., and Wang, Z. (2007). *Empirical likelihood ratio test for the change-point problem*. Statistic & Probability Letters, 77(4), 374-382. doi: 10.1016/j.spl.2006.08.003.